2018•2019
Faculteit Industriële ingenieurswetenschappen
master in de industriële wetenschappen: biochemie

Masterthesis

Prediction of therapy response in lung cancer patients:
Optimization and evaluation of a robust 1H-NMR metabolomics
protocol

PROMOTOR :
ing. Liesbet PAULS

PROMOTOR :
Prof. dr. Peter ADRIAENSENS

COPROMOTOR :
Prof. dr. Michiel THOMEER

BEGELEIDER :
Mevr. Elien DERVEAUX

Kenny Schraepen
Scriptie ingediend tot het behalen van de graad van master in de industriële wetenschappen: biochemie

Gezamenlijke opleiding UHasselt en KU Leuven

▶▶ UHASSELT   KU LEUVEN

▶▶ UHASSELT   KU LEUVEN

2018•2019
Faculteit Industriële ingenieurswetenschappen
master in de industriële wetenschappen: biochemie

Masterthesis

Prediction of therapy response in lung cancer patients:
Optimization and evaluation of a robust 1H-NMR metabolomics
protocol

PROMOTOR :
ing. Liesbet PAULS

PROMOTOR :
Prof. dr. Peter ADRIAENSENS

COPROMOTOR :
Prof. dr. Michiel THOMEER

BEGELEIDER :
Mevr. Elien DERVEAUX

Kenny Schraepen
Scriptie ingediend tot het behalen van de graad van master in de industriële wetenschappen: biochemie

▶▶ UHASSELT KU LEUVEN

Foreword

I got the chance to perform my internship in a study of the hospital 'Ziekenhuis Oost Limburg' in collaboration with the UHasselt. This study is called the Prolung study. During these four months I got to assist with the starting phase of the ^{1}H-NMR measurements at the UHasselt.

My thesis was especially the optimization and evaluation of a robust ^{1}H-NMR metabolomics protocol. Additionally, I performed multivariate statistical analyses on previously acquired data.

Firstly, I want to thank Phd. student Derveaux Elien for a chance to perform my internship at the Prolung study. Also, I want to thank her for the support, adjustment and improvement of my work that I delivered.

Secondly, I want to thank Professor doctor Adriaensens Peter for the support and revise of my work about nuclear magnetic resonance and Professor doctor Thomeer Michiel for the support and revise of my work about lung cancer.

Thirdly, I want to thank ing. Pauls Liesbet for supporting me and coaching me through this period. Also, I like to thank her for adjusting and improving my thesis.

Thereafter, I want to thank ing. Reekmans Gunther and Van Vinckenroye Koen for the support and help they gave me in this internship.

Also, I like to thank academic bachelor student chemistry - biochemistry Philippens Nele for the support during the first two months.

After this, I would like to thank Smets Gwendolien and Vos Stijn for proofreading and improving my thesis and to give a lot of support during my internship.

At last, I want to thank my family for the support that they gave me through this stressful period.

Table of content

List of tables

List of figures

Abstract (in English)

Metabolomics, the study of small metabolites, can be used to discover biomarkers. Such biomarkers can provide information about the biochemistry of different biochemical pathways. To better understand and evaluate cancer-related aberrations in biochemical pathways, the metabolite peaks in the proton nuclear magnetic resonance (^{1}H-NMR) spectrum must be identified in a correct and quantified manner.

The peak positions in an ^{1}H-NMR spectrum are sensitive to changes in the protocol. Therefore, this thesis covers the optimization of a robust ^{1}H-NMR metabolomics protocol for human blood plasma. Moreover, this thesis performs multivariate statistical analyses of previously acquired metabolomics data to determine if the metabolite profile can discriminate between smokers, ex- and non-smokers.

During this optimization, several parameters were evaluated namely buffer components, buffer capacity, plasma concentration, total sample volume, temperature, power of water suppression and total echo delay in the Carr-Purcell-Meiboom-Gill pulse sequence. The statistical analyses were performed in SIMCA, which created a model that differentiates patients based on their smoking status. This model was then tested for possible confounders.

The optimization results in a robust ^{1}H-NMR metabolomics protocol for human blood plasma. The multivariate statistical analysis shows that a discrimination can be made between smokers and ex- or non-smokers based on the metabolite profile. And no possible confounders for the smoking status model are found.

Abstract (in Dutch)

Metabolomics, de studie van kleine metabolieten, kan gebruikt worden om biomerkers te ontdekken. Dergelijke biomerkers leveren informatie over de biochemie van verschillende biochemische paden. Om kanker gerelateerde afwijkingen in biochemische paden te evalueren, moet een proton nucleaire magnetische resonantie (^{1}H-NMR) spectrum in staat zijn om plasmametabolieten op een correcte en kwantificeerbare manier weer te geven.

De piek posities in een ^{1}H-NMR-spectrum zijn gevoelig aan veranderingen in het protocol. Daarom werkt deze thesis de optimalisatie van een robuust ^{1}H-NMR-metabolomics protocol uit voor humaan bloedplasma. Bovendien worden ook multivariate statistische analyses uitgevoerd op eerder verkregen metabolomics data om te bepalen of het metabool profiel kan discrimineren tussen rokers, ex- en niet-rokers.

Bij de optimalisatie worden verschillende parameters geëvalueerd zoals buffercomponenten, buffercapaciteit, plasmaconcentratie, totaal volume, temperatuur, vermogen van water onderdrukking en Carr-Purcell-Meiboom-Gill pulssequentie. De statistische analyses worden uitgevoerd in SIMCA, waarbij een model gebaseerd op rokersstatus wordt gemaakt. Dit model wordt dan getest op mogelijke confounders.

De optimalisatie resulteert in een robuust ^{1}H-NMR metabolomics protocol voor humaan bloedplasma. De multivariate statistisch analyse toont aan dat er een onderscheid kan gemaakt worden tussen rokers en ex- of niet-rokers gebaseerd op het metabool profiel. Ook zijn er geen confounders gevonden voor het rookstatus model.

1 Introduction

Lung cancer is one of the most common cancer types, as well as one of the leading causes of cancer-related death worldwide. The high mortality is due to a difficulty in early detection of lung cancer [1], [2]. The most common type of lung cancer is non-small cell lung carcinoma (NSCLC) that represents 85% of all lung cancers. Lung cancers are categorized in stages as well as assigned a grade during diagnosis. If the patient is diagnosed with a stage I-IIIA, the currently prevalent treatment is surgical removal of the tumor [3]. However, 30-50% of those patients have a poor prognosis [4].

Metabolomics, which can be defined as the study of chemical processes of the metabolism involving small, low weight molecules in living organisms, can be used to study biofluids and as such provide much information about a disease. With this information, researchers aim to comprehend the metabolism of cancer cells and develop better treatment strategies for patients with lung cancer. It is already known that the metabolism of cancer cells differs from that of normal cells [5], [6]. First, cancer cells divide uncontrollably, which requires an increase in the anabolic activities of the cell. All mammalian cells break down glucose through the glycolysis and the tricarboxylic acid cycle (TCA) also known as the citric acid cycle or the Krebs cycle. After the TCA, these cells require oxygen to generate energy through the oxidative phosphorylation [7]. Secondly, Otto Warburg found that the cancer cells rely more on the glycolysis for energy production than on the oxidative phosphorolysis, even in the presence of oxygen. This phenomenon is called the Warburg-Effect [1], [7]–[9]. For example, it was determined that cancer cells take up more glucose and glutamine for energy production and have a higher secretion of lactate as a metabolite to regenerate the co-enzyme NADH from the glycolysis.

Metabolites are already measured with gas chromatography, liquid chromatography and mass spectrometry. Dr. J. Nicholson is one of the pioneers for the use of nuclear magnetic resonance (NMR) spectroscopy for the analysis of biofluids [10]. NMR spectroscopy has already been used in several other study areas like in breast cancer, toxicology and coronary heart diseases [9], [11], [12]. The use of this technique for the metabolic profiling of lung cancer by plasma analysis is innovative. The technique has already proven to be valuable for the detection of lung cancer but not yet for the evaluation of a metabolic profile over time. It provides a non-invasive identification and quantification of metabolites in a complex mixture, requires minimal or no sample preparation, and is sensitive and reproducible [13]–[15].

The Prolung study with NCT number NCT02024113 is a new study at the Hospital 'Ziekenhuis Oost-Limburg (ZOL)' in cooperation with UHasselt and funded by 'Kom op tegen kanker'. In this study, patients with non-small cell lung cancer (NSCLC) who undergo surgery to remove the primary lung tumor with curative intent are investigated. From these patients, plasma samples are collected at different times before and after surgery. These samples will be examined with proton nuclear magnetic resonance (^{1}H-NMR) spectroscopy [16]. The set-up of the Prolung study is based on the problem that for 30 to 55% of patients with a stage I-IIIA NSCLC undergoing a surgery (i.e. lobectomy), a relapse occurs [4]. The main goal of this study is to use the metabolic profile of the individual to predict whether the surgical removal of the lung tumor will have a positive effect or if there will be a relapse in a time span of one year.

The inclusion of patients and collection of multiple plasma samples is currently ongoing in different hospitals. The ^{1}H-NMR analyses of the plasma samples still need to start.

Before these analyses can start, the optimal understanding of the spectrum is needed. Therefore, the protocol should be carefully evaluated and optimized. This thesis will contribute in the optimization and evaluation of the ^{1}H-NMR protocol and in addition perform statistics on previously acquired data.

The metabolomic experiments that will be performed are very sensitive to changes in the protocol. The optimization of the ^{1}H-NMR protocol includes revision of different analytical parameters. This protocol is of uttermost importance since a very robust and reproducible measuring method is desired to analyze patient plasma samples.

Moreover, a dataset from a previous study was analyzed. These raw data includes information of the metabolic profile from smokers and non-smokers [8]. Thus, an additional hypothesis is to search for differences in the metabolic profile between smokers and non-smokers.

2 Lung cancer

Like stated in the introduction lung cancer is one of the most appearing cancer types worldwide [17]. The 5-year survival rate is about 16% [3]. It is the most important cancer-related death cause in industrialized countries and accounts for one out of three cancer deaths in men [3]. Around 2013 there were 8196 new diagnosis of lung cancer of which 70% were males and 30% were females in Belgium [18].

2.1 Cancer cells behavior

Cancer cells metabolism relies on an irreversible, uncontrollable growth. Therefore, the cell division or mitosis will be increased. This means that the cell needs more DNA, proteins and lipids, which translates in increased anabolic activities of the cell. Because of this increased mitosis the distance from the cells to the blood vessels grows larger. This distance decreases the supply of nutrients and oxygen thus the biosynthetic processes and energy supplies decreases as well. These cancer cells will have undergone a metabolic transformation which results in a reprogrammation in multiple ways such as a modified metabolism. Therefore, there is an in- and decrease in the uptake of several metabolites [7].

2.2 Types of lung cancer

Lung cancer can be divided into four major histologic tumor types: adenocarcinoma, squamous cell carcinoma, small cell carcinoma and large cell carcinoma. Those four groups can be grouped into two broad classes, NSCLC and small cell lung carcinoma (SCLC) [3]. Around 85% of all lung cancer patients will be diagnosed with NSCLC [17]. Unlike NSCLC, SCLC mainly have metastases at the time of diagnosis and is not curable with surgery [3]. These tumor types can further be typified into stadia and grades.

2.3 The different stadia and grading

When a patient is diagnosed with lung cancer, a grade and stage will be allocated according to the TNM classification, from which currently the eighth version is ongoing [19]. First, 'T' describes the size of the primary tumor. Secondly, 'N' defines if there is metastasis of a regional lymph node. And last, 'M' accounts for possible metastasis on a greater distance from the primary tumor [7]. The TNM score is based on expertise in multiple fields including clinical- and radiographic examination, imaging with computed tomography (CT) and magnetic resonance imaging (MRI) scans and in some cases a surgical exploration is necessary [3]. The lung tumor will thus be given a code in order of increasing anaplasia; I, II, III, IV and subclasses like A or B [3].

3 Metabolism of normal and cancer cells

The metabolism of the cell is very complicated. The production of energy of normal cells is mainly covered by the breakdown of glucose. This process consists of several steps and hereby different metabolites are produced as shown in Figure 1 and Figure 2.

Firstly, glucose is transported in the cell by glucose transporters and undergoes the glycolysis where it will be converted to pyruvate as shown in Figure 1.
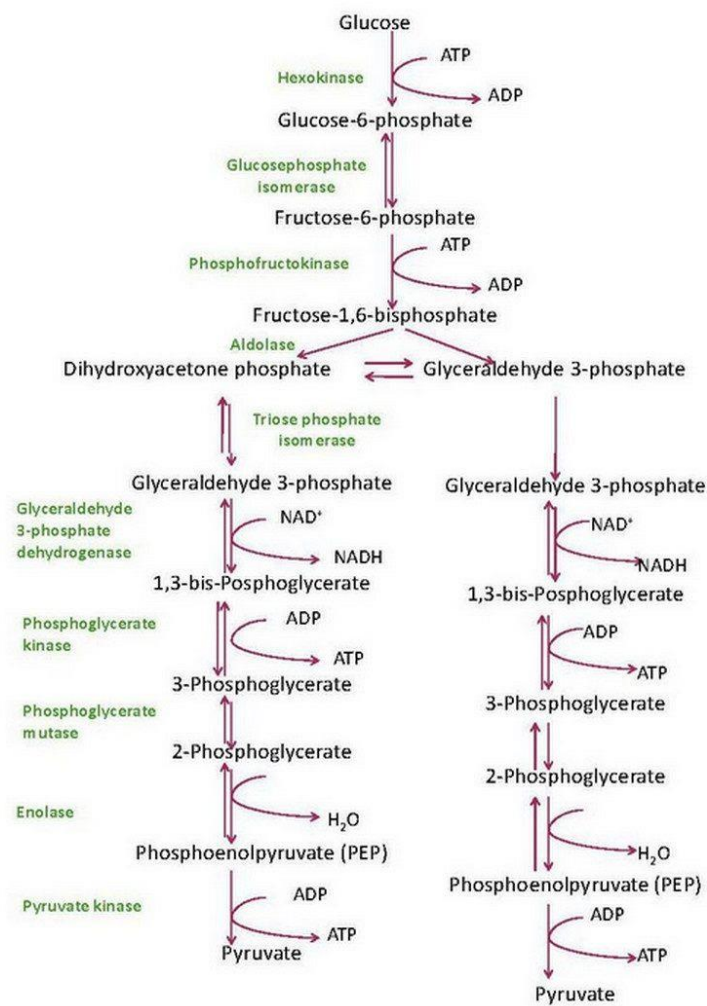
Figure 1: Glycolysis of normal cells [20]

Pyruvate will be transformed in acetyl-coA in an oxidative decarboxylation step and further acetyl-coA will be oxidized inside the TCA producing reduced co-enzymes as shown in Figure 2.
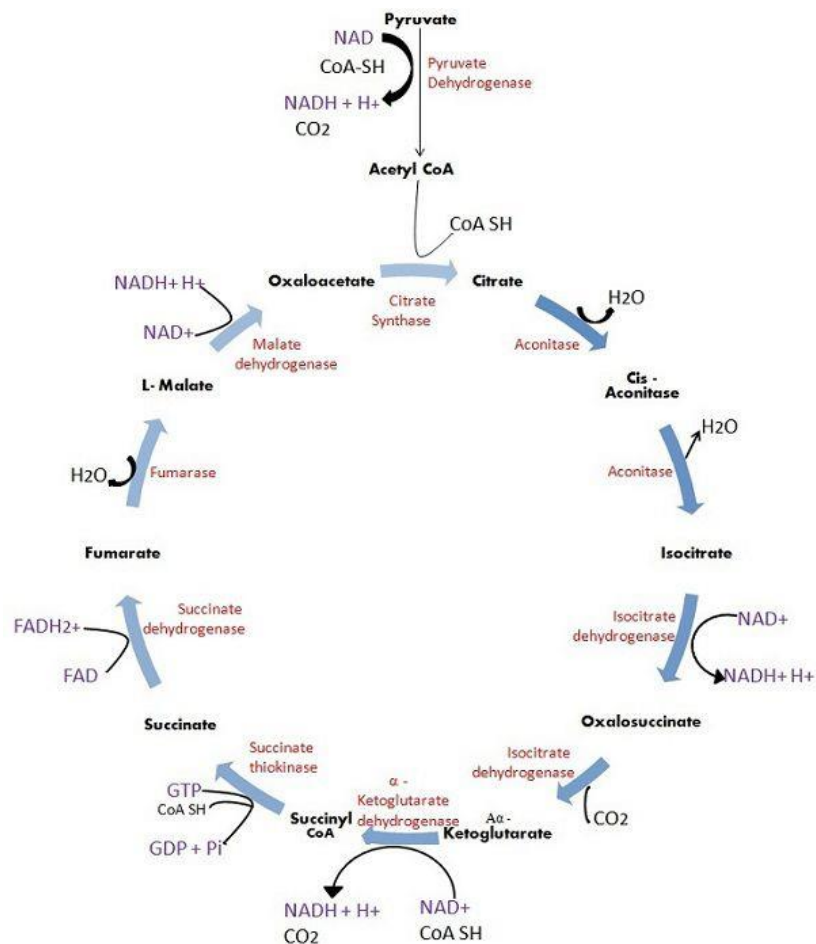
Figure 2: TCA cycle of normal cells [21]

The regeneration of the co-enzymes, that were used during the TCA namely nicotinamide adenine dinucleotide (NADH) and flavin adenine nucleotide ($FADH_2$), happen through the electron transport chain (ECT) where energy will be produced in the form of adenosine triphosphate (ATP) molecules. The glucose will then be fully degraded to carbon dioxide (CO_2), water (H_2O) and energy [22].

Metabolomics is the study of these small, low molecular weight metabolites in biofluids such as blood plasma which is used in this thesis. It enables the evaluation and quantification of individual metabolites that are expressed by cells. These metabolites can then be used to get a better understanding of the metabolic pathways of certain cells like lung cancer cells. Other clinical related metabolomics goals might include differentiation of the tumor type, or tumor stage and response prediction to different drug treatments of several diseases [17].

Through genetic mutation the cancer cells undergo a metabolic change which will cause them to grow uncontrollably [23]. When those normal cells turn cancerous, the level of some metabolites will in- or decrease inside the tumor. Those metabolites can on their turn be released into the bloodstream. On the other hand when they release metabolites into the bloodstream certain metabolites will be taken from the bloodstream into the cells for proliferation, therefore decreasing those in the bloodstream [17]. Proliferating cells take up more nutrients for bioenergy and shunt metabolites into pathways that support biosynthesis. This supply of the glycolysis and TCA is called anaplerosis [17]. This cancer metabolism will be further explained in the discussion part.

One of the first discovered difference between malign and benign cell metabolism inside tumor cells is called the Warburg-effect.

18

3.1 The Warburg-effect

In 1920, the scientist Otto Warburg discovered that cancer cells have a different cellular oxidation and oxygen usage [24]. For this discovery Otto Warburg received the Nobel prize in 1931 [7]. Warburg found that proliferating tumor cells consume more glucose at higher rates compared to normal cells. Most of the glucose derived carbons were found in large amounts of excreted lactate. Even in the presence of oxygen there is an increased amount of lactate observed. This effect is called the Warburg-effect as shown in Figure 3.
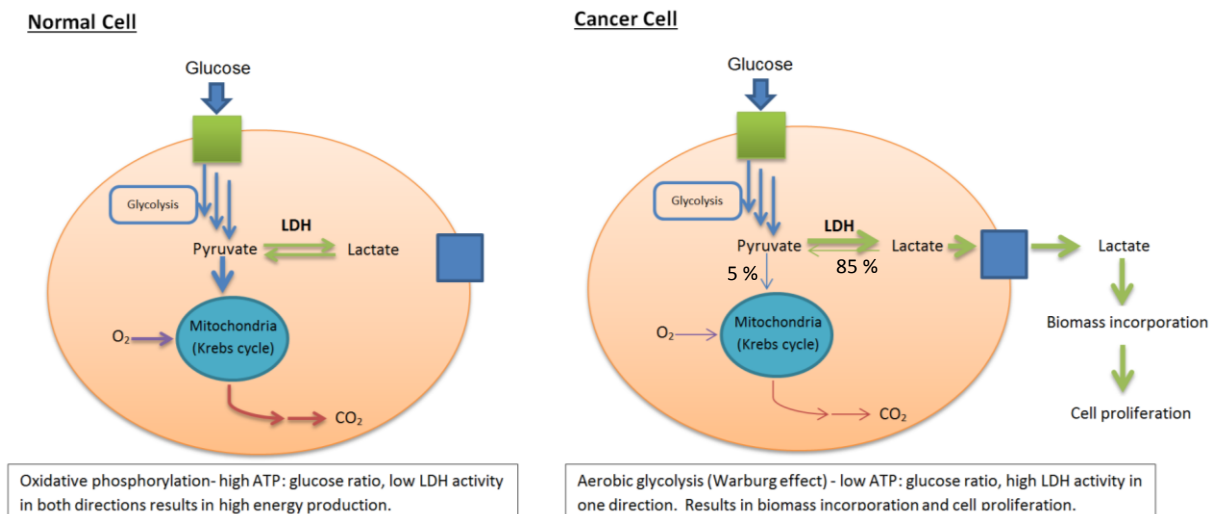
Figure 3: Difference in glucose breakdown between normal cells and cancer cells [25]

The conversion of pyruvate to lactate through lactate dehydrogenase (LDH) with NADH recuperation is faster than the recuperation through the oxidative phosphorylation. This means that the glycolysis becomes more important in cancer cells compared to the full oxidative phosphorylation [7], [26].

When the lactate molecules are released, the extra cellular environment will acidify. This pH lowering will lower the survival of normal cells and is therefore a disadvantage for the immunologic cells, like the T-cells. This will increase the invasiveness of the tumor cells [7], [23].

3.2 Metabolite levels from cancer patients

The detection of metabolites from lung cancer patients can be investigated on different types of samples. The four most studied samples for metabolite detection are blood plasma, blood serum, lung tissue and urine.

3.2.1 Blood plasma

As previously mentioned, the uptake and the release of small metabolites will cause a change of the metabolite levels in the blood stream. Rocha *et al.* and Duarte *et al.* also did research on blood plasma from patients with stage I-IIIA NSCLC. They found that these patients had higher HDL, VLDL, LDL, lactate and pyruvate concentrations. But they also found a decreased amount of glucose, formate, acetate, methanol and several amino acids (alanine, glutamine, histidine, tyrosine and valine). Those increased

19

and decreased metabolites can be linked to the enhanced glycolysis, glutaminolysis and gluconeogenesis and the suppression of the TCA and a reduced lipid catabolism [27], [28]. The suppression of the TCA happens because of the Warburg effect. When the pyruvate is converted to lactate instead of acetyl-coA [7], [26].

3.2.2 Blood serum

Blood serum difference from blood plasma. Sera is derived from blood that has been coagulated, so no coagulation factors are present in the sera [29]. Plasma is derived from blood that had an anti-coagulants, like EDTA or heparin added to the blood tube [29]. Therefore, the concentration of some metabolites differs from plasma.

Zhang *et al.* researched the blood serum of lung cancer patients and found that there was an increase in glutamine levels. Glutamine enters the cell and can be converted to glutamate. If a deamination finds place the glutamate converts to α-ketoglutarate, a TCA cycle intermediate, through the glutaminolysis. This intermediate can further be converted by several enzymes to aspartate, malate and pyruvate. Therefore, glutamine is an important amino acid for cancer cell growth and proliferation [30].

Puchades *et al.* found decreased levels of glutamine levels in sera from lung cancer patients. But this difference can be related to the population used for the study. Zhang *et al.* used early stage lung cancer patients whereas Puchades *et al.* used all stages [30], [31]. These findings could indicate that disease progression demands for more glutamine. Therefore, lower levels of glutamine are observed in the sera from lung cancer patients.

Another characterization of lung tumor discovered by Chen *et al.* are lower serum levels of histidine and threonine. These are being used in the glycine-, serine-, threonine- and pyrimidine pathways that are upregulated in lung tumors. Also, the levels of choline are often lowered in lung cancer patients. Choline is a precursor to membrane phospholipids, those membrane phospholipids are in highly demanded in proliferating tumors [32]. Other metabolite changes include a decrease in 2,3,4,-trihydroxybutyric acid and an increase in α-hydroxyisobutyric acid. These changes in those metabolites occur both before and after surgical resection of the tumor [32].

Hori *et al.* found that the levels of the following metabolites were significantly changed: the levels of lactic acid, fumaric acid and malic acid were increased in the lung cancer patients. Also, an increase in the serum level of some amino acids were observed, these include sarcosine, proline, threonine and 4-hydroxyproline. On the contrary, the levels of aspartic acid, 5-oxoproline and l-glutamic acid were decreased.

Also some other metabolites changed in level compared to healthy volunteers. An increase of level was observed of 2-hydroxyisobutyric acid, 3-hydroxybutyric acid, 2-hydroxyisovaleric acid, malonic acid, benzoic acid, octanoic acid, 2-hexenedioic acid, 2-propyl-glutaric acid and uric acid. A decrease of level was observed of glycerol, phosphoric acid, glyceric acid and lauric acid [1].

20

3.2.3 Lung tumor tissue

Whereas blood plasma and serum metabolomics studies provides information about metabolites in the blood stream, studies using lung tumor tissue samples describe the metabolite concentrations inside the cancer cell.

Increased levels of alanine are observed in most tumors, including lung tumors, due to increased glycolysis. There is also an increased uptake of valine, leucine and isoleucine by lung tumors as these are required to fuel the production of Krebs cycle intermediates. Glutamine levels are usually found to be lower in tumor tissues, as a large amount of glutamine is metabolized in tumor tissues [24].

In contrast, Hori *et al.* found higher glutamine levels in lung tumor tissue compared to normal tissue. A possible explanation might be that glutamine levels are usually higher in smokers and a large percentage of lung cancer patients in the above study were chronic smokers [1].

Hori *et al.* found that the levels of lactic acid, succinic acid, fumaric acid and malic acid were increased. In the lung tissue, the levels of the following amino acids were increased: alanine, valine, leucine, isoleucine, proline, glycine, serine, threonine, aspartic acid, methionine, 5-oxoproline, cysteine, glutamic acid, phenylalanine, N-acetylaspartic acid, glutamine, ornithine, tyrosine and tryptophan. The levels of propanoic acid, glycolic acid, oxalic acid, 2-hydroxybutylic acid, butanoic acid, ethanimidic acid, malonic acid, phosphoric acid, aminomalonic acid, thiodiglycolic acid, 2-hydroxyglutaric acid, 4-hydroxyphenylacetic acid, aminoadipic acid, 4-hydroxyphenyllactic acid, gluconic acid and stearic acid in the tumor tissue were also increased, and the levels of lauric acid and myristic acid were decreased. Regarding nucleotides, the levels of inosine, adenosine, guanine, pyrimidine, uracil and 9H-purine in the tumor tissue were increased, although no nucleotides were observed in serum [1].

3.2.4 Urine

Carrola *et al.* found in urine samples of lung cancer patients an increase in: β-hydroxyisovalerate, α-hydroxyisobutyrate, N-acetylglutamine, creatinine and citrate. But a decrease in hippurate, trigonelline, trigonellinamide and phenylacetylglycine was observed [33].

4 Nuclear magnetic resonance spectroscopy

The measurement of these metabolites is performed with an NMR spectrometer. NMR spectroscopy is generally used for the study of magnetic properties of a nucleus from different atoms and molecules. Different atoms can be used such as hydrogen, carbon, nitrogen, or others. The technique used in this thesis is based on the magnetic rotation of hydrogen atoms and is called ^{1}H-NMR. This will be used to measure the concentration of those small metabolites in biofluids, in our case human blood plasma [17]. This robust technique requires little or no sample preparation. In addition, NMR has a high reproducibility, gives good results for quantification and provides extensive structural information [13]–[15].

4.1 Principle of ^{1}H-NMR

The general principle of ^{1}H-NMR is to change the core spin rotation of the hydrogen atom. All hydrogen nuclei are electrically charged and spin around their own axis. This angular momentum P and the charge will together create a magnetic moment µ [34].

$$\mu = P * \gamma \qquad (1)$$

The γ in the equation (1) represents the magnetogyric ratio and its value is constant for each type of nucleus. Figure 4 shows the angular momentum P and magnetic momentum µ. Both are represented as a vector with a specific magnitude and direction.
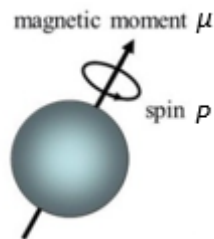
Figure 4: The angular momentum P (spin) and the magnetic moment µ of a nucleus [35]

All the magnetic moments are randomly orientated as shown in Figure 5. When hydrogen nuclei or protons are placed in a strong external magnetic field B_0, all the random orientated magnetic moments of the protons will align with this field as shown in Figure 5.
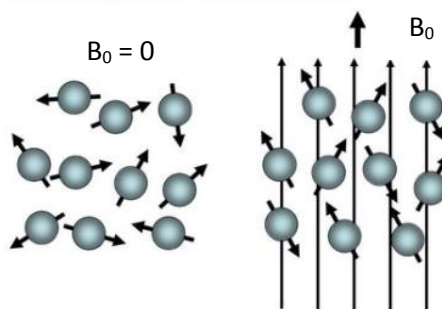
Figure 5: Spinning nuclei in an external magnetic field B_0 [35]

Some of them will point in the same direction of the magnetic field lines, this is called parallel or the α-state. The other will point in the opposite direction and are called anti-parallel or β-state. There will always be more magnetic moments directed in parallel compared to antiparallel, because protons with parallel orientated magnetic moment are at a lower energy level [34], [36].

These protons and their magnetic moments in the magnetic field B_0 will be simply represented by a vector in a coordinate system, whereas the B_0 field will always be in the positive direction of the Z-axis as shown in Figure 6.
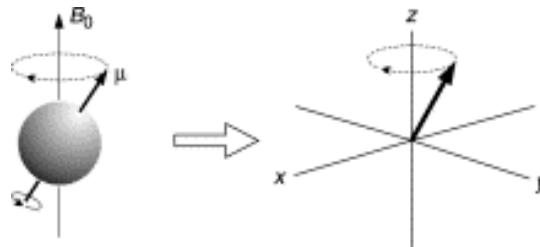
Figure 6: The transformation of a proton to a vector [29]

The net magnetization M_0 representation of all the nuclear magnetic moments in the Z-axis and is therefore pointed in the same direction as B_0 since more magnetic moments are parallel orientated with the B_0 as previously explained [34]. Figure 7 represents M_0 by one vector in the Z-axis.
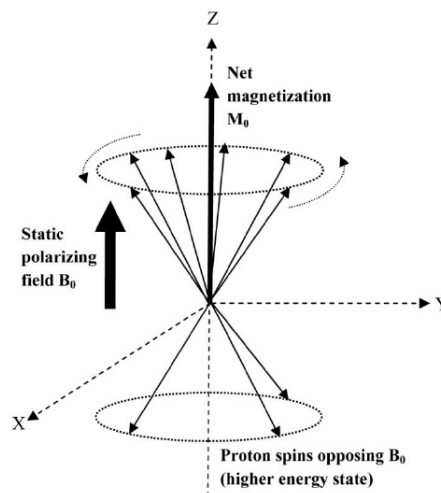
Figure 7: Spinning nuclei in an external magnetic field with a net magnetization M_0 [37]

Through the effect of the static field on μ, a torque on μ is created, which therefore traces a circular path around the applied field, this is known as precession. So, the Larmor frequency or the Larmor precession is the movement of the nuclei around the direction of the B_0 field as shown in Figure 8 [34].
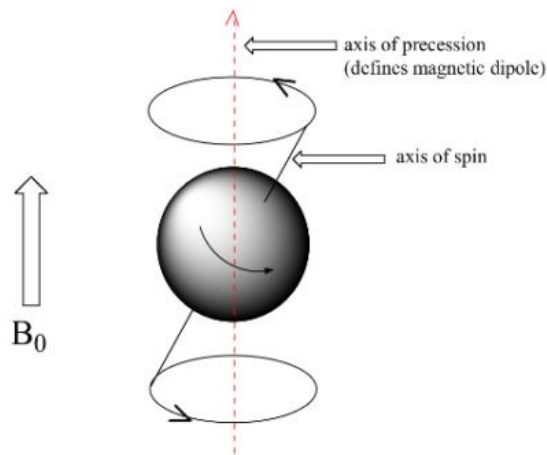
Figure 8: Axis of spin and precession [35]

After the alignment of the magnetic moments in the Z-axis, a well-defined radio frequency (RF) wave or a so-called pulse is transmitted to the protons. This frequency is the Larmor frequency [34]. This pulse will cause the protons to absorb energy and go in phase coherence, this is called excitation as shown in Figure 9 and Figure 12 [34]. This phase coherence is when all the magnetic moments from the protons are directed to one side. When the total vector M_0 reached the Y-axis, it represents a 90° pulse as shown in Figure 9.
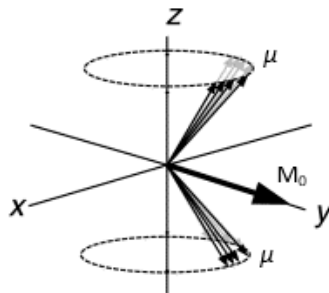
Figure 9: All magnetic moments after the 90° pulse [34]

This short RF 90° pulse is accurately set till a new state of resonance is formed equalizing the population of the α- and β-state, thus were the number of protons with parallel orientated magnetic moments is equal to those in anti-parallel and there is no net Z magnetization anymore. The net magnetization M_0 is now in the X-Y plane. At this point the previously stated phase coherence is forced upon them by the radio frequency pulse [34].

Immediately after the 90° pulse is given, the total net magnetization M_0 is at maximum in the Y-axis. Then the net magnetization in the Y-axis will decrease while it will increase in the Z-axis until there is no net magnetization in the X-Y plane left. This process is called T_1 relaxation as shown in Figure 12. Also, when the 90° pulse stops, all the different magnetic moments will start to fan out in the X-Y plane as shown in Figure 10 [34].
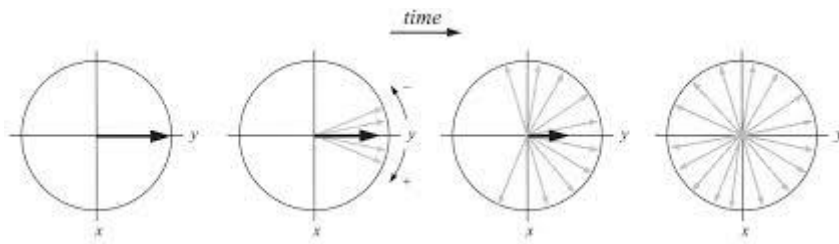
Figure 10: Fanning out of fast and slow magnetic moments in a rotating coordinate system [34]

To better explain this fanning, one must assume the coordinate system is rotating clockwise at the speed of the Larmor frequency. All the magnetic moments precess clockwise, but at slightly different speeds. The magnetic moments that precess faster than the Larmor frequency are displayed on the bottom half of the circle, whereas the moments that precess slower than the Larmor frequency are represented on the top half of the circle. As a consequence of fanning, all magnetic moments lose their phase coherence or are so-called dephasing [34].

Thus, because of the signal decrease in the Y-axis, the NMR signal decays with time and produces a Free induction decay (FID). This decay happens by two relaxation contributions as shown in Figure 12, T_1 and T_2. T_1 is the recovery of the magnetization along the Z-axis by losing excess of energy. This happens until the equilibrium population will be re-established. T_2 is the decay of the magnetization in the X-Y plane, due to the loss of phase coherence (fanning out), i.e. T_2 relaxation, and this is measured as the FID. This decay is linked to the emission of specific electromagnetic radiation emitted by each proton. The magnitude of the emission is dependent on the chemical environment of the proton. This will be measured in time by certain receiver coils directed along the Y-axis to get a signal, this result is like previously stated a FID as shown in Figure 11 [34].
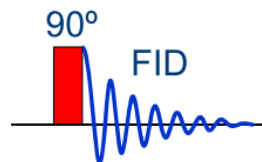
Figure 11: FID formation after 90° pulse

When the measurement is complete there will be no net magnetization in the X-Y plane and maximum M_0 along the Z-axis [17], [34], [36].
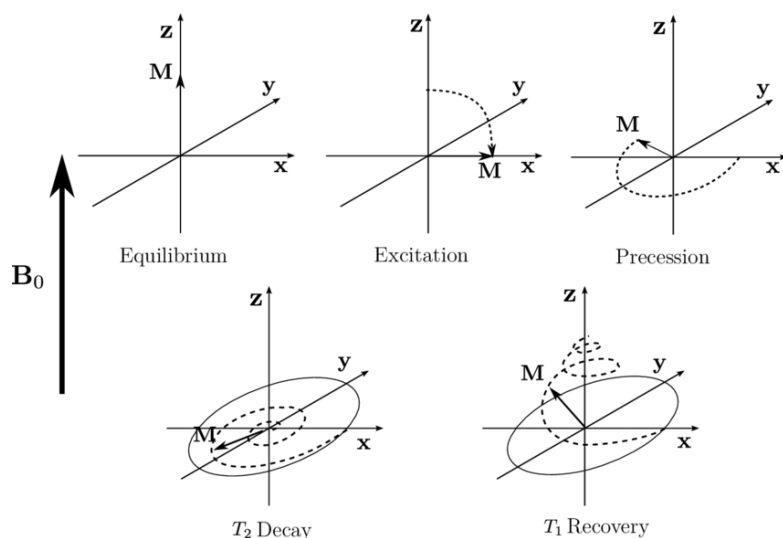
Figure 12: Different states when applying a specific radiofrequency [38]

Furthermore, the signal needs to be transformed from a time scale to a frequency scale. This is obtained by a Fourier transformation (FT). Figure 13 shows how the FT is being implemented on the FID of different molecules, whereas Figure 14 shows the FT of an FID from a mixture of molecules, which is the case in biofluids [17], [34], [36].
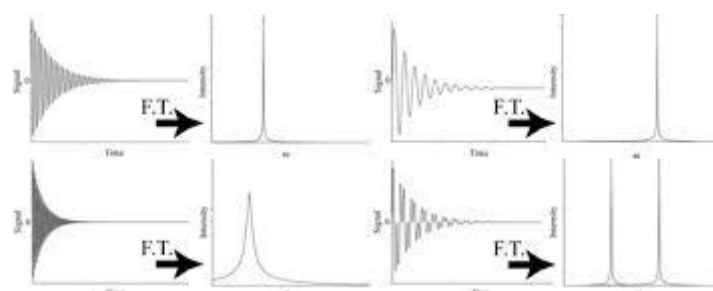
Figure 13: Different Fourier transformations of FID from simple molecules [39]

FID   **FT**
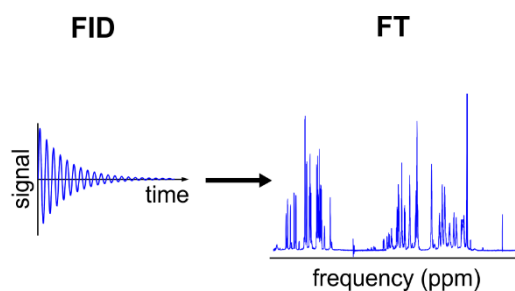
Figure 14: Fourier transformation of a biofluid FID [40]

After this transformation a typical ^{1}H-NMR spectrum is obtained as shown in Figure 15, whereas the Y-axis shows the abundance and the X-axis is converted from Hz to a magnetic field independent scale expressed in parts per million (ppm).
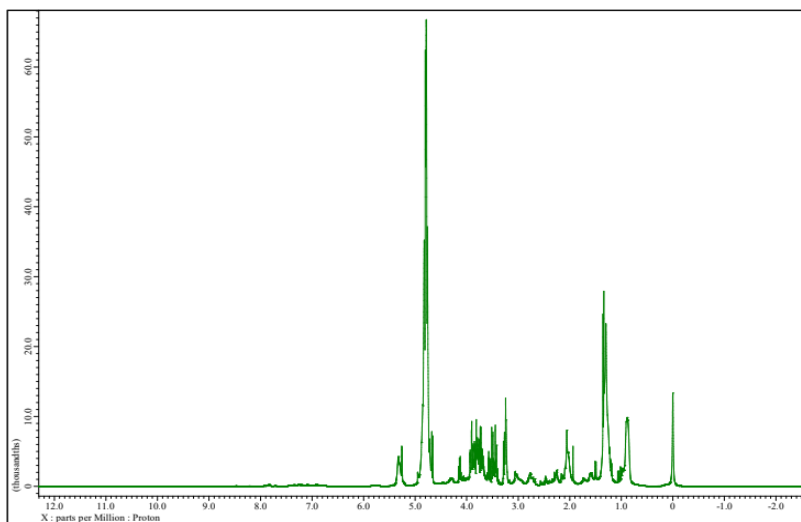
Figure 15: Example of ^{1}H-NMR spectrum

This spectrum provides information about the chemical environment and integral values of the molar amount of different molecules or mixtures of molecules. Consequently, in most of the mixture there is an overlap of proton resonance which leads to a specific peak in the spectrum that can be assigned to different molecules [17], [34], [36].

This thesis is about the measurement of human blood plasma, which is full of different molecules. To obtain a spectrum providing as much information as possible, a specific pulse sequence is required to suppress the macromolecules but not the metabolites in the sample.

4.2 Carr-Purcell-Meiboom-Gill sequence

The editing of the spectrum to filter out the macromolecules is called T_2-editing. This finds place if the difference in the T_2 relaxation times of the metabolites and macromolecules are sufficiently large enough. The faster relaxing molecules, the macromolecules mainly proteins, can be preferentially be reduced in intensity with the Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence, while the slower relaxing molecules, the metabolites, are decreased by a neglectable amount. This means the spectrum can be crudely edited according to molecule size retaining the resonance of smaller molecules at the expense of the more massive ones [34]. This also means that the larger molecules typically display broader resonances than smaller ones since they possess a shorter T_2 spin relaxation times as shown in Figure 16.
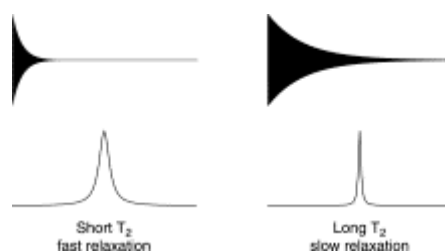
Figure 16: FT of a short T_2 - and long T_2 relaxation [34]

To better understand this, Figure 17 shows the T_2 decay of molecules. Those with a short T_2 represent the macromolecules and these with long T_2 represent the metabolites [34].

28
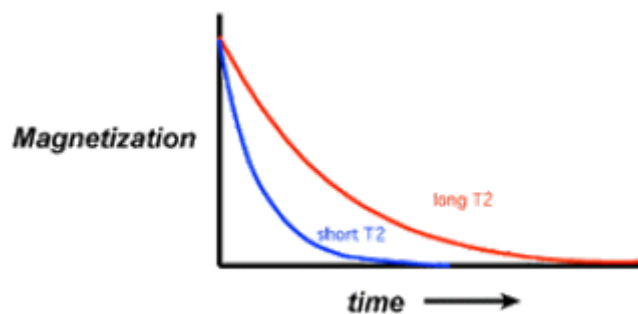
Figure 17: T_2 decay for metabolites (long T_2) and macromolecules (short T_2) [33]

Thus, for this T_2-editing a specific pulse sequence, the CPMG pulse sequence, is required to suppress the macromolecules for better visibility and interpretation of the small molecule metabolites [16], [27], [34]. The CPMG pulse sequence is a sequence of pulses, which can be represented by a diagram as shown in Figure 18.
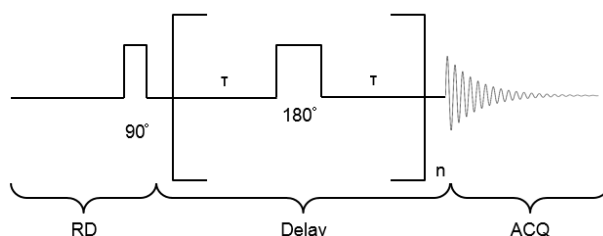
Figure 18: CPMG pulse sequence

This diagram is translated to the sequence of RD-90°-(τ-180$_Y$°-τ)n-ACQ. This sequence starts with the relaxation delay (RD), then a 90° pulse followed by a loop of 180° pulses along the Y-axis and evolution delays τ. The number of loops is the loop number (n) and after those loops the acquisition (ACQ) begins for obtaining the FID [34].

If the total echo time (n.2.τ), from the diagram in Figure 18, is too short then the macromolecules will not be suppressed properly. Consequently, if it is too long, the metabolites will also be suppressed. An optimum can be found where the metabolites are almost untouched in the ^{1}H-NMR spectrum with a good suppression of the macromolecules at the same time [34].

Now that all the terms of the diagram have been explained, the following paragraph covers the explanation of the 180° pulse of the CPMG pulse sequence whereas the 90° pulse is already been explained in a previous part.

The 180° pulse is a little more complicated to explain. First, after the 90° pulse takes place the magnetic moments fan out in the X-Y plane as shown in Figure 19. Then, the 180° pulse mirrors the µ vectors around the Y-axis [34].
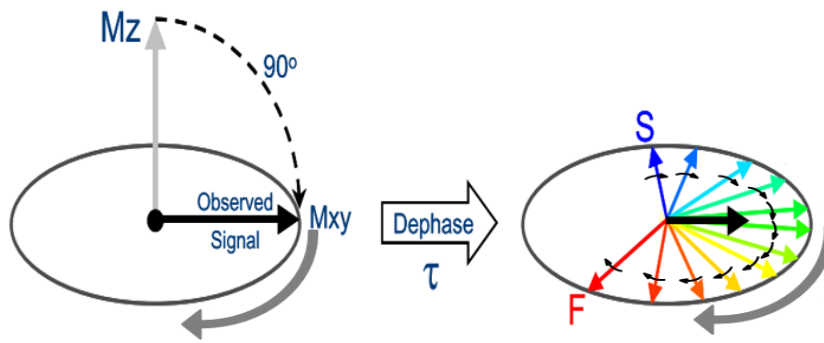
Figure 19: 90° pulse and dephasing in a clockwise rotating coordinate system [41]

To explain the 180° pulse in an understandable manner, Figure 20 shows runners in a circle which represents a circle around the Y-axis. Some of those runners will go faster than others. After a time period τ, a 180° pulse is given, which mirrors the position of the runners along the Y-axis. The fastest are now behind the slowest and all will come back together again at the start after a time period 2τ [34], [41].
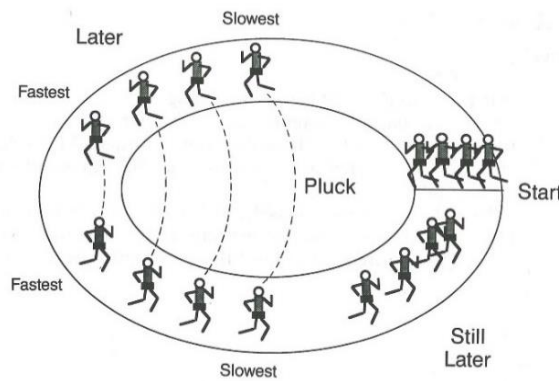
Figure 20: Display of fast and slow runners that when a 180° pulse takes place

Because of the explanation with the runners, the interpretation of the following figures might be better to understand. Figure 21 shows when the 180° pulse is given, all the magnetic moments change sides. The coordinate system is still rotating clockwise, but now the faster precessing ones are behind the slower ones. This results in the faster precessing magnetic moments to catch up to the slower ones and finally rephase at a maximum along the Y-axis [34], [41].
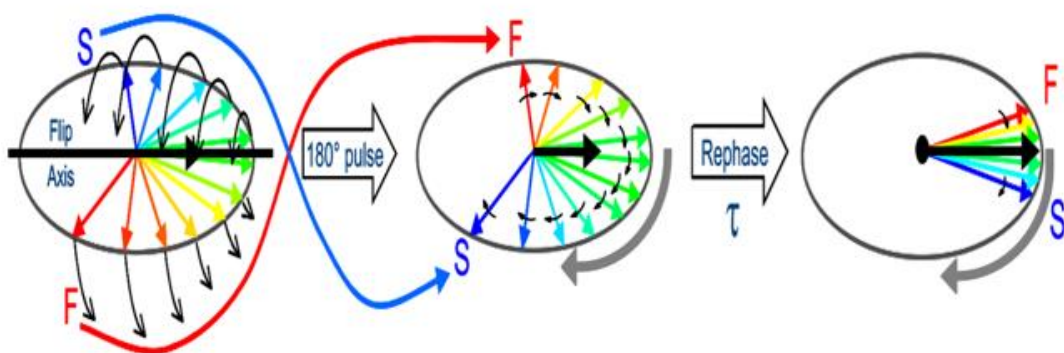
Figure 21: Changing direction after 180° [41]

In Figure 22 the CPMG diagram is added to better visualise what the 180°pulse does together with the 90° pulse and obtaining the FID.
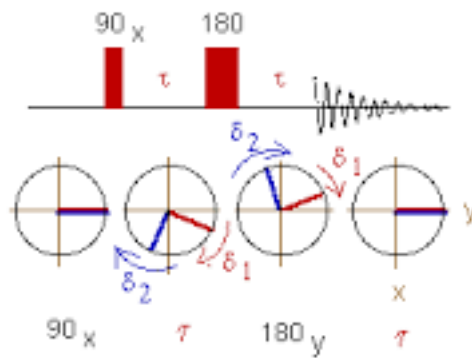
Figure 22: The 180° pulse with the CPMG diagram [42]

When the 90° pulse is given, all the magnetic moments are in phase. After this pulse the moments start to dephase, because some precess faster than others. As a result of the rotating coordinate system, δ_1 and δ_2 precess both faster than the Larmor frequency, whereas δ_2 precesses faster than δ_1. When the 180° pulse is given, the magnetic moments are being flipped around Y-Z plane and they start to get into phase again. Thus, the magnetic moments that precess faster than the Larmor frequency will catch up to the rotating coordinate system. The rotating coordinate system will now in its place catch up to the slower precessing magnetic moments [34], [41].

The CPMG pulse sequence also has a secondary effect: the sequence takes care of inhomogenities in the field [34]. Moreover, the loop number is always an even number to erase errors from the pulse length imperfections as shown in Figure 23 [34].
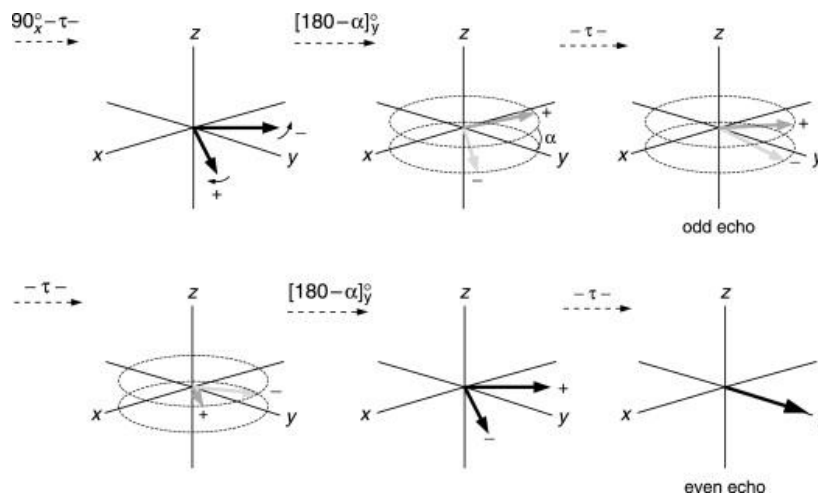
Figure 23: 180° pulse length imperfections are cancelled after each period of 2τ [34]

5 Prolung study

As stated in the introduction, the Prolung study with NCT number NCT02024113 is a recent longitudinal study at the Hospital 'Ziekenhuis Oost-Limburg (ZOL)' in cooperation with UHasselt. In the next phase, other Belgian hospitals will participate in the inclusion process in order to recruit as many patients as possible in a multicentric study design.

The Prolung study investigates if early disease recurrence in patients with NSCLC who will undergo surgery with curative intent, can be predicted through metabolic phenotyping by ^{1}H-NMR spectroscopy.

5.1 Study population of the Prolung study

Patients with a stage I-IIIA NSCLC who undergo complete surgical resection of the tumor qualify for the study, but they need to measure up to several inclusion and exclusion criteria and they need to be willing to provide a written informed consent. These subjects will be recruited among patients who are referred to the pulmonology departments of the hospital 'Ziekenhuis Oost-Limburg'. The inclusion and exclusion criteria are being represented in Table 1.

Table 1: The inclusion and exclusion criteria for the Prolung study

Inclusion criteria	Exclusion criteria
Patients with a stage I-IIIA NSCLC tumor who undergo complete surgical resection	No fasting starting from 22:00h the day prior to blood sampling
Patients who provide a signed written informed consent	Medication intake in the morning of the blood sampling
	Fasting blood glucose concentration higher than 200 mg/dl in the morning of the blood sampling
	History of cancer during the past five years
	Treatment for cancer during the past five years

5.2 Goals of the Prolung study

These goals are derived from the study protocol of the Prolung study, which is currently ongoing. From the population, it will be documented whether the disease recurs within one year or not. Then, the group of patients will be divided into two different groups: one group contains the patients with a disease relapse within one year and the other group contains those without a relapse.

The primary objective of the study is to search for differences in the plasma metabolic profile before surgery between these two groups. This grouping is performed so a specific metabolic profile can be linked with a possible relapse. In case a significant difference can be demonstrated, the metabolite profile can serve as a predictive biomarker for early disease relapse after surgery.

The secondary objective of this study is to compare the metabolic profile before surgery with the metabolic profile at different timepoints after surgery. This comparison is useful to check whether there will be changes in the individual profile over time. Thus, as a result of all this information from

the objectives a prediction can be made based on the metabolic profile whether a surgery is going to be successful or not. Therefore, this prediction needs to be as accurate as possible.

Some other objectives of the study include a different type of investigation where it will be examined if the degree of change in circulating tumor DNA (ctDNA) mutations in plasma before and after surgery correlates with the risk for early disease recurrence. Furthermore, it will be examined if a specific metabolic profile can be correlated to the presence or absence of certain hotspot genetic mutations. Also, it will be investigated if the tissue metabolic profile of the primary tumor correlates with the metabolic profile in plasma.

5.3 Blood sample collection

If a patient agrees to participate in the study, fasting venous blood samples will be collected at the day of the inclusion, the day of the surgery, week 1, 4, 6, 12 and week 52 after surgery or earlier when disease progression is diagnosed. The fasting venous blood sample will be collected in one lithium-heparin (LH) blood tube of 6 ml and one or two dipotassium ethylenediaminetetraacetic acid (K_2EDTA) blood tubes of 10 ml, depending on the timepoint.

All collected tubes will be put on crushed ice and kept on 4 °C within five to ten minutes after blood drawing. Within eight hours after blood collection, the blood samples will be centrifuged at room temperature [16]. After centrifugation, plasma should be carefully collected such that the pipette tip stays away from the buffy coat for at least 0.5 cm, as shown in Figure 24.
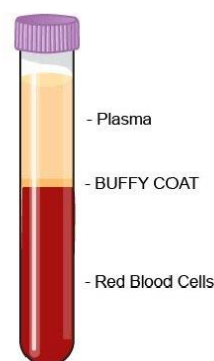
Figure 24: Blood sample after centrifugation [43]

Utmost care should be taken not to disturb the buffy coat. From the LH-coated sample, which is intended for metabolic analysis, four plasma aliquots of 400 µl will be transferred into sterile 500 µl cryovials. The plasma of the other K_2EDTA-coated blood samples is intended for ctDNA mutational analysis. The K_2EDTA plasma will be pooled and an additional centrifugation step for 10 minutes (6000 g) will be performed. In multiple 2 ml tubes, 1.7 ml plasma supernatant will be collected. This plasma is intended for ctDNA mutational analysis.

Afterwards, all aliquots will be stored at -80 °C. If patients give permission to store their biological material, all aliquots of 2 mL and three of the four 500 µl plasma aliquots will be stored at the University Biobank Limburg (UBiLim) for future biomedical research purposes.

If the subjects give consent to take a fourth fasting venous blood sample and store their biological material at UBILim. This blood will then be collected in a K_2EDTA 5 ml tube and centrifuged for 10

minutes at 3400 g. After collection of the plasma, it will be centrifuged again for 10 minutes at 1940 g. Afterwards, three aliquots of 500 μl will be transferred into sterile 1 ml cryovials and stored in the -80 °C freezer at UBILim. These samples will be used to isolate and identify free circulating microRNAs (miRNAs) [16]. The function of certain miRNA's (e.g. miRNA-34) is tumor growth and metastasis [26], [44]. This miRNA can also leak out of the cancer cell and migrate to the bloodstream. The identification of these miRNA in the collected sample will tell if the miRNAs were originated from lung cancer cells [26], [44], [45].

5.4 Plasma sample preparation and measurement

5.4.1 Sample preparation and measurements for nuclear magnetic resonance

The plasma sample preparation and measuring conditions with a Jeol 600 Megahertz (MHz) ^{1}H-NMR spectrometer needs to be optimized. The optimization of the ^{1}H-NMR protocol covers a large part of this thesis and will be extensively explained in the results section.

5.4.2 Sample preparation and measurements for ctDNA measurements

Liquid biopsy assays will be used to identify hotspot genetic mutations in the ctDNA of the blood plasma which was collected in K_2EDTA tubes. These assays typically use 1-2 mL of plasma and generate results in a completely automatic manner within 120 min. Since the study includes stage I-IIIA tumors, ctDNA levels are expected to be lower as compared to stage IV samples.

Liquid biopsies protocols are currently in need for plasma sample preparation that would allow a higher sensitivity which would be more suitable for early stage lung cancers. Since the current sensitivities are already based on the detection of one to a few mutant copies per polymerase chain reaction (PCR), sensitivity gains can only be achieved by departing from a higher sample volume. Therefore, it is expected that tests using up to 10 ml of plasma input will be needed for the analysis of this cohort. At present, sensitive liquid biopsy tests for KRAS and EGFR, optimized for stage IV tumors, are available [46]. Since this concerns a prospective collection over the next two years, it is proposed to use the most informative and sensitive test available at that time.

6 Thesis objectives

In this section the objectives of this thesis will be explained.

6.1 Optimization and evaluation ^{1}H-NMR metabolomics protocol

During the optimization period, the effect of several parameters (e.g. buffer components, buffer capacity, plasma concentration, total sample volume, measuring temperature, power of water suppression and CPMG pulse sequence) on the spectra will be determined. All peaks from the spectrum should be as sharp and separated as possible. This means a good resolution and signal to noise ratio are required. These tests will be performed in order to obtain a robust protocol. If possible, the measurement time should be relatively short, but this shortening may not affect the robustness of the protocol.

6.2 Statistical analysis of a former metabolomics dataset

In this thesis there will be multivariate statistical analyses of data acquired from a previous study. These data contain the metabolic profile of 233 patients with lung cancer in the training cohort and 97 in the validation cohort. It also contains the data from 238 controls in the training dataset and 84 in the validation dataset [8]. From each participant the smoking status, the number of pack years and other medical data have also been documented. The statistical analyses will be performed on the control group which contains a total of 322 subjects. These analyses investigates if the metabolic profile of an individual can be linked with a specific smoking status or possible other confounders.

7 Materials and methods

7.1 Optimization and evaluation ^{1}H-NMR metabolomics protocol

The optimization of the ^{1}H-NMR protocol was performed on a JEOL Resonance 400 MHz (JEOL Ltd, Tokyo, Japan) with a CPMG pulse sequence [9]. This NMR machine with autosampler attached is shown in Figure 25.

Figure 25: JEOL Resonance 400 MHz NMR spectrometer with autosampler [47]

The methods that are shown in following sections are not completed one after another, the research was rather performed in a loop. The parameters that need to be investigated were the buffer components, buffer capacity, plasma concentration, total sample volume, measuring temperature, power of water suppression and CPMG pulse sequence. The measuring temperature, power of water suppression and CPMG pulse sequence were regulated with the JEOL software interface. Each time an optimal value was reached, the next parameter was determined with the previous optimal values. Additionally, if the measurement time decreased without affecting the robustness of the spectrum, the value of the investigated parameter was maintained

The optimization of the protocol was investigated and performed on a large pool of blood plasma samples from breast cancer patients [9]. From this large pool, samples of each 500 µl (Eppendorf, 1.5 ml microtubes) were made and stored in a freezer at -80 °C until further analysis. The usage of this large pool allowed to measure with the exact same sample composition each time thus changes in the spectrum were not due to the sample composition.

7.1.1 Sample preparation

When a sample was taken out of the freezer, it was constantly stored on ice until the actual measurement. This action was performed to prevent any metabolic activities inside the sample. Then the sample was centrifuged (fixed roto Eppendorf centrifuge 5415 R, Hamburg, Germany) for 4 minutes at 13000 rpm at 4 °C. Figure 26 shows the workflow after the centrifugation of the sample.
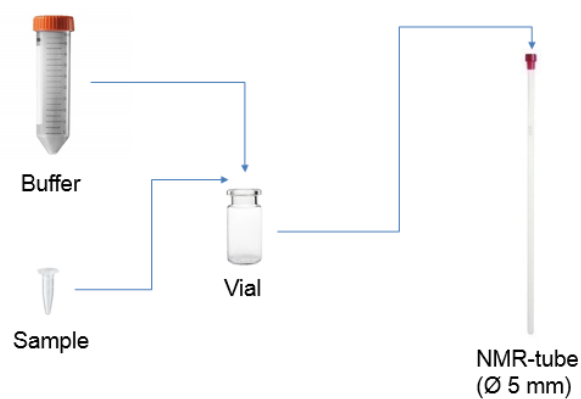
Figure 26: Workflow sample preparation

A certain amount of sample was then transferred to a vial and mixed with buffer (Sarstedt 50 ml tubes). This buffer contained sodium-3-trimethylsilyl propionate (TSP) as an internal standard reference. This mixture was then transferred into a 5 mm NMR-tube.

Now the sample was ready to be introduced into the NMR spectrometer. The NMR-tube was placed in the autosampler of the JEOL 400 MHz spectrometer. The sampler automatically inserted the NMR-tube in the right place inside the spectrometer. After this, the system setup settings needed to be adjusted.

7.1.2 System setup

After the NMR-tube was placed correctly inside the NMR spectrometer, four additional steps needed to be taken before the measurements could start. These four steps included the selection of a CPMG pulse sequence, suppression of the large water peak, spinning of the sample and checking of the shims. To better visualize and compare the spectra, the TSP peak was placed on 0 ppm and served as an internal chemical shift reference.

Firstly, the CPMG pulse sequence was selected. This pulse sequence was required to filter out the signal of macromolecules. Figure 27 shows the difference between spectra when the CPMG pulse was activated (B) or switched off (A).
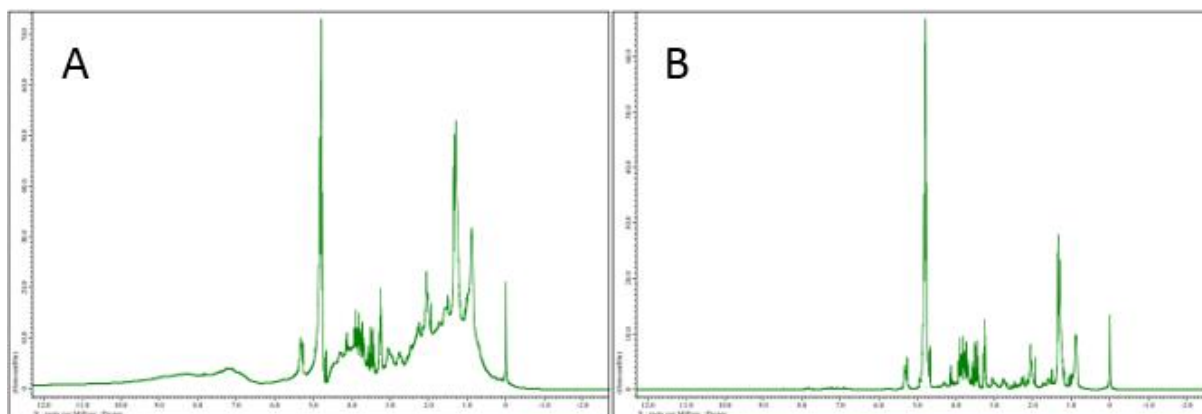
Figure 27: ^{1}H-NMR spectrum of blood plasma without CPMG (A) and with CPMG (B)

This CPMG pulse ensured that a lot of the underlying area was faded out. This means that the CPMG pulse sequence was necessary for the visualization and correct interpretation of metabolite signals from blood plasma.

Secondly, suppression of the water peak was necessary because blood plasma consists mainly of water, this is called the water suppression. If the water signal was not suppressed it would dominate the spectrum as shown in Figure 28.
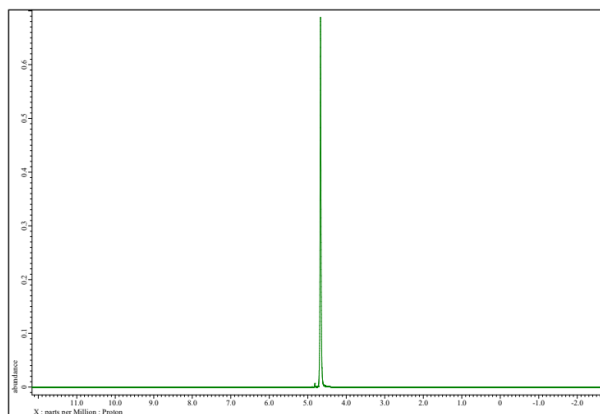
Figure 28: ^{1}H-NMR spectrum of blood plasma without water suppression showing mainly the water peak

When the water peak was partly suppressed it can show up differently in the spectra. Figure 29 shows the two main possibilities.
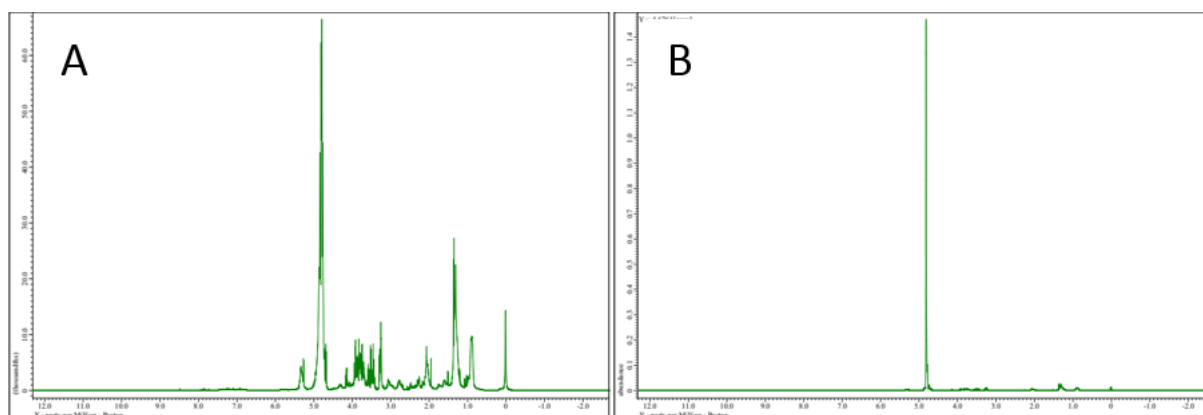
Figure 29: ^{1}H-NMR spectrum of blood plasma with a good water suppression (A) and with a failed water suppression (B)

When the peak was well suppressed, it resembled spectrum A. When the water peak was poorly suppressed, the spectrum could look like the example of spectrum B.

Thirdly, the sample needed to start spinning at 15 Hz. This could be achieved via the software as shown in Figure 30.

Figure 30: JEOL-software setup interface

Finally, the peaks around 1.5 ppm were analyzed to ensure if the shims were correctly performed. These shims needed to be as correct as possible because they indicate if the creation of a homogeneous B_0 field was succeeded. A homogeneous B_0 field was necessary, otherwise, each molecule in the sample would experience a different magnetic field. The peak width at half height of these specific peaks may not exceed the threshold value of 1.4 Hz as shown in Figure 31. If this was the case, the shims created an acceptable homogeneous B_0 field to perform further measurements. Also, the line shape of the peaks were analyzed to be symmetrical (Lorentzian line shape).
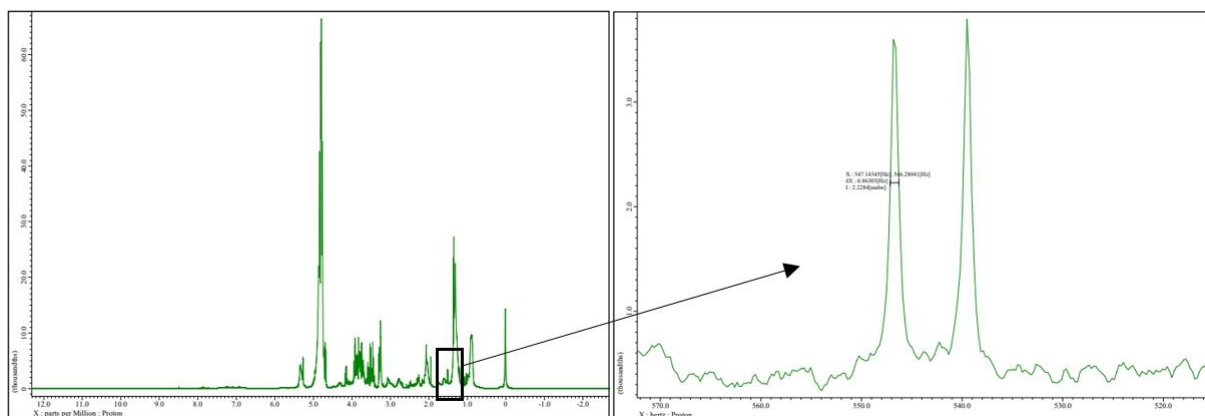
Figure 31: Evaluation of shims on metabolite peaks at a specific region in the ^{1}H-NMR spectrum

At last, the number of scans was set at 96 for the final measurements instead of 4 during the system setup. Figure 32 shows the difference in spectra obtained with a different number of scans.
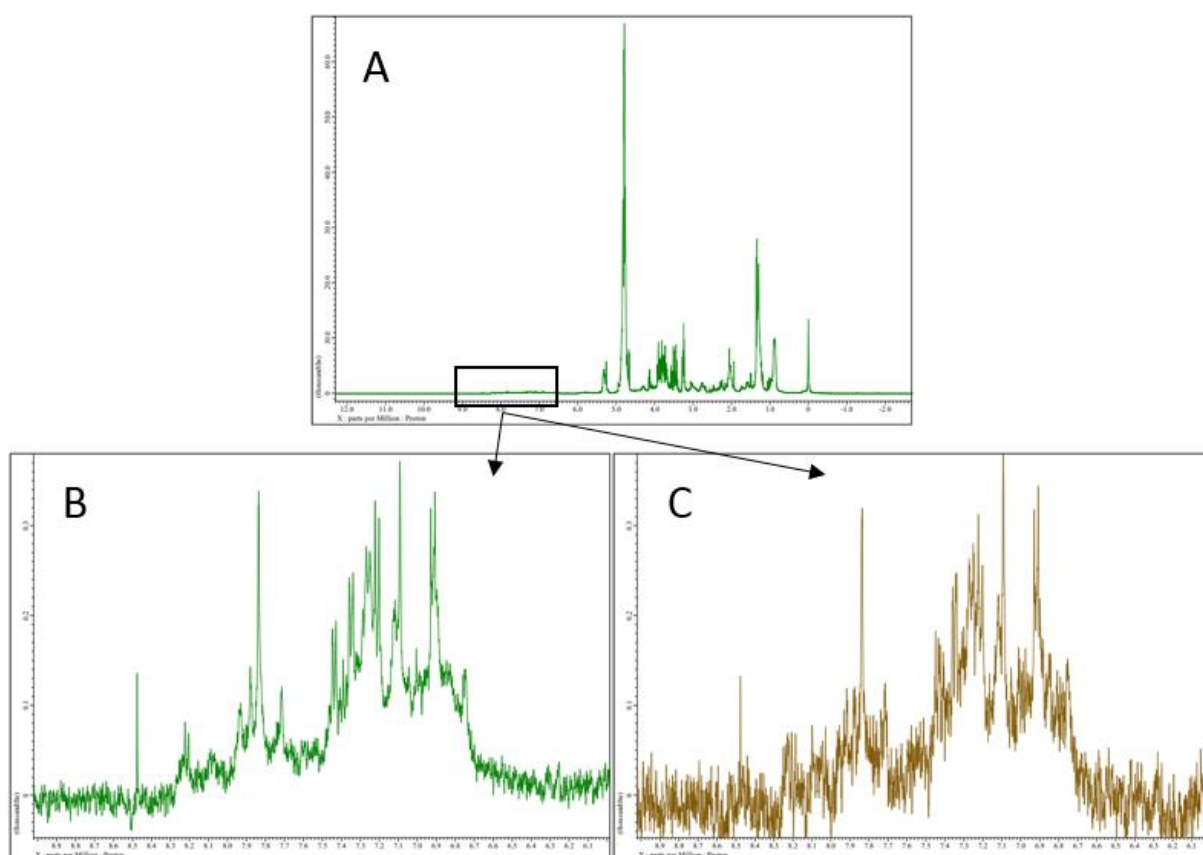
Figure 32: Different number of scans A: full spectrum B: zoom with 96 scans C: zoom with 4 scans

The spectrum obtained with only 4 scans showed a lower signal-to-noise ratio compared to the spectra obtained with 96 scans.

7.1.3 Buffer components and buffer capacity

The first parameter that was tested is the type of buffer used in sample preparation. In a previous study, a solution of pure deuterated oxide (D_2O, 99.9 %, Cambridge Isotope Laboratories Inc, Andover, USA) and TSP (98 %, Cambridge Isotope Laboratories Inc., Andover, USA) was used [9]. Like early stated a lot of water was present in the plasma sample, therefore D_2O was used to minimalize the use of water. The following selected buffers were compared to the spectrum obtained with a D_2O solution. The first measurement used a 1:3 sample/buffer ratio, total volume of 800 µl and room temperature, based on a previous study of our research group [9].

Based on recent literature the two buffers which were tested are 0.15 M K_2HPO_4 and 0.14 M Na_2HPO_4 [48], [49]. The Na_2HPO_4 (Acros Organics, 99 %, Belgium) was made by following the protocol provided by the research group [49]. When preparing the 0.15 M K_2HPO_4 (VWR BDH chemicals, 99.5 %, Germany) buffer, a change to the protocol was made by using more TSP (0.3 g/l) [48]. The chemicals were weighed on an analytical balance (Sartorius La instruments GmbH, Goettingen, Germany). The two buffers were adjusted to pH 7.4 with an 0.1 M HCl solution. The D_2O solution, 0.15 M K_2HPO_4

buffer and 0.14 M Na_2HPO_4 buffer were compared by using a sample to buffer ratio of 1:1 and a total volume of 800 µl.

Another buffer that was tested is an 0.15 M K_2HPO_4/KH_2PO_4 buffer. This buffer was tested because by weighing both salts in a specific ratio, the pH was exactly 7.4 and no additional acidification was required. Therefore, no extra water was added to the buffer solution. The calculation of the necessary quantities of the two salts K_2HPO_4 and KH_2PO_4 (Sigma-Aldrich Chemie GmbH, 99.5%, Germany), could not be made with the classical formulas) and (3), since the solvent that was used is not H_2O but D_2O.

$$pH = pKa + \log\left(\frac{[Base]}{[Acid]}\right) \tag{2}$$

$$Concentration\ of\ buffer = [Base] + [Acid] \tag{3}$$

The pH of D_2O differs from the pH of H_2O. Therefore, these formulas did not calculate the pH correctly [50]. The actual calculation followed a different approach by using volumes of stock solutions of 1 M K_2HPO_4 and 1 M KH_2PO_4. The research group of DeAngelis K. M. measured the volume of each buffer needed to achieve a final pH value of 7.4 [51]. For this buffer, the two stock solutions were made in D_2O. From these stock solutions, a 'trial and error' was performed to eventually become the exact volumes of each stock solution needed to obtain a final pH of 7.4. The spectrum of the 0.15 M K_2HPO_4/KH_2PO_4 buffer was compared to the 0.15 M K_2HPO_4 buffer spectrum. This was analyzed with a different setup than the previous buffers, here a sample to buffer ratio of 1:1 and a total volume of 700 µl were used.

The pH of the mixture buffer and sample may not greatly differ from the pH of the plain buffer. Therefore, the buffer capacity of each buffer was confirmed by comparing the pH's of plain buffer and the mixture of buffer with sample (Consort C860 Multi-parameter analyzer, Belgium). The pK_a of the phosphate buffer was 7.21. All mixtures were measured in an equal ratio sample:buffer of 1:1. If the buffer concentration with the ratio sample:buffer of 1:1 provided an acceptable buffer capacity, higher buffer concentrations would also provide an acceptable buffer capacity.

7.1.4 Plasma concentration and total sample volume

The second parameter which was changed is the sample to buffer ratio. For these measurements, a total volume of 800 µl and the K_2HPO_4 buffer was used at a temperature of 25 °C. In a previous study, 200 µl sample and 600 µl buffer was used [9]. In other literature, sample to buffer ratios of 200/600 (1:3), 265/535 (± 1:2) and 400/400 (1:1) were often used [48], [49], [52].

The third parameter was the total sample volume in the NMR-tube. Hereby, the volumes of 600 µl, 700 µl and 800 µl were tested with a sample to buffer ratio of 1:1 using the K_2HPO_4 buffer. In a previous study, a total volume of 800 µl was used [9].

7.1.5 Temperature

The fourth parameter was the temperature, ranging from 22.5 °C (room temperature) to 37.5 °C (body temperature) [8], [31], [49], [52]. Firstly, the temperature was changed in steps of 2.5 °C. After this, a fine-tuning experiment was performed with a closer range from 22.5 °C to 27.5 °C in smaller steps of 0.5 °C. In a previous study, a temperature of 21.6 °C (room temperature) was used [9]. This parameter was measured using the K_2HPO_4 buffer, 1:1 sample/buffer ratio and 700 μl total volume.

7.1.6 Power of water suppression and CPMG pulse sequence

Power of water suppression:

The fifth parameter to change was the power of water suppression, as shown in the diagram of Figure 33, which influences the signal of the water peak. The range during measurements covered 63 dB to 69 dB. This range was chosen because the default setting on the JEOL software for ^{1}H-NMR measurements was 67 dB. The power value was machine dependent. Therefore, it could not be compared to other powers from different machines found in the literature. The parameter was measured using the 0.15 M K_2HPO_4 buffer, 1:1 sample to buffer ratio, 600 μl total volume and a temperature of 22.5 °C.

CPMG pulse sequence:

The last parameter was the CPMG pulse sequence. This pulse consisted of three parts: the relaxation delay (RD), spin echo delay (delay) and acquisition (ACQ) as shown in Figure 33.
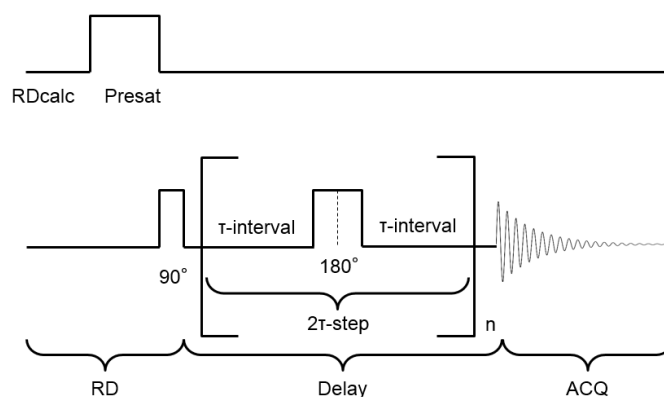
Figure 33: CPMG pulse sequence with water suppression

The total relaxation time for the magnetic moments, shown in formula (4), consisted of the RD time and the ACQ time. At the ACQ time, the measurement takes place and also provided relaxation of the magnetic moments.

$$Total\ relaxation\ time\ of\ magnetic\ moments = \ RD + ACQ \tag{4}$$

The spectra of each parameter were compared to the spectrum obtained with the settings which Louis *et al.* used [9].

The first part of the CPMG pulse sequence contained four parameters: the RD, the water suppression or presaturation (presat), the RD calculated (RDcalc) and the 90° pulse. Firstly during the water suppression or presaturation (presat), a certain amount of power was given at a specific location, which suppressed the signal of the hydrogen atoms from H_2O. Because of the long saturation, the irradiation was frequency selective. Secondly, the 90° pulse occurred right after the presat with a duration of 6.6 µs. The RDcalc was then calculated with formula (5).

$$RDcalc = RD - presat - 90° \, pulse \tag{5}$$

The RD and presat were adjusted because those two parameters were interdependent. Louis *et al.* used an RD of 3 s and a presat of 3, whereas the RDcalc was 0 s [9]. In other literature the RD ranged from 2 – 4.5 s [7], [52], [53]. Since no literature mentioned presat values, a range from 1.5 – 4 s was chosen. Table 2 shows all combinations that were tested.

Table 2: Experimental runs for the relaxation delay and presaturation time

Test run	1	2	3	4	5	6	7	8	9
Relaxation delay (s)	2	2.5	3	3.5	3.5	3.5	4	4.5	4.5
Presaturation (s)	1.5	2.5	3	2	3	3.5	3	3	4

During the second part of the CPMG pulse sequence, the delay was adjusted to suppress the signal of the macromolecules. The loop number was held constant on 80 or 160. Formulas (6) and (7) are used to calculate the remaining times.

$$\tau\text{-}step = 2 * \left(\tau\text{-}interval + \left(\frac{180°}{2}\right)\right) \tag{6}$$

$$Delay = Spin\ echo\ delay = \tau\text{-}step * n \tag{7}$$

The duration of the 180° pulse is a constant time of 13.2 µs. Louis *et al.* used a delay of 32 ms, 2τ-step of 0.2 ms and a loop number of 160 [9]. In other literature other times were found and some needed to be calculated with formulas above to know the exact times [15], [49], [52], [54]. Table 3 gives the combinations that were tested.

Table 3: Experimental runs for the spin-echo delay

Test run	1	2	3	4	5	6	7	8
Spin-echo delay (ms)	10	32	32	64	64	100	128	210
2τ-step (ms)	0.0625	0.4	0.2	0.8	0.4	0.625	0.8	1.31
Loop number	160	80	160	80	160	160	160	160

In the last part of the CPMG pulse sequence, the ACQ was examined for the collection of the result. Louis *et al.* used an ACQ of 1.36 s [9]. This parameter was more difficult to adjust on the JEOL software because it could not be changed directly. To adjust the ACQ, the X-points and X-sweep needed to be changed. The X-sweep was the range in ppm were the spectrum was obtained and the X-points was

the number of data points in this range. Because of these two parameters, an ACQ of 1.36 s was not easily reproduced. Table 4 shows all combinations that were tested. With these results, more information was gained by comparing the FIDs rather than the spectra. Thus, these FIDs were compared to an FID obtained by using standard values of an X-points of 16384 and an X-sweep of 15 ppm. When the X-points and X-sweep were adjusted, it resulted in a calculated ACQ on the JEOL software, the ACQ was slightly lower during experiments.

Table 4: Experimental runs for the acquisition time

Test run	1	2	3	4	5
X-point	16384	16384	16384	16384	32768
x-sweep (ppm)	14	15	17,5	20	20
ACQ calculated (s)	2.9	2.7	2.34	2.05	4.09
ACQ reality (s)	2.33	2.18	1.87	1.64	3.27

The total measurement time needed for analyzing one blood plasma sample is being calculated with formula (8). Hereby, the number of scans was constant set on 96 scans and pre-scans on 16.

$$Total\ measurement\ time = (RD + Delay + ACQ) * (scans + pre\text{-}scans) \tag{8}$$

7.1.7 Robustness tests

At the end of the optimization, a robustness test was performed which tested the robustness of the protocol using all the optimal values. This was performed on five plasma samples of the breast cancer pool and two plasma samples of a healthy subject, each of these samples was measured five times.

7.2 Statistical analysis of a former metabolomics dataset

Former data collected during the study NCT02024113 was assembled in Microsoft Excel. This consisted of metabolic and clinical data of 233 patients with lung cancer in the training cohort and 97 in the validation cohort. It also contained the data from 238 controls in the training dataset and 84 in the validation dataset [8]. Only the control group was used for the statistical analyses in this thesis. The raw data from each subject consisted of 110 variables collected from ^{1}H-NMR measurements and clinical data (age, sex, length, weight, body mass index (BMI), smoking status, packyears, chronic obstructive pulmonary disease (COPD), diabetes and medications intake). The plasma ^{1}H-NMR spectra of each subject were divided into several integral regions. Each of these integral regions represented one of the 110 variables from the data set.

The general data overview was created with Microsoft Excel and SPSS to determine, if possible, the number of subjects (n), mean, standard error (St. error), maximum (Max) and minimum (Min) values for each clinical parameter.

Multivariate statistical analyses were performed with the software program SIMCA (SIMCA-P+ version 15, Umetrics, Umea, Sweden). In this program, the data was analyzed with various statistical formulas to search for a possible link between the metabolite profile of an individual and their smoking status. The smoking status was divided into three groups: the active smokers, the ex-smokers and the non-

smokers. The subjects were classified as ex-smokers when they stopped smoking for at least six months. The smoking status groups were not evenly distributed, the group with active smokers was the smallest only consisting of 63 subjects. After exclusion of outliers in the group of active smokers 60 subjects remained. Therefore, the groups were equalized randomly for a better comparison.

Also, several possible confounders were investigated. Other clinical variables that might influence the resulting discriminating model based on one other specific variable, thus complicating the prediction model. The length and weight were not analyzed separately since both parameters were covered by the BMI, which was calculated with equation (9).

$$BMI = \frac{weight\ (kg)}{length\ (m)\ x\ length\ (m)} \tag{9}$$

All possible confounders were grouped using a threshold value based on the median. Another score-plot for the BMI was made where it was categorized into three separate groups since these values were clinically relevant. Those groups ranged from 0-20, 20.01-30 and 30.01-100.

To visualize a multivariate dataset, an orthogonal partial least square discriminating analysis (OPLS-DA) score-plot was trained. OPLS-DA could be used to discriminate between groups and thus was an interesting tool to use for testing biomarkers candidates or comparing/classification studies. Thus, OPLS-DA was a prediction and regression method that finds information in data that was related to known information. When a prediction model was made, the program chooses the right number of axes needed to make the best possible model. These axes consisted of a prediction component (P) and an orthogonal component (O). Those axes were projected in 2D, which formed the OPLS-DA score-plot. When clusters were visible in the score-plot, the data of one group could be discriminated from the data of the other group.

The ^{1}H-NMR spectra consisted of several large and smaller peaks. But the data extracted from those larger peaks should not dominate the data belonging to the smaller ones. To adjust for this, Pareto scaling was used. Formula (10) describes the calculation of Pareto scaling.

$$Par = \frac{x - \bar{x}}{\sqrt{SD}} \tag{10}$$

Par = Pareto scale
x = Observed values
$\bar{x}$ = Mean of the observed values
SD = Standard Deviation

Figure 34 shows an OPLS-DA score-plot that was trained to create a prediction model based on the kind of fruit.
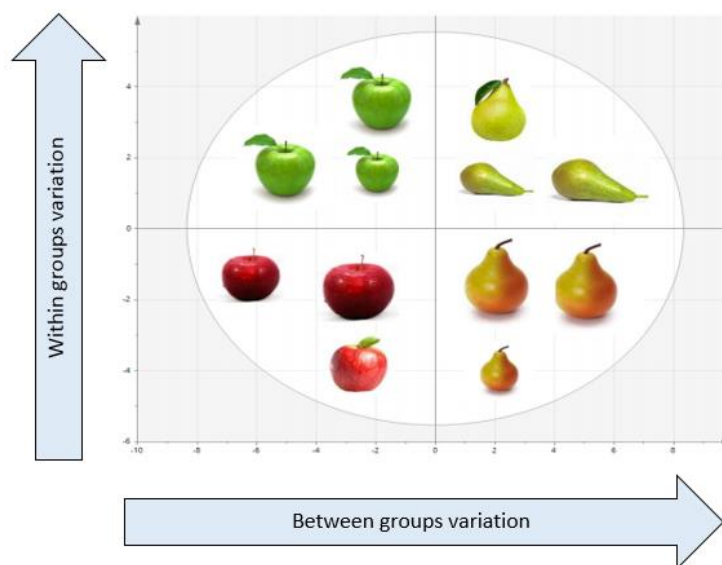
Figure 34: Example of an OPLS-DA score-plot of two different fruits (apples and pears)

In this OPLS-DA score-plot, a discrimination could be made between apples and pears, this was the variation between groups. The discrimination between the green fruit and red fruit is the variation within a group. This could be implemented on our dataset of the metabolite profile of the smoking status, whereas the smoking status represented the kind of fruit. The red and green fruits discrimination represented the confounders.

These OPLS-DA models also gave an R^2X, R^2Y and Q^2 value. R^2X explained the variation inside a group, R^2Y represented the variation between groups and the Q^2 value expressed how good the prediction model was. R^2X was best when close to 0.9. A Q^2 value close to 0.5 meant that the predictability of the model was good. Also, the sensitivity and specificity of the model were determined by analyzing a misclassification table.

8 Results

8.1 Optimization and evaluation ^{1}H-NMR metabolomics protocol

The ^{1}H-NMR spectra, that are obtained by systematically varying each parameter, are evaluated at four regions as shown in Figure 35a and Figure 35b.
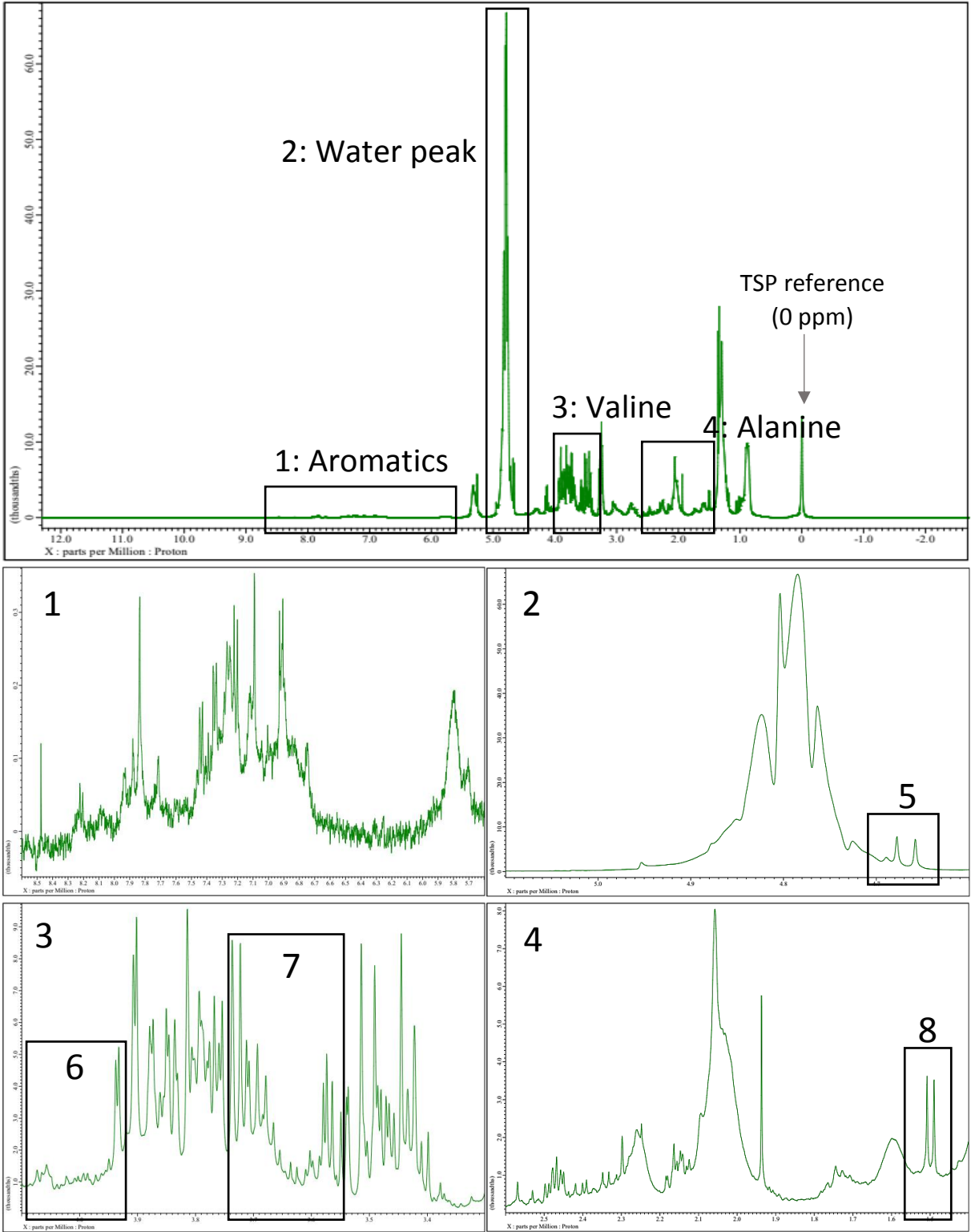
Figure 35a: The four evaluation regions of a ^{1}H spectrum of blood plasma
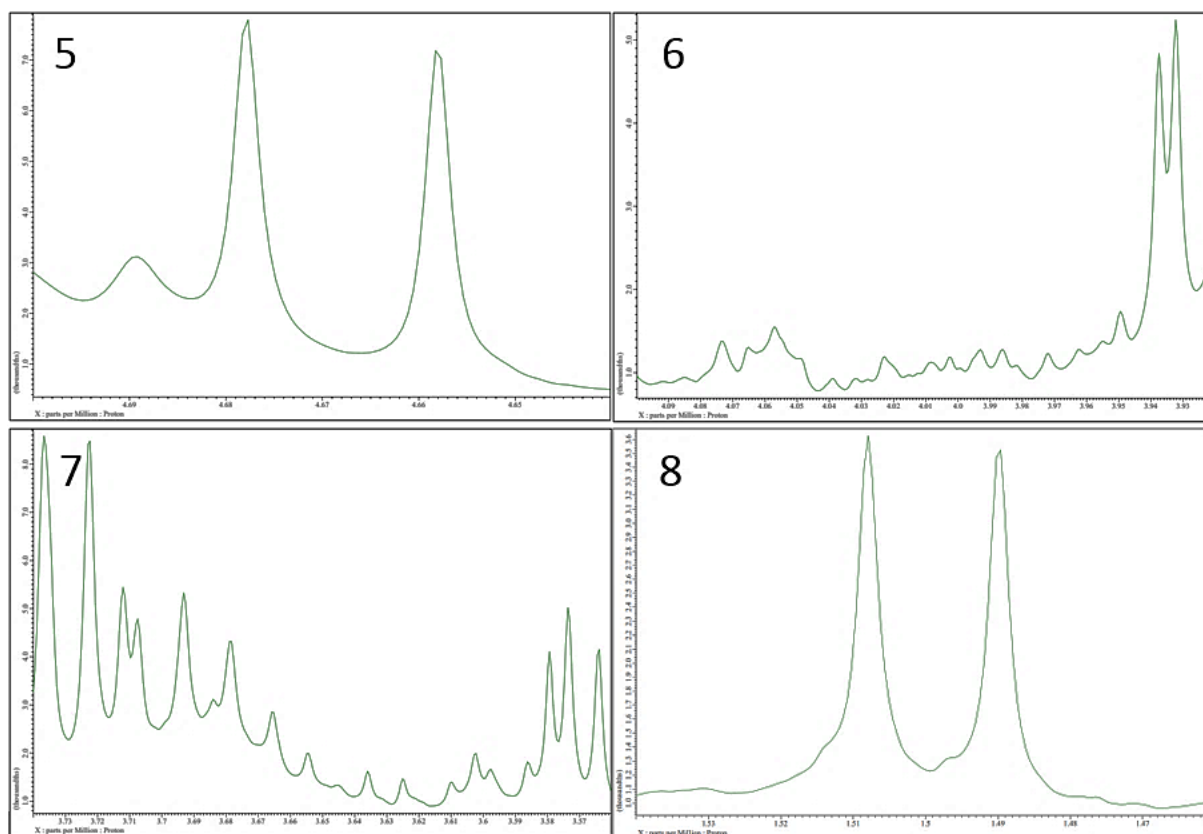
Figure 35b: Zoom in of the four evaluation regions

The measurements are performed on blood plasma samples from the large pool of samples collected from breast cancer patients. All spectra of the following figures are arranged in the same order as Figure 35, unless indicated otherwise. On the spectra, the Y-axis shows the abundance and the X-axis is expressed in ppm. Also the TSP peak is referenced on 0 ppm in each spectrum.

The full spectrum for each parameter is not displayed in this thesis, although the full spectra are closely evaluated during the practical analyses. Therefore, four important different regions are chosen to demonstrate in this thesis, as shown in Figure 35a. Also, several peaks in these regions are zoomed-in to be better visualized, as shown in Figure 35b. The first region is the region of the aromatics (5.6 – 8.6 ppm). The second region (4.6 – 5.1 ppm) is the water peak (4.7 – 4.9 ppm) with two important peaks at the right side (4.65 – 4.69 ppm), region five (4.64 – 4.70 ppm) is a zoom-in of those peaks. The third region (3.3 – 4.0 ppm) is what we call the valine region and the fourth region (1.4 – 2.5 ppm) is the region we call the alanine region, region seven (3.56 – 3.74 ppm) and eight (1.46 – 1.54 ppm) are a zoom-in of those peaks.

Each hydrogen atom has a different chemical environment; thus each atom feels a different magnetic field, which cause them to spin at different frequencies. The proton peaks originating from the alanine (± 1.5 ppm) and valine (± 3.63 ppm) molecule are displayed in Figure 36.
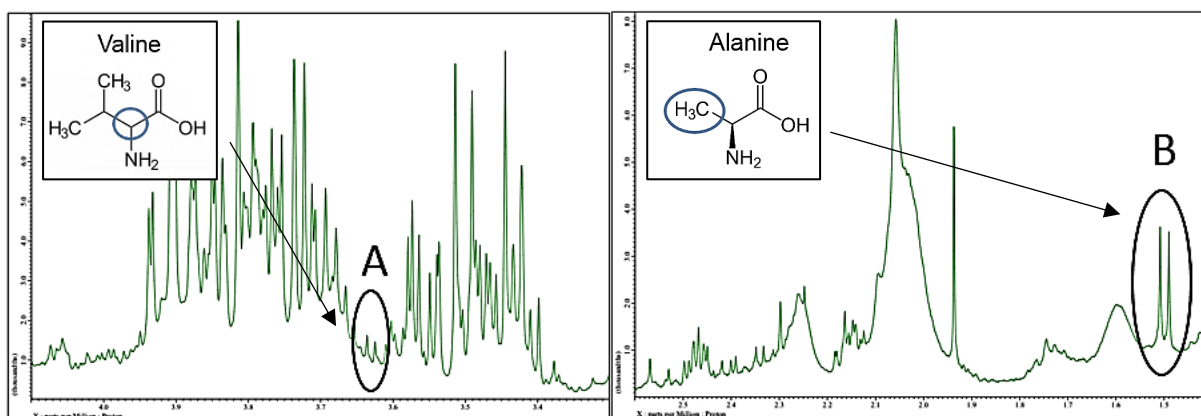
Figure 36: The peaks in the ^1H spectrum originating from specific protons of the valine (A) and alanine (B) molecules

8.1.1 Buffer components and buffer capacity

Buffer capacity:

Table 5 shows the confirmation of the buffer capacity of each tested buffer and sample/buffer mixture. The $H_2PO_4^-/HPO_4^{2-}$ buffer has a pK_a of 7.21.

Table 5: Confirmation of Buffer capacity of the mixture of sample with buffer

pH	D$_2$O solution		0.15 M K$_2$HPO$_4$ (adjusted with HCl)	0.14 M Na$_2$HPO$_4$ (adjusted with HCl)	0.15 M K$_2$HPO$_4$/KH$_2$PO$_4$
Buffer	8.50		7.41	7.43	7.40
Mixture (ratio sample:buffer)	8.80 (1:3)	8.66 (1:1)	7.37 (1:1)	7.44 (1:1)	7.33 (1:1)

These measurements confirms that the pH does not change when plasma is added to the buffer. The plain buffers, like stated before, measure approx. pH 7.4 and the pure D$_2$O solution a pH of 8.5. This pH difference of the D$_2$O solution is because it is not acidified. The difference in pH value of the D$_2$O solution creates a different spectrum then the spectrum with the buffer, because the ^1H-NMR spectrum is pH dependent.

Firstly, the spectra of D₂O solution, 0.15 M K₂HPO₄ and 0.14 M Na₂HPO₄ are compared as shown in Figure 37.
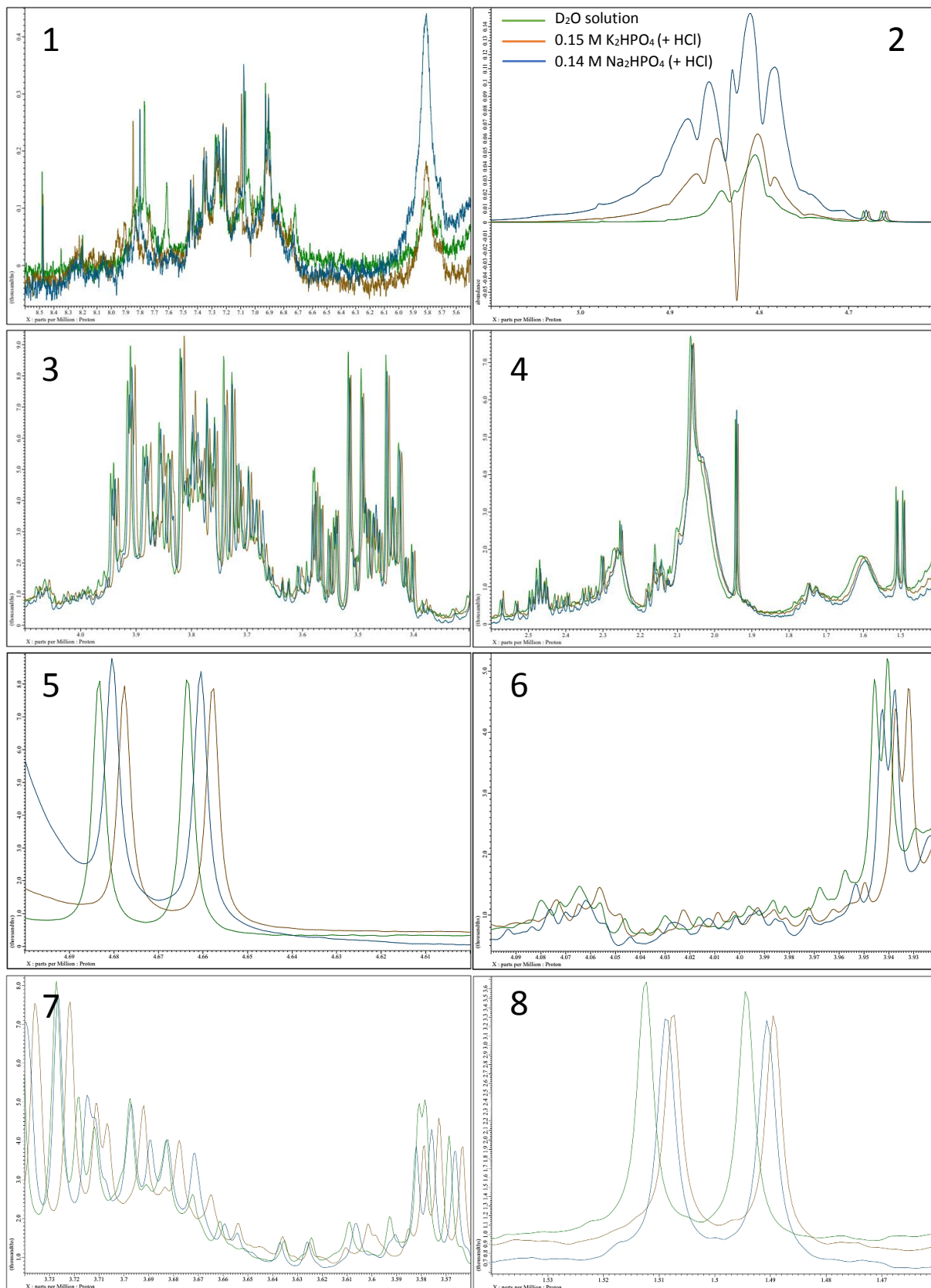
Figure 37: The resulting ¹H-NMR spectrum of D₂O solution, 0.15 M K₂HPO₄ and 0.14 M Na₂HPO₄ buffers

These spectra show that the buffers give a similar spectrum and only using D_2O gives a bigger change in chemical shift, like in Figure 36 spectrum 8 (green line). This shift is due to the pH of D_2O, thus working with only D_2O holds the risk of not always providing the same pH environment.

The 0.15 M K_2HPO_4 buffer and 0.14 M Na_2HPO_4 buffer have the same pH, approx. 7.4, but in these spectra a little chemical shift is observed. This shift is due to the little molar concentration difference.

From the three buffers the 0.15 M K_2HPO_4 buffer (adjusted with HCl) is withheld. The first reason is the uncertainty of the D_2O solution due to the pH difference. The second reason is the preparation of the 0.14 M Na_2HPO_4 buffer, which is solved in water with a little D_2O. Therefore, the 0.15 M K_2HPO_4 buffer is chosen, which is fully solved in D_2O and has a good buffering capacity.

This 0.15 M K_2HPO_4 buffer adjusted with HCl is then compared to the 0.15 M K_2HPO_4/KH_2PO_4 buffer, having the same buffer concentration but prepared by using two salts as shown in Figure 38.
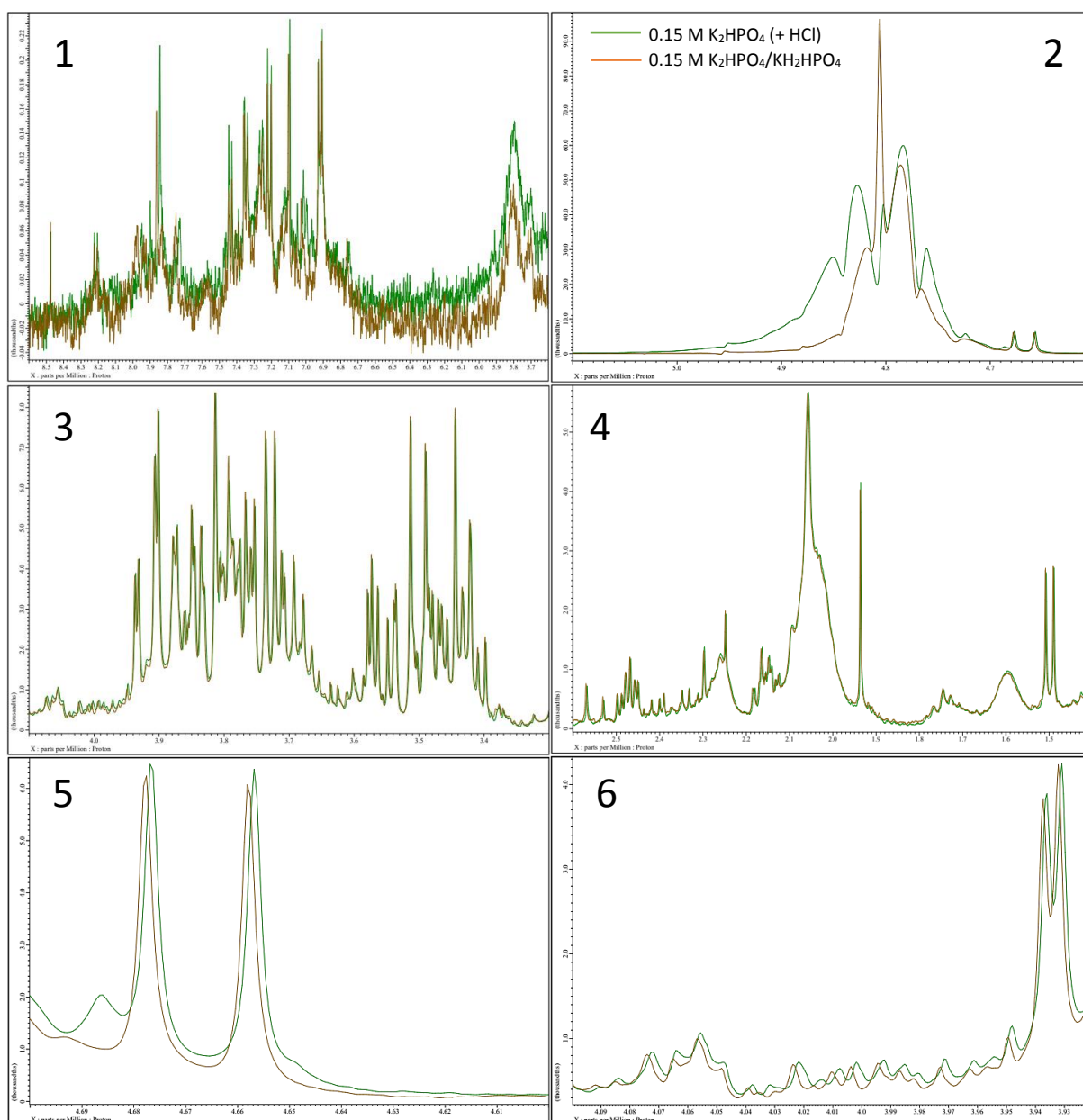
Figure 38: The resulting ^1H-NMR spectrums of 0.15 M K_2HPO_4 and 0.15 M K_2HPO_4/KH_2PO_4 buffers
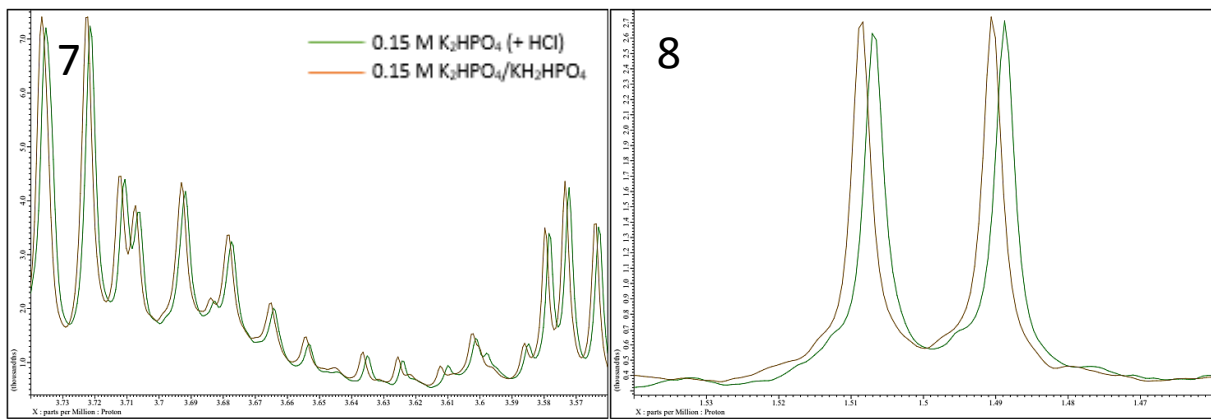
Figure 38: Continued

The 0.15 M K_2HPO_4/KH_2PO_4 buffer shows the same high resolution as the 0.15 M K_2HPO_4 buffer, but for the protocol the 0.15 M K_2HPO_4/KH_2PO_4 buffer is chosen. Because in the preparation of the buffer no extra HCl or NaOH must be added. The 0.15 M K_2HPO_4/KH_2PO_4 buffer is also easier to prepare namely by weighing than the one that needs to be acidified.

8.1.2 Plasma concentration and total sample volume

Plasma concentration:

Figure 39 shows the results of the influence of the ratio sample:buffer variations on the ^{1}H-NMR spectrum, leading to different sample concentrations.
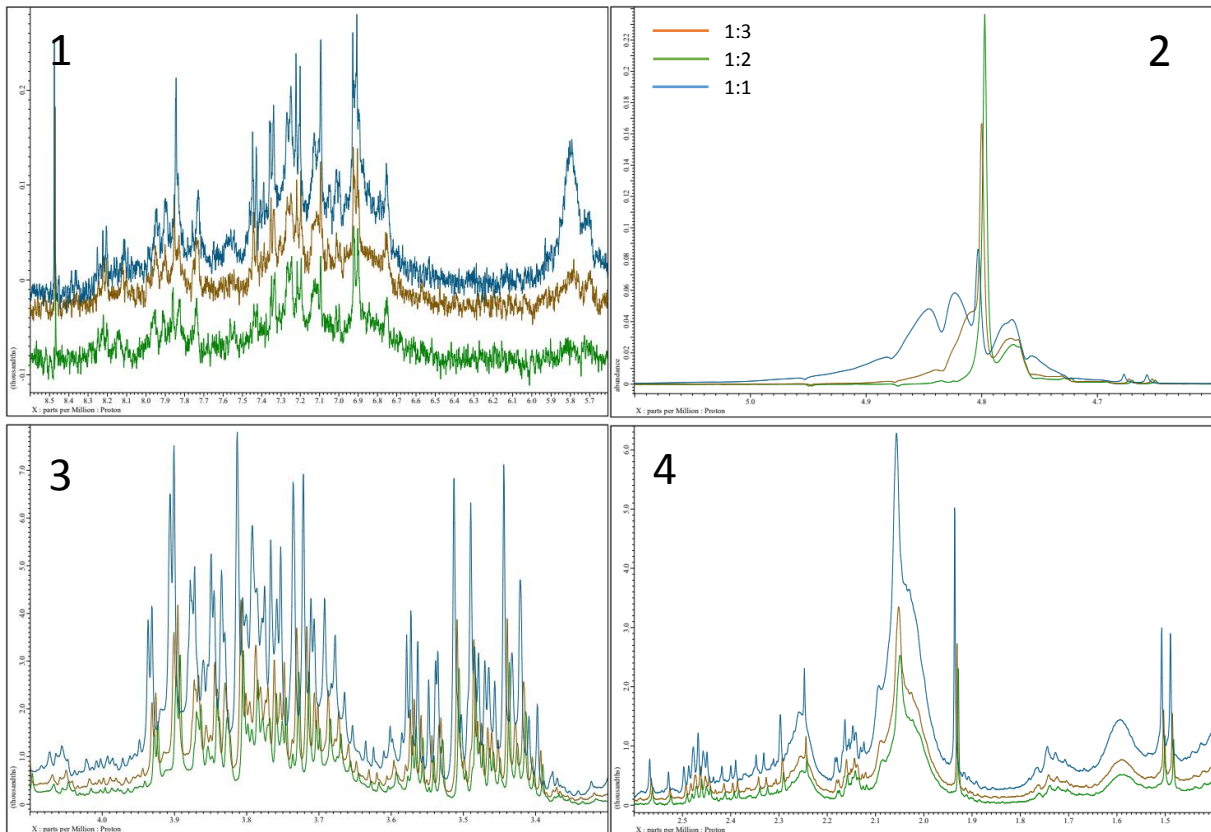
Figure 39: Influence of different concentration ratios sample:buffer on the ^{1}H-NMR spectrum of blood plasma
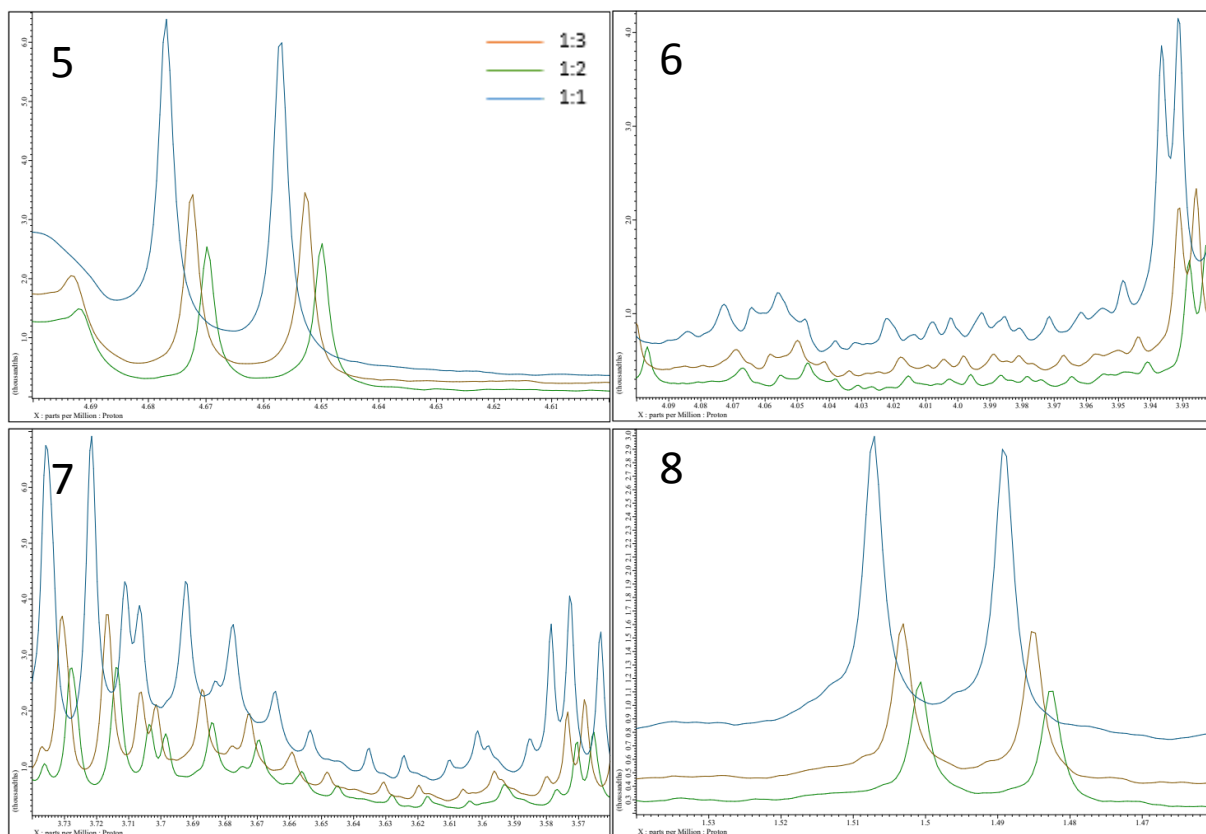
Figure 39: Continued

From these results an optimal sample concentration ratio sample:buffer of 1:1 is used in the protocol. The ratio sample:buffer of 1:1 gives a spectrum in which there is no loss of resolution due to peak overlap and in which the peaks are better visible due to the higher concentration of the sample resulting in a better signal-to-noise ratio. If a low sample:buffer ratio of 1:3 is used, the resolution of some peaks decreases as shown in the spectrums of Figure 39.

Total sample volume:

For the total sample volume, the 600, 700 and 800 µl gave exactly the same results. Thus, the total volume between 600 and 800 µl does not cause the spectrum to change. For the protocol an optimal value of 700 µl is used.

8.1.3 Temperature

For this comparison, not all temperatures are displayed, because otherwise the spectrum is unreadable. Therefore, only the temperatures of 22.5 °C, 25 °C, 30 °C and 37.5 °C are displayed in Figure 40.
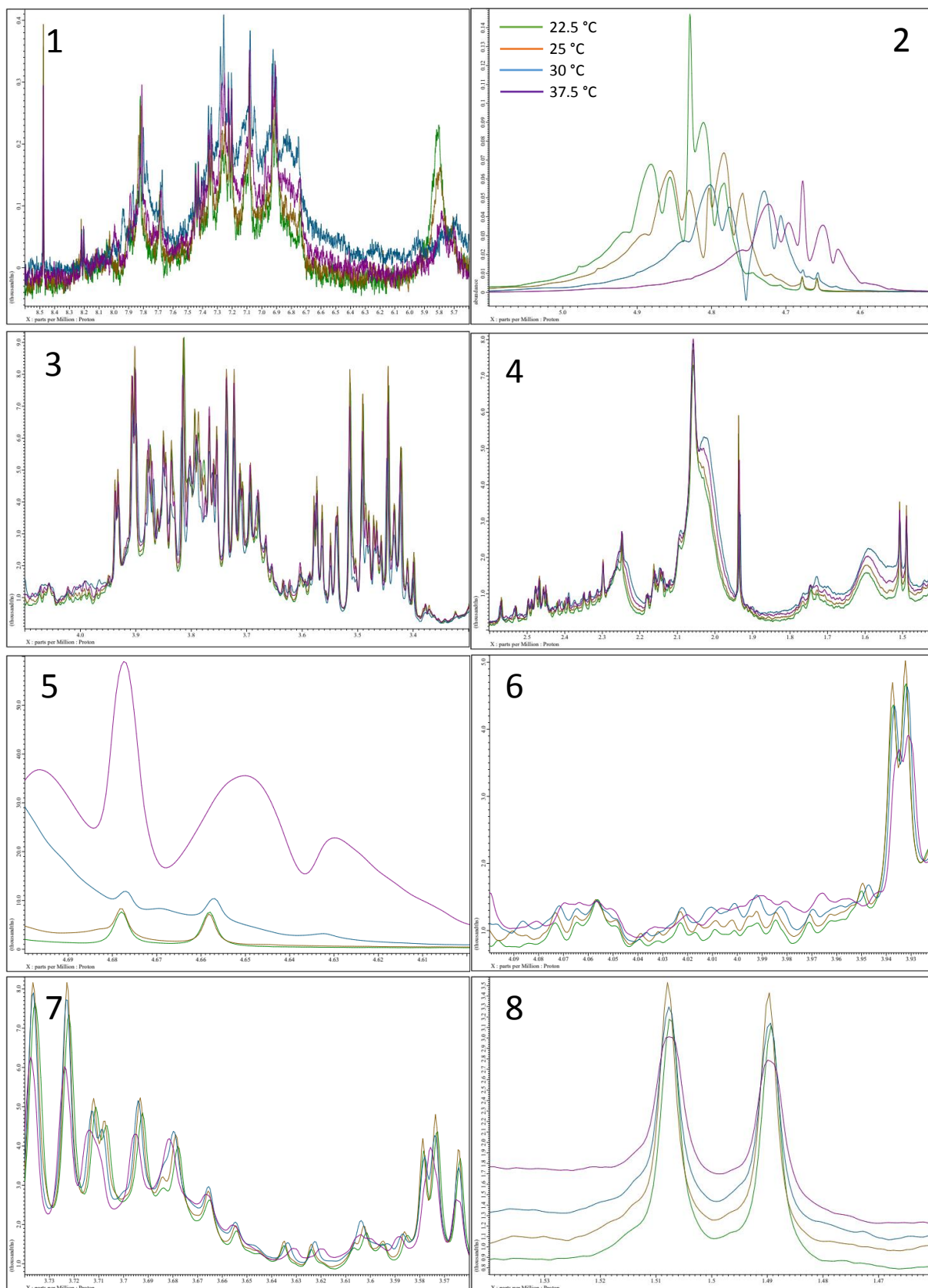
Figure 40: Influence of different temperatures on the ^{1}H-NMR spectrum of blood plasma

These results show that when the temperature during measurements is higher, more loss of resolution appears. This can be best seen in spectra 6 and 7 with a temperature of 37.5 °C (purple line), which does not create separate peaks. The last spectrum of Figure 40 shows that the temperature of 25 °C

(orange line) has the sharpest peaks, which means it has a good resolution and there is an optimum around this temperature. Therefore, the temperature of 25 °C was chosen.

Fine tuning:

To find the optimal measuring temperature, a fine tuning is performed. The results of the fine tuning in the next figure does not include all tested values, because the figure would be unreadable. Therefore, only the spectra of 24 °C, 25°C and 26°C are being displayed in Figure 41.
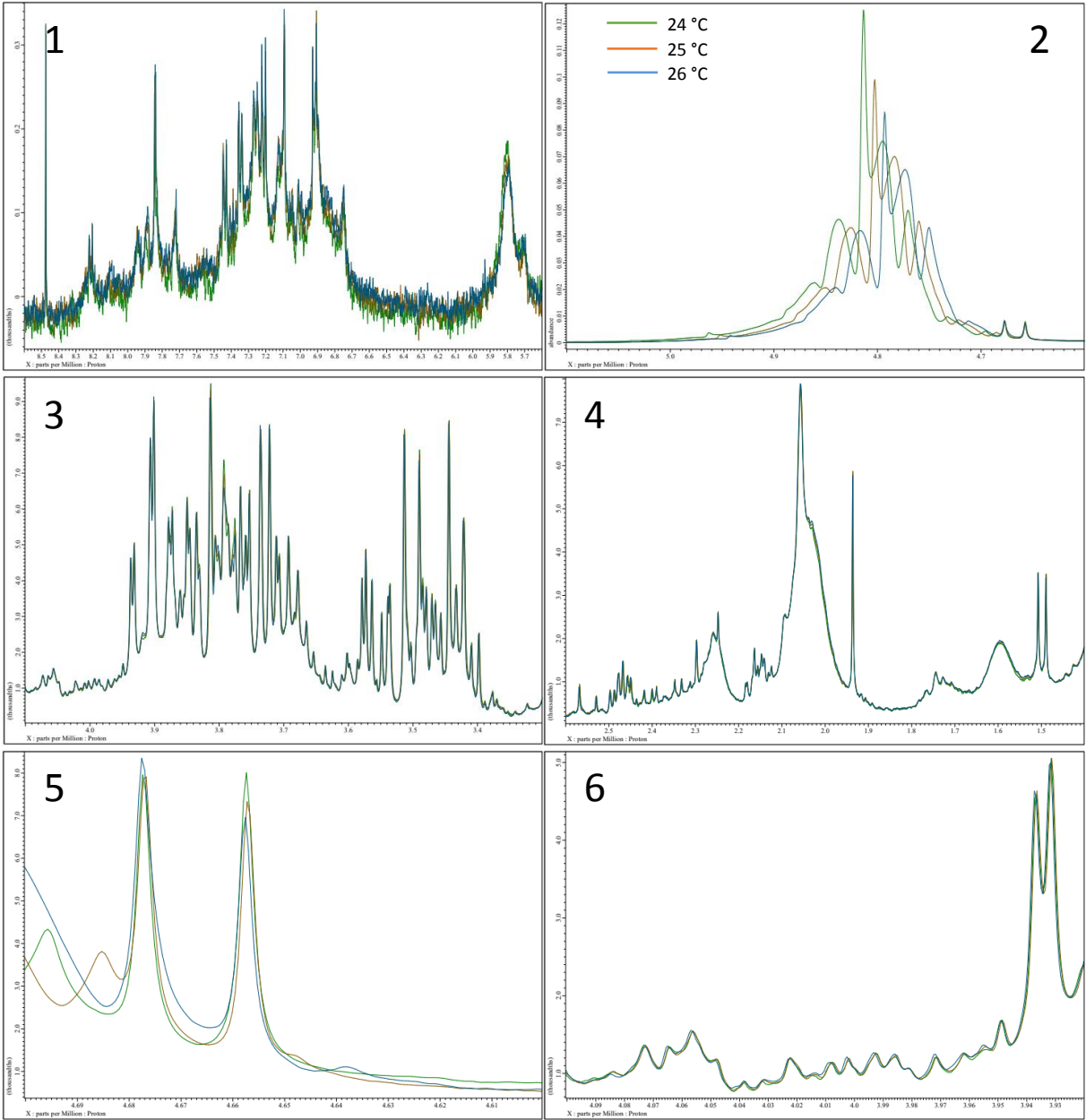
Figure 41: Fine tuning of the measuring temperature
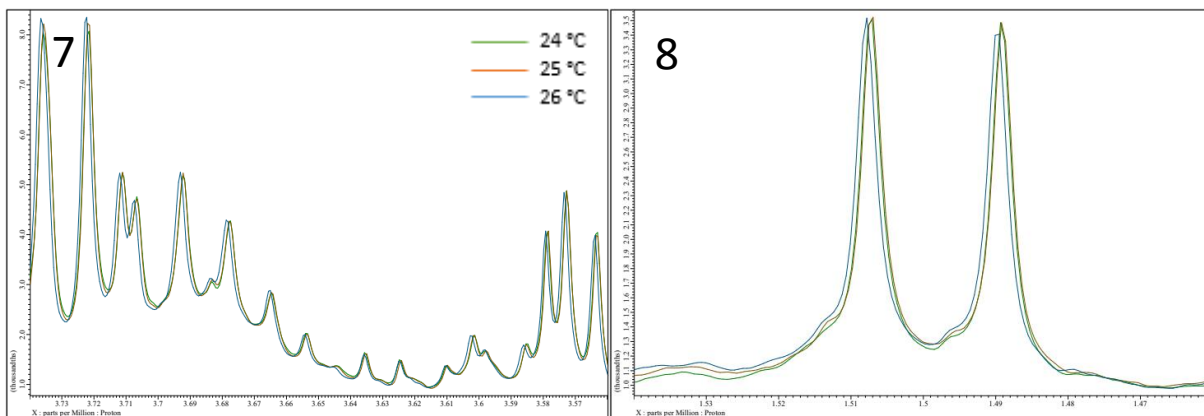
Figure 41: Continued

Between these spectra no difference is visible. This means that small temperature changes do not affect the spectrum, but large changes, like in Figure 40, change the spectrum. Therefore, a measuring temperature of 25 °C is chosen to be used in the protocol.

8.1.4 Power of water suppression and CPMG pulse sequence

Power of water suppression:

Figure 42 shows the results of the power measurements: 63 dB, 65 dB, 67 dB and 69 dB are tested. In this figure only a few regions are displayed, because the power only affected the water peak signal and the signal of two peaks at the right side of the water peak.
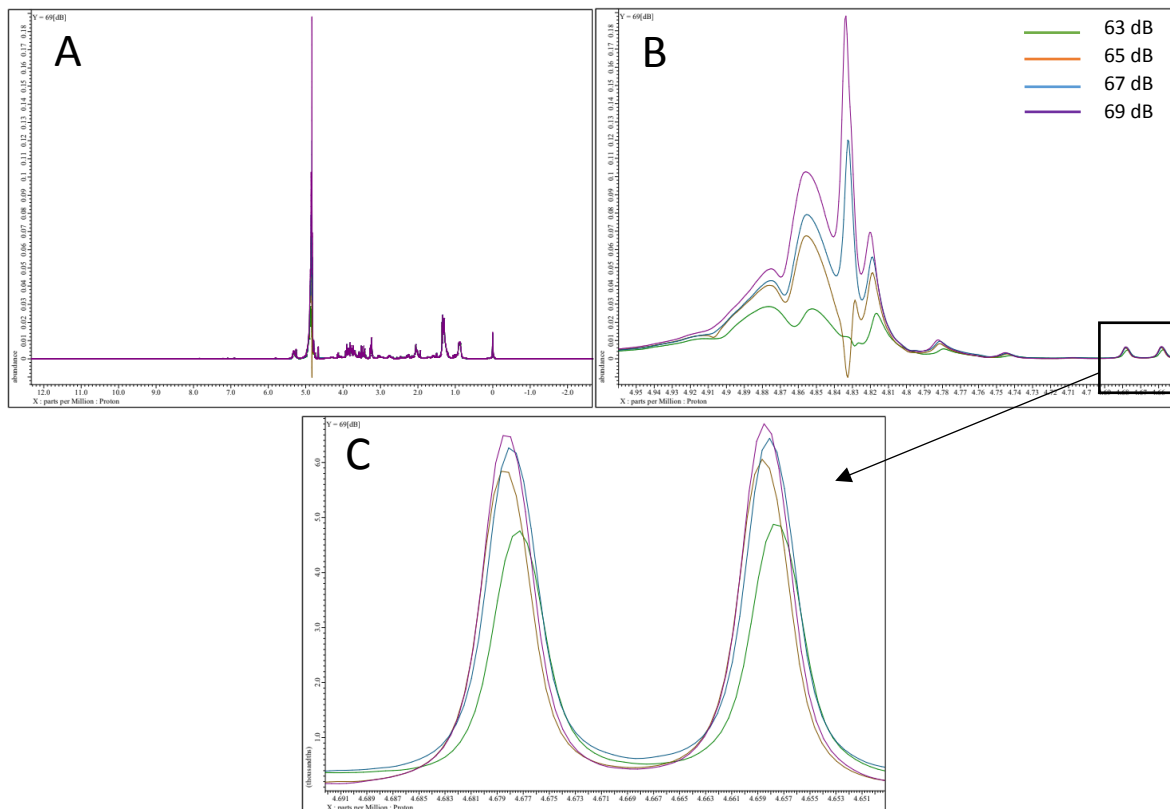
Figure 42: Influences of different powers of water suppression on the ^{1}H-NMR spectrum A: full spectrum, B: water peak

When the power of water suppression is set to 63 dB (green line), much saturation power is given to the water. This suppresses the water peak very strongly, but with the drawback that two peaks at the right side of the water peak (Figure 42 C) are also being suppressed. Therefore, an optimal value of 67 dB is chosen, because it provides a good suppression of the water peak signal and it does not affect the intensity of two peaks at the right side of the water peak.

CPMG pulse sequence:

Relaxation delay and presaturation:

From the tested RD and presat not all test runs are being displayed in Figure 43.
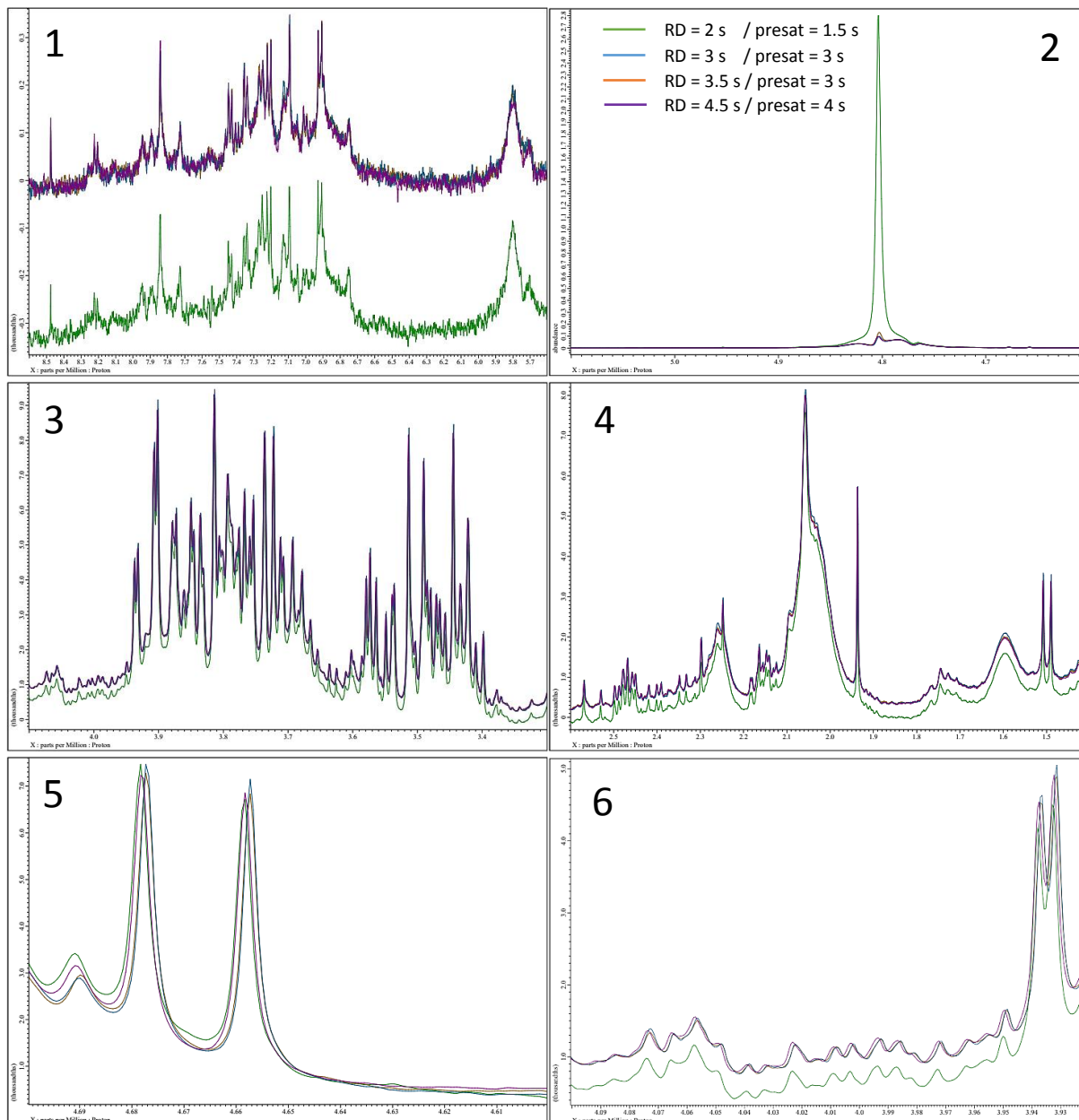
Figure 43: Influence of different relaxation delays and presaturation values on the ^{1}H-NMR spectrum of blood plasma
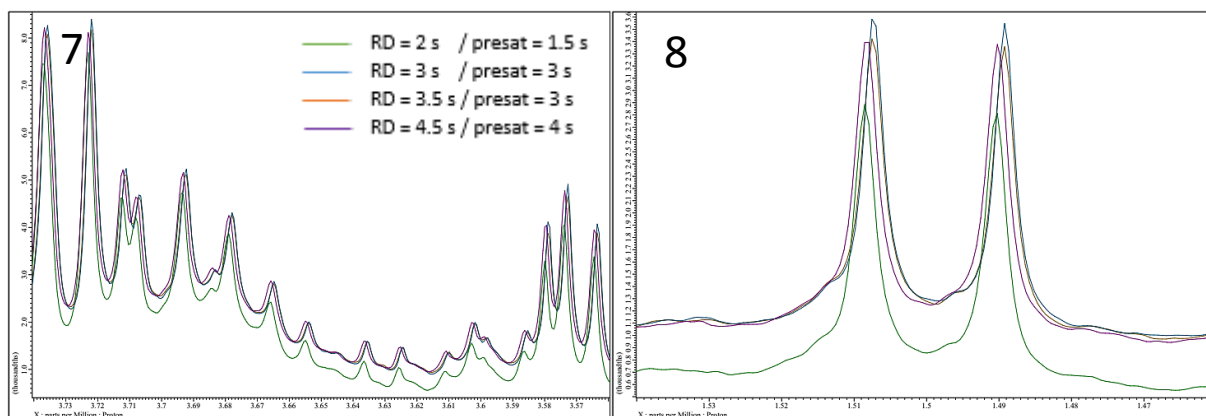
Figure 43: Continued

The tested RD and presat values only influenced the water peak. When these parameters are too low, for example an RD of 2 s and a presat of 1.5 s (green line), the spectrum cannot be processed correctly afterwards. This means that a different integral value under each peak will be found.

Thus, when the RD and presat are too low this phenomenon occurs. Therefore, a proposition for a decrease in measurement time is made. The optimal values chosen for the RD is 3 s and for the presat 3 s, because if the presat is lowered, the water peak would not be suppressed as good, therefore complicating the processing of the spectra afterwards.

Spin echo delay:

The results show that when the loop number is 80 or 160 it did not affect the spectrum. Firstly, the total spin echo delay of 10, 32 and 210 ms with loop number 160 are being displayed in Figure 44.
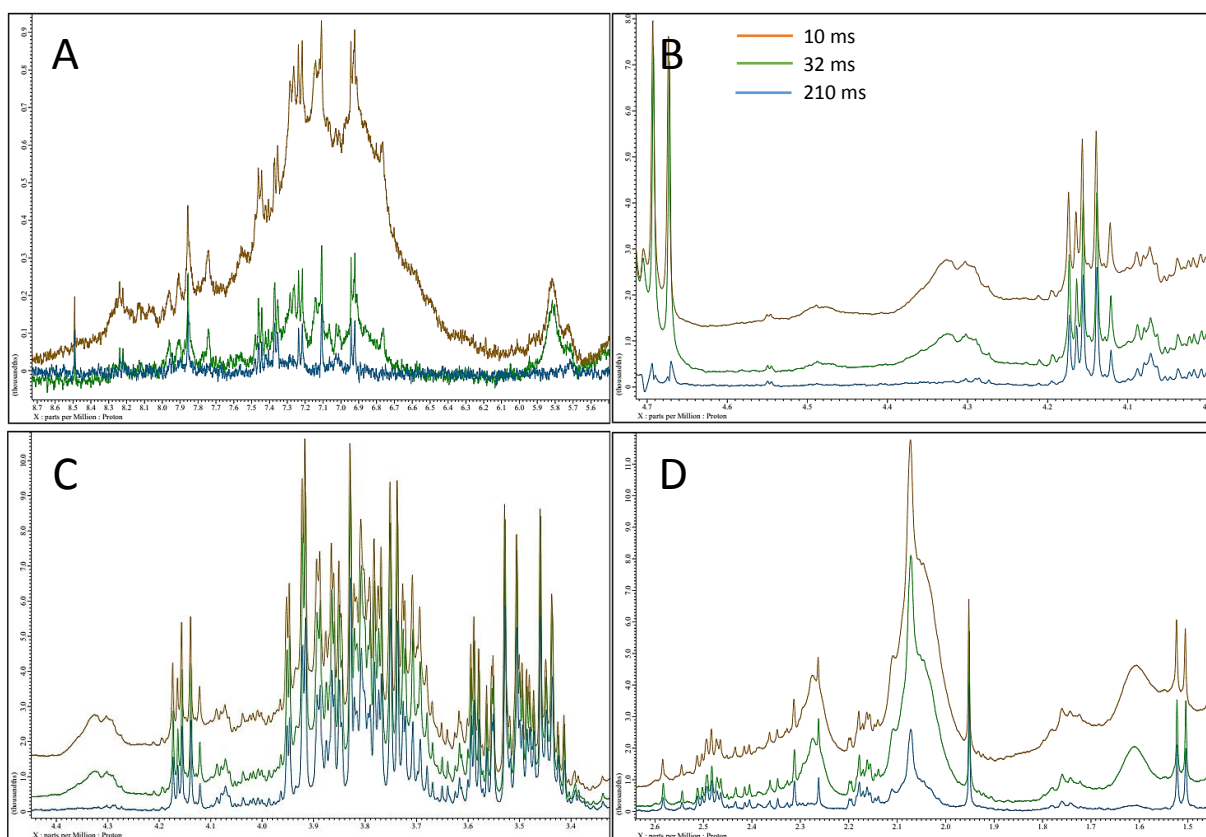
Figure 44: Influence of different spin echo delays on the ^{1}H-NMR spectrum of blood plasma

This figure shows that when the total spin echo delay (n.2.τ) increases, the peaks decrease. Some of those peaks even disappear as seen in the spectra of 210 ms (blue line), this means the sample is being suppressed too long. Therefore, not only the signal of macromolecules are suppressed, but also the signal of some metabolites. When the total spin echo delay is too low, the spectrum becomes larger as seen in the spectrum with a delay of 10 ms (orange line). Thus, the signal of the macromolecules is not suppressed enough.

In the 10 to 210 ms range, a fine tuning is performed. The total delays of 32, 64, 100 and 128 ms with a loop number of 160 are displayed in Figure 45.
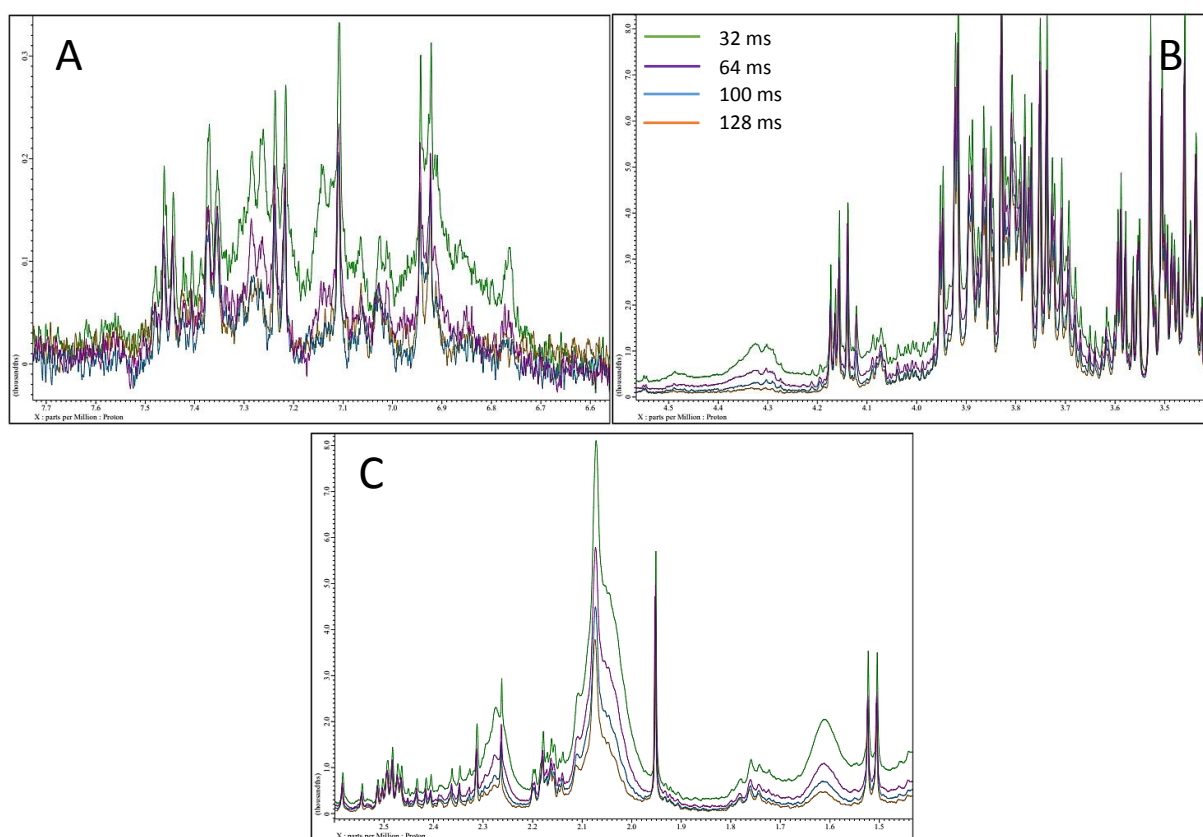
Figure 45: Fine tuning of the spin echo delay

In this fine tuning can be seen that a delay of 32 ms (green line) still has a lot of macromolecule signals and the delay of 100 ms and 128 ms suppressed more of metabolite signals than the delay of 64 ms. Therefore, an optimal value of 64 ms with 160 loops is chosen for the protocol, which provides a good suppression of the signal of macromolecules and did not affect the signal of metabolites too much like the delay of 100 ms, 128 ms and 210 ms.

Acquisition time of FID:

For the comparison of the ACQ, all the obtained FID's from the test runs are overlaid as shown in Figure 46.
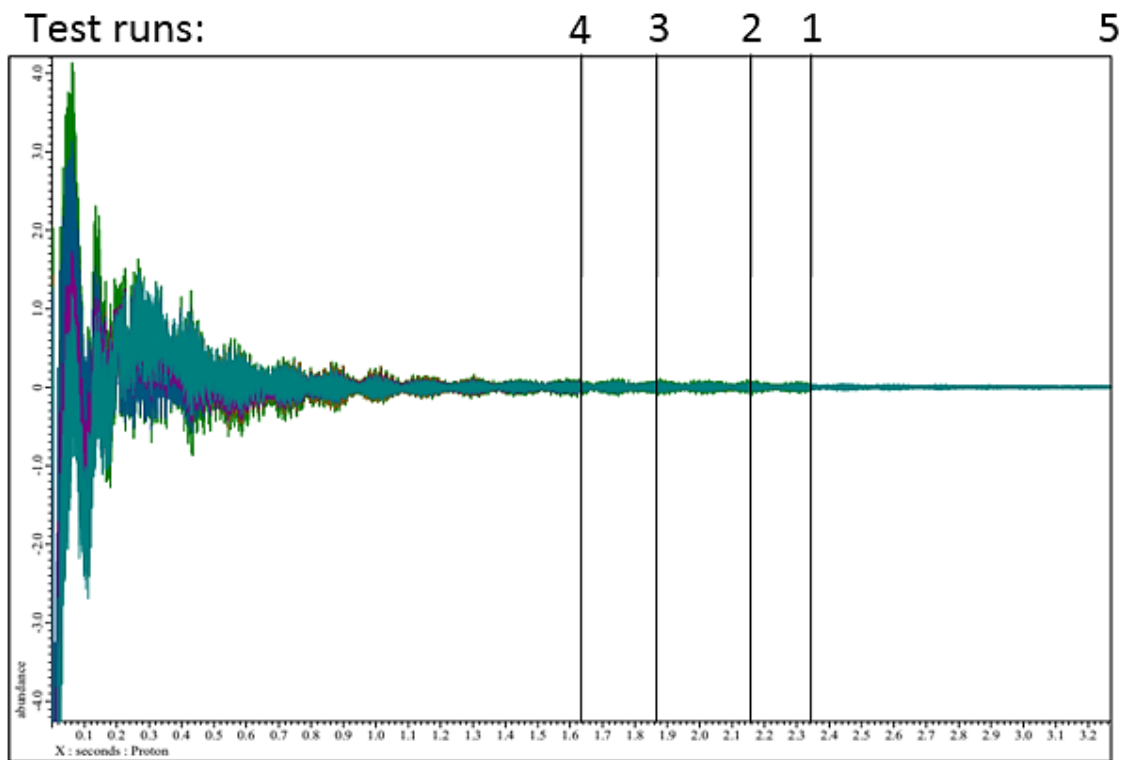
Figure 46: Results of different acquisition times

These results give the same ^{1}H-NMR spectra, since even for test run four. In the ^{1}H-NMR spectra, no peak truncation are observed, meaning that the AQC time is long enough. Moreover, to be on the safe side, an AQC time of 2.18 s is used.

8.1.5 Robustness test

The spectra from the robustness test are identical when compared to each other. Thus, the optimization delivered a robust ^{1}H-NMR protocol for blood plasma.

8.1.6 Summary of parameters

Table 6 gives a summary of all the optimal parameters values for a robust ^{1}H-NMR metabolomics protocol when measuring blood plasma on an JEOL NMR spectrometer.

Table 6: Summary of all optimal parameters for a robust ^{1}H-NMR metabolomics protocol on the JEOL 400 MHz spectrometer

	Optimal values
Buffer (pH)	0.15 M K$_2$HPO$_4$/KH$_2$PO$_4$ buffer (7.4)
Plasma concentration (ratio sample:buffer)	1:1
Total sample volume	700 µl
Measuring temperature	25 °C
Power of water suppression	67.19 dB
Relaxation delay	3 s
Presaturation	3 s
Spin-echo delay	64 ms
2τ-step (=2 x τ)	0.4 ms
Loop number	160
Acquisition (calculated)	2.7
Acquisition (reality)	2.18
Measurement time	9 min 50 sec

Figure 47 displays the final spectrum, like Figure 35, that is obtained with all the optimized parameters.
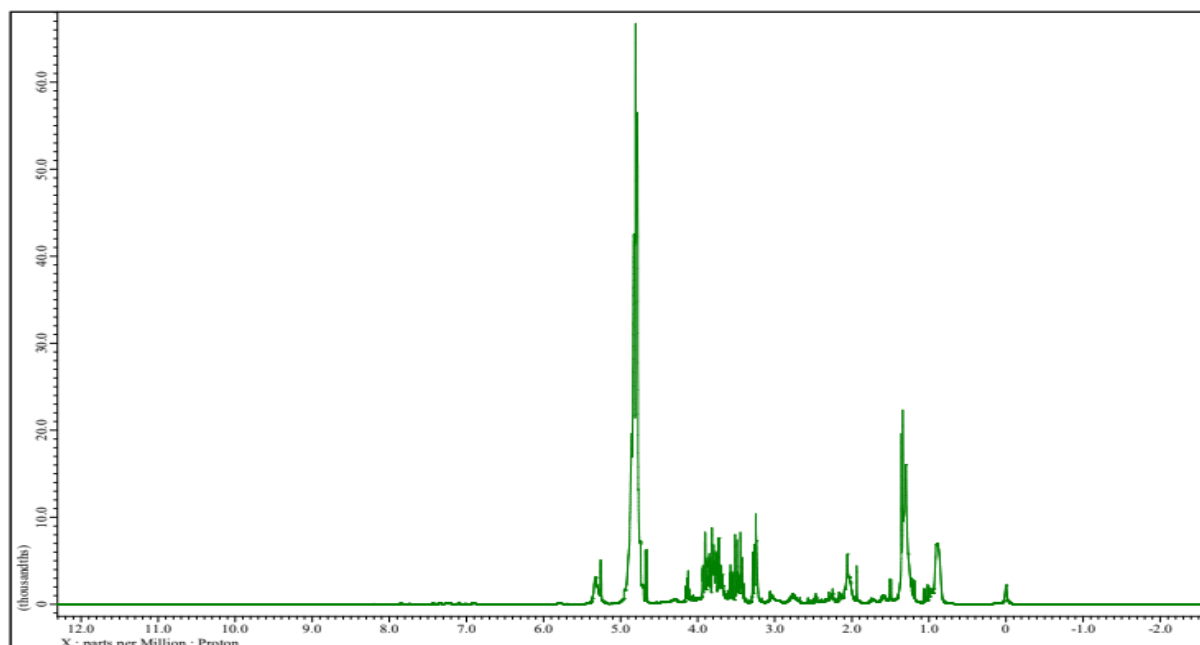
Figure 47: Final spectrum of blood plasma with the optimal values
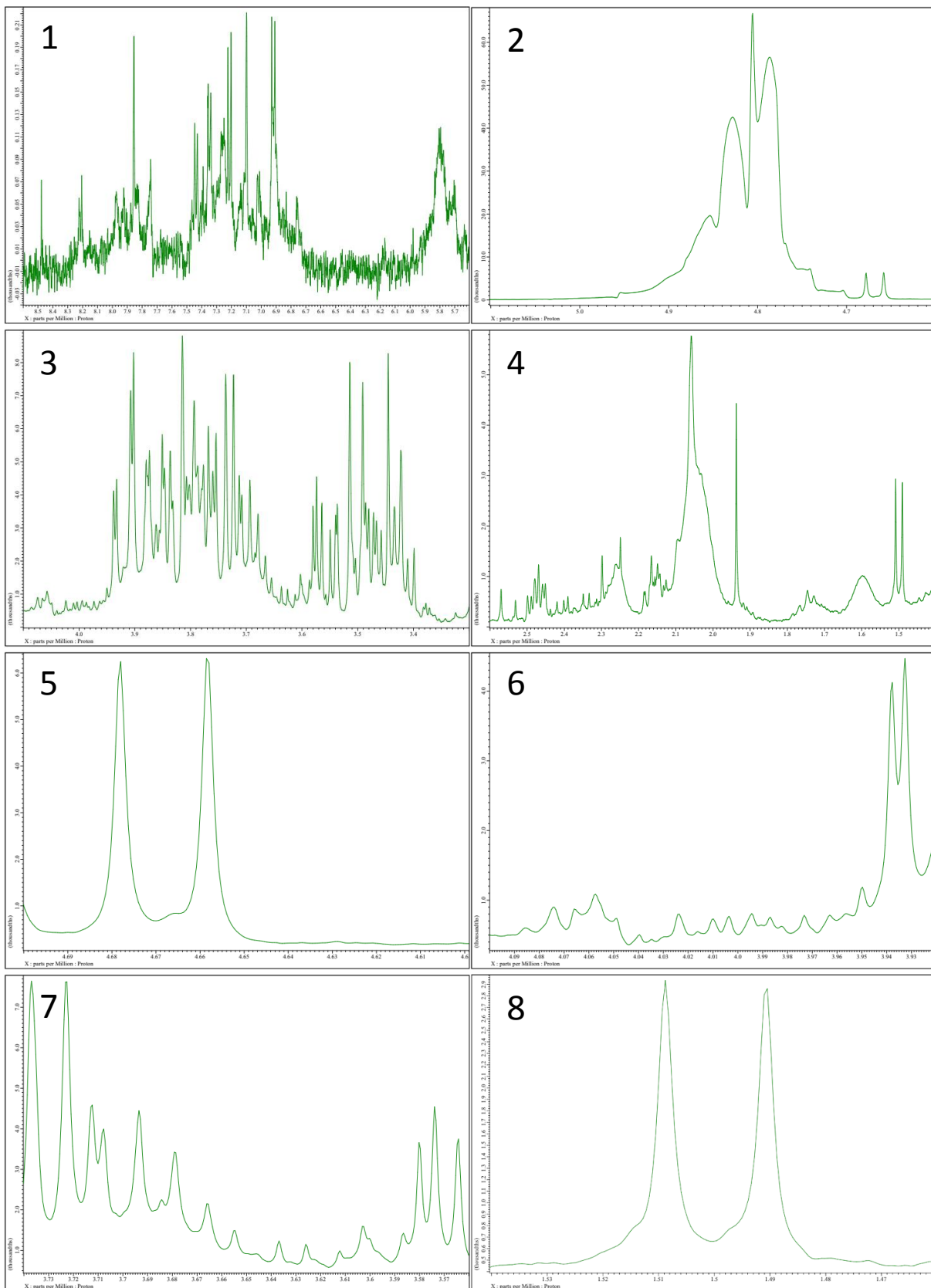
Figure 47: continued

8.2 Statistical analysis of a former metabolomics dataset

8.2.1 Data overview

Table 7 describes the general overview of the dataset of the control group, which is divided into three smoking status groups. These data is used for the multivariate statistical analyses.

Table 7: Data overview of the former dataset based on the smoking status

| | | Control group | | |
		Smokers	Ex-smoker	Non-smoker
Total number subjects		63	140	119
Age (years)	**Mean**	59.76	68.28	70.61
	St. error	1.23	0.82	0.94
	Min	38	39	41
	Max	83	88	89
Sex	**Male, n (%)**	41 (65.08)	94 (67.14)	31 (26.05)
	Female, n (%)	22 (34.92)	46 (32.86)	88 (73.95)
Packyears	**Mean**	32.62	20.21	0
	St. error	3.54	1.76	0
	Min	1	0.5	0
	Max	175	125	0
Weight (kg)	**Mean**	77.65	80.76	75.17
	St. error	2.04	1.44	1.23
	Min	43	48	51
	Max	125	136	117
Length (m)	**Mean**	1.69	1,69	1.62
	St. error	0.01	0.01	0.01
	Min	1.43	1.41	1.44
	Max	1.86	1.95	1.94
BMI	**Mean**	27.26	28.3	28.71
	St. error	0.69	0.42	0.49
	Min	16	19.2	18.7
	Max	52	46.7	46.6
Diabetes	**n (%)**	6 (9.52)	27 (19.29)	26 (21.85)
COPD	**n (%)**	10 (15.87)	18 (12.86)	5 (4.20)
High cholesterol	**n (%)**	31 (49.21)	92 (65.71)	61 (51.26)
Malfunctioning thyroid	**n (%)**	3 (4.76)	8 (5.71)	16 (13.45)
High blood pressure	**n (%)**	43 (68.25)	104 (74.29)	90 (75.63)
Coagulation problems	**n (%)**	36 (57.14)	100 (71.43)	69 (57.98)
Cardiac arrhythmias	**n (%)**	4 (6.35)	9 (6.43)	9 (7.56)

		Control group		
		Smokers	**Ex-smoker**	**Non-smoker**
Medication	**n (%)**	51 (80.95)	126 (90.00)	110 (92.44)
Oral antidiabetics medication	**n (%)**	4 (6.35)	21 (15.00)	21 (17.65)
High cholesterol medication	**n (%)**	31 (49.21)	91 (65.00)	59 (49.58)
Malfunctioning thyroid medication	**n (%)**	3 (4.76)	8 (5.71)	16 (13.45)
Medication high blood pressure	**n (%)**	43 (68.25)	102 (72.86)	89 (74.79)
Anticoagulation medication	**n (%)**	36 (57.14)	100 (71.43)	69 (57.98)
Cardiac arrhythmias medication	**n (%)**	4 (6.35)	9 (6.43)	9 (7.56)

8.2.2 Smoking status

The main goal of the statistical analysis is to investigate if there is a discrimination possible between smokers, ex- and non-smokers based upon the metabolite profile. Figure 48 shows a score-plot where all three groups are compared.
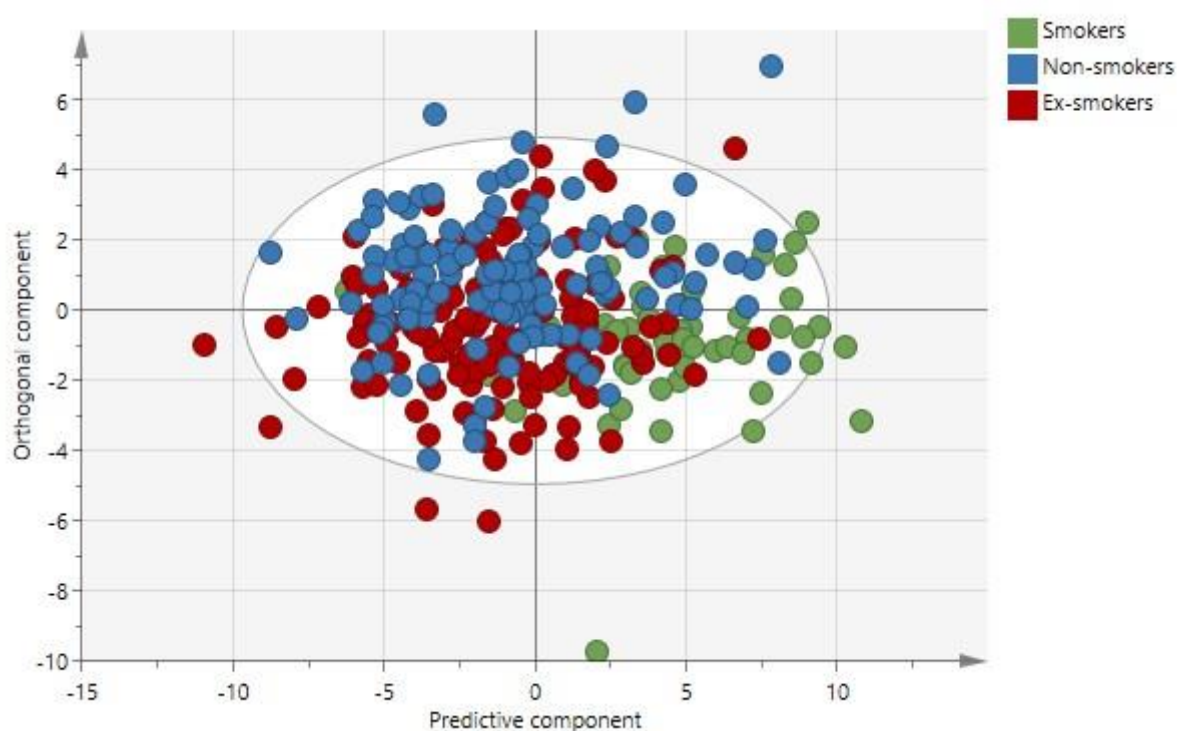
Figure 48: OPLS-DA score-plot of all smokers, all ex-smokers and all non-smokers

In this model a small discrimination from the smokers-group can be observed, thus new models are made in pares to better visualize the discrimination. Figure 49 shows a model when all the smokers are compared to all the non-smokers.
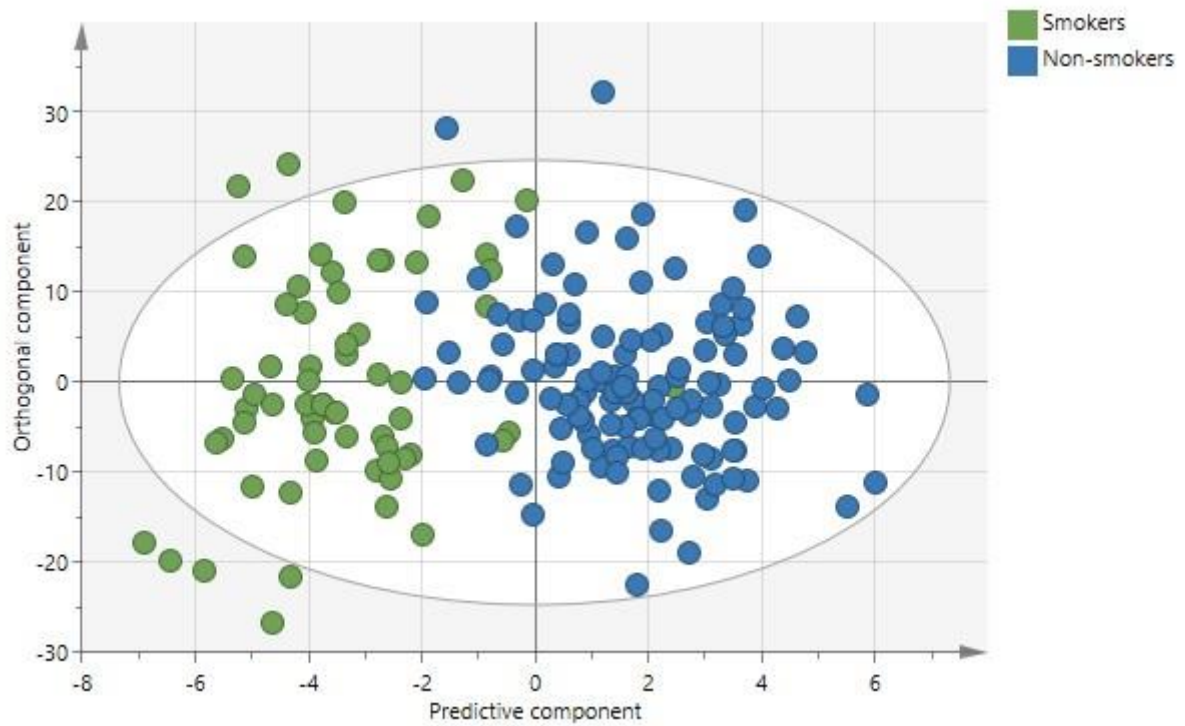
Figure 49: OPLS-DA score-plot of all smokers and all non-smokers

This model shows a clear discrimination between the two groups, but the groups are not equal in numbers. The distribution is 63 smokers and 119 non-smokers. When outliers of the smokers are removed, 60 smokers remain. Therefore, a new model is made with equal groups, thus 60 non-smokers are chosen randomly as shown in Figure 50.
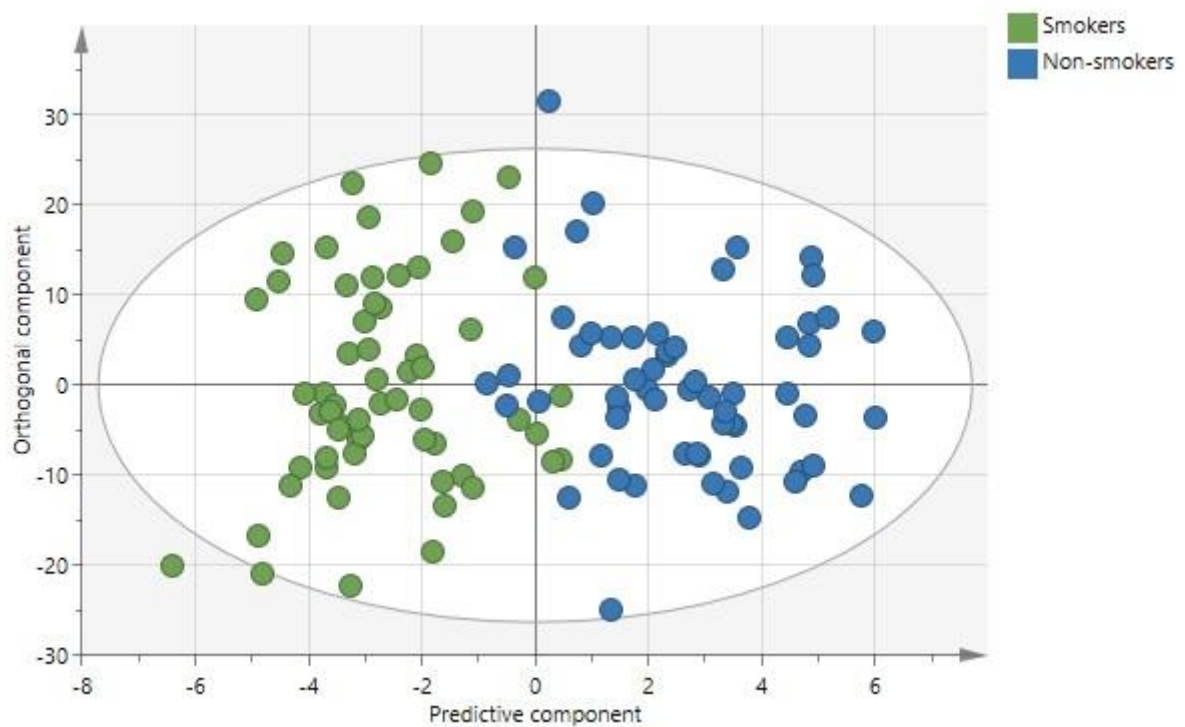
Figure 50: OPLS-DA score-plot of equal groups of smokers and non-smokers

Also, this OPLS-DA diagram model shows a clear discrimination of the metabolite profile between the two groups. The statistical analyses in Table 8 show that the model is a good prediction based on the Q^2 because this is close to 0.5.

Table 8: Statistical analyses for equal groups of smokers and non-smokers

Number of axis (P+O)	1+6
R^2X	0.907
R^2Y	0.733
Q^2	0.468
Misclassification:	
Smokers	55 of 60 are correctly classified (91.67 %)
Non-smokers	56 of 60 are correctly classified (93.33 %)

Figure 51 shows the resulting score-plot in which all smokers are compared to all ex-smokers.
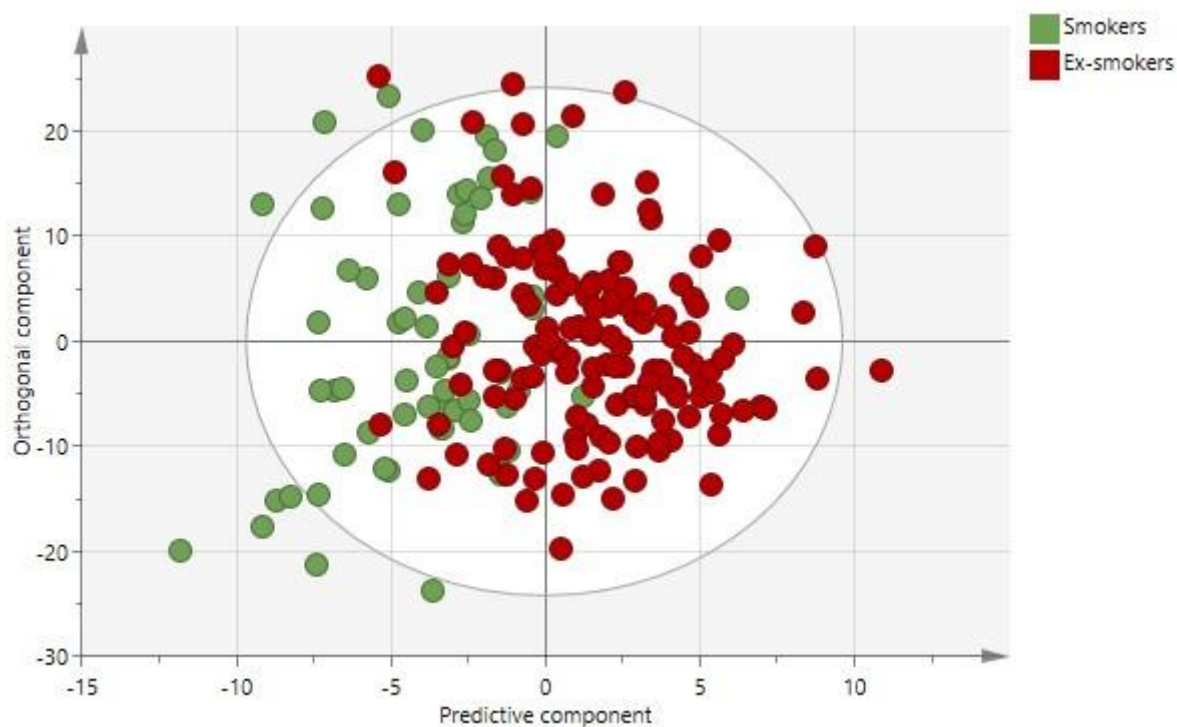
Figure 51: OPLS-DA score-plot of all smokers and all ex-smokers

In this OPLS-DA score-plot, a little discrimination can be observed between the two groups, but also these groups are not equal in numbers. The distribution is 63 smokers and 140 ex-smokers. Thus, the three outliers are again removed. Therefore, a new model is made with equal groups, thus 60 ex-smokers are chosen randomly as shown in Figure 52.
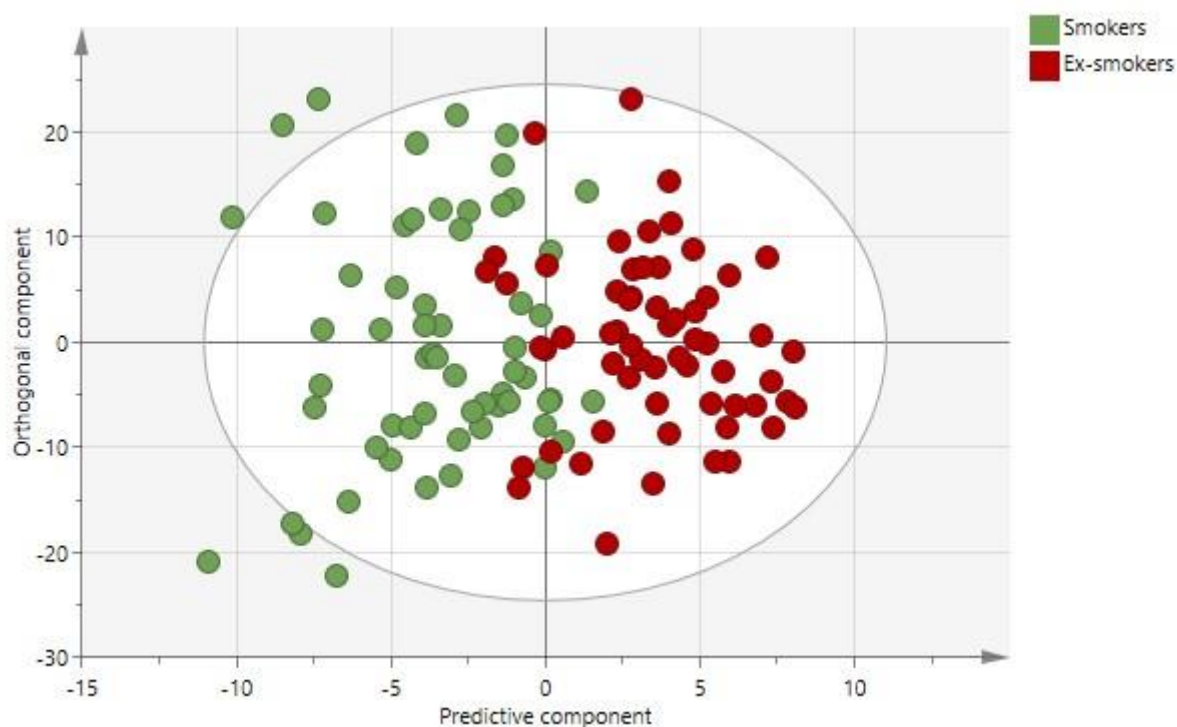
Figure 52: OPLS-DA score-plot of equal groups of smokers and ex-smokers

This OPLS-DA score-plot shows a clearer discrimination of the metabolite profile between the two groups. The statistical analyses in Table 9 shows that also this model is a good prediction based on the Q^2. But this model is not as good as the model made for the group of the smokers and non-smokers.

Table 9: Statistical analyses for equal groups of smokers and ex-smokers

Number of axis (P+O)	1+5
R^2X	0.891
R^2Y	0.624
Q^2	0.364
Misclassification:	
Smokers	54 of 60 are correctly classified (90.00 %)
Ex-smokers	52 of 60 are correctly classified (86.67 %)

When the ex-smokers are compared to the non-smokers, no acceptable OPLS-DA model could be trained by the software program. This means there is no discrimination found between the metabolite profile of ex-smokers and non-smokers.

8.2.3 Possible confounders

Also, several possible confounders to the model of smokers and non-smokers (Figure 50) are tested to investigate if the model made for the smoking status is not influenced by other clinical parameters. The possible confounders include age, sex and BMI. The model of smokers and non-smokers (Figure 50) was colored for these three clinical variables. Due to the low number of patients with diabetes, COPD and the subjects with no medication intake, none of these clinical variables could be tested for the smoking status model.

Age:

For this possible confounder a threshold of 69 years is set to divide the patients into two groups based on their age. The OPLS-DA score-plot of Figure 53 shows no clustering inside the smokers or the non-smokers group. Therefore, age is no confounder of the smoking status model.
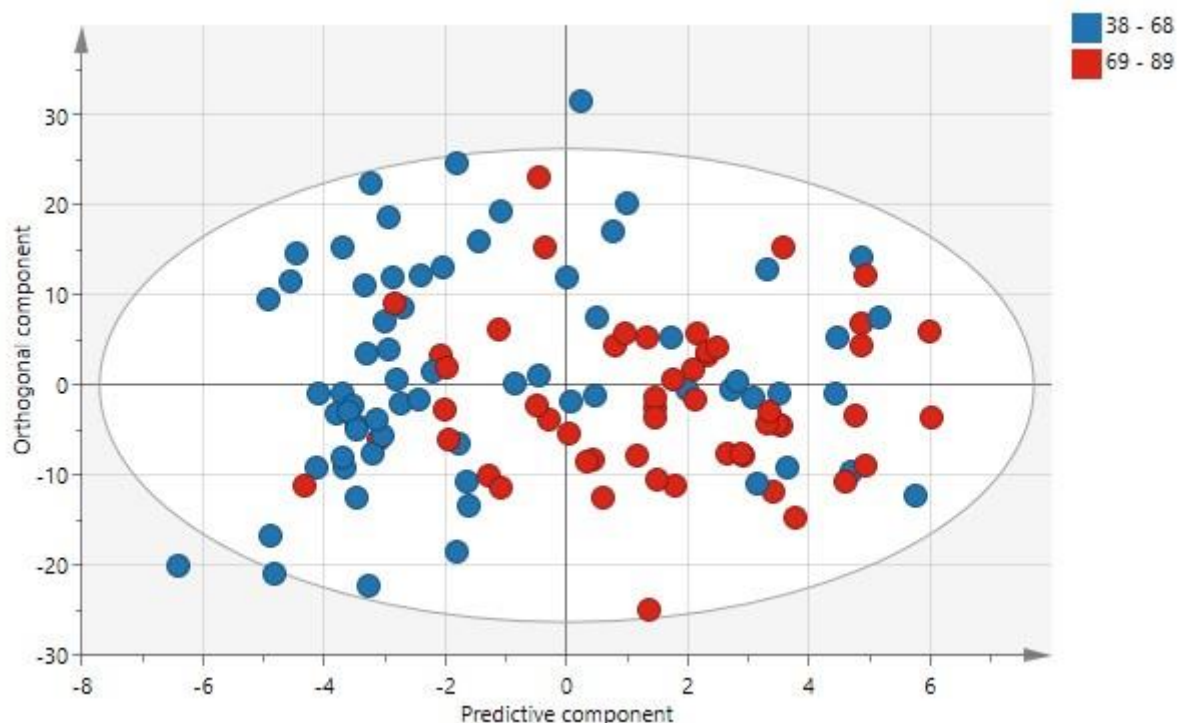
Figure 53: OPLS-DA score-plot of smoking status colored on possible confounder 'age' based on median

Sex:

Figure 54 shows the OPLS-DA score-plot based on the sex of the subjects. In this model no clustering is visible inside the smokers or the non-smokers group. Therefore, sex is no confounder of the smoking status model.
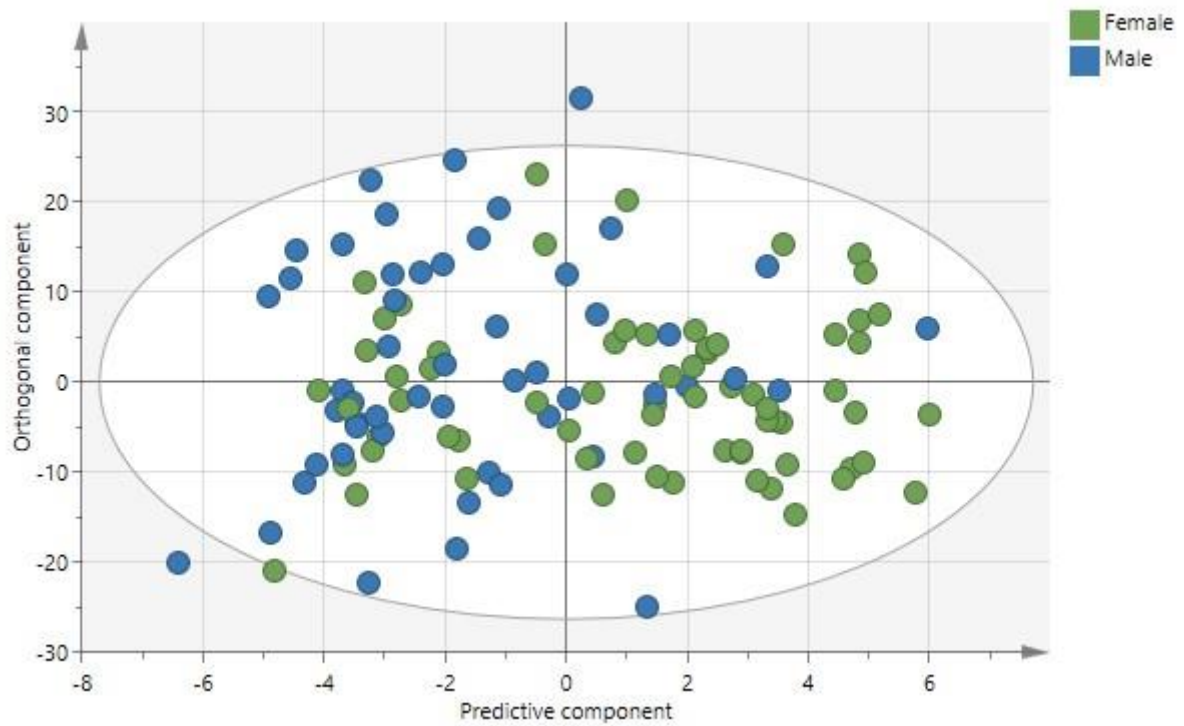
Figure 54: OPLS-DA score-plot of smoking status colored on possible confounder 'sex'

<u>BMI:</u>

Figure 55 shows the OPLS-DA score-plot in which the subjects are divided in two groups based on the median of their BMI. This score-plot from the smoking status model shows no clustering of the BMI inside the smokers or the non-smokers group.
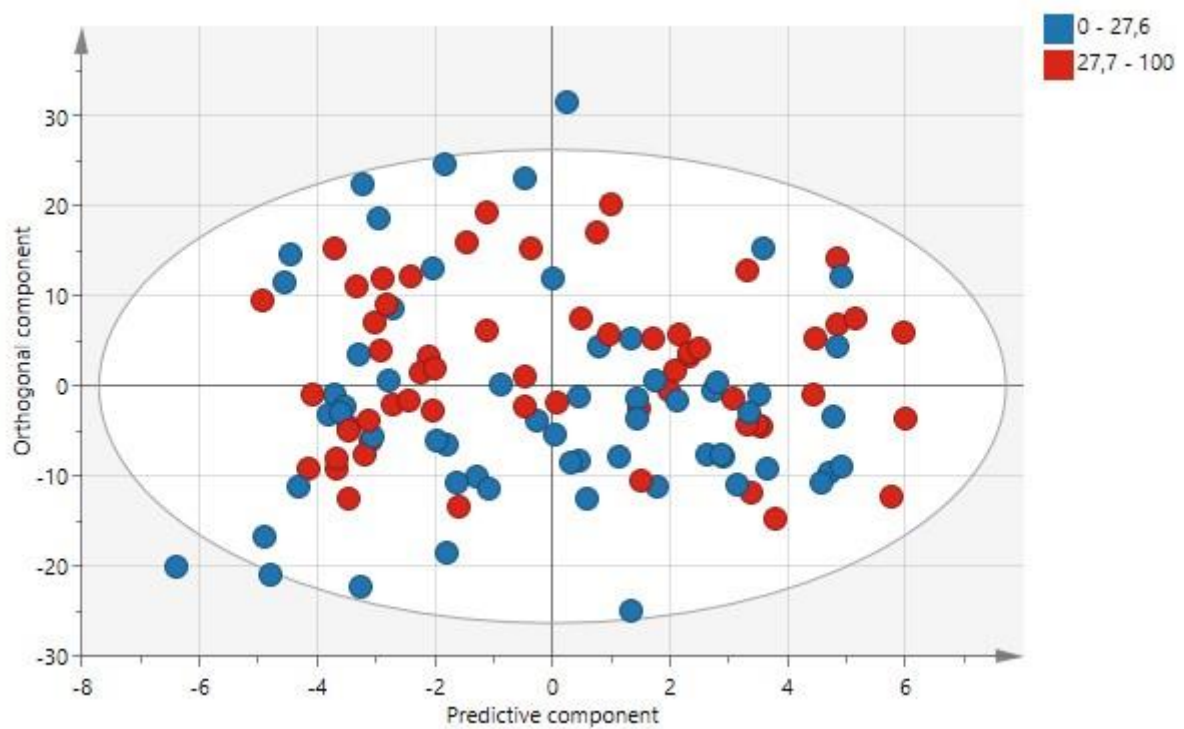
Figure 55: OPLS-DA score-plot of smoking status colored on possible confounder 'BMI' based on median

Figure 56 shows the result of the OPLS-DA score-plot based on the smoking status model, but with the subjects divided into three groups based on their clinically relevant BMI values.
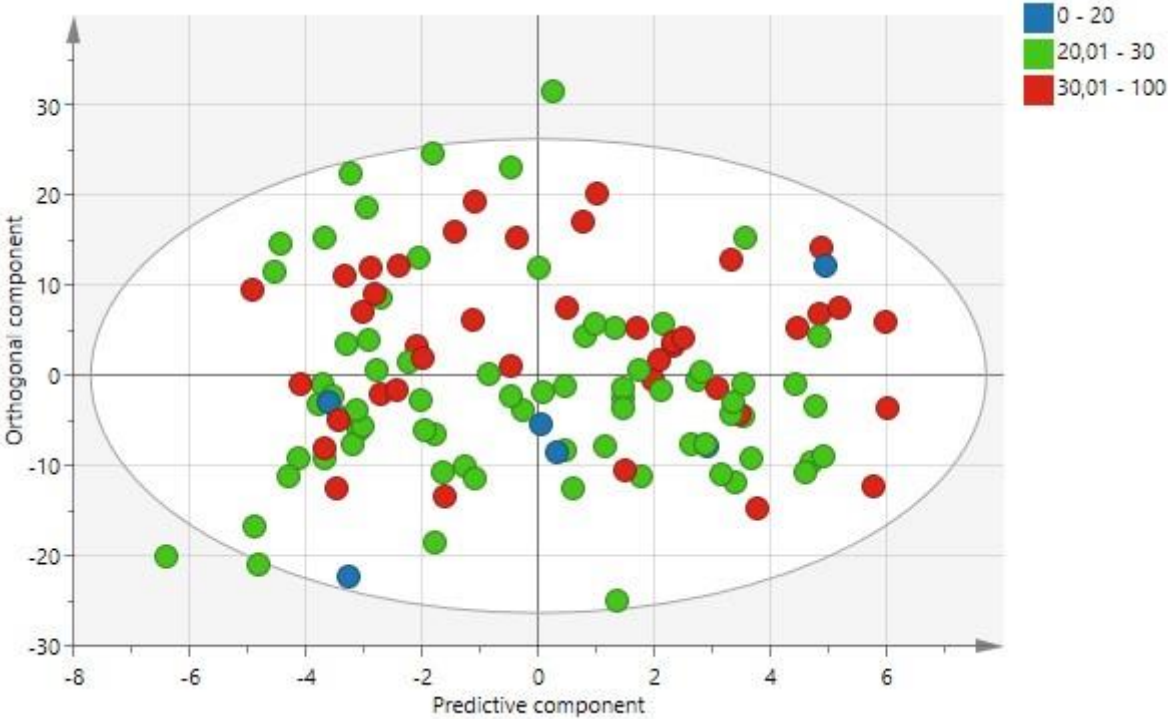
Figure 56: OPLS-DA score-plot of smoking status colored on possible confounder 'BMI' based on the medical grouping

Also, in this OPLS-DA score-plot no clustering is visible inside the smokers or non-smokers group. Therefore, based on Figure 55 and Figure 56, BMI is no confounder of the smoking status model.

9 Discussion and future goals

During the optimization process, different parameters were evaluated to observe how they influenced the peak positions (chemical shifts in ppm) and peak intensities in a ^{1}H-NMR spectrum, its robustness and total measurement time. The tested parameters are previously summed up and explained in the result section. Firstly, one of the important parameters is the total spin echo delay of the CPMG pulse sequence as shown in Figure 18. It is required to suppress the signal of macromolecules in the spectrum, but if the delay is too long, not only the macromolecules will be suppressed but also the signal of metabolites. When the delay is too short the macromolecules will not be suppressed enough and therefore the spectrum will deteriorate. The optimal values found in this study for the CPMG pulse sequence are as followed: a RD of 3 s, a presat of 3 s, a total spin echo delay of 64 ms and an AQC of 2.18 s. Secondly, also the sample concentration is important, because if it is too low some peaks might possibly be missed in the spectrum. The optimal value found in this study for the concentration is a sample to buffer ratio of 1:1. The third parameter that influenced the spectrum is the temperature during analysis. If the temperature is too high, like approx. body temperature, it is not recommended to evaluate the peaks for the study of biochemical pathways. The body temperature will start some metabolic processes whereby the hydrogen (H) atoms will be exchanged more between molecules. Therefore, the resolution of some peaks will decrease. The optimal value found in this study for the temperature is 25 °C. The last parameter is the total sample volume. When the volume is kept between 600 and 800 µl, it does not change the spectrum. The optimization of all these parameters has resulted in a robust ^{1}H-NMR metabolomics protocol for human blood plasma.

The statistical data analyses have shown that smokers and non-smokers can be discriminated based on their metabolite profile. These analyses confirm that age, sex and BMI are no confounders for the smoking status model. To better validate these findings, it needs to be performed again but on a greater scale and with an independent cohort.

The Prolung study will provide data of 200 early staged NSCLC patients. Each of those patients will donate seven samples of blood plasma at different times. This results at 1400 samples that will be measured. Afterwards, multivariate statistics will search for discriminations in the metabolite profile between patients with and without progression-free survival within one year. Thus, these results will provide more insight about the reaction of the body after the removal of a tumor. Thereafter, it will be studied how the metabolite profile will change over time between the two groups. Thus, the main goal of the Prolung study is to develop a prognostic biomarker that can predict if there will be an early relapse after surgery. The participants of the Prolung study can hereby also provide additional data for the validation of the discriminations in the metabolite profile between smokers and ex- or non-smokers.

The next steps of the research consist of spiking experiments and measurements of blood plasma from lung cancer patients. The spiking experiments will be required for quantification of plasma metabolites, to show the exact position of metabolites in the ^{1}H-NMR spectrum. When the position is known, the quantification can take place and the link with the metabolism of cancer cells can be made. For these spiking experiments reference samples from healthy individuals will be used. A well-defined concentration of a specific metabolite will be added to the sample, thereafter the sample is measured with ^{1}H-NMR spectroscopy. Observing the spectrum, it is expected that some regions will be increased. Those regions then can be assigned to that specific metabolite. The chosen metabolites that will be spiked belong to biochemical pathways such as the glycolysis, TCA cycle and amino acids that feed the

TCA by anaplerosis. Thus, based on the spiking experiments, the assignment of metabolites in a ^{1}H-NMR spectrum that Louis *et al.* performed will be confirmed and possible other relevant metabolites will be spiked for a better understanding of the spectrum

Louis *et al.* found that the concentration of glucose is increased while lactate and phospholipids concentrations are decreased in blood plasma of lung cancer patients [8]. Glucose and lactate are part of the glycolysis and/or gluconeogenesis and/or TCA cycle. The phospholipids are an important element for the production of a cell membrane. A plausible hypothesis is that those observations are a result of the metabolism of the body that strives for a homeostasis instead of the metabolism of a growing tumor [55]. When the tumor is having an increased uptake of glucose, the body will produce more glucose as a counter-reaction to achieve this equilibrium. Hart *et al.* and Ricard *et al.* measured these increased glucose values in breast cancer plasma samples [53], [55]. Xie *et al.*, Moreno *et al.*, Chen *et al.* and Hori *et al.* measured a decrease in glucose levels in lung cancer tissue, which connects to this hypothesis since the metabolism in cancer cells itself might be in contrast with the one observed in plasma [1], [56]–[58].

However, around this hypothesis is a lot of discussion, because there are a lot of different hypotheses on the metabolism of cancer. One of those other possible hypothesis stated that a growing tumor will take a lot of metabolites out of the bloodstream and therefore the metabolites, such as glucose, will be lower in the bloodstream [27]. Rocha *et al.* found these decreased levels of glucose in blood plasma of lung cancer patients [27]. Puchades-Carrasco *et al.* and Zhang *et al.* also measured a decrease in glucose levels, but in blood serum from lung cancer patients [30], [31].

The normal metabolism of the cell will convert glucose to pyruvate and pyruvate to acetyl-coA whereas this acetyl-coA fuels the TCA-cycle. The cancer cells need a lot of ATP on a short time to grow, thus pyruvate will be converted to lactate through the enzyme lactate dehydrogenase (LDH), as shown in Figure 57, to get the needed energy in a fast way [24], [59]. Therefore, there will be less acetyl-coA to fuel the TCA-cycle. Nevertheless, the cancer cell can still fuel the TCA, but with a different source such as glutamine [6], [26]. The usage of glutamine, through the glutaminolysis, will serve as a different fuel source for the TCA-cycle. This anaplerotic pathway, as shown in Figure 57, starts with glutamine entering the mitochondria where it converts to glutamate through the enzyme glutaminase (GLS1). This glutamate can then be secreted extracellular or further converted to the TCA intermediate α-ketoglutarate through the enzyme glutamate dehydrogenase (GLUD1). The α-ketoglutarate coverts in a few reactions in malate and can further in pyruvate what then leads to acetyl-coA or lactate depended on which metabolite the cancer cell needs on that specific moment [22], [26], [60].

Metabolomics can thus discover possible biomarkers, which can provide information about different biochemical pathways. This information can be used to find possible pharmaceutical targets, such as the enzyme GLUD1, for several diseases.
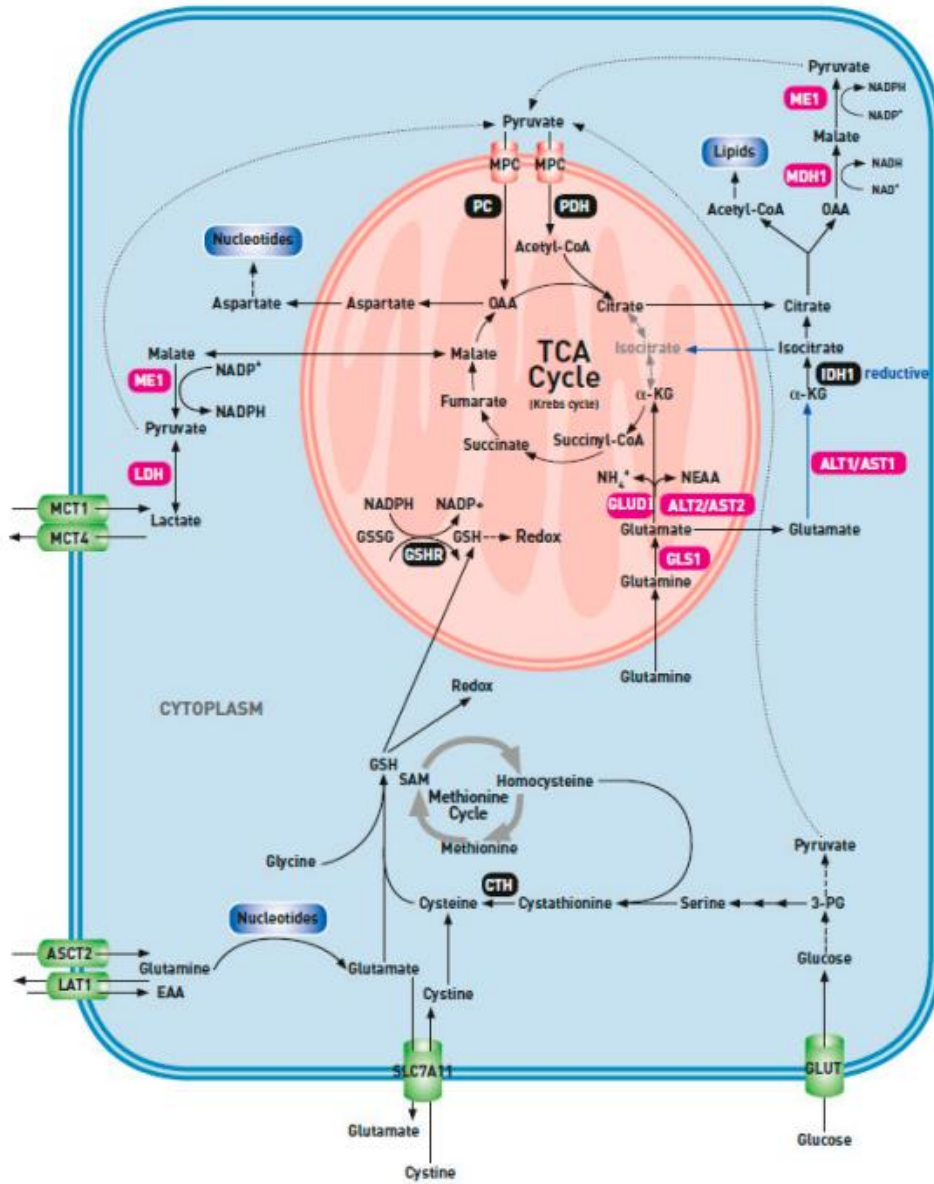
Figure 57: Anaplerotic pathway of glutamine in cancer cells, which can fuel the TCA cycle through conversion to α-ketoglutarate [26]

10 Conclusion

During the optimization different parameters were researched to observe how they influenced a metabolite ^{1}H-NMR spectrum, its robustness and measurement time. Some of those parameters had a great influence on the spectrum. The optimization has successfully delivered a robust ^{1}H-NMR metabolomics protocol for blood plasma. With this protocol plasma metabolites can be represented correctly and in a quantified manner. This protocol will be used for future spiking experiments and final measurements of patients from the Prolung study. With the information that will come from these measurements, the hypotheses of the Prolung study wants to be confirmed and expand to get more insights in the biochemical pathways of cancer cells.

The statistical analyses of a former metabolomics dataset have shown that there is a difference in the metabolite profile between smokers and non-smokers, but also between the metabolite profile between smokers and ex-smokers. This analyses also found that sex, age and BMI are no confounders for the smoking status model. To get more insights of these findings more research is required on a greater scale. The Prolung study will provide more patients and thus deliver a more representable statistical analysis.

Bibliography

[1]   S. Hori *et al.*, "A metabolomic approach to lung cancer," *Lung Cancer*, vol. 74, no. 2, pp. 284–292, 2011.

[2]   M. Burotto, A. Thomas, D. Subramaniam, G. Giaccone, and A. Rajan, "Biomarkers in early-stage non-small-cell lung cancer: Current concepts and future directions," *J. Thorac. Oncol.*, vol. 9, no. 11, pp. 1609–1617, 2014.

[3]   V. Kumar, A. Abbas, and J. Aster, *Robbins Basic Pathology*, Ninth. Philadelphia: Elsevier Saunders, 2013.

[4]   H. Uramoto and F. Tanaka, "Recurrence after surgery in patients with NSCLC," *Transl. lung cancer Res.*, vol. 3, no. 16, pp. 242–249, 2014.

[5]   J. H. G. M. van Beek, "The dynamic side of the Warburg effect: glycolytic intermediate storage as buffer for fluctuating glucose and O2 supply in tumor cells," *F1000Research*, vol. 7, no. 1, p. 1177, 2018.

[6]   R. J. DeBerardinis, J. J. Lum, G. Hatzivassiliou, and C. B. Thompson, "The Biology of Cancer: Metabolic Reprogramming Fuels Cell Growth and Proliferation," *Cell Metab.*, vol. 7, no. 1, pp. 11–20, 2008.

[7]   M. Bracke, F. Lardon, P. Vandenberghe, and K. Vanderkerken, *Kanker biomedisch onderzoek*, First. Berchem: De Boeck, 2015.

[8]   E. Louis *et al.*, "Detection of lung cancer through metabolic changes measured in blood plasma," *J. Thorac. Oncol.*, vol. 11, no. 4, pp. 516–523, 2016.

[9]   E. Louis *et al.*, "Phenotyping human blood plasma by 1H-NMR: a robust protocol based on metabolite spiking and its evaluation in breast cancer," *Metabolomics*, vol. 11, no. 1, pp. 225–236, 2015.

[10]   J. C. Lindon, J. K. Nicholson, and J. R. Everett, "NMR Spectroscopy of Biofluids," *Annu. Reports NMR Spectrosc.*, vol. 38, no. C, pp. 1–88, 1999.

[11]   M. Coen, E. Holmes, J. C. Lindon, and J. K. Nicholson, "NMR-based metabolic profiling and metabonomic approaches to problems in molecular toxicology," *Chem. Res. Toxicol.*, vol. 21, no. 1, pp. 9–27, 2008.

[12]   J. T. Brindle *et al.*, "Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics," *Nat. Med.*, vol. 8, no. 12, pp. 1439–1444, 2002.

[13]   N. V. Reo, "NMR-based metabolomics," *Drug Chem. Toxicol.*, vol. 25, no. 4, pp. 375–382, 2002.

[14]   J. L. Markley *et al.*, "The future of NMR-based metabolomics," *Curr. Opin. Biotechnol.*, vol. 43, no. 1, pp. 34–40, 2017.

[15]   A. Zhang, H. Sun, and X. Wang, "Serum metabolomics as a novel diagnostic approach for disease: A systematic review," *Anal. Bioanal. Chem.*, vol. 404, no. 4, pp. 1239–1245, 2012.

[16]   E. Derveaux *et al.*, "Diagnosis of Lung Cancer: What Metabolomics Can Contribute," *Lung Cancer - Strategies for Diagnosis and Treatment*. pp. 80–94, 2018.

[17]   S. Bamji-Stocke, V. van Berkel, D. M. Miller, and H. B. Frieboes, "A review of metabolism-associated biomarkers in lung cancer diagnosis and treatment," *Metabolomics*, vol. 14, no. 81,

pp. 1–16, 2018.

[18]   L. Van Eycken, "Cancer Burden in Belgium 2004-2013," Brussels, 2015.

[19]   P. Goldstraw *et al.*, "The IASLC Lung Cancer Staging Project : Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classi fi cation for Lung Cancer," *J. Thorac. Oncol.*, vol. 11, no. 1, pp. 39–51, 2015.

[20]   M. Shayan, "Enzymes in chemical reactions of glycolysis are located where?," 2017. [Online]. Available: https://www.quora.com/Enzymes-in-chemical-reactions-of-glycolysis-are-located-where. [Accessed: 05-Jun-2019].

[21]   C. Rachna, "Difference Between Glycolysis and Krebs (citric acid) Cycle," 2017. [Online]. Available: https://biodifferences.com/difference-between-glycolysis-and-krebs-citric-acid-cycle.html. [Accessed: 05-Jun-2019].

[22]   J. M. Berg, J. L. Tymoczko, G. J. J. Gatto, and L. Stryer, *Biochemistry*, Eighth. New York: Kate Ahr Parker, 2015.

[23]   I. Martínez-Reyes and N. S. Chandel, "Waste Not, Want Not: Lactate Oxidation Fuels the TCA Cycle," *Cell Metab.*, vol. 26, no. 1, pp. 803–804, 2017.

[24]   O. Warburg, "On the origin of cancer cells," *Bull. At. Sci.*, vol. 123, no. 3191, pp. 309–314, 1956.

[25]   "LDH activity- normal vs canceous cells."

[26]   K. Vanhove *et al.*, "Glutamine Addiction and Therapeutic Strategies in Lung Cancer," *Int. J. Mol. Sci.*, vol. 20, no. 2, p. 252, 2019.

[27]   C. M. Rocha *et al.*, "Metabolic Signatures of Lung Cancer in Biofluids: NMR-Based Metabonomics of Blood Plasma," *J. Proteome Res.*, vol. 10, no. 9, pp. 4314–4324, 2011.

[28]   I. F. Duarte, C. M. Rocha, and A. M. Gil, "Metabolic profiling of biofluids: Potential in lung cancer screening and diagnosis," *Expert Rev. Mol. Diagn.*, vol. 13, no. 7, pp. 737–748, 2013.

[29]   Z. Yu *et al.*, "Differences between human plasma and serum metabolite profiles," *PLoS One*, vol. 6, no. 7, pp. 1–6, 2011.

[30]   X. Zhang, X. Zhu, C. Wang, H. Zhang, and Z. Cai, "Non-targeted and targeted metabolomics approaches to diagnosing lung cancer and predicting patient prognosis," *Oncotarget*, vol. 7, no. 39, pp. 63437–63448, 2016.

[31]   L. Puchades-Carrasco *et al.*, "Serum metabolomic profiling facilitates the non-invasive identification of metabolic biomarkers associated with the onset and progression of non-small cell lung cancer," *Oncotarget*, vol. 7, no. 11, pp. 12904–12916, 2016.

[32]   Y. Chen *et al.*, "Metabolomic profiling of human serum in lung cancer patients using liquid chromatography/hybrid quadrupole time-of-flight mass spectrometry and gas chromatography/mass spectrometry," *J. Cancer Res. Clin. Oncol.*, vol. 141, no. 4, pp. 705–718, 2015.

[33]   J. Carrola *et al.*, "Metabolic Signatures of Lung Cancer in Biofluids : NMR-Based Metabonomics of Urine research articles," *J. Proteome Res.*, vol. 10, no. 1, pp. 221–230, 2011.

[34]   T. D. W. Claridge, *High-Resolution NMR Techniques in Organic Chemistry*, Second. Oxford: Elsevier Ltd, 2009.

[35]   "A more precise measurement of the proton's magnetic moment," 2019. [Online]. Available: https://resonance.is/precise-measurement-protons-magnetic-moment/. [Accessed: 05-Apr-2019].

[36] E. J. Blink, "MRI: Principes," pp. 10–20, 2004.

[37] L. J. Bond, "Evaluation of Non-Nuclear Techniques for Well Logging : Final Report PNNL-20831 Evaluation of Non-Nuclear Techniques for Well Logging : Final Report," no. June, 2011.

[38] A. Hazra, "Numerical Simulation of Bloch Equations for Dynamic Magnetic Resonance Imaging," Georg-August-Universität Göttingen, 2016.

[39] I. Bolotin, D. McElheny, and Y. Ishii, "Fourier Transform (FT) NMR and Determination of Molecular Diffusion Coefficients by Pulse Field Gradient (PFG) Experiment for Unknown Samples," 2011. [Online]. Available: https://pdfs.semanticscholar.org/538a/414cbe74db8715dccbf607bce735455c01ce.pdf. [Accessed: 11-Apr-2019].

[40] "Theoretical introduction to Nuclear Magnetic Resonance." [Online]. Available: https://dnangelica.com/dnangelica/index.php/2018/08/14/what-about-nuclear-magnetic-resonance/. [Accessed: 05-Apr-2019].

[41] S. R. Hinks, "Spin Echo," *MR Physiscs Tech. Clin. Spin Echo*, pp. 1–7, 2015.

[42] "Nuclear magnetic resonance (NMR)." [Online]. Available: http://alma.karlov.mff.cuni.cz/hamrle/teaching/lectures/hamrle_nmr.pdf. [Accessed: 09-Apr-2019].

[43] K. Ruotsalo and M. S. Tant, "Buffy Coat Examination for Mast Cells." [Online]. Available: https://vcahospitals.com/know-your-pet/buffy-coat-examination-for-mast-cells. [Accessed: 05-Apr-2019].

[44] N. H. Kim *et al.*, "A p53/miRNA-34 axis regulates Snail1-dependent cancer cell epithelial-mesenchymal transition," *J. Cell Biol.*, vol. 195, no. 3, pp. 417–433, 2011.

[45] N. K. Altorki *et al.*, "The lung microenvironment: an important regulator of tumour growth and metastasis," *Nat. Rev. Cancer*, vol. 19, no. 1, pp. 9–31, 2019.

[46] E. Heitzer, I. S. Haque, C. E. S. Roberts, and M. R. Speicher, "Current and future perspectives of liquid biopsies in genomics-driven oncology," *Nat. Rev. Genet.*, vol. 20, no. 2, pp. 71–88, 2019.

[47] JEOL, "JNM-ECZS series FT NMR." .

[48] S. Kostidis, R. D. Addie, H. Morreau, O. A. Mayboroda, and M. Giera, "Quantitative NMR analysis of intra- and extracellular metabolism of mammalian cells: A tutorial," *Anal. Chim. Acta*, vol. 980, no. 1, pp. 1–24, 2017.

[49] O. Beckonert *et al.*, "Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts.," *Nat. Protoc.*, vol. 2, no. 11, pp. 2692–2703, 2007.

[50] K. A. Rubinson, "Analytical Methods Practical corrections for p (H, D) measurements in mixed H2O/D2O biological buffers," *R. Soc. Chem.*, pp. 1–7, 2017.

[51] K. M. DeAngelis, "Phosphate Buffer," *Cold Spring Harbor Protocols*. 2007.

[52] E. Jobard *et al.*, "A systematic evaluation of blood serum and plasma pre-analytics for metabolomics cohort studies," *Int. J. Mol. Sci.*, vol. 17, no. 12, pp. 1–12, 2016.

[53] C. D. Hart *et al.*, "Serum Metabolomic Profiles Identify ER-Positive Early Breast Cancer Patients at Increased Risk of Disease Recurrence in a Multicenter Population," vol. 23, no. 5, pp. 1422–1432, 2017.

[54] S. Deja *et al.*, "Metabolomics provide new insights on lung cancer staging and discrimination

[36] E. J. Blink, "MRI: Principes," pp. 10–20, 2004.

[37] L. J. Bond, "Evaluation of Non-Nuclear Techniques for Well Logging : Final Report PNNL-20831 Evaluation of Non-Nuclear Techniques for Well Logging : Final Report," no. June, 2011.

[38] A. Hazra, "Numerical Simulation of Bloch Equations for Dynamic Magnetic Resonance Imaging," Georg-August-Universität Göttingen, 2016.

[39] I. Bolotin, D. McElheny, and Y. Ishii, "Fourier Transform (FT) NMR and Determination of Molecular Diffusion Coefficients by Pulse Field Gradient (PFG) Experiment for Unknown Samples," 2011. [Online]. Available: https://pdfs.semanticscholar.org/538a/414cbe74db8715dccbf607bce735455c01ce.pdf. [Accessed: 11-Apr-2019].

[40] "Theoretical introduction to Nuclear Magnetic Resonance." [Online]. Available: https://dnangelica.com/dnangelica/index.php/2018/08/14/what-about-nuclear-magnetic-resonance/. [Accessed: 05-Apr-2019].

[41] S. R. Hinks, "Spin Echo," *MR Physiscs Tech. Clin. Spin Echo*, pp. 1–7, 2015.

[42] "Nuclear magnetic resonance (NMR)." [Online]. Available: http://alma.karlov.mff.cuni.cz/hamrle/teaching/lectures/hamrle_nmr.pdf. [Accessed: 09-Apr-2019].

[43] K. Ruotsalo and M. S. Tant, "Buffy Coat Examination for Mast Cells." [Online]. Available: https://vcahospitals.com/know-your-pet/buffy-coat-examination-for-mast-cells. [Accessed: 05-Apr-2019].

[44] N. H. Kim *et al.*, "A p53/miRNA-34 axis regulates Snail1-dependent cancer cell epithelial-mesenchymal transition," *J. Cell Biol.*, vol. 195, no. 3, pp. 417–433, 2011.

[45] N. K. Altorki *et al.*, "The lung microenvironment: an important regulator of tumour growth and metastasis," *Nat. Rev. Cancer*, vol. 19, no. 1, pp. 9–31, 2019.

[46] E. Heitzer, I. S. Haque, C. E. S. Roberts, and M. R. Speicher, "Current and future perspectives of liquid biopsies in genomics-driven oncology," *Nat. Rev. Genet.*, vol. 20, no. 2, pp. 71–88, 2019.

[47] JEOL, "JNM-ECZS series FT NMR." .

[48] S. Kostidis, R. D. Addie, H. Morreau, O. A. Mayboroda, and M. Giera, "Quantitative NMR analysis of intra- and extracellular metabolism of mammalian cells: A tutorial," *Anal. Chim. Acta*, vol. 980, no. 1, pp. 1–24, 2017.

[49] O. Beckonert *et al.*, "Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts.," *Nat. Protoc.*, vol. 2, no. 11, pp. 2692–2703, 2007.

[50] K. A. Rubinson, "Analytical Methods Practical corrections for p (H, D) measurements in mixed H2O/D2O biological buffers," *R. Soc. Chem.*, pp. 1–7, 2017.

[51] K. M. DeAngelis, "Phosphate Buffer," *Cold Spring Harbor Protocols*. 2007.

[52] E. Jobard *et al.*, "A systematic evaluation of blood serum and plasma pre-analytics for metabolomics cohort studies," *Int. J. Mol. Sci.*, vol. 17, no. 12, pp. 1–12, 2016.

[53] C. D. Hart *et al.*, "Serum Metabolomic Profiles Identify ER-Positive Early Breast Cancer Patients at Increased Risk of Disease Recurrence in a Multicenter Population," vol. 23, no. 5, pp. 1422–1432, 2017.

[54] S. Deja *et al.*, "Metabolomics provide new insights on lung cancer staging and discrimination

from chronic obstructive pulmonary disease," *J. Pharm. Biomed. Anal.*, vol. 100, pp. 369–380, 2014.

[55]   V. Richard, R. Conotte, D. Mayne, and J. Colet, "Does the 1 H-NMR plasma metabolome reflect the host-tumor interactions in human breast cancer ?," vol. 8, no. 30, pp. 49915–49930, 2017.

[56]   H. Xie *et al.*, "Targeting lactate dehydrogenase-A inhibits tumorigenesis and tumor progression in mouse models of lung cancer and impacts tumor-initiating cells," *Cell Metab.*, vol. 19, no. 5, pp. 795–809, 2014.

[57]   P. Moreno and C. Jim, "Metabolomic profiling of human lung tumor tissues – nucleotide metabolism as a candidate for therapeutic interventions and biomarkers," vol. 12, pp. 1778–1796, 2018.

[58]   W. Chen *et al.*, "Study on metabonomic characteristics of human lung cancer using high resolution magic-angle spinning 1H NMR spectroscopy and multivariate data analysis," *Magn. Reson. Med.*, vol. 66, no. 1, pp. 1531–1540, 2011.

[59]   J. Lu, M. Tan, and Q. Cai, "The Warburg effect in tumor progression: Mitochondrial oxidative metabolism as an anti-metastasis mechanism," *Cancer Lett.*, vol. 356, no. 2, pp. 156–164, 2015.

[60]   R. Beger, "A Review of Applications of Metabolomics in Cancer," *Metabolites*, vol. 3, no. 1, pp. 552–574, 2013.