universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Establishment of a Shotgun Lipidomics Platform“

verfasst von / submitted by

Florian Prodinger BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2016 / Vienna 2016

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	A 066 862
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Chemie
Betreut von / Supervisor:	Univ.-Prof. Dr. Christopher Gerner

Table of Contents

List of Figures

List of Tables

Formulaic directory

List of Abbreviations

DMEM *Dulbecco's Modified Eagle Medium*
DPPC *Dipalmitoylphosphatidylcholine*
ESI *Electrospray ionization*, *Elektrospray-Ionisation*
LOD *Limit of detection*
LPC *Lysophosphatidylcholine*
MFQL *Molecular Fragmentation Query Language*
MS1 *(Precursor) mass spectrum*
MS^2 *(Product) mass spectrum*
MTBE *Methyl-tert-butyl-ether*
mzML *mass spectrum file format*
PA *Phosphatidic acid*
PC *Phosphatidylcholine*
PC-O *Phosphatidylcholine with one ether bond*, *Phosphatidylcholine with one ether bond*
PE *Phosphatidylethanolamine*
PE-O *Phosphatidylethanolamine with one ether bond*
PG *Phosphatidylglycerol*
PI *Phosphatidylinositol*
PS *Phosphatidylserine*
SAGM *small airway growth medium*
SFC *supercritical fluid chromatography*
SM *Sphingomyelin*
SOP *Standard Operating Procedure*
txt *Text file format*
VBA *Visual Basic for Applications*

1 Introduction

The analysis of the lipidome under the name of lipidomics has started at the beginning of the 21st century. Before that time was the human genome project, which was completed in 2003 and offered information about several illnesses, which are used to aid drug discovery. However soon the scientific community realized that the knowledge of genes would not suffice for a holistic view of the complicated system of the human body. [19]

The next step was to identify the whole human proteome, which is an ongoing undertaking and does not, like the human genome project, have a date on which the list of human proteins will be completed, as it is more complicated. [27]

The understanding of proteins will again help to understand illnesses, however there is still a major component missing to really grasp a cellular pathway: The metabolome. When checking a random Pathway on the KEGG database, it becomes apparent that the small molecules, which are studied in the field of metabolomics, are indeed an important part of understanding cellular behavior. [20]

Especially in recent years the study of lipids, which are considered metabolites, has seen a significant rise. This shows the interested of the scientific community! In the first ten years of the term lipidomics, there has been a similar number of papers published, than in the last three and a half years. [30]

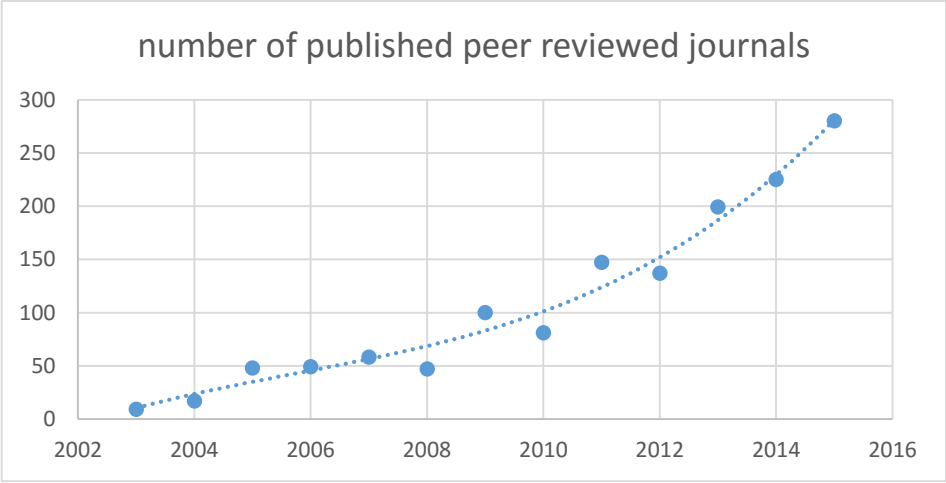

Figure 1: number of published articles during the years. [30]

Lipids are usually studied with a mass spectrometric approach, like the relatively new technology of orbitrap mass spectrometers. This technology allows to generate *gb*s of data

in relatively short time and shifts the complex and difficult part of analysis away from wet chemistry and measuring. Again we need to take a look at the history of 'omics' approaches; while bioinformatics and databases are not even mentioned in a short explanation of the human genome project, these computer scientific methods are one of the three pillars of the proteomics study of the human proteome project. [27, 19]

It goes without saying that lipidomics will be highly dependent on computer science and bioinformatics, all while being very hard to combine with these methods, for the chemical diversity of the analytes. [6, 4]

In this thesis we were asked to analyze a series of amniotic tissue supernatant samples and outline its similarities to a second sample, which was lung surfactant. Preliminary studies on this topic were done by researches of the Ludwig Boltzmann Institute for Experimental and Clinical Traumatology, they found DPPC in amniotic tissue supernatant in not yet published experiments. DPPC or PC [16:0 / 16:0] is the most abundant lipid in mammalian lung lavage. This challenge is especially interesting, because we had to setup the extraction of the lipids from the different samples, the measurement method and also the interpretation. We promptly used this task to not only to apply our skills in the laboratory, but also to implement an automated identification software called LipidXplorer.

1.1 Workflow and process

Also we set up our workflow including sample preparation methods and Thermo Scientific LTQ Velos orbitrap as well as Advion TriVersa NanoMate settings by studying papers published by the group, who created the LipidXplorer software. [13,17]

In this thesis we will mainly explore the phospholipid content of different samples and compare the 'phospholipidome' of these samples. The focus on phospholipids arise from our experimental setup, which is ideal for the task given by our partners: Angela Lemke, Susanne Wolbank and Heinz Redl.

After setting up a method using literature and lipid standards our partners provided us not only with lavage samples in which the phospholipid composition was determined, but also with more complex samples which have not been analyzed by a lipidomics approach. Hence

we are the first scientists to examine the secreted lipids of amnion cells via mass spectrometry.

Our partners were interested in estimating the similarity of the surfactant of the lung and the surfactant produced by amnion cells. Hence we first used literature to get an overview of secreted lipids. As previous studies have shown, the mentioned surfactant is usually composed of mainly phospholipids. Especially Phosphatidylcholines with PC [16:0 / 16:0] being the most abundant phospholipid. [12,2,16]

The method was established by studying literature of similar approaches, especially the setting of the ESI source and the mass spectrometer stated by a leading lipidomics group. [17] However the fine tuning of our method was done via trial and error and consulting a technician. Hence our settings changed over the course of the master thesis; all experiments are reproducible nevertheless.

1.2 LipidXplorer

The biggest challenge of our assignment was the data interpretation. This has several reasons. Firstly, even though the MS1 spectrum is highly reproducible, the masses chosen for a MS^2 experiment are not. This impediment resulted in a more complex problem; we had to discuss at what point we identified a detected lipid with a 100% certainty. Some groups tended to assume a correct 'mass' in a high resolution MS1 experiment is already an identified hit, since the elemental composition can be estimated from the correct mass. [18]

However, the papers which introduce and verify this approach are already several years old and with an advancement in hardware the criteria for a hit became stricter, hence we also had to rely on a more stringent approach to produce serious results. By not only relying on the highly reproducible MS1 but also the less reproducible MS^2 patterns, we created a method for data interpretation. It was difficult to assess which MS experiment provided a MS^2 spectrum, which also fulfills our set criteria.

Here we implemented an automated data interpretation software called LipidXplorer, a software tailored for shotgun lipidomics. After working out the settings we quickly were able to streamline the whole process. In the end we were able to extract samples, measure them

roughly 50 to 60 times, convert those measurements to a master scan file and produce a crude lipidomics overview in a single day.

2 Theoretical background

This chapter will give an overview of the analytes, used principles of mass spectrometry and the software used for data interpretation. I will carefully explain the details of LipidXplorer, so that no 'black boxes' remain of how we obtained knowledge from our measurements.

2.1 Lipids

Lipids can be divided in 8 categories, which can be further divided in a vast amount of classes and species. The categories are:

Fatty acids

Glycerolipids

(Glycero)phospholipids

Sphingolipids

Saccaharolipids

Polyketides

Sterol lipid

Prenol lipids

Each of these categories has different biological function, chemical properties and a high variety of derivates (e.g. reduced forms, isoforms, inserted heteroatoms, rings…). [21, 7: p. 6-10]

In this thesis we will mostly explore the phospholipid content of our samples. These lipids feature a glycerin backbone, two of the three hydroxyl groups are linked as an ester to a

fatty acid. The third hydroxyl group is bound to a phosphate which is further linked to a specific group. This group divides the lipid category phospholipids in several classes and can take the form of serine, choline, inositol, etc. [1: p. 118]

The fatty acids on the phospholipid usually feature a length between 14 and 22 carbon atoms, this is most likely due to the function of the phospholipids, they are basically the building blocks of membranes and the length of the fatty acids determines the thickness of the membrane. [14: p. 24] The thickness is usually around 4nm. [1: p. 119]

The synthesis of fatty acids, which are part of certain lipid groups, usually produce an even numbered carbon atom chain. This is because the most prominent pathways of lipid synthesis build lipids out of two carbon atom building blocks called acetyl. [25] However, even oddly numbered fatty acids can occur, for instance a 17 carbon atom chain with a double bond Even though there is evidence of oddly numbered fatty acids, they are usually a minority in the lipidome. [18, 14: p. 24]

Lipidomics is relatively new field of the 'omics' analysis, which, combined with the previously stated chemical variance of lipids demotes software development. In this thesis we will use and (thereby) test freely accessible software called LipidXplorer, which also has a growing community and a very personal trouble shooting care. Furthermore, several papers with detailed settings and MFQL scripts were published. [9, 8]

2.2 Surfactants

All of the biological samples, which are analyzed in this work, are surfactants. Either lung surfactants or a surfactant produced in the amniotic sack between the sack and the fetus.

In the lung the surfactant has an important role, to stabilize the network of air filled channels, which are called alveoli, during exhalation. Without the surfactant all the air would be forced out of the alveoli, which results in a collapse of the lung. It is debated that the surfactant, which mainly consists of PC [16:0 / 16:0] or DPPC, has more tasks than just stabilization, however the details about the lipid monolayer in the lung are not yet completely understood. The stabilization feature of the lung surfactant is not open for debate, because there are several medical conditions, which cause a low in surfactant, which results in a collapse of the lung. One interesting case are prematurely born infants; Lung

surfactant is only formed in the late stages of pregnancy. If a child is born early, the act of breathing is energetically demanding and also dangerous for the risk of lung to collapse. [14: p. 114]

A review of the current literature describing lung surfactant revealed a rather variable composition. [12, 2, 16]

The differences in these lung surfactants may suggest that the specific fatty acids in the described lipids might play a minor role, as opposed to certain classes of phospholipids, which are always present. This indicates that the function of the lung surfactant mainly depends on a high concentration of PC [16:0 / 16:0] and other saturated phospholipids, and not on a very specific, highly preserved combination of classes of phospholipids featuring specific fatty acids.

The amniotic cell surfactant has not yet been studied in a shotgun Lipidomics approach. By analyzing the composition of both, amniotic tissue supernatant and pig lung lavage, and discovering a similarity in the surfactant lipid wise, we might be able to aid illnesses accompanied by a loss of lung surfactant. This might sound farfetched, but depending on our findings we may be able to argue that amniotic surfactant can be used as a surrogate for lung surfactants.

2.3 LTQ Oribitrap Velos

The LTQ orbitrap Velos is a hybrid mass spectrometer consisting of a LTQ Velos and an orbitrap analyzer, both are sold by the vendor Thermo Scientific. The mass spectrometer was coupled with an ESI source called the TriVersa NanoMate pro, which will be introduced in a later chapter.

2.3.1 Schematics and working principle

Here is a rough draft of the used mass spectrometer. It is a so called ‘LTQ orbitrap Velos’ and consists of an LTQ Velos and an orbitrap analyzer, this is why the schematic is split in two parts.

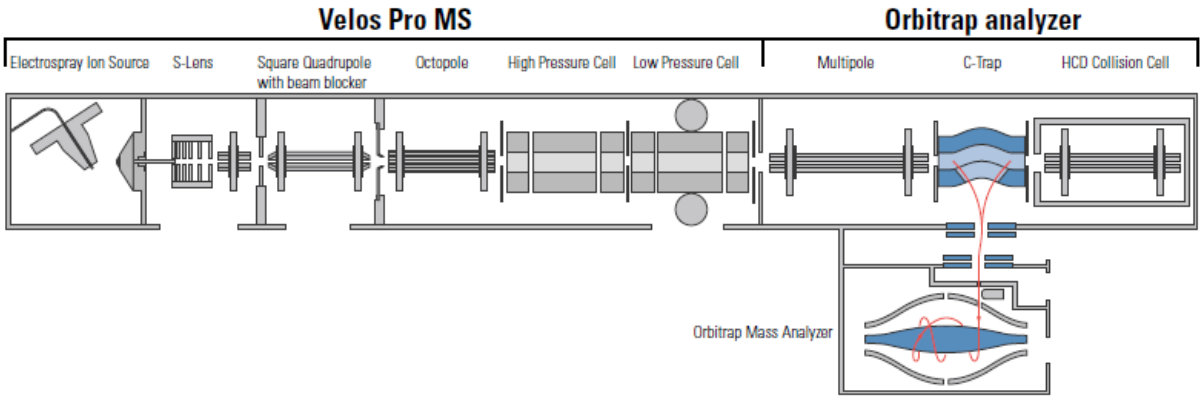

Figure 2: Schematics of the LTQ orbitrap Velos. [28]

The LTQ Velos is a mass spectrometer on its own, it was released in 2008. [24: p. v] Its mass analyzers are two ion traps [24: p. 43], which are called high pressure chamber and low pressure chamber in the schematics. The orbitrap analyzer is an add-on and was released later. It represents an enormous upgrade to the system through its C-trap and orbitrap mass analyzer

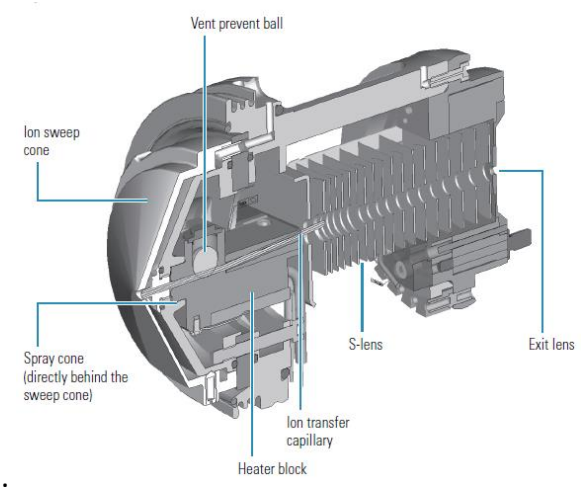

Figure 3: The metal cone and S-lens prevent staining of the mass spectrometer and keep the ions on track. [24: p. 38]

An ESI source is utilized to deliver the sample to the mass spectrometer.

The ionized molecules pass through a metal cone, which acts as first separation. It helps to keep droplets and unionized molecules out of the system. After the cone the so called S-lens follows. This part consists of a number of rings, which guide the ions through a bent path, again this is done to keep uncharged particles out of the system and to avoid staining of the inner parts of the mass spectrometer.

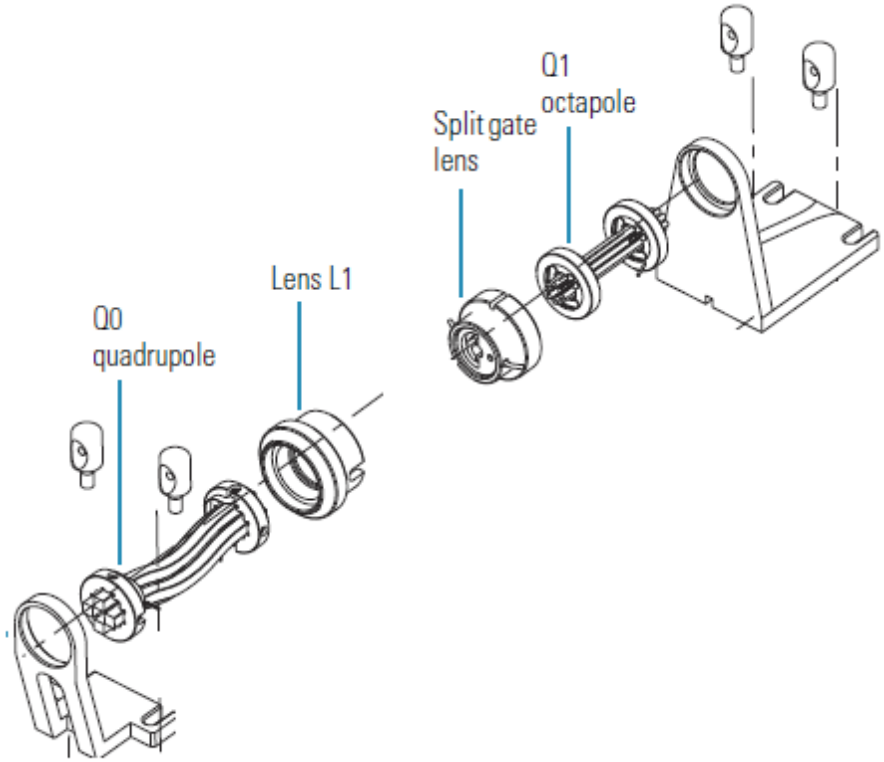

Figure 4: Setup of the ion optics of the LTQ Velos. [24: p. 134]

After the S-Lens follows an array of lenses and quadrupole, which focus the ion beam and guide it to the mass analyzer. [24: p. 40-45] In the original Velos the mass analyzer consisted of two ion traps, however after updating the machine the new mass analyzers are the C-trap and the orbitrap. [26: p. 38]

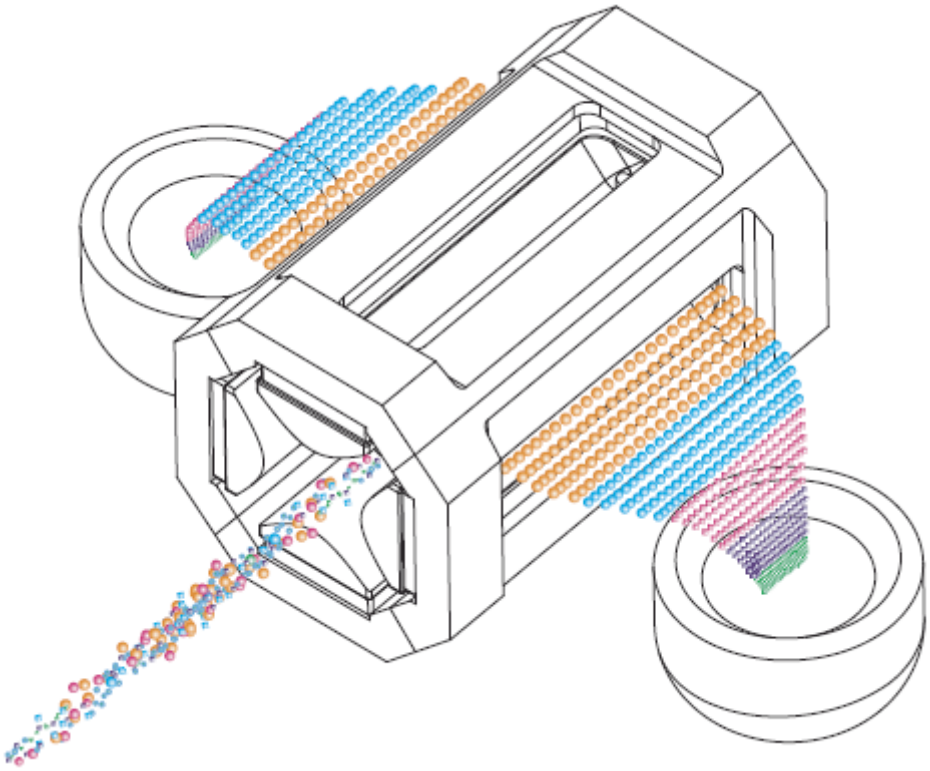

Figure 5: Visualization of the mass detector of the original Velos. [24: p. 46]

The orbitrap mass analyzer works completely differently than the older ion traps. While ion traps and the C-trap are not analyzers per se and always need a detector, the simplest one being a conversion dynode and a photo multiplier. [24: p. 46]

The linear traps work both as a filter and can store the ions or push them in different directions. The two traps are used to store and eject all ions or just a selected mass or mass range. If the ions are to be analyzed with the orbitrap, they are passed to the C-trap. [26: p. 38]

The C-trap (curved linear trap) can also store and eject ions, for this the inserted ions are bathed in a collision gas to negate their kinetic energy. Through an applied voltage to the aperture at the front and back end of the C-trap, the ions can be stored in the middle of the trap and by ramping up the voltage the ions can be concentrated in a thread along the axis. This ion beam can be directed in the orbitrap for further analysis or on a detector. [24: p. 39]

The orbitrap does not work like a trap, a filter combined with an unselective detector, which needs physical contact with the detected ion. The orbitrap utilizes the electronic field of the moved ions. This means that the orbitrap does not need physical contact which keeps the mass spectrometer clean and regular cleaning is no longer a necessity. It also means that the

orbitrap can detect multiple ions at once. Every *m/z* ratio has its own trajectory and frequency, which can be read out by the orbitrap. This can be visualized by simplifying the 3D function to two dimensions and imagining a different cycle frequency, which is described by:

$$\omega = \sqrt{\frac{z * k}{m}}$$

Equation 1: Cycle time and mass to charge ratio.

While *z* is the charge, *m* the mass, *k* a instrumental constant and ω is the cycle frequency. [26: p. 41]

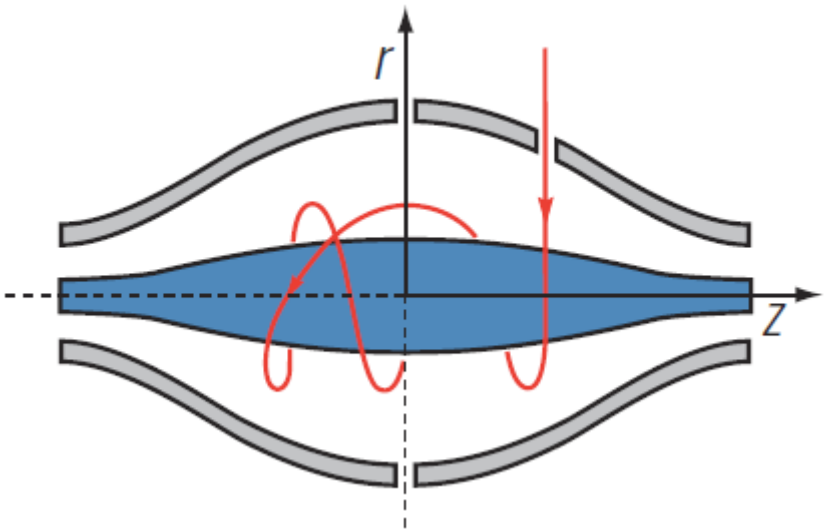

Figure 6: Stable ion trajectory in the orbitrap. [26: p. 40]

Imagine having four different types of ions with different concentrations in the orbitrap. The different m/z values cause different orbiting frequencies, which can be depicted as four simple *sin* functions:

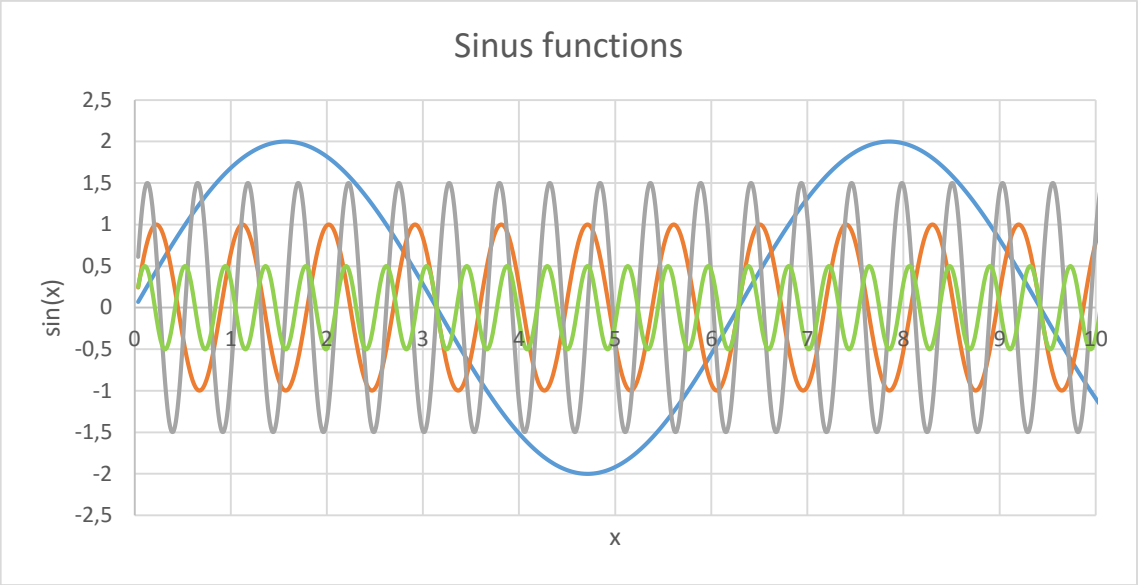

Figure 7: Trajectories of four different ions.

The frequency depends on the *m/z* value and the intensity depends on the number of ions in the orbitrap. In reality the signal from the four different ions merge to a single signal. This signal naturally decays over time and might look like this:

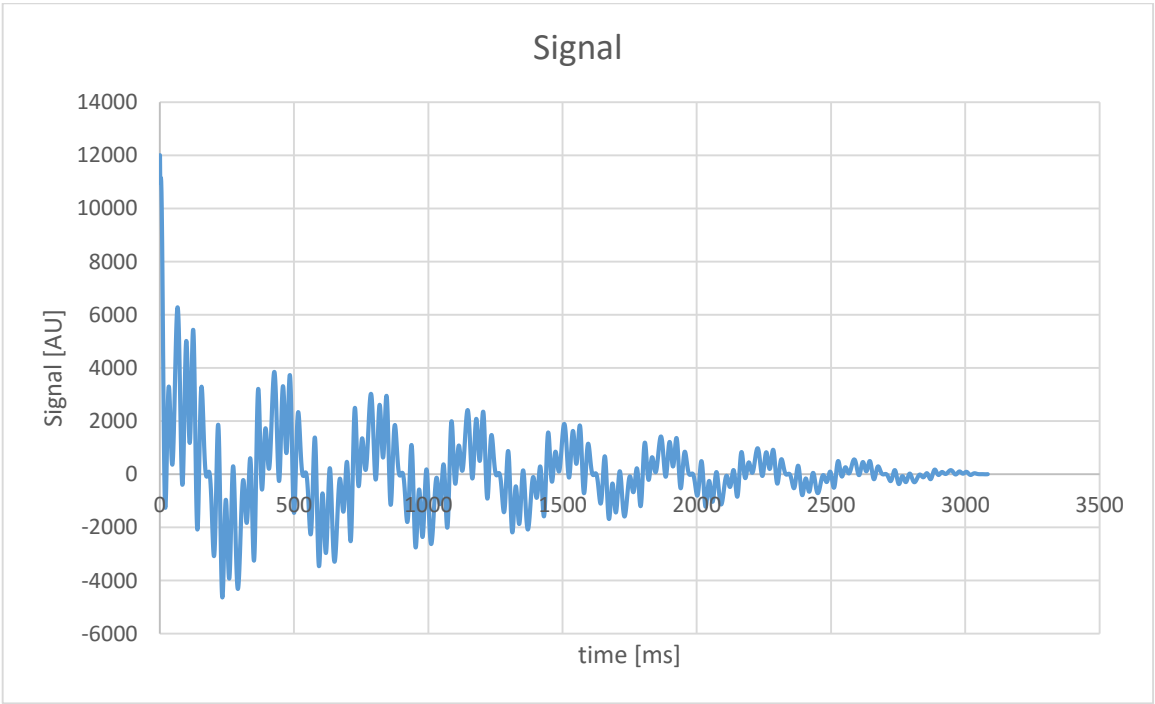

Figure 8: Decaying signal received by the mass detector.

Through a Fourier Transformation the information about the mass to charge ratio can be obtained:

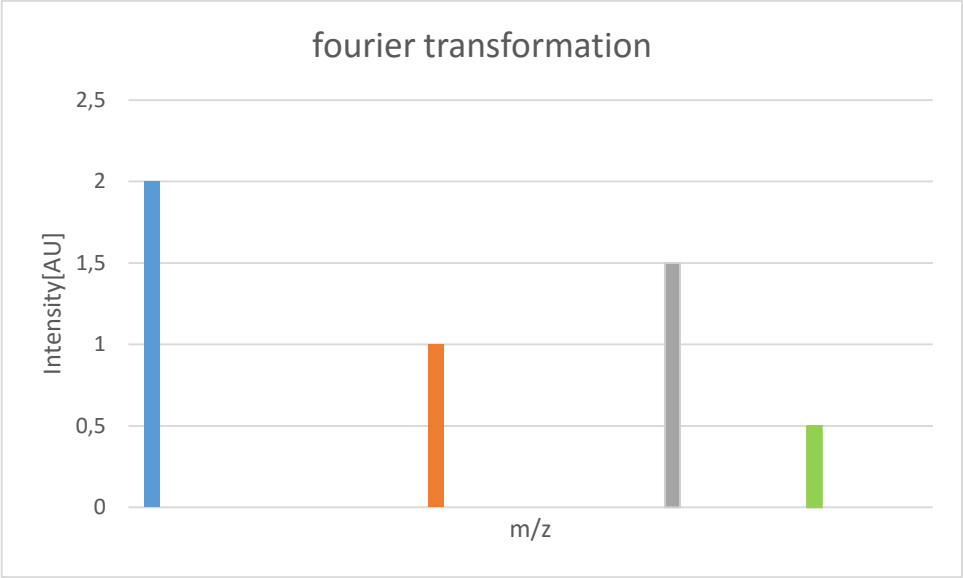

Figure 9: Prompted data after Fourier Transformation.

As hinted in the untransformed signal graph, the resolution of the orbitrap is depending on its acquisition time. While data at the beginning of the scan is very dense, the information density decays over time, so there is no need to measure for longer than a certain time.

2.4 Advion TriVersa NanoMate Pro

The Nanomate is a robotic chip based ESI Ion source. It is very different to a regular syringe based ESI, because of its single-use nozzles and its auto-sampler, which must not be underestimated in practical application!

2.4.1 Schematics and working principle

The samples are placed in a well plate and can be sealed with adhesive aluminum foil. From there the robotic arm with an attached pipette can pre-pierce the aluminum foil, prepare a tip and aspirate the sample.

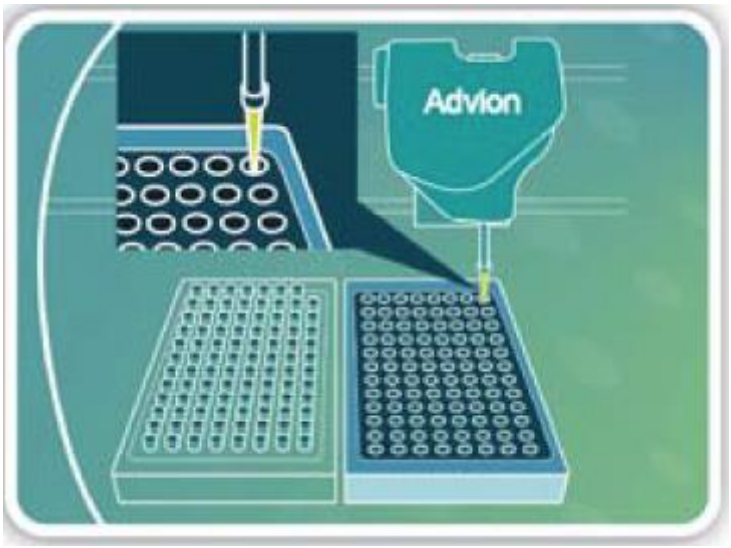

Figure 10: Cartoon shows the tip tray, the robotic arm and the well plate. [29]

The robot can tilt and move the single use conductive tip, here shown in yellow. The tip is pressed against a chip. On this chip are 400 nozzles from which the sample is sprayed one at a time.

Figure 11: Depiction of the spray-nozzle with a diameter of roughly 5µm. [29]

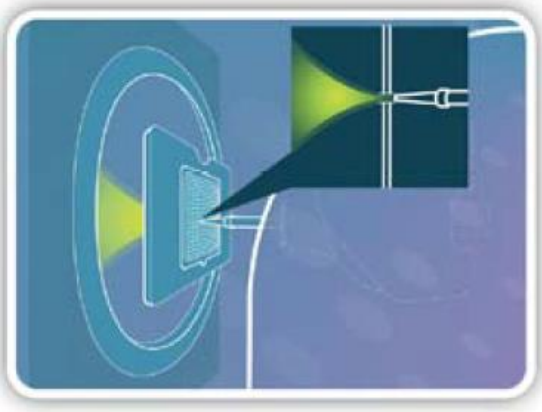

Figure 12: Depiction of the ESI. [29]

The robot can be controlled and programmed by using a computer with respective software called 'ChipSoft'. It is easy to program; a long queue and with the machines ability to cool samples even overnight runs are possible.

2.5 LipidXplorer

After our first experiments it became very clear that a manual data interpretation would be time consuming and subjective, which is of course not scientific. It was obvious and easy to pick this software for data manipulation and interpretation, for it was also used in the papers we were trying to recreate. In this chapter I want to give an overview of the software and its working principle and explain the software settings.

2.5.1 Intro

LipidXplorer is a software created solely to interpret shotgun lipidomics data. The software is freely available and has personal trouble shooting care, however, like most software used in a scientific background it is not free of bugs and not very user friendly. Like mentioned before the data we created when running our experiments was hard to interpret, especially because we wanted to use MS^2 spectra for identification and not only a correct MS1 mass. There we already developed problems: We created several technical replicates of one sample; however, the data dependent acquisition is (not reliably) biased towards peaks with higher intensity; a lower abundant mass will less likely be selected. This leads to the problem that while lipids with high abundance are easily confirmed, lower abundant lipids need many measurements to trigger the fragmentation. Without software it is further impossible to know in which run the desired MS^2 spectrum, of a mass of interest, is.

We then assumed that a certain mass, which was fragmented in one of the many runs and assigned to lipid with the MS^2 spectrum, is present in all samples, if the correct mass appears in a MS1 spectrum, even though it was not fragmented in this specific run. This seems obvious, why should there be a different molecule with the same mass in the same sample? After all these experiments are technical replicates. However, can this be assumed for a different biological replicate or even a different sample? This and many other questions concerning false positives and false negatives usually boil down to a simple statement: Either

it is accepted that some lipids might be false positives (these have to be manually confirmed or refuted) or the employed mindset is so strict, it can only make obvious (and there not helpful) statements.

This is true with manual and automated data interpretation, therefore the LipidXplorer is just a tool, the same outcome can be achieved be just investing an infinite amount of time in a manual, highly regulated data interpretation.

2.5.2 Theory of de novo data interpretation: How LipidXplorer works

The main reason why we chose LipidXplorer for automated data interpretation is that the software does not use a spectral library. In Lipidomics the employed strategy can change, depending on the category of lipids, different protocols are needed to successfully identify an analyte: for instance, polarity of the ESI, fragmentation energy, e.g. hence there are no uniform workflows or SOPs for the whole lipidome. This might not even be chemically possible, for the chemical heterogeneity of lipids.

Further it is challenging to build a holistic library, which is not biased. There are companies and research groups trying to setup libraries and SOPs for Lipidomics, however even if the offered solution seems reasonable the monetary expenditure does usually not. The libraries which are free of usage are usually not holistic, further more it is not how trustworthy the data is. So obviously there has been an effort to create libraries, however a complete and free library has not yet been composed.

The simple solution to this problem is to create an *in silico* library. This is made possible by LipidXplorer, through an easy to use programming language and basic chemical knowledge of the investigated lipids. Basically the user defines the chemical space in terms of estimated elemental composition and double bond equivalents, then the program creates all possible masses and searches the spectra for those created masses. The spectra recorded by the mass spectrometer are **not** used to create this library, they however are used to create a so called master scan.

The program uses several MS experiments and creates one matrix, the so called master scan. This master scan is then searched against the defined in silico database and hits are registered in an automatically created text file.

2.5.3 Creating a master scan

After acquiring mass spectra on the mentioned machines the data is converted from a vendor format in this case .raw to a more accessible file format called .mzML. [3] Then the spectra are summed to a master scan, which is basically a matrix consisting of all the relevant data. Which data is considered relevant is depending on the import options.

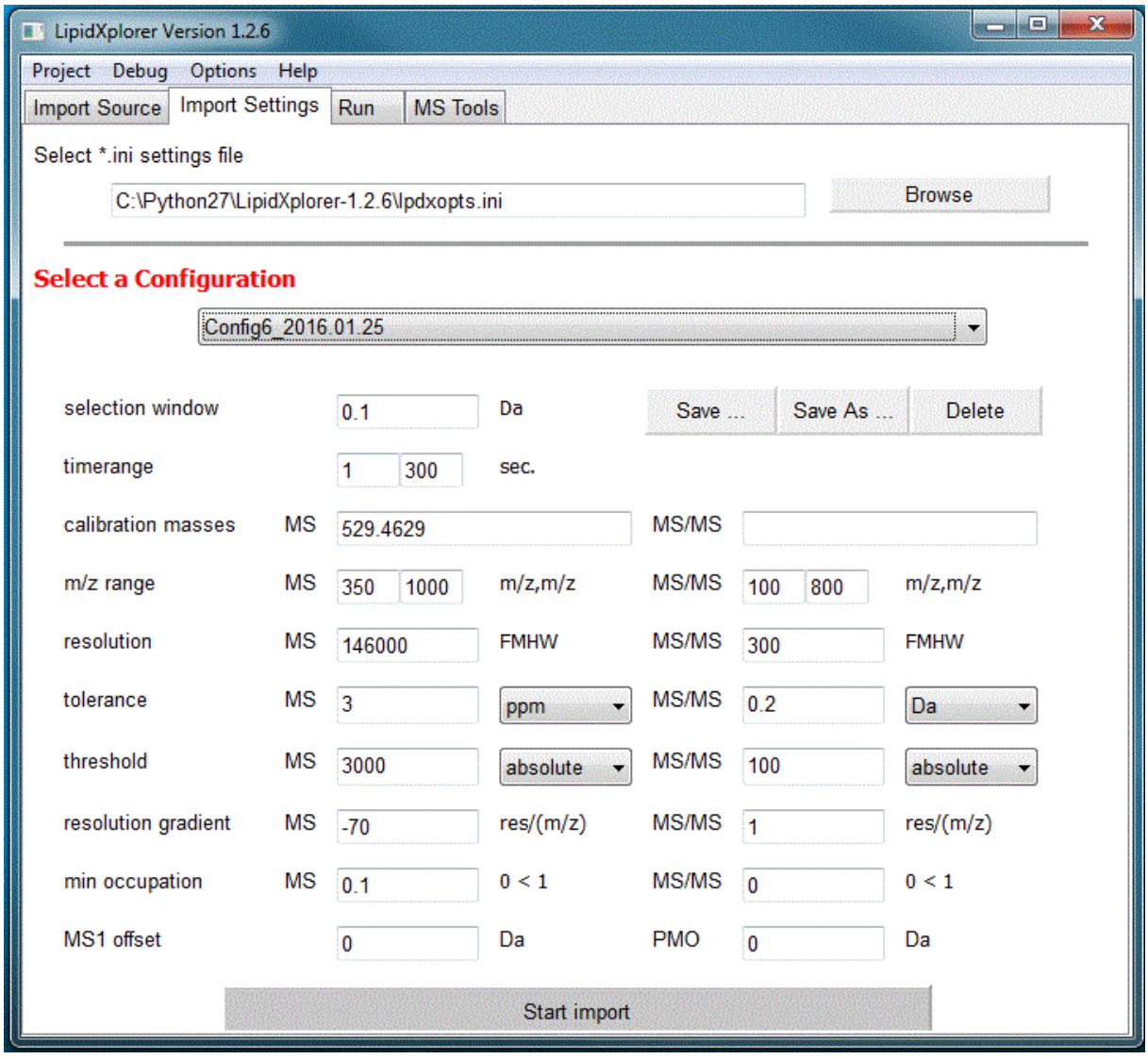

Figure 13: Screenshot showing the employed import settings.

The shown options are usually self-explanatory. For more information, the mouse can be moved to the option and a short explanation will pop up. All of the options are explained in detail in the cited source. The most important options are:

Import Settings:

resolution

tolerance

resolution gradient

threshold

min occupation

The master scan is not only a matrix consisting of the used runs, it reduces the available information drastically. Usually the master scan file is smaller in data volume than some of the files of a single MS-experiment used to create the master scan. This happens because the creation of a master scan is also an averaging and thus contains less data. For the averaging the program matches peaks with the same mass from different spectra. This is a seemingly easy task, which becomes a sophisticated problem at further inspection. The masses of the spectrum are usually specified on six internal decimal places, the last few digits vary and hence two technical replicates are likely to have not a single exact mass in common. This was proven by writing a VBA script which compared spectra and using it on several technical replicates of an arbitrary sample.

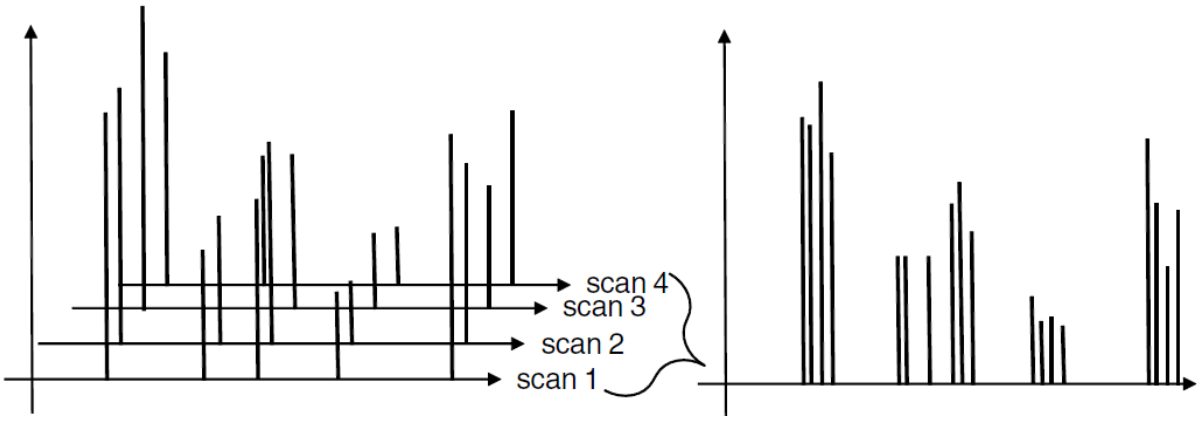

Figure 14: Creating a master scan: All selected MS experiments are combined into one list, which is further processed in additional steps.[9]

First step of a LipidXplorer based analysis is to average spectra and match peaks which belong to the same m/z value (which will further be called mass for simplification). To decide which masses are equal, the program uses the stated resolution and resolution gradient to calculate the resolution of every mass and assumes a Gaussian distribution.

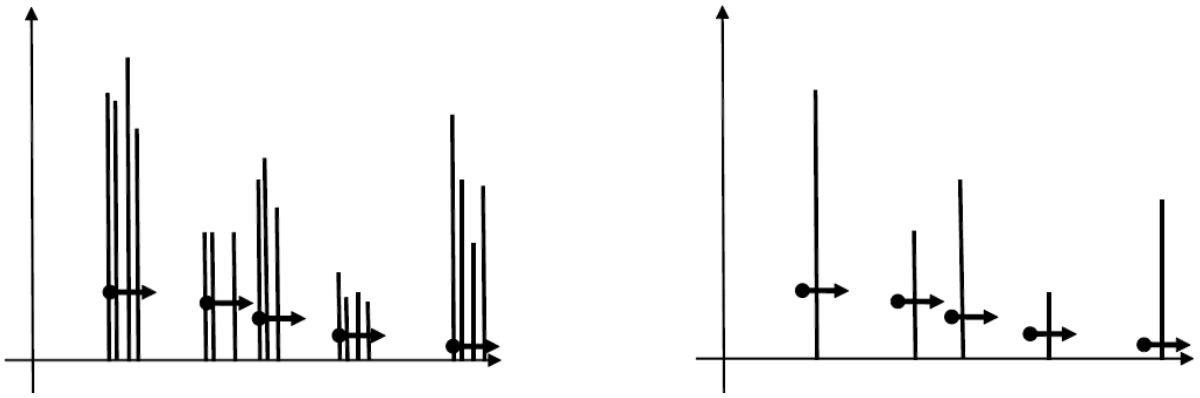

Figure 15: Creating a master scan: The resolution for each mass is calculated, if two masses are not separable by the respective resolution they are averaged. The averaging is weighted depending on the intensity of the peak. This is repeated twice. [9]

It creates a list of all masses and averages all masses which fall into the range of

$$m \mp \frac{m}{R(m)}$$

Equation 2: mass and deviation form a 'bin'.

and their intensities. [9]

The resolution gradient can easily be calculated and visualized by plotting the mass against the resolution and determining the slope via a spreadsheet-program like MS Excel or OpenOfficeCalc.

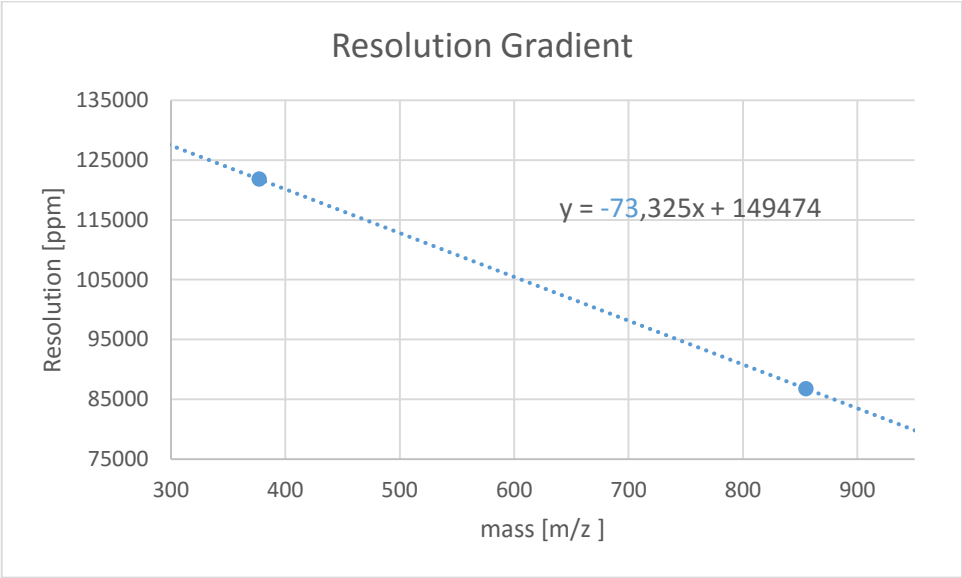

Figure 16: Visualization of the resolution gradient. The resolution gradient was usually calculated in between measurements, while the workflow was still set up. Calculations were done by using one low mass and one heavier mass.

The gradient varies from experiment to experiment most likely because of the condition of the mass spectrometer, so only a rough estimation is possible. Several mass spectra were used to create an average of the resolution gradient, which was then used for all analysis. The tolerance section is a helpful tool, while creating a master scan. Intensities lower than the tolerance value will only be visible in the resulting file if the intensity was higher in at least one experiment. I used 3000[] as a threshold. This lets the user set a threshold and thereby a stringent 'found' or 'not found' criterion. The values for our experiments were carefully selected, using lower values results in many more very uncertain detections, setting higher values will only display a few abundant lipids.

The minimum occupation is also a filtering tool; it simply lets the user specify a percentage of spectra in which a peak has to be present. If a peak is not present in the specified amount of experiments the peak will not be added to the master scan. I set 10%, which might seem redundant but reduced the false positive detection rate. [8]

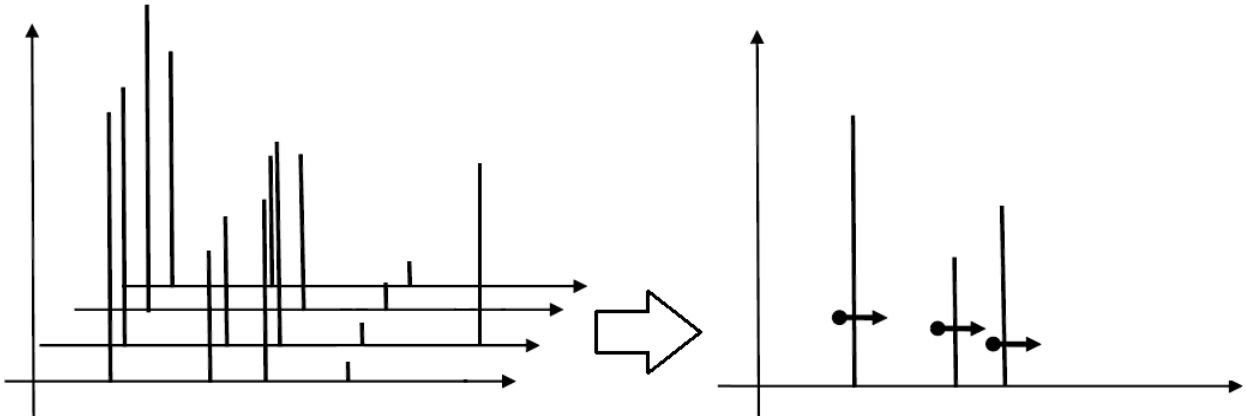

Figure 17: Example for combination of min occupation and tolerance. Two peaks are not present in the master scan, one peak because of its repeatedly low intensity, the other one for its low frequency. [8]

These two simple filters can be combined and are a powerful tool. Further combined with a sufficient amount of technical replicates it helps reduce noise and reinforces the certainty of a detected analyte.

Another handy tool for trouble shooting is the 'dump master scan' option. It saves the mentioned list of masses with matched lipids from the in silico database of the MFQL queries, if any MFQL queries are added. This can be used to prevent systematic mistakes by adapting the import settings and the used MFQL queries. [10]

2.5.4 MFQL – scripts

The molecular fragmentation query language or MFQL is used to program a query and search the generated master scan for defined analytes, usually lipids. Several MFQL scripts were used to setup the workflow of our data interpretation. [9]

The Code in the script is printed at the end of the thesis. All MFQL queries are setup the same way, here I will show two examples and explain the code.

Screening MFQL for Phosphatidylcholines:

```
QUERYNAME = PhosphatidylcholineEven;

DEFINE
PR = 'C[38..54] H[30..130] O[10] N[1] P[1]' WITH DBR = (2.5,9.5), CHG = -1;

IDENTIFY
PR IN MS1-

SUCHTHAT
isEven(PR.chemsc[C])

REPORT
MASS = "%4.4f" % "(PR.mass)";
CHEMSC = PR.chemsc;
ERROR = "%2.2fppm" % "(PR.errppm)";
NAME = "PC [%d:%d]" % "((PR.chemsc)[C] - 10, (PR.chemsc)[db] - 2.5)";
PRECURINTENS = PR.intensity;
;
```

Identification MFQL for Phosphatidylcholines:

```
QUERYNAME = PhosphatidylcholineEven;

DEFINE PR = 'C[38..54] H[30..130] O[10] N[1] P[1]' WITH DBR = (2.5,9.5), CHG = -1;
DEFINE headPC = 'C[3] H[6] O[2]' WITH CHG = 0;
DEFINE FA1 ='C[14..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FA2 ='C[14..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;

IDENTIFY

PR IN MS1- AND
FA1 in MS²- AND
FA2 in MS²- AND
headPC in MS²-

SUCHTHAT

isEven(PR.chemsc[C]) AND
isEven(FA1.chemsc[C]) AND
isEven(FA2.chemsc[C]) AND
FA1.chemsc + FA2.chemsc + 'C10 H21 P1 O6 N1' == PR.chemsc

REPORT

MASS = "%4.4f" % "(PR.mass)";
CHEMSC = PR.chemsc;
ERROR = "%2.2fppm" % "(PR.errppm)";
```

```
NAME = "PC [%d:%d]" % "((PR.chemsc)[C] - 10, (PR.chemsc)[db] - 2.5)";
SPECIES = "PC [%d:%d / %d:%d]" % "(FA1.chemsc[C], FA1.chemsc[db] - 1.5,
FA2.chemsc[C], FA2.chemsc[db] - 1.5)";
PRECURINTENS = PR.intensity;
NLSPIS = headPC.intensity;
FAS = sumIntensity(FA1.intensity, FA2.intensity);
;
```

As shown above the query are two short codes, which feature several segments. Those segments are specialized parts of the matching identification and notation of masses and corresponding lipids.

A few general rules apply in a MFQL script, firstly every line and query has to end with a semicolon (;). Also a number sign (#) can be used to insert comments, because lines starting with a number sign are ignored by the software. [23]

These rules however are very general, for instance there are no need to set a semicolon in the SUCHTHAT or IDENTIFY Section. In these sections lines and indentations are purely cosmetic, the string can be written in a single line and must not have a semicolon at the end. Also the number sign cannot be set anywhere; if used in the wrong places it might render the MFQL script unusable.

2.5.4.1 QUERYNAME

QUERYNAME = PhosphatidylcholineEven;

As shown this is the first line in the queries (without a number sign at its beginning). The number sign is used to add notes in the code, however it might produce an error and should be used with caution. In both codes the name is ‘PhosphatidylcholineEven’. This is because the code will look for masses typical for phosphatidylcholines and only show even numbered fatty acids of the species. Adding spaces or underlines in the name of the query might result in problems, hence the two words are only separated with an upper case at the beginning of each word.

2.5.4.2 DEFINE

Screen:

```
DEFINE PR = 'C[38..54] H[30..130] O[10] N[1] P[1]' WITH DBR = (2.5,9.5), CHG = -1;
```

Identification:

```
DEFINE PR = 'C[38..54] H[30..130] O[10] N[1] P[1]' WITH DBR = (2.5,9.5), CHG = -1;
DEFINE headPC = 'C[3] H[6] O[2]' WITH CHG = 0;
DEFINE FA1 ='C[14..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FA2 ='C[14..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
```

In this segment the analyte is defined. The number of carbon atoms and hydrogen atoms in Phospholipids vary because of different fatty acids in the molecule. This issue is solved by not defining a certain number but an upper and lower limit of said atoms. However, the number of oxygen atoms, the nitrogen atoms and phosphorus atoms does not vary. This enables the program to generate a library of possible sum compositions, which are named by the user, in this case the precursor mass is called `PR`. In the identification query a few more sum compositions are defined: a neutral loss fragment is called `headPC` and the fatty acids are called `FA1` and `FA2`. These fragments are later looked for in the MS^2 spectra. All fragments which will be looked for in any MS^2 spectra of the right mass must be defined here. The program looks for all those masses in the master scan. [8] Because of the high resolution of the used mass spectrometer and the specific exact mass the exact mass of the sum composition can be used as a first lead for lipid identification. Further it is possible to distinguish chemically similar lipids by assuming that every lipid category ionized by a distinct way, for instance by adding an acetate ion or losing a proton. [18]

2.5.4.3 IDENTIFY

Screen:

```
PR IN MS1-
```

Identification:

```
PR IN MS1- AND
FA1 in MS²- AND
FA2 in MS²- AND
headPC in MS²-
```

In this segment the interaction between the master scan and the defined sum composition happens. Here it is possible to let the program search for masses and match possible sum compositions. It is further an option to not only search MS1 masses, but also fragments in

the MS^2 spectra. The mathematical symbols after the MS1 or MS^2 are used to tell the program the charge of the searched ions. [9]

The screening script only looks for masses of analytes in the MS1, this is especially helpful when the background of medium of cultured cells needs to be determined or when only a few spectra are available. The identification script will only accept MS1 peaks with matching mass as hit, if a MS^2 spectrum with all defined fragments are present with a defined frequency and intensity.

This is one of the most powerful features of the program. As explained before MS^2 spectra are triggered data dependently which can seem random at some times. The mentioned feature finds single MS^2 spectra out of several hundred runs. Further the intensity of the defined fragments can be printed in a resulting text file, which makes it easy to distinguish useful and noisy or void spectra.

The 'define' and 'identify' section are the main part of the script.

2.5.4.4 SUCHTHAT

Screen:

```
isEven(PR.chemsc[C])
```

Identification:

```
isEven(PR.chemsc[C]) AND
isEven(FA1.chemsc[C]) AND
isEven(FA2.chemsc[C]) AND
FA1.chemsc + FA2.chemsc + 'C10 H21 P1 O6 N1' == PR.chemsc
```

This part of the script is optional and can be left out at first when composing a new script and then added later, for it is not a necessity. [8, 9] Here further constraints for the identification are applied. In the screening script the only constraint is that the number of carbon atoms in the molecule is even, which excludes phosphatidylcholines containing a single oddly numbered fatty acid. When using more constraints in the identification script, it is very important to write the script in a way that will sum up all found fragments and compare it with the precursor mass, else a lot of false positives are found. This is done in the last line of the SUCHTHAT section; the detected fatty acids combined with an invariable part of the molecule must produce the sum composition of the precursor. Thus the double equal

signs (like used in the Python language) which result a TRUE if correct. The PR.chemsc calls the function chemsc is applied on the PR, which again returns the chemical composition.

Further a function can be used to only show evenly number fatty acid chains. Especially in a script that only uses MS1 masses for quick screening of possible analytes many sum compositions will be prompted, which are illogical or unlikely. Therefore, the function isEven, which is a simple function and checks if a number is even and returns a TRUE if it is. The different restrictions are combined with logical operators of the Boolean algebra. When connecting all statements with an AND all of them must be TRUE to pass the test.

In this segment the identification can be tuned and made more stringent. This is necessary, because in the DEFINE segment the carbon atoms, hydrogen atoms and double bond equivalents might be defined loosely and without further filtering many chemically illegitimate structures might be prompted. [15]

2.5.4.5 REPORT

Screen:

```
MASS = "%4.4f" % "(PR.mass)";
CHEMSC = PR.chemsc;
ERROR = "%2.2fppm" % "(PR.errppm)";
NAME = "PC [%d:%d]" % "((PR.chemsc)[C] - 10, (PR.chemsc)[db] - 2.5)";
PRECURINTENS = PR.intensity; ;
```

Identification:

```
MASS = "%4.4f" % "(PR.mass)";
CHEMSC = PR.chemsc;
ERROR = "%2.2fppm" % "(PR.errppm)";
NAME = "PC [%d:%d]" % "((PR.chemsc)[C] - 10, (PR.chemsc)[db] - 2.5)";
SPECIES = "PC [%d:%d / %d:%d]" % "(FA1.chemsc[C], FA1.chemsc[db] - 1.5,
FA2.chemsc[C], FA2.chemsc[db] - 1.5)";
PRECURINTENS = PR.intensity;
NLSPIS = headPC.intensity;
FAS = sumIntensity(FA1.intensity, FA2.intensity);;
```

This section does not influence the found hits in any way and is only used to prompt the filtered data. However, if notes are added by using the number sign between the lines, the MFQL query might not work. It is important to name all the desired information alike in every MFQL script. If this is not done, the resulting txt file will be confusing. The resulting comma separated values file will be in a chart, in which the columns represent the respective runs while the lines are the prompted data defined in the REPORT section:

MASS: mass of the analyte.

CHEMSC: the chemical composition.

ERROR: the average error of the analyte in ppm, this might depend on the searching options used later on and the prompted value is usually too high. For this reason, we accepted a deviation of up to 7.5ppm.

NAME: The name of the lipid with the sum of the carbon atoms of the fatty acids and double bonds.

SPECIES: Similar to name, but the individual fatty acids with their number of double bonds are prompted.

PRECURINTENS, NLSPIS, FAS: the intensity of the precursor in the MS1, the neutral loss fragment and the sum of the two fatty acids of the MS^2. Here it is important to note that the MS1 intensity is an integral, hence the value obtained cannot be compared by opening the spectra and estimating the peak height. The MS^2 peaks are centroid, which means their integral is essentially the same as their height, hence these peaks can easily be manually compared.

The way the data is processed is similar to the python programming language, for instance:

```
DATA =“Number of x: %4.4f” % “x”
```

This string will insert the number, which is represented by `x` in the text that will be stored in the `DATA` variable. This variable will later be prompted in the resulting txt file. So in the first line of the code for instance, the number stored in the variable `PR.mass` is saved in the variable `MASS` and printed with four floating numbers.

After the txt file is created it can easily be opened with a word processing program like WordPad.

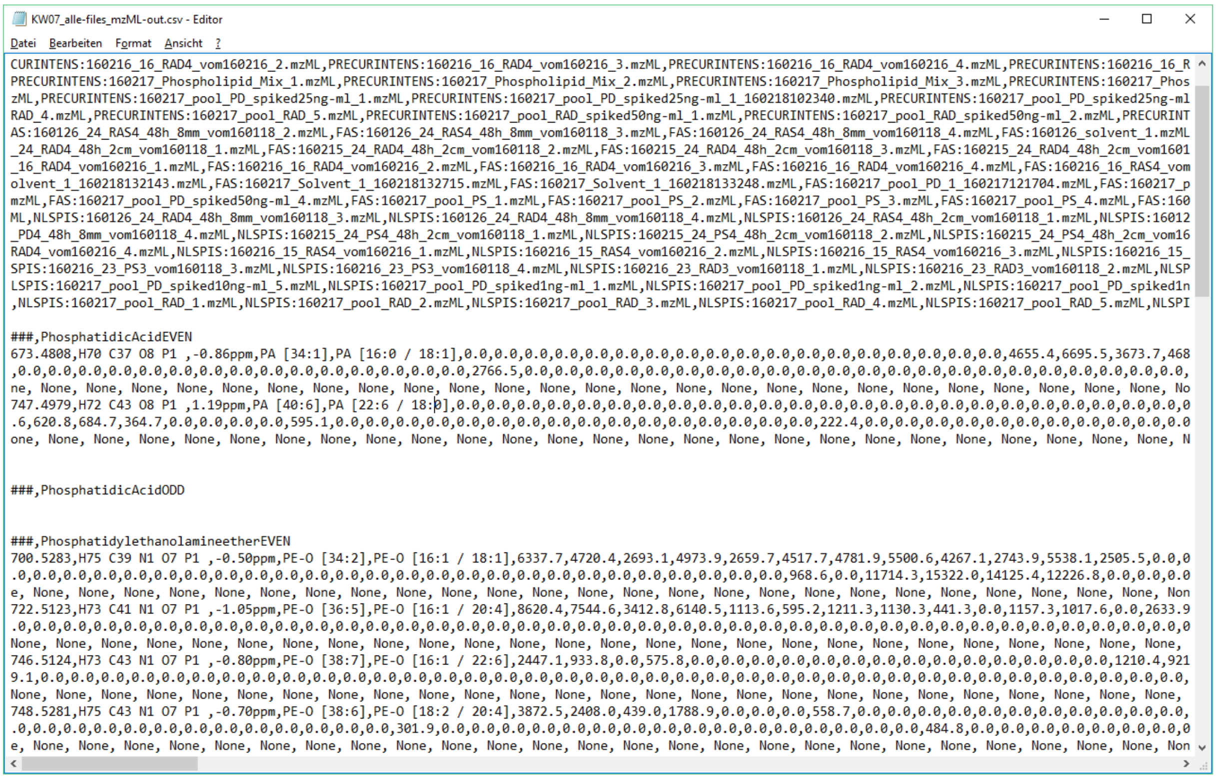

Figure 18: Output file of LipidXplorer opened with WordPad.

The same file can be dragged and dropped in any spreadsheet program or imported with respective options in the program, where the data can be displayed more clearly.

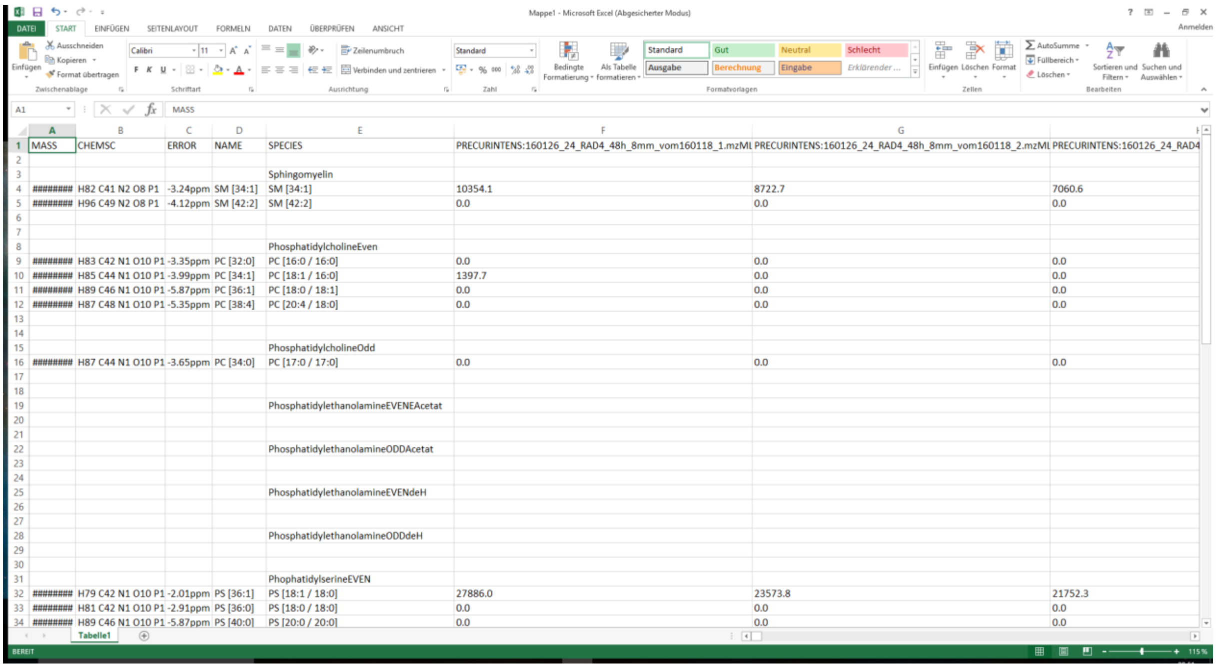

Figure 19: Output file in a spreadsheet program.

2.5.5 Source and creation of the MFQL files

When downloading and installing LipidXplorer a variety of MFQL scripts are already present. If at all only minor modification are needed to make the MFQL script run on LipidXplorer. Here we firstly measured MS1 and MS^2 spectra of our standards and interpreted MS1 spectra and MS^2 fragment peaks manually. Usually one to three fragments of a correct mass in high resolution measurements were chosen as an identification criterion. The already downloaded MFQL scripts were then used directly or modified on measurements of the standards. The used standards are bought from Avanti:

12 LPC

14 PE

17 PC

17 PG

17 PS

These were measured in our first measurement mixture of 70% methanol like mentioned before. After spraying, measuring and creating a master scan with the shown settings a list of lipids was attained. The detailed list and interpretation will be discussed in another chapter. We were able to find all standards with the identification script. Only the 12 LPC standard was not found with the identification script, however the screening script found said standard too. [23]

2.5.6 Critical remarks

LipidXplorer is a useful program, if used correctly it saves time and offers a solution to interpret and manipulate many spectra at once. Our assignment would have not been as successful without LipidXplorer. However, after working for several months with the program I noticed several shortcomings.

Firstly, there are just a few rough bugs, which can be worked around. Some examples are:

- The program would identify a peak for a lipid and using an MS^2 from another run where said peak was not present. In this case there is one run with a correct MS1 mass and one run with a usually noise MS^2.
- The built in MFQL editor crashes easily and was not really used to edit the MFQL queries.
- Changing the QUERYNAME or the REPORT section sometimes renders an MFQL unusable, even though it should be possible to change this parts of the program. For a person without any computer science background, it can be completely inapprehensible why and how the program or the MFQL stops working.
- The deviation of a peak in the MS1 from the correct mass is calculated depending on the accepted deviation. This is the reason I used a startling 7.5 ppm as mass accuracy threshold, the LTQ Velos orbitrap is supposed to measure accurately up to 1ppm at 400 Th. [26: p. 30] However, peaks which might be shown in LipidXplorer with a deviation of 5ppm would usually not deviate over 1.5 ppm in spectra. This might be related to the 'profile data type' of the MS1 spectra. Also when a master scan is searched twice with different mass precision threshold, the LipidXplorer would sometimes find the same lipids in both searches with a different mass deviation.

I already mentioned that the only general rules according the number sign and the semicolon are not really general. This information took me days or even weeks to find out, because after reading the tutorial I was relying on this information. Compiling a MFQL script can be difficult, not because the language is hard to understand or use, but because there is no advanced and user friendly editor and compiler. If there is a mistake or error in the code, the error message prompted by LipidXplorer does usually not explain the mistake or where it occurred. User friendliness seems to be a general issue. I would have denied this statement. However, I was not able to write a MFQL script for not only screening but identifying LPC. I was very confused, because the MFQL script would find 12 LPC in some samples but not in the standard. I frequently spent several days every month, hoping my knowledge would have advanced. I contacted the author of the program, he helped me by explaining the 'dump master scan' option which was helpful. [10] However I was not able to find or create a reliable MFQL script for LPCs.

Unfortunately, it is also not trivial to setup the program, python needs to be installed and several partly unofficial packages for python need to be downloaded and installed. This is

difficult enough for some users to take LipidXplorer not even into account when trying to setup a lipidomics method. Another Problem might be that knowledge about chemistry and mass spectrometry is crucial to use the program. The advantage of defining a spectral library out of estimated fragments comes with the price, to be able to estimate where molecules might break after activation.

However, I enjoyed working with the program very much. Compared to modern proteomics software like Thermos Proteome Discoverer or MaxQuant, LipidXplorer offers less 'black boxes' and fewer options. Hence it is very easy to follow the programs processes, especially through the 'debugging window' which always shows what the program is calculating in real time. This is also the reason why the explained bugs and shortcomings were not too bad, there is always an option to look into a problem and find the source. Even if this is not possible, at least the problem can be identified and taken into account or worked around.

3 Materials and Methods

This chapter describes the details of our shotgun Lipidomics approach, which features an MTBE extraction. We further optimized spray conditions for our ESI, both in means of settings and spray solvent and settings for the LTQ Velos orbitrap.

3.1 Amniotic cells and cultivation

Amniotic cells were obtained from four Donors, two tissue samples were collected, one at the placenta associated area and one at the 'reflected area'. The biopsies were respectively cultivated 3-4 days with 2 different media, SAGM and DMEM. DMEM is Dulbecco's Modified Eagle Medium by Sigma, while SAGM is small airway growth medium purchased from Lonza.

3.2 Standards

Standards were obtained from Avanti Lipids:

- [17:0 / 17:0] Phosphatidylcholine
- [17:0 / 17:0] Phosphatidylglycerol

- [17:0 / 17:0] Phosphatidylserine
- 14:0 Lysophosphatidylethanolamine
- 12:0 Lysophosphatidylcholine

3.3 MTBE Extraction

The MTBE extraction was chosen because it is, alongside the Folch Method, a gold standard of lipid extraction. Several papers were published, which proofed that the extraction efficiency is well above 90%, which is sufficient, given that we were only aiming to provide a semi quantitative approach. Being a gold standard we did not further investigate the extraction efficiency, especially after the pig lavage measurements turned out as predicted in other studies.[13, 16]

3.3.1 Workflow

The MTBE extraction is performed by pipetting 400µl of supernatant of the amniotic cells in a 2ml Eppendorf tube. 320µl of 97% MeOH is added to the sample, after adding 1ml of MTBE the mixture is incubated for 1h at 24°C in a thermomixer at 1400rpm. 250 µl of MS grade water are added to induce the phase separation. After another incubation of 10 min with the same settings as bevor, the mixture is centrifuged for 1 min at 4000rpm. The 800µl of the upper layer organic phase was collected carefully.

The sample was measured using a Triversa Advion Nanomate coupled to a Thermo LTQ Velos Orbitrao. 8µl of sample were mixed with 80µl of measurement mix in the wells of a 384 well plate bevor the ESI-MS-MS analysis. Later an Eppendorf 96 well plate in combination with adhesive aluminum foil was used.

3.3.2 Used Equipment

Here is a list of the used wet-chemistry equipment:

- DMEM by Sigma
- SAGM by Lonza
- 97% MeOH:

1. Water: MilliQ Integral5+ with an attached Millipak Express 40 filter
2. Methanol: HiPervSolv (LC-MS-grade)

- 1ml MTBE
    1. MTBE: Riedel de Haën
- Thermomixer: Eppendorf Thermomixer comfort
- One minute of centrifugation at 4000rpm: Centrifuge: Eppendorf Centrifuge 5424
- Eppendorf well plates:
    1. 396 well plate
    2. 96 well plate

3.4 Measurement Mixture

The measurement mixture was empirically determined to be an adequate solvent to assist the electron spray ionization. It consists of 70% MeOH, 30% $CHCl_3$ and 7,5mmol NH_4-$OOCCH_3$.

For the determination we consulted literature: [17]

Isopropanol / methanol / chloroform in a mixture of 4 / 2 / 1 and a concentration of ammoniumformate of 7.5 mmol.

However, we were not able to reproduce the method and did not achieve a stable spray. A colleague advised us to only use a mixture of methanol and chloroform. We still adopted the concentration of ammonium formate but used ammonium acetate instead. We prepared several mixtures of methanol and chloroform:

Chloroform [%]	Methanol [%]
0	100
10	90
20	80
30	70
40	60
50	50
60	40

70	30
80	20
90	10
100	0

Table 1: Chloroform and methanol mixtures.

Resulting in the following spray stabilities:

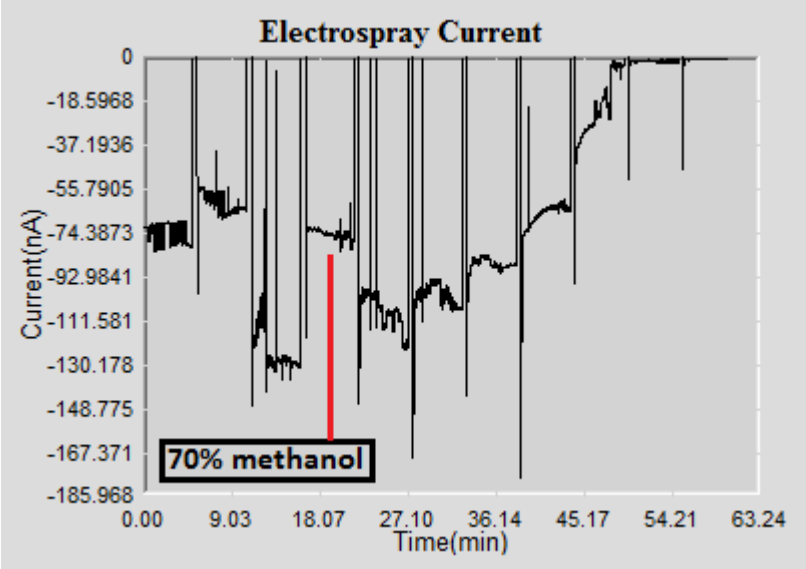

Figure 20: Spray-current and spray-time show spray stability.

It is obvious that the different mixtures result in a different spray stability, as shown in the graphic. The spray stability experiments are separated by a termination of the current, followed by a peak in current. A high current is not a requirement for a good ionization, however a stable spray is. Further chloroform is known to extract additives from the used plastic tools and also the well plate in which the sample is pipetted before spraying, a low chloroform concentration is an advantage. Considering these criteria, we decided to use

70% methanol 30% chloroform mixture with 7,5mmol of ammonium acetate

for our measurements.

Later on we had problems with our ESI source, the Advion Triversa Nanomate, which will be described later on. We contacted a technician, who was able to further improve our spray stability. He helped us fine tune the settings for the Triversa Nanomate, which will be discussed in detail in another chapter. We learned that 7.5mmol were too many ions for negative mode, concentrations in this area are usually associated with positive ionization

methods, which we did not employ. This was the reason for not achieving a stable spray with said 4/2/1 three component mixture. Further the technician told us to treat our measurement mixture with ultrasonic to eliminate spray instabilities caused by air bubbles. Another practical feature is the pre piercing, to write long sequences it is possible to apply an adhesive aluminum-foil onto the well plates, which were also not perfect for our method. One of our most intense noise peaks (473,22Th) was dissolved from the well plate, after switching the well plate and using the aluminum foil the peak was mostly gone. This foil can be pierced by the robotic arm bevor a tip sucks the sample, this conserves the sample and minimalizes the evaporation of solvent. Another useful feature is the cooling function of the Nanomate; by cooling the samples evaporation of solvent is further reduced which renders longer sequences with more runs possible, here we were advised to cool to 10°C instead of lower temperatures to avoid water condensation in the respective wells. By far the most useful improvement was the knowledge about a more suitable solvent mixture. We followed the advice of the technician and changed our mixture:

Isopropanol / methanol / chloroform in a mixture of 4 / 2 / 1 and a concentration of ammoniumacetate 0.75 mmol.

This did not change our results dramatically, the mass spectra with our empirically devised mixture of 70% methanol were not easily distinguished from the 4/2/1 three component mixture. However, the spray stability was increased.

Here are graphs representing the change in spray stability. First the stability was not good:

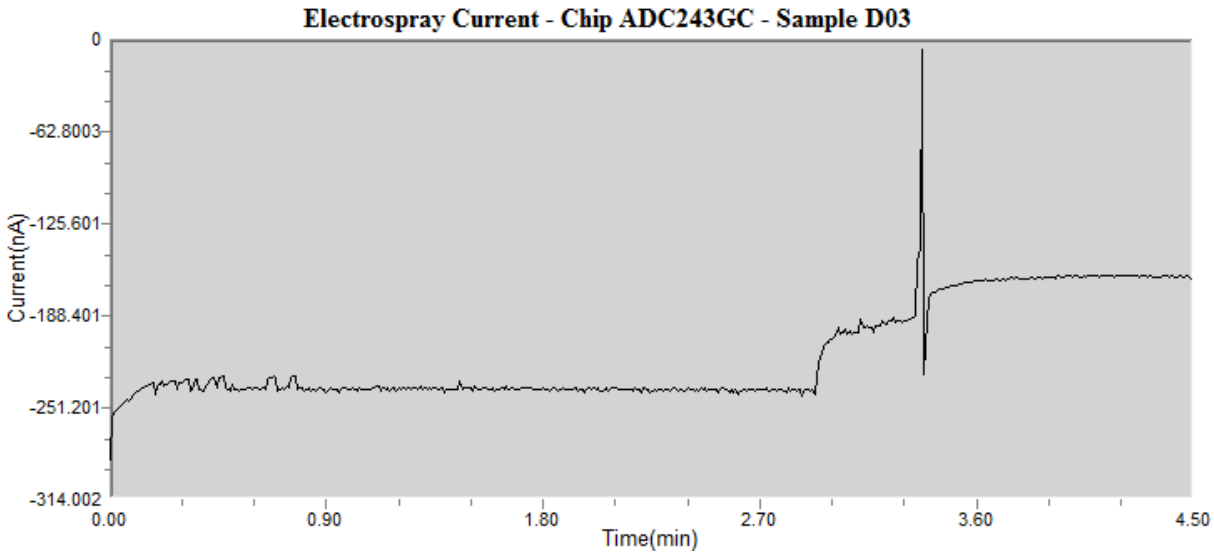

Figure 21: This sprays stability is representative for most measurements. Instabilities were frequent and a current of over 200nA was usually observed.

The settings here were 0.3 psi backpressure and a spray voltage of 1.4kV. The used solvent was the 70% methanol, 30% chloroform and 7,5 mMol ammonium acetate mixture.

Here is a graph showing a typical experiment after the improvements:

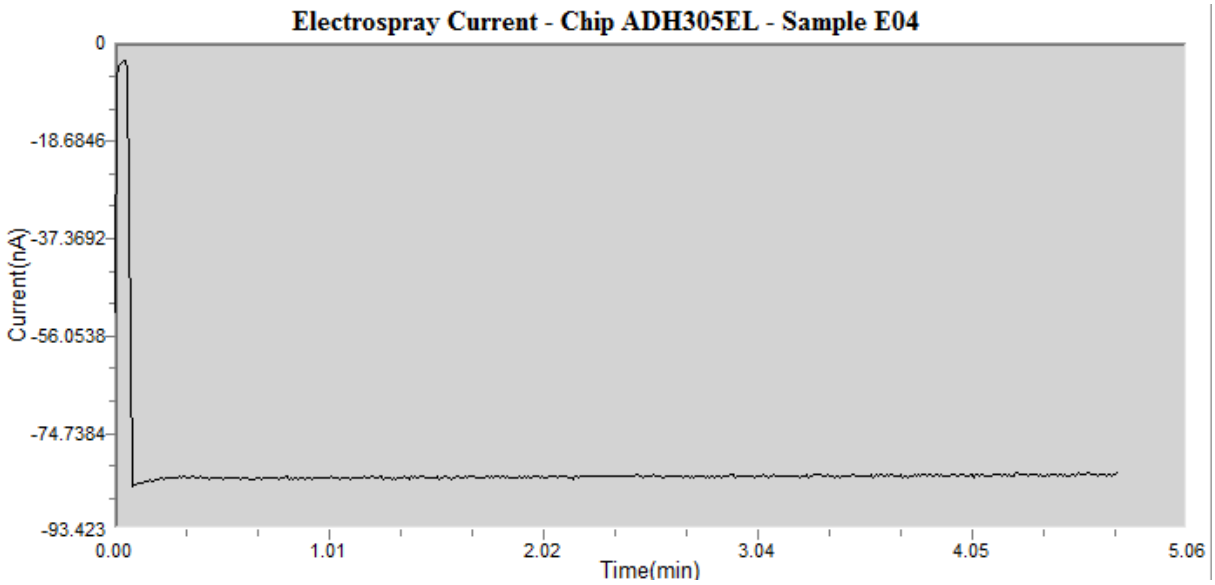

Figure 22: After applying the changes the spray was not only more stable, the current was also lower.

The settings changed to 1.1 psi and 1.1 kV, while a 4/2/1 mixture of isobutanol, methanol and chloroform with a concentration of 0,75mMol ammonium acetate was employed.

We used the empirically devised 70% methanol mixture for most of our measurements. Every mixture was prepared by only using glass equipment. When storing the mixture for at the most 3 days at 4°C an Erlenmeyer flask with a grinded preferably tight bottleneck was additionally sealed with Para-film.

For the measurement, 8µL of the MTBE extract were mixed with 80µL of the measurement-solvent-mix within the well plate. A thorough mix was achieved by pipetting up and down multiples times after adding the sample extract.

3.5 Mass spectrometer settings and how settings changed our amount of detected lipids

Here I will describe our negative-mode-only MS^2 method. We used a data dependent acquisition, because our approach was to screen for a variety of unknown lipids. The MS

spectrum was acquired using the orbitrap, the window was set from 350 Th to 1000 Th, the Normalized Collision Energy was 35.0[] and the isolation width was 1 Th.

This is might be why even though masses of frequent fatty acids (16:0 at 255 Th and 18:0 at 283 Th) are present in many MS^2 spectra, the isolation window is rather broad and there is a good chance that lipids with similar masses exist near lipids which do not feature the mentioned fatty acids.

The MS^2 settings took some time for finalization. It is important to fit the settings to the analytes, when we started measuring the dynamic exclusion was set badly. The window of the masses excluded was too big, hence we got only a few MS^2 spectra per run. After narrowing the window to 0.5 Th we got more MS^2 spectra per run, which resulted in a rise of detected lipids by the employed software LipidXplorer. We also wrote a parent mass list and enabled the '*most intense peak if no parent mass found*' option, however the parent mass list was usually ignored. The way the mass spectrometer selects masses for fragmentation can be seemingly arbitrary, our approach to counter its randomness was to produce a lot of technical replicates. We only analyzed single charged ions and fragments, for our analytes are not multiply charged.

The data was recorded with a profile data type in MS1 and a centroid data type in MS^2.

We tried to change the settings of the mass spectrometer several times, we tried different voltages for the S lens, different acquisition times, programmed a parent mass list, shortened and lengthen the time for dynamic exclusion, changed the collision energy and even tried a delayed data acquisition because of the instability of the ESI source.

The only real difference was achieved by narrowing the exclusion mass list-window, like mentioned before. Another way to improve mass precision is to use a freshly cleaned mass spectrometer; we usually did not mind when our measurements were delayed by the technician, because after she cleaned the mass spectrometer, we got nicer spectra and more MS^2 spectra, which are critical for the automated interpretation software.

3.6 ESI method development

As explained before, we were working with three different settings. Our first method featured measurement mixture, which is a mix from several organic solvents with a set ion concentration to aid the sample ionization. The sample is solved in MTBE and mixed with the measurement mixture in the well plate. The first measurement mixture was a methanol – chloroform mix, which not only evaporated fast but also introduced ghost peaks, through the chloroform and the inevitably used plastic components (Eppendorf-pipette-tips and the well plates). The spray was unstable and sometimes stopped abruptly, which terminated the measurement and wasted time. The 30sec of contact closure (starting the measurement) was meant to aid spray stability, however as we discovered did nothing of that sort. There was no pre piercing, hence the method was not fit for longer measurement queues.

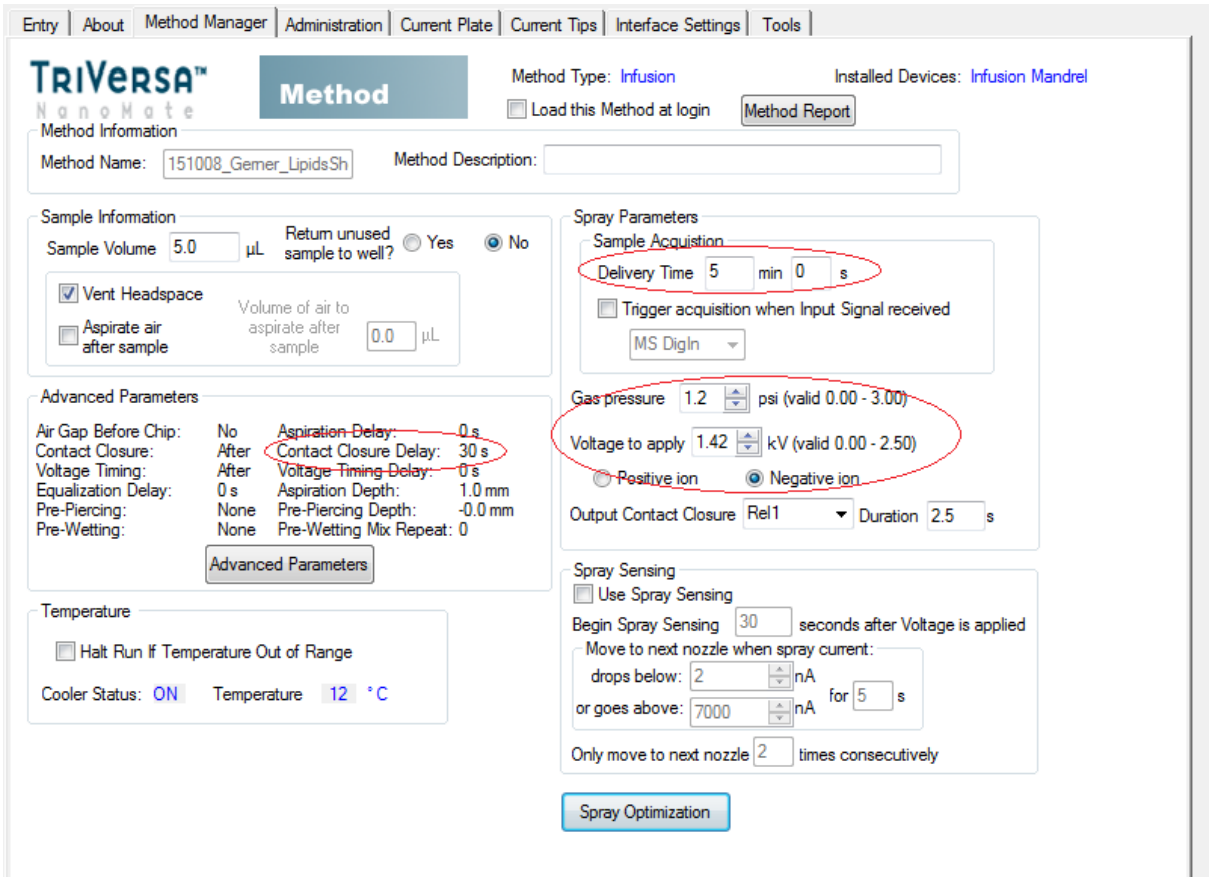

Figure 23: The first method, it was empirically derived.

After a vacation we tried to measure, however our method would not yield any results, because the spray was not stable anymore. We then derived our second method:

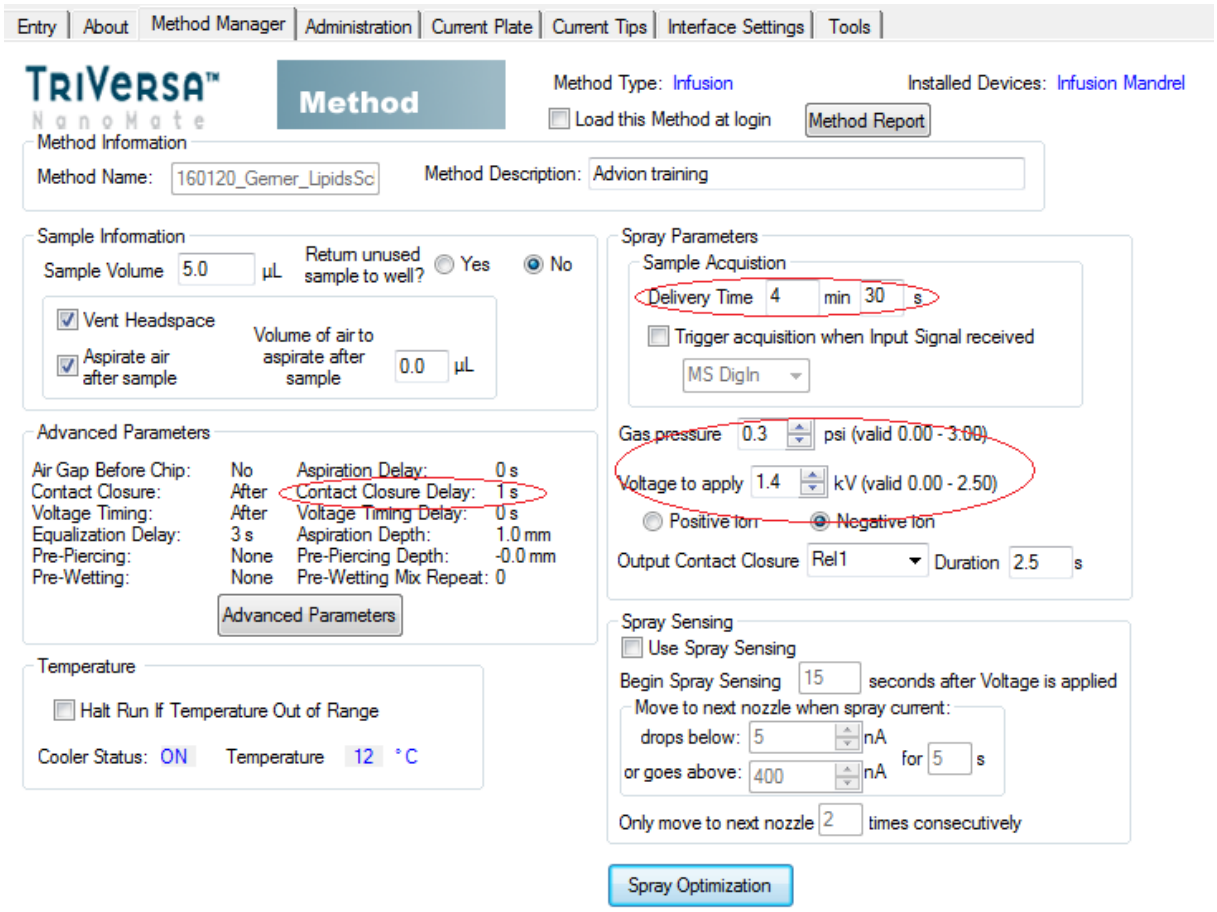

Figure 24: Our second method was only employed for a short time.

The gas pressure and voltage was changed until the spray was stable again. We also changed the useless spray stabilization time and shortened our measurement time. The fact that the first method worked fine for several weeks and just stopped working suddenly had us worried, a technician was contacted. Said technician knew the paper we tried to recreate and worked with the research group. He then helped us by creating a method tailored for our samples:

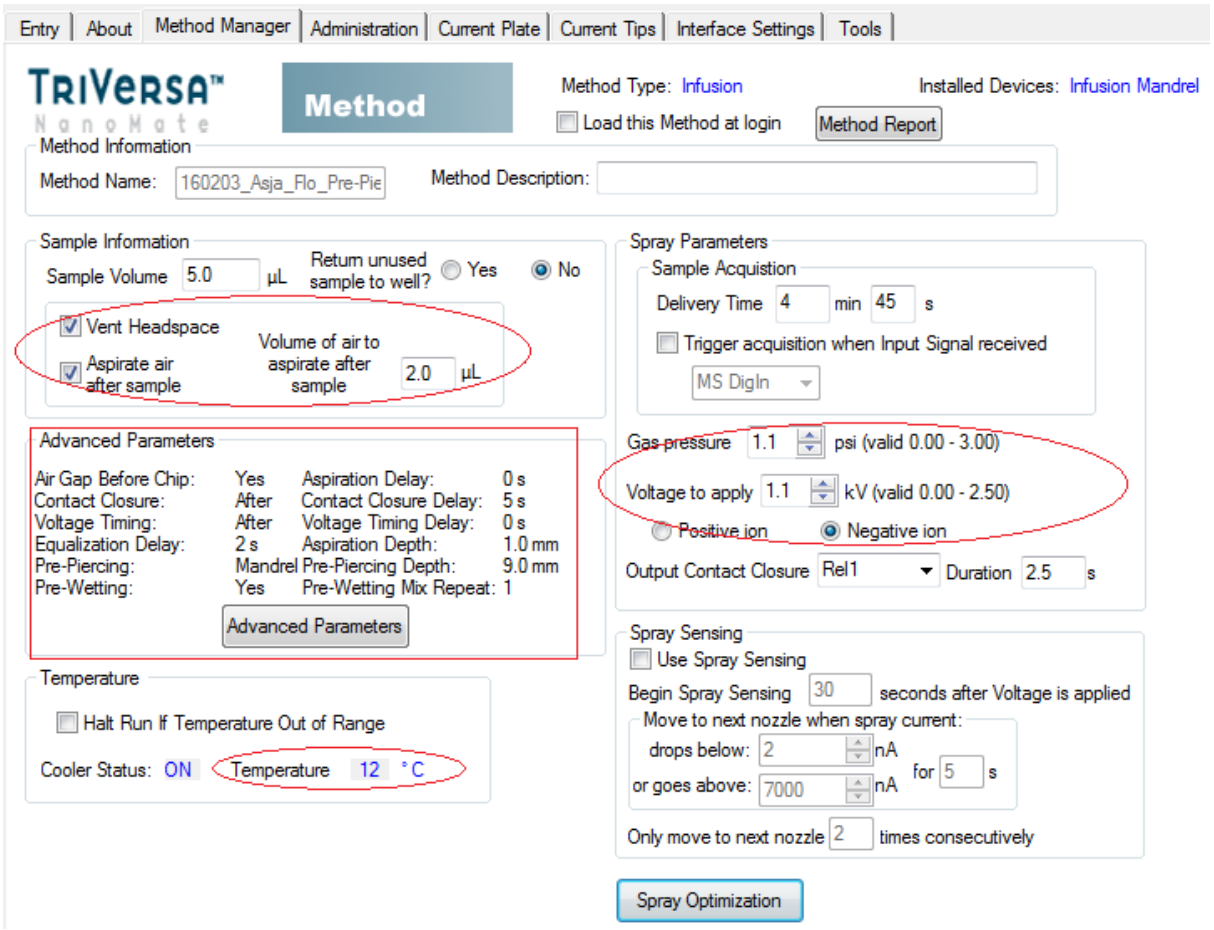

Figure 25: Third and final method.

After switching to a better measurement mix, which was described before, and a more suited well plate, we also optimized the method. The spray parameters were changed again, the method became longer, and the gas pressure and voltage were adjusted. Further improvements were the 'Vent Headspace' and 'Aspirate Air after sample' which helped to avoid peak voltages at the beginning and end of the measurement. Especially helpful were the settings in the 'Advanced Parameters' section. The technician showed us the aluminium foil and the pre pierce option; by telling the program to use the 'Infusion Mandrel' (the pipette on the robot) to pierce the foil, we were able to write long queues for measuring. We also introduced pre wetting to be sure that the mixture in the well is mixed and that a reproducible amount of liquid was taken from the well.

The changes resulted into a highly reproducible method, which not only eliminated prematurely ended MS experiments due to spray instability but also enabled us to combine the machines features to write several hour long queues. The first method was time

consuming and it took a week to generate enough replica to create meaningful results. In the end, we were able to extract samples and measure those samples repeatedly in a single day.

4 Results and Discussion

In this chapter I will present our results. Firstly, I want to show how we used standards with well-defined chemical structures to gain MS^2 spectra. The knowledge of the chemical structure, the way of ionization and the pattern of fragmentation was later used to write a species specific MFQL-code. We then critically assessed our method by measuring well known samples of pig lavage. We tried to be as critically as possible!

Lastly we used our derived method on unknown samples of amniotic-cell supernatant. All of the MFQL-scripts, which were deemed fit for purpose were used to discover the content of our samples.

4.1 Shotgun lipidomics of pig lavage

After we got to know the extraction method, the electro spray, the mass spectrometer and the program, we needed to test our method. While we were already measuring amnion samples we were not able to find literature on the lipid content of amniotic fluid, hence we focused on qualitative measurements of the phospholipidome of pig lavage samples.

These samples are easier to interpret, because their lipid content is higher. This is because the lung lavage is directly obtained from a living creature; the surfactant of the lung is washed out in a high concentration and volume. The amnion sample were received by cultivating tissue and using their supernatant. Obviously the amniotic tissue was kept in medium which made the analysis more complicated for us, for the analyte is diluted and the matrix is more complex.

For this reason, we firstly measured standards, used our method and after detecting most standards in various concentrations we analyzed the pig lavage. We then compared our findings to literature and proceeded to measure the amniotic supernatant.

4.1.1 Measuring Standards

Here I will show how our standards ionized and fragmented and how we used that information to compose MFQL queries. The fragmentation is highly dependent on fragmentation energy, hence we always used the same fragmentation energy for all measurements.

4.1.1.1 17 PC

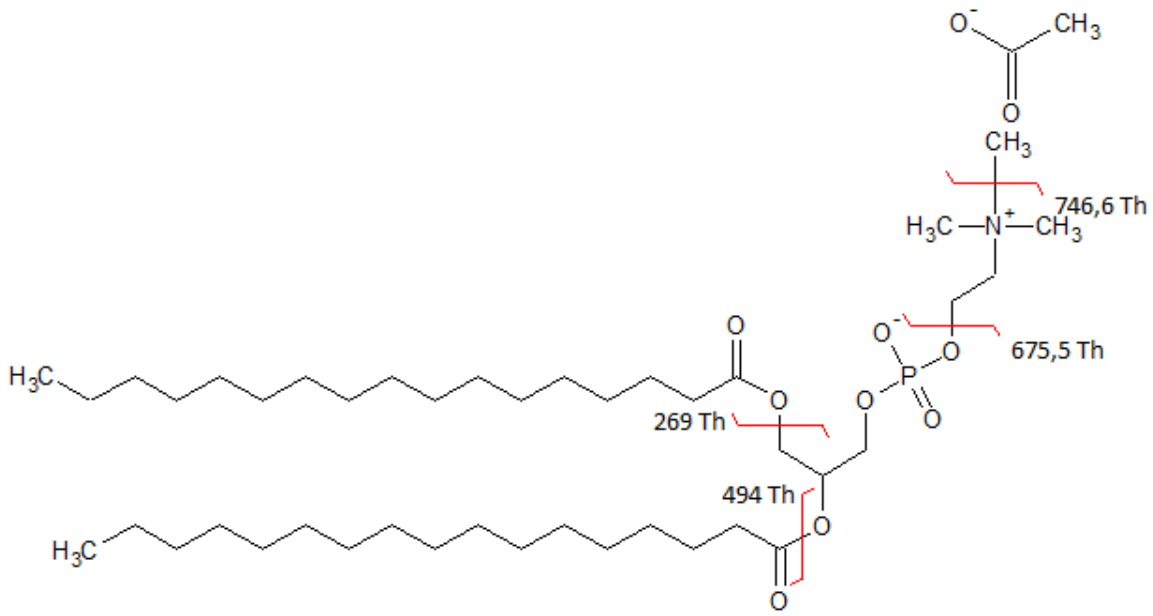

Figure 26: Structure of 17 PC.

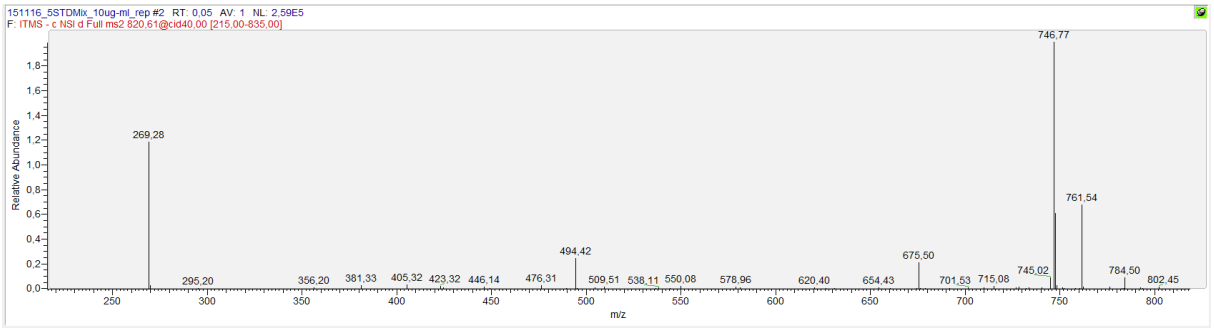

Figure 27: MS² spectrum of 17 PC.

Here the structure of 17PC-ion and its MS² spectrum are depicted. As shown the PC ionizes with an acetic acid adduct. It is important to note that the MS² spectrum is zoomed in, the

base peak (747 Th) is 50 times more intense. This mass consists of the precursor without the acetic acid adduct which is neutralized by reacting with one of the methyl groups of the choline to an ester. When the spectrum is shown, it is the only apparent peak. After zooming in, several other peaks become visible. The mass of 269 Th are the segregated fatty acids, this is the second most intense peak, and also the severed fatty acids are a strong piece of information to deduce the molecule composition. Since the precursor mass is known and the two fatty acids are the only varying part of the molecule it is important to find at least one fatty acid in the spectrum. Hence the most intense peak and the fatty acid peak were used to identify lipids in the master scan.

MFQL code is built to use the precursor mass, the neutral loss of methyl-acetate and the severed fatty acid:

```
DEFINE PR = 'C[38..54] H[30..130] O[10] N[1] P[1]' WITH DBR = (2.5,9.5), CHG = -1;
DEFINE headPC = 'C[3] H[6] O[2]' WITH CHG = 0;
DEFINE FA1 ='C[14..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FA2 ='C[14..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
```

The other shown peaks are different fragmentation patterns. For instance, the 494 Th peak is the remaining PC after one fatty acid is separated as a ketene, this peak could have also been used to identify the structure of the PC, however this peak is not intense and might not be visible in spectra of samples with matrix and a lower analyte concentration. The peak with a mass of 675 Th is the PC molecule without the choline group. [22]

4.1.1.2 12LPC

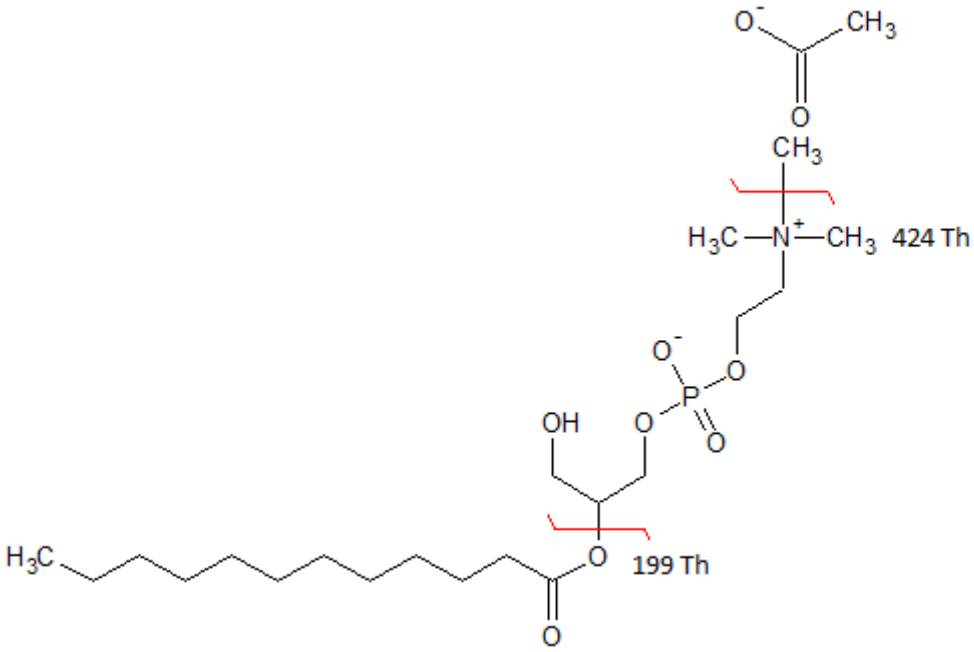

Figure 28: Structure of 12 LPC.

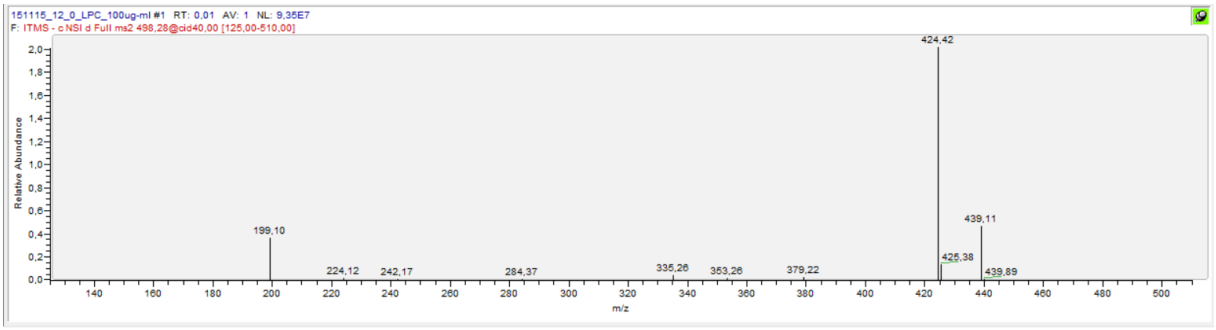

Figure 29: MS² spectrum of 12 LPC.

Similar to the PC the LPC ionizes via an acetic acid adduct, also the MS² spectra are similar; the intensity of the base peak requires a zoom in the spectra, again all peaks except the 424 Th peak are 50 times smaller. The obvious peaks are again the leaving methyl-acetate and the fatty acid. [22]

MFQL code is again using the precursor mass, the neutral loss of methyl-acetate and the severed fatty acid:

```
DEFINE PR = 'C[18..32] H[20..80] O[9] N[1] P[1]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE headPC = 'C[3] H[6] O[2]' WITH CHG = 0;
DEFINE FA1 ='C[10..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
```

The peak with a mass of 439 Th might be an ionized LPC after the acetic adduct leaves as neutral loss by reacting with a proton from the molecule. The spectrum is zoomed in by a factor of 50.

4.1.1.3 17 PG

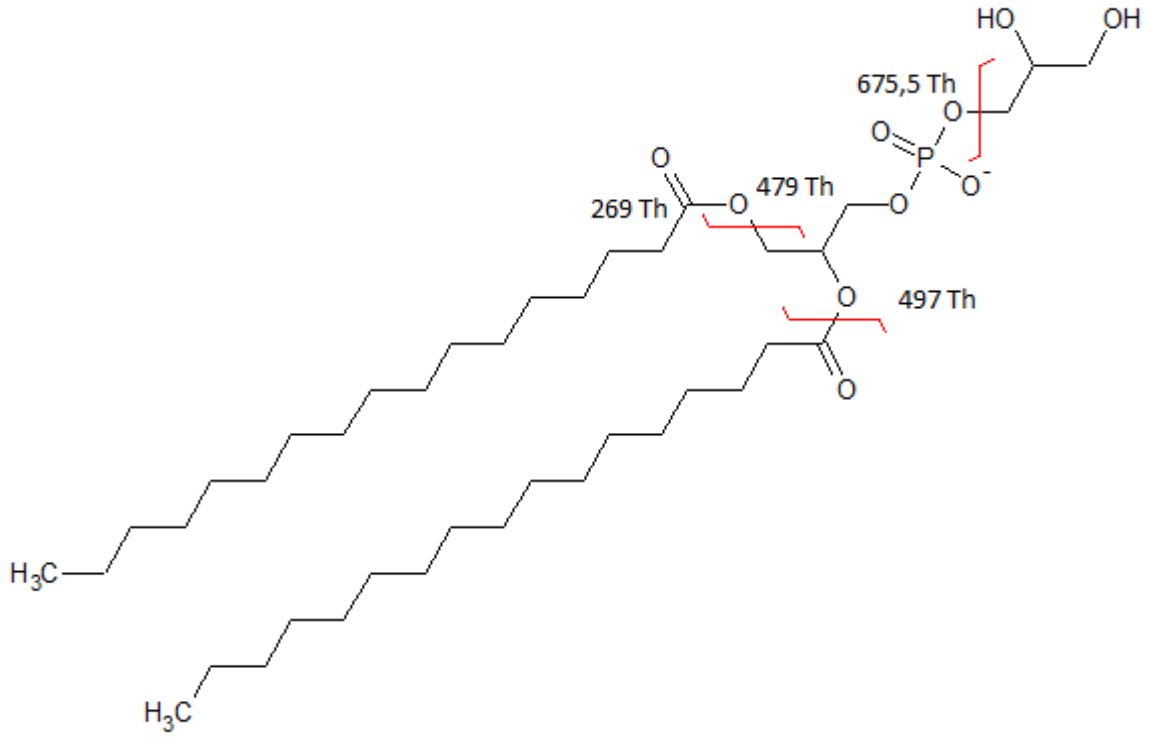

Figure 30: Structure of 17 PG.

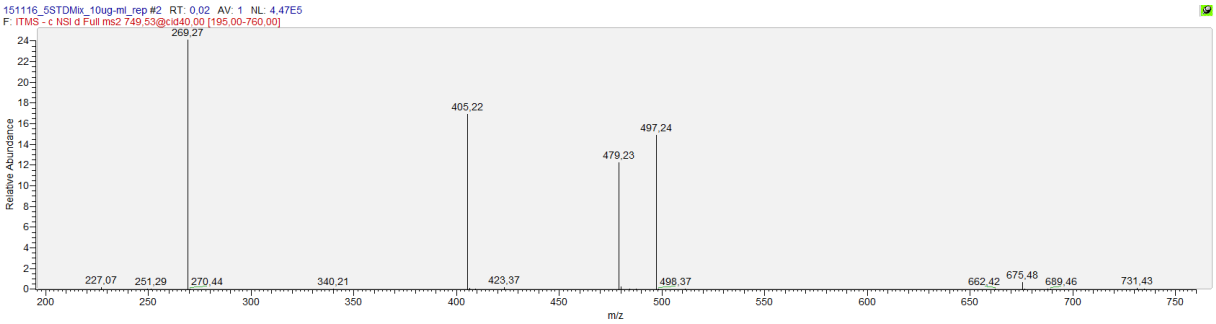

Figure 31: MS² of 17 PG.

The PG offers several fragmentations. The base peak is the separated fatty acid at 269 Th, all other peaks are magnified by a factor of four.

The MFQL relies only on finding the fatty acids. There were plans of inserting more features, however there has hardly ever been a false positive PG and because all other peaks are less intense it did not seem logical to make the script more stringent:

```
DEFINE PR = 'C[34..50] H[30..120] O[10] P[1]' WITH DBR = (2.5,8.5), CHG = -1;
DEFINE FA1 ='C[14..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FA2 ='C[14..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
```

The fragment of 405 Th is an ion which lost not only the fatty acid but also the glycerol. The peak at 423 Th is similar, here the ion lost the glycerol and the fatty acid as a ketene. The peaks at 479 Th and 497 Th are the respective less fragmented ions, here the fatty acid leaves as an acid or ketene, and however the glycerol was not severed. The minor peak at 675.5 Th is a neutral loss peak of the glycerol. [22]

4.1.1.4 17 PS

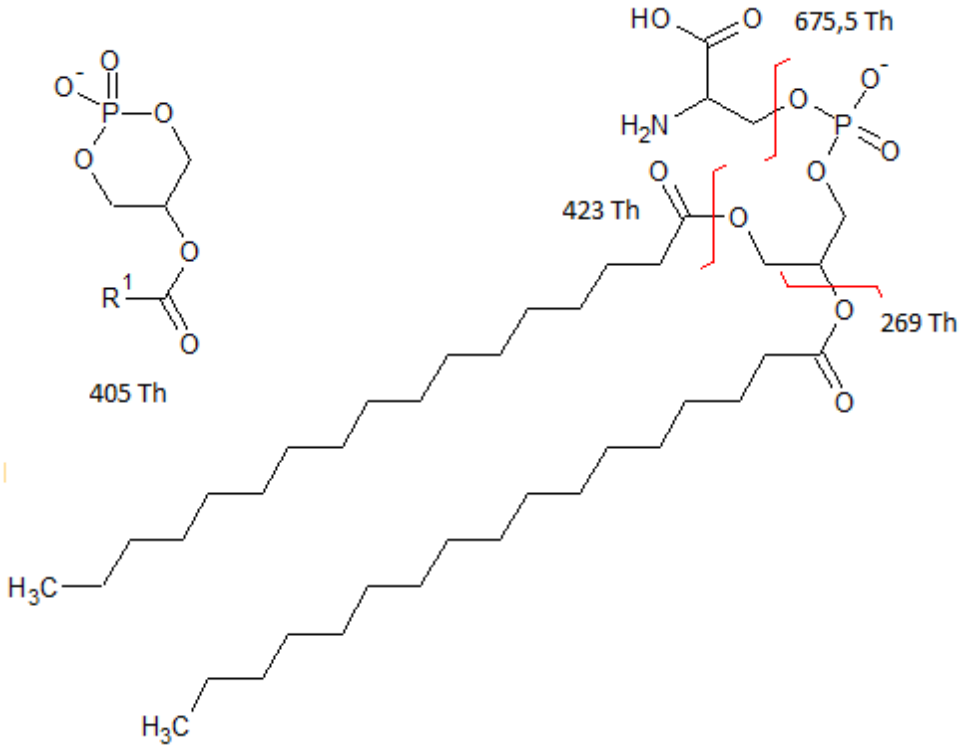

Figure 32: Structure and several fragments of 17 PS.

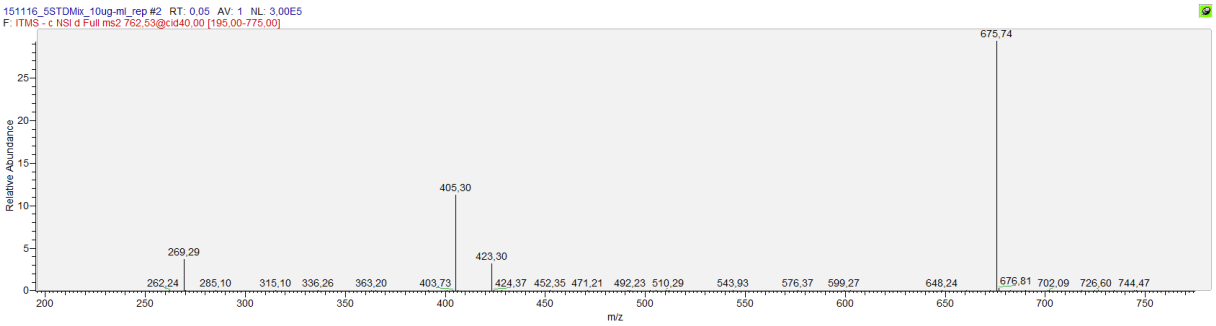

Figure 33: MS² spectrum of 17 PS.

It was a necessity to enhance the most peaks by a factor of four in this spectrum, the base peak of 675.5 Th is a neutral loss of serine and along with the fatty acid peaks at 269 Th used for the MFQL script for identification:

```
DEFINE PR = 'C[30..50] H[30..120] O[10] N[1] P[1]' WITH DBR = (3.5,10.5), CHG = -1;
DEFINE headPS = 'C[3] H[5] O[2] N[1]' WITH DBR = (0,2) , CHG = 0;
DEFINE FA1 = 'C[12..22] H[23..43] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FA2 = 'C[12..22] H[23..43] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
```

The peak at 405 Th is the resulting ion after losing one of the fatty acids as a neutral acid, while also losing the serine group. 423 Th is a similar fragment, the fatty acid is lost as a ketene, and the serine is also severed. [22]

4.1.1.5 14 LPE

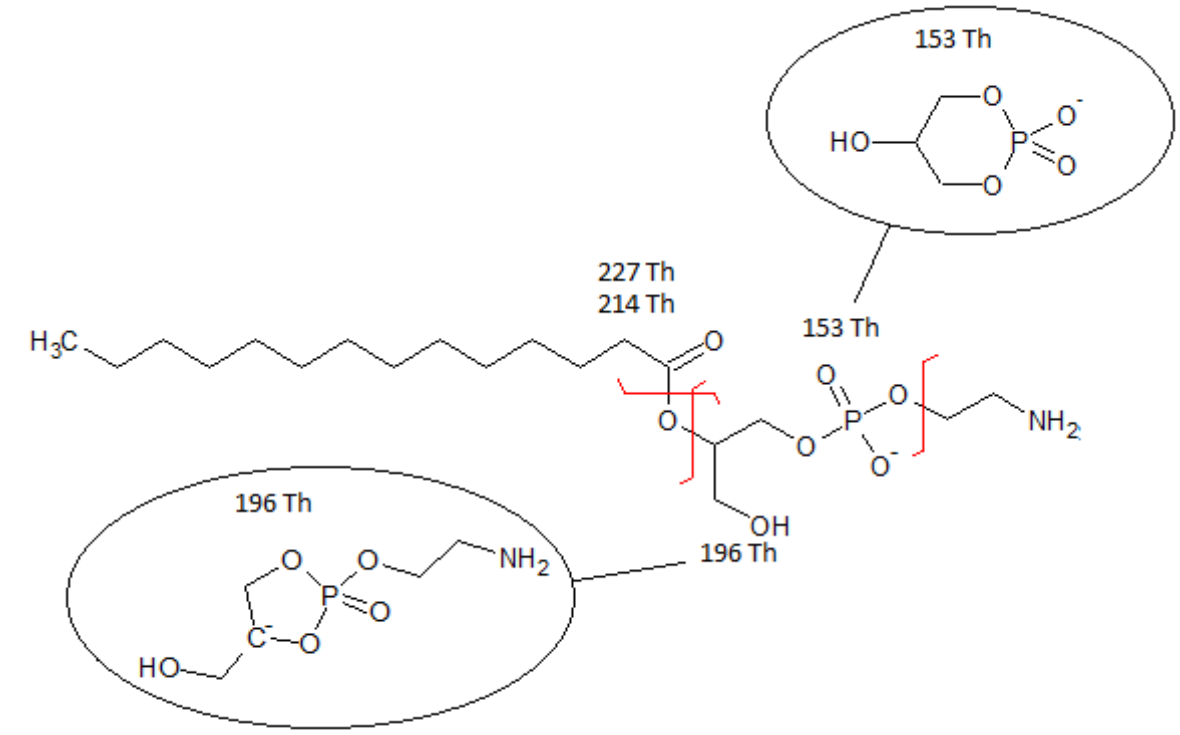

Figure 34: Structure and several fragments of 14 LPE.

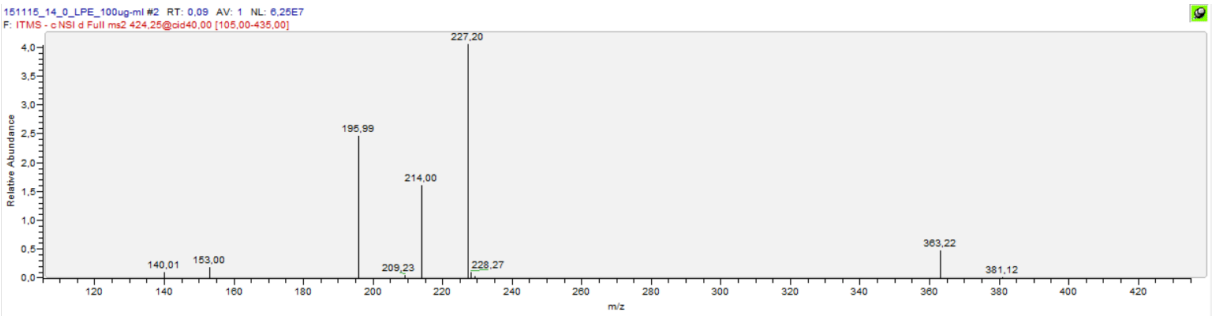

Figure 35: MS² spectrum of 14 LPE.

Even though there are several peaks visible in the MS² spectrum of LPE, only the base peak of the fatty acid was chosen as an identification criterion. This has two reasons, firstly because it is the most intense peak, all other peaks are magnified 25 times and also because there were hardly any hits for LPE, hence a stricter MFQL script would not have made much sense:

```
DEFINE PR = 'C[19..27] H[30..70] O[7] N[1] P[1]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FA1 ='C[14..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
```

The peak at 196 Th is a neutral loss of the fatty acid. The following peak at 214 Th is a loss of the fatty acid as a ketene. The minor peaks of 153 Th is the resulting ion after losing the fatty acid and the ethanol amine. The even smaller peak at 140 Th, which is not included in the drawing is an ion resulting from the neutral ethanolamine being split from the charged phosphate group. [22]

4.1.2 Data preparation of the standard measurements and reliability of the method

After measuring all standards in measurement mixture separately and without any matrix, we created a mix of the standard and measured different concentrations. The data was used to create a master scan file. I used the shown settings and used all of the described MFQL scripts. This was done to check for false positives, because only 5 lipid classes were present and only these lipid classes should be found. However not only the standards were found, but also several other lipids. At first it was assumed that those are false positives, however after manual check it became apparent that the standards were obviously not completely homogenous or even decaying over time. This induced two big questions which were hard for us to answer, for we had hardly any lipidomics experience:

1. What is a certain hit and at which point is it an uncertain hit? How can we use the hints from the mass spectrometer to give a yes or no answer, if we do not get such a straight forward answer from the mass spectrometer and the program?

This question arose early in our work, right after measuring the standards in a simple matrix. We came to the conclusion that our experiments were screening approaches and as such there obviously is a margin of error. Hence we heavily relied on the used standards for certainty. The other lipid classes we found, like certain Phosphatidylethanolamine-hits and lipids from a different category called Sphingomyelin are less likely to be actual hits, because we could not use standards to check whether our devised workflow was able to detect these lipids or not. However, it is important to mention that there are no standards for reduced Phosphatidylethanolamine available and the standards for Sphingomyelin are very expensive.

Later we measured our samples (amniotic samples and pig lavage) and used all of the mentioned MFQL scripts. Some of the scripts would never return a result like PI or PC-O, even though these scripts were used on all measurements. Hence, there is no way of telling whether the scripts are just not functional or the concentrations of respective lipid classes are below the LODs in our samples. Some of the MFQL scripts would return a detected lipid. Mostly PA, PE, PE-O and SM. All of the hits were checked manually. PA and PE-O for instances usually return several hits, which mostly consist of false positives of the same mass. This becomes apparent because either the MS^2 is noise and the program interpreted the noise or the MS^2 spectrum is decent, but the fatty acids are oddly number which is unlikely. [25]

Even though oddly numbered fatty acids exist in the mammalian lipidome [18], the approach to scan for oddly numbered fatty acids to detect false positives is valid. Imaging obtaining several hits for oddly numbered fatty acids and only a few for only evenly numbered fatty acids. This outcome indicates a flaw in the system, because it is unlikely to find a variety of seldom detected analytes and only few bona fide abundant lipids. Hence all found lipids featuring oddly numbered fatty acid chains are deemed biologically unlikely until proven through a 'proper' MS^2 spectrum.

Hence we achieved the highest certainty of detection by using standards and visually compare the standard and the analyte MS^2 spectra. The next step of certainty are the several PE-O and SM hits. There MS^2 spectra always look alike and the fragments are chemically plausible.

The lowest level of certainty are hits, which may already be false positives. For instance, some of the hits with oddly numbered fatty acid chains, which have a decent MS^2 spectrum but are unlikely due to the fact that odd fatty acids are usually regarded as not easily detected in mammals.

2. Why can we detect more lipids in our standards? Why are we not able to work quantitatively?

After we measured our standards and used the software on our data we not only found 4 of our 5 standards, we also found some other lipids.

PA [17:0 / 17:0]
LPE [14:0]
LPE [16:0]
PG [18:0 / 18:0]
PG [17:0 / 17:0]
PG [18:0 / 17:0]
PG [19:0 / 17:0]
PC [17:0 / 17:0]
PS [17:0 / 17:0]

Table 2: List of lipids found in the standard mix. Bold lipids are the standard lipids.

As mentioned before all MFQL queries were used, hence we also acquired some seemingly false positive hits:

PA [17:0 / 17:0]
LPE [16:0]
PG [18:0 / 18:0]
PG [18:0 / 17:0]
PG [19:0 / 17:0]

Table 3: Assumed false positive lipids.

However, after manual check of the spectra, these findings seem to be true. The PA [17:0 / 17:0] standard was only found in measurements in which the PS [17:0 / 17:0] standard was added, further the MS1 peak is well above noise level and the MS^2 spectrum looks acceptable. Hence it is possible that this finding is not a false positive, but an impurity of the standard. Likewise, the LPE [16:0] which was only found in the measurements containing the LPE [14:0] standard. The several PG lipids were only found in measurements in which the PG [17:0 / 17:0] standard was present.

It was assumed that these lipids are byproducts of the lipid synthesis, however later measurements on a different system revealed that our standards were already decaying, which explains why PA [17:0 / 17:0] was found in a solution containing solely PS [17:0 / 17:0] standard.

4.1.3 Identification of Lipids without Standards

As described before all the MFQL queries were used on the measurements containing standards. This was done to check if the MFQL scripts will produce false positives. However as shown before this did not happen. Hence even the MFQL scripts which were not tested by using them on standards were used on our samples. If the LipidXplorer was able to find Lipids from categories of which no standard was present, the respective MS^2 experiments were checked manually. If the MS^2 spectrum of several lipids of one category (like SM [42:1] and SM [38:1]) was similar to another assumed lipid of that category, these lipids were generally assumed as present. Similarity is in this case given when the same neutral loss group or fragmentation pattern also has a similar intensity distribution. Further standards of these lipids are hard to come by like for PE-O, or just very expensive like SM.

4.1.3.1 PA

Even though no standard was available, the PS [17:0/17:0] standard contained a contamination which most likely is PA [17:0 / 17:0]. The MS1 mass is precise and the MS^2 only contains fragments typical for PA. [22]

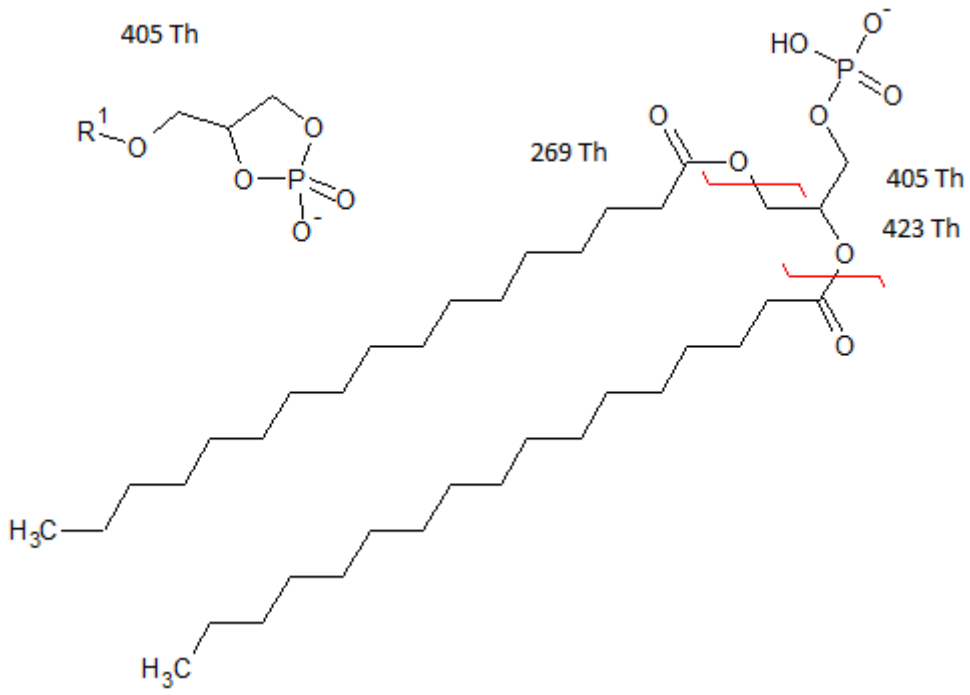

Figure 36: Structure of 17 PA and its fragmentations patterns.

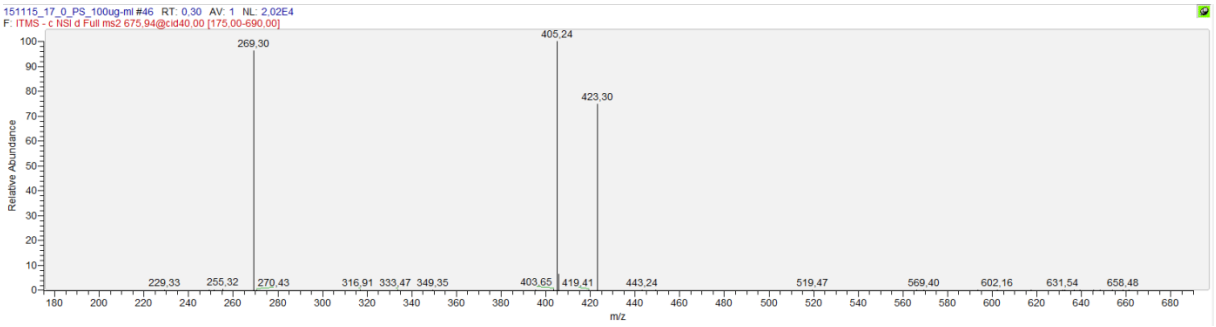

Figure 37: MS² of PA [17:0 / 17:0]

The spectrum of the PA is very clean, only masses are shown that are interpretable by comparing the MS² spectrum with the suggested hits from lipid maps. The 269 Th peak are the fatty acids, 423 Th is the precursor with a lost fatty acid, where the fatty acid leaves as ketene, while the 405 Th peak is basically the same fragmentation, with the oxygen staying at the fatty acid. The MFQL script relies on the correct mass in MS1 and the fatty acids peaks in the MS² for detection:

```
DEFINE PR = 'C[31..47] H[30..120] O[8] P[1]' WITH DBR = (2.5,9.5), CHG = -1;
DEFINE FA1 ='C[12..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FA2 ='C[12..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
```

4.1.3.2 PE

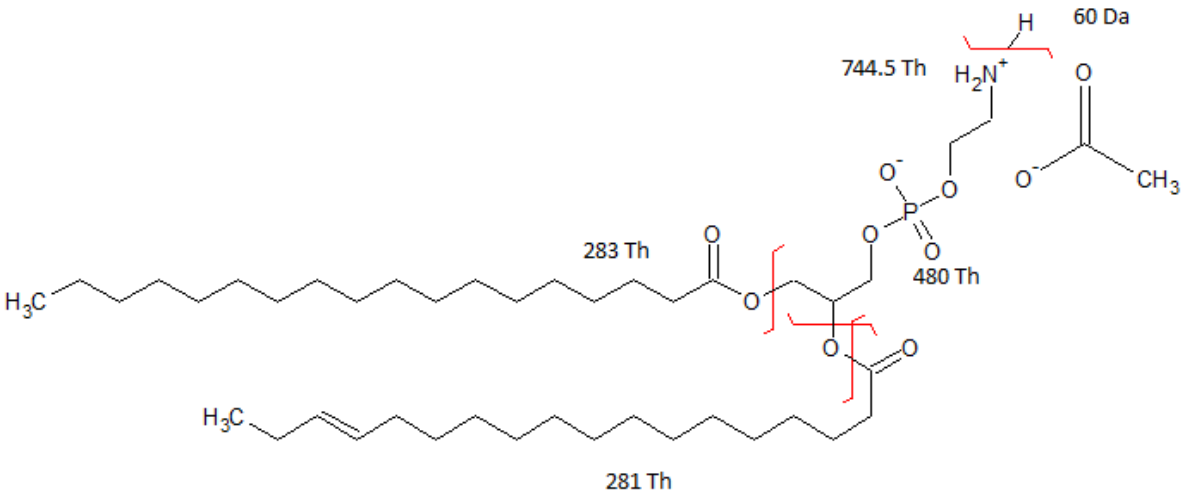

Figure 38: Structure and fragments of PE [18:0 / 18:1].

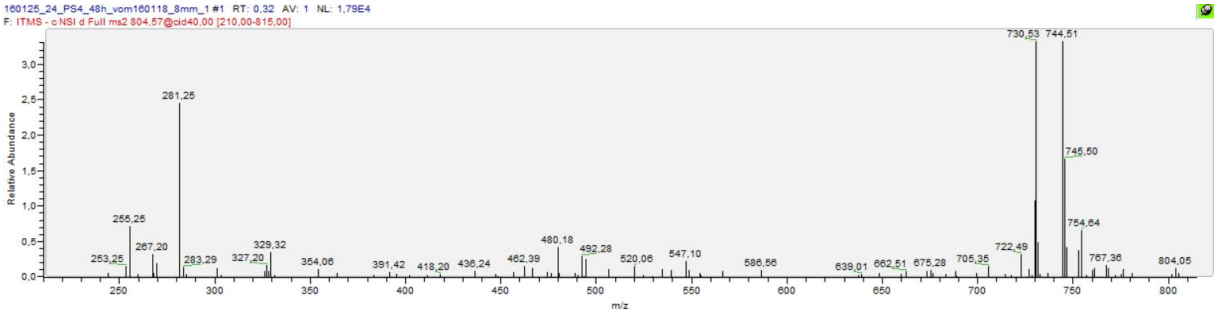

Figure 39: MS² of PE, which is magnified 30 times.

Here is one of the more promising PE hits. It is a PE [18:0 / 18:1] from one of the amnion samples, the precursor was ionized via addition of acetate and hence the most abundant peak in the MS² is the neutral loss peak of 744 Th. Every other peak is magnified by an order of 33. Also visible is the peak for the unsaturated fatty acid at 281 Th (of course the location of the double bond is not known.) 480 Th is the loss of one fatty acid as ketene, the unsaturated equivalent is not seen. The 462 Th peak which is hardly distinguished from the noise is again the remaining molecule after the neutral loss of a fatty acid. 281 Th is the unsaturated fatty acid, the saturated form is hardly visible. [22] Spectra like this are very common, there are two peaks, one is easily charged and discriminates the other fragments. After enhancement of the spectrum several fitting peaks appear, however they hardly seem to reach an S/N above 3.

The MFQL script relies on detecting the precursor and the fatty acids in the MS²:

```
DEFINE PRACETAT = 'C[33..49] H[50..100] O[10] N[1] P[1]' WITH DBR = (2.5,9.5), CHG =
-1;
DEFINE FA1 ='C[12..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FA2 ='C[12..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
```

4.1.3.3 SM

Here I will show the assumptions made for the interpretation of sphingomyelin. The detected class of lipids share a common structure [11]:

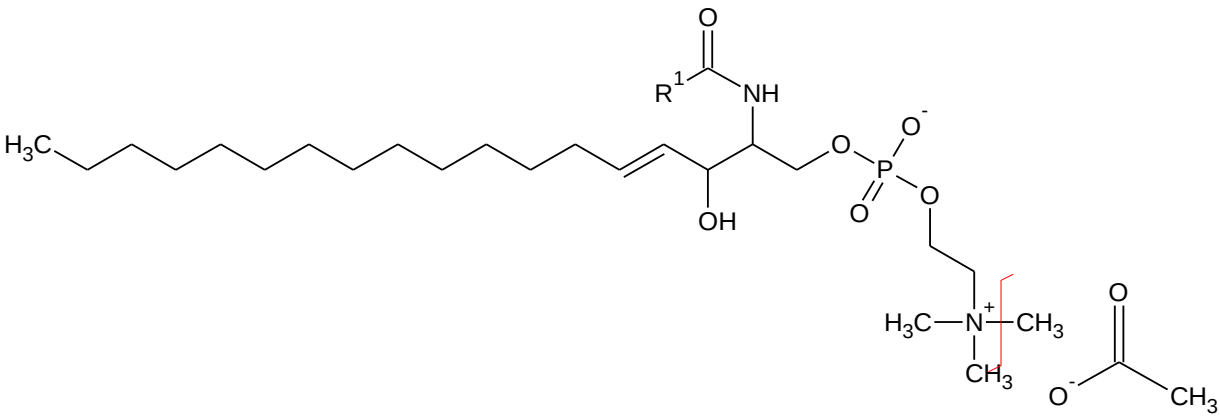

Figure 40: Typical structure for SM.

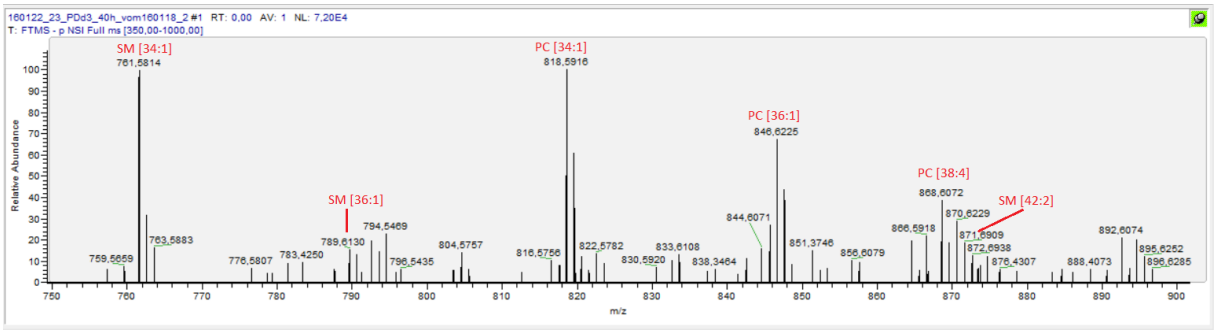

Figure 41: MS1 spectrum showing the masses of detected SMs and PCs.

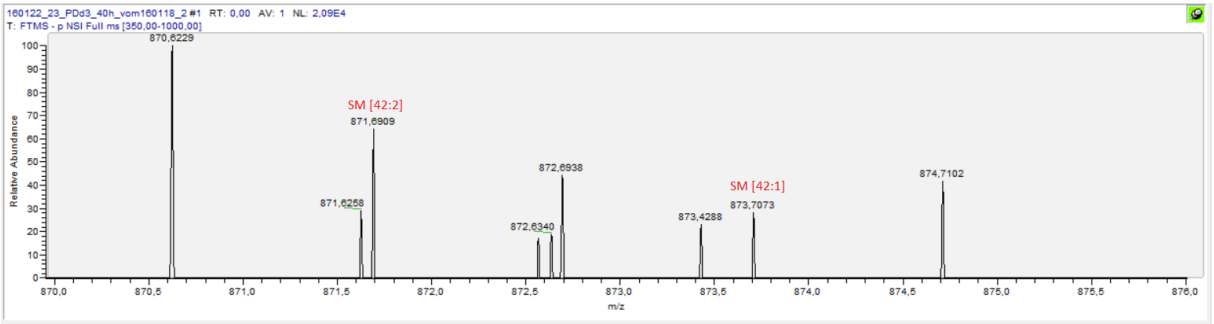

Figure 42: A magnified MS1 spectrum showing the less abundant SMs.

The choline group is basis for the detection. While we were working with the PC standard and later detected PCs in our samples we realized that fragmenting a phosphatidylcholine will always yield a neutral loss product with the methyl-acetate leaving the charged precursor molecule. In fact, this fragment dominates the whole Ms^2 spectrum with its

intensity. This knowledge gave us confidence in using the downloaded MFQL script for detection. Several sphingomyelins were found, their MS^2 spectra look similarly.

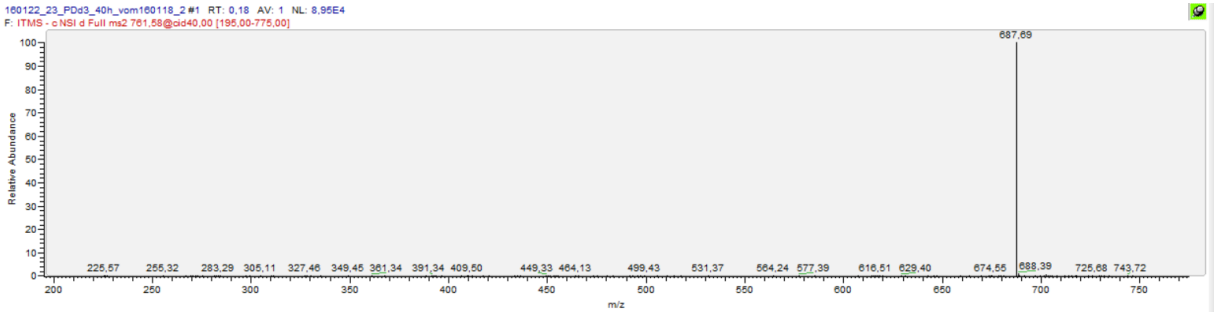

Figure 43: MS^2 spectrum of SM [34:1] with a precursor mass of 761.6 Th.

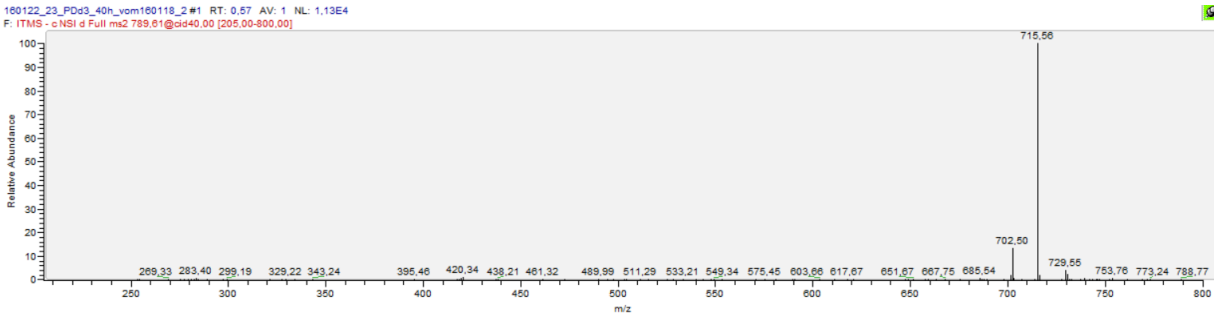

Figure 44: MS^2 spectrum of SM [36:1] with a precursor mass of 789.6 Th.

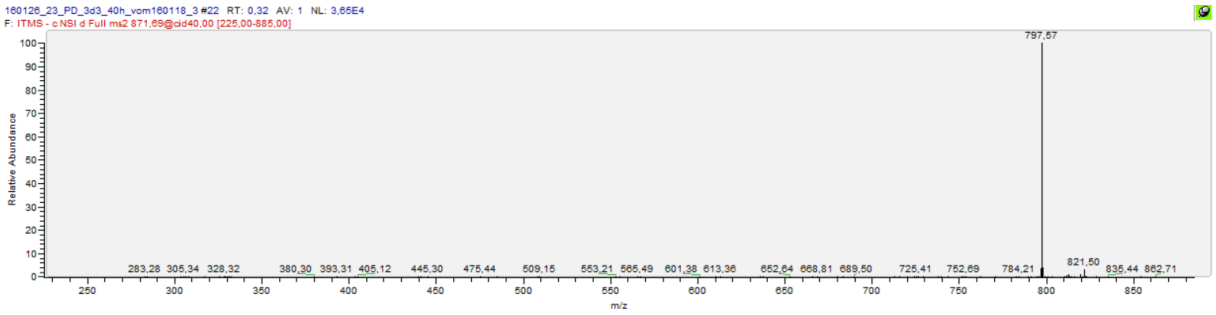

Figure 45: MS^2 spectrum of SM [42:2] with a precursor mass of 871.6 Th.

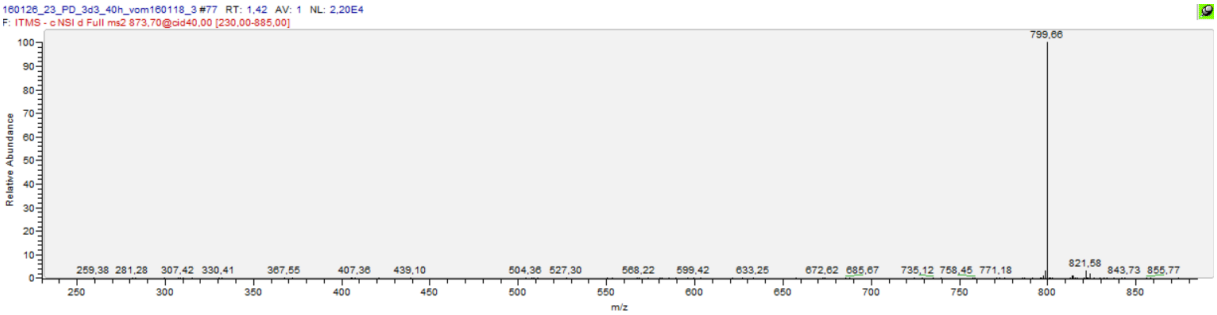

Figure 46:MS² spectrum of SM [42:1] with a precursor mass of 873.6 Th.

All of the spectra, which are also similar in intensity only feature one peak. Even if magnified 50 times there are hardly any hints for the composition of the fatty acids. This was used in the MFQL script, which looks for only one peak in the MS²spectrum:

```
DEFINE PR = 'C[39..49] H[30..130] O[8] N[2] P[1]' WITH DBR = (1.5,3.5), CHG = -1;
DEFINE headSMPC = 'C[3] H[6] O[2]' WITH CHG = 0;
```

The ceramide part of the sphingomyelin structure is usually consisting of 18 C atoms and one double bond, it is safe to assume that only the alkyl chain, which is connected to the ceramide structure via an amid bond, varies. [11]

Hence the fatty acid attached to the invariable part are: 16 C atoms, 18 C atoms, and 24 C atoms with none to one double bond.

4.1.3.4 PE-O

PE-O stands for Phosphatidylethanolamines which feature a fatty acid, which is reduced to an alcohol. Even though, they look similar to a non-reduced PE, they chemical properties are different, hence the fragmentation pattern is not predictable by using non-reduced PE as standard.

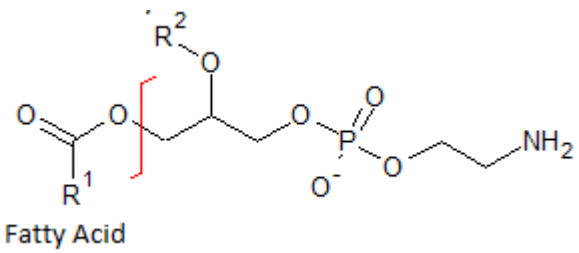

Figure 47: Assumed fragmentation of PE-O.

If our assumptions are correct, then they are usually two intense peaks. These fragments are shown here: the lighter fragment is the non-reduced fatty acid. The heavier fragment is the reduced fatty acid and the invariable part of the PE. Because of this the interpretation of the spectrum and peak assignment is easy. The MFQL script looks for these respective fragments in the MS^2 spectrum:

```
DEFINE PR = 'C[33..49] H[50..100] O[7] N[1] P[1]' WITH DBR = (1.5,8.5), CHG = -1;
DEFINE FA1 ='C[12..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FAO ='C[17..27] H[20..80] O[6] N[1] P[1]' WITH DBR = (0.5,6.5), CHG = -1;
```

Here is the MS1 spectrum of an experiment featuring all PE-O hits.

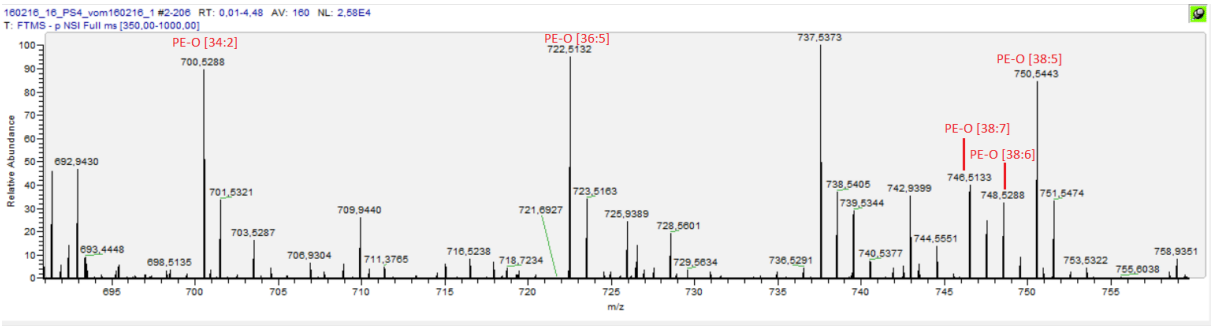

Figure 48: MS1 spectrum schowing several bona fide PE-O hits.

Several technical replicates were needed to trigger the data dependent fragmentation, hence the shown spectra come from different amnion-experiments:

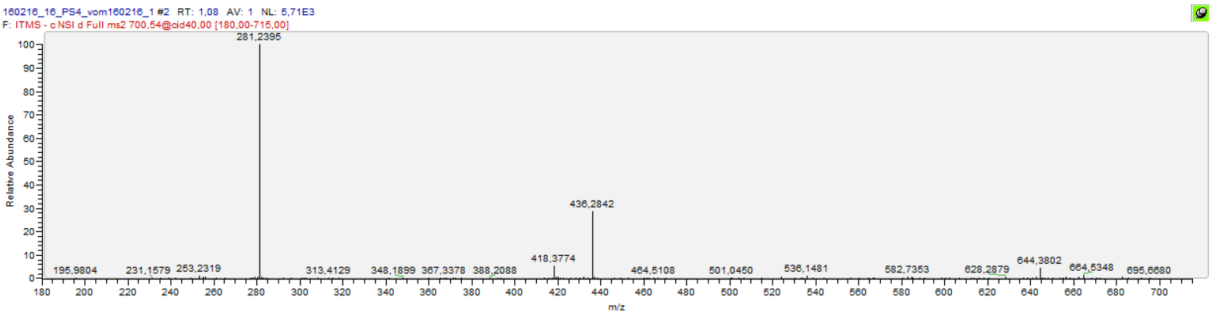

Figure 49: MS^2 of PE-O [16:1 / 18:1] with a precursor mass of 700.5 Th.

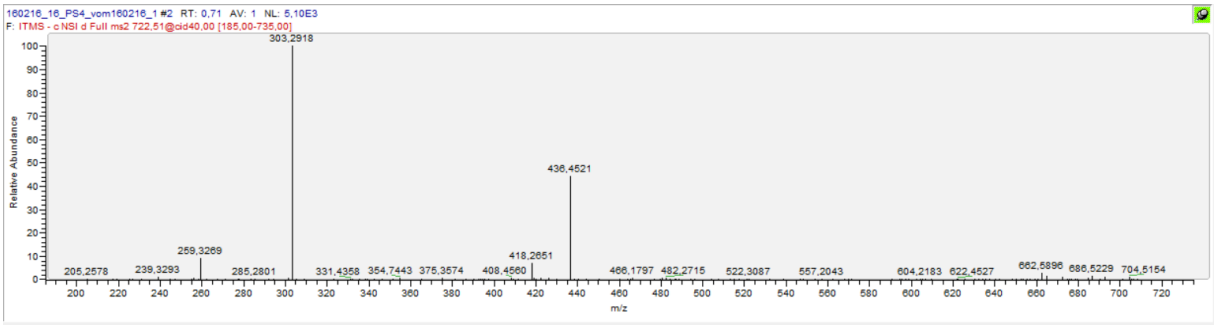

Figure 50: MS² of PE-O [16:1 / 20:4] with a precursor mass of 722.5 Th.

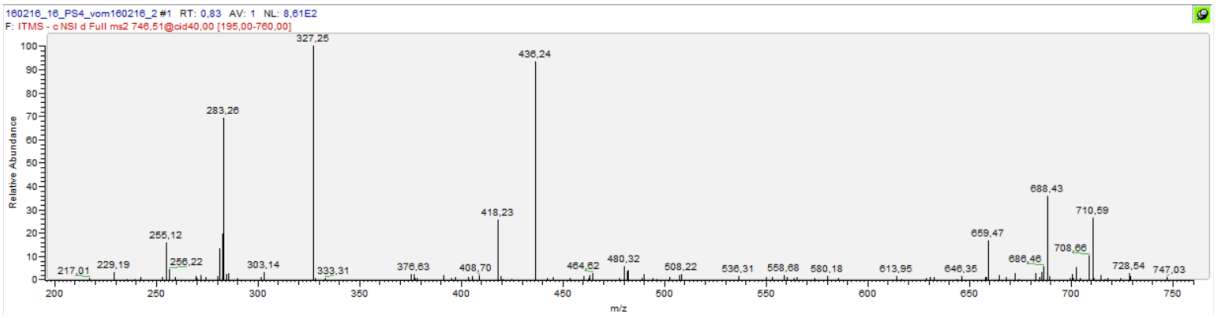

Figure 51: MS² of PE-O [16:1 / 22:6] with a precursor mass of 746.5 Th.

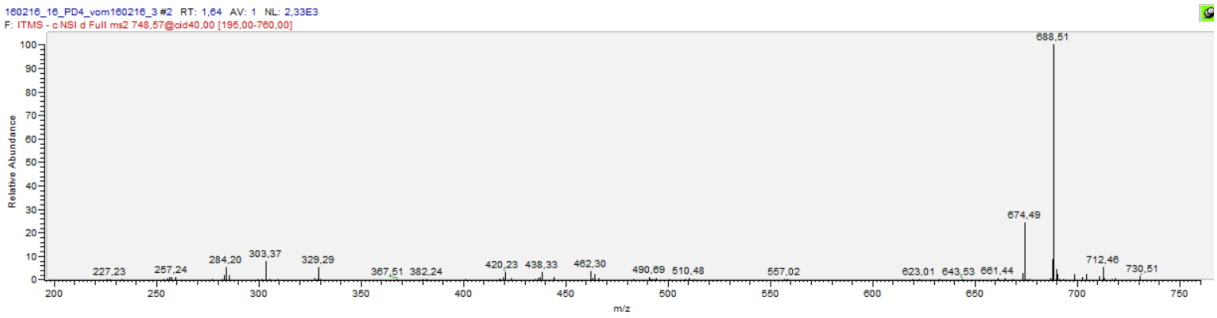

Figure 52: MS² of PE-O [18:2 / 20:4] with a precursor mass of 748.5 Th

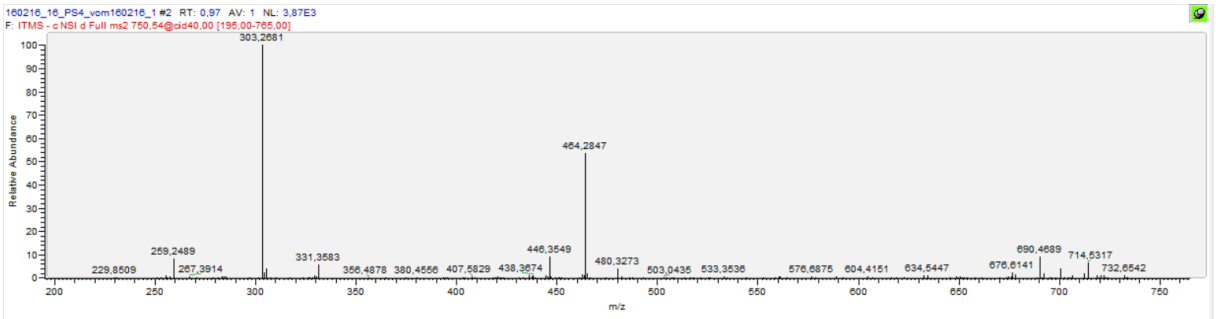

Figure 53: MS² of PE-O [18:1 / 20:4] with a precursor mass of 750.5 Th.

It is important to note, that the first mentioned fatty acid, is in its reduced form and bound as an ether. The spectra of PE-O [18:1 / 20:4], PE-O [16:1 / 18:1] and PE-O [16:1 / 20:4] are very similar. Their intensity is in the range of 4*10^3, the fatty acid is the base peak and the second most intense peak is the reduced fatty acid, which is still linked to the invariable part of the molecule. Also the MS1 intensities of these fatty acids are similar.

The spectra of PE-O [16:1 / 22:6] and PE-O [18:2 / 20:4] differ from the other PE-O spectra. This can be explained by looking at the MS1 spectrum, these lipids ionized worse or are simply less concentrated. The spectrum of PE-O [16:1 / 22:6] looks noisier, this might be a result of the fewer ionized molecules. The 327 Th peak is the fatty acid with 22 C Atoms, while the peak at 436 Th is the reduced fatty acid with the invariable part. The other intense peaks might be a neutral loss of H_2O of the 438 Th fragment at 418. The peak at 283 hints a 18 C atom fatty acid, while the peak at 255 Th is typical for a 16 C atom fatty acid, which might be a fragments of lipids with similar mass, which were not separated because of the broad window of 1 Da of the C-trap. The other spectrum of PE-O also features the neutral loss peak, however because the intensity of the base peak of the PE-O [16:1 / 22:6] spectrum is only around 8*10^2 noise is more visible.

After setting up this method we were able to detect several lipid classes from two lipid categories: phospholipids and sphingolipids. However due to the different way of deducing a MFQL script, certain scripts offer a higher confidence than others, which were used without testing on a standard or failed to identify the standard.

Lipid class	Standard available	Standard found	Mode of ionization	Number of fragments used for identification	Subjective level of confidence
PC	Yes	yes	$[M+Ac]^-$	3	high
LPE	yes	yes	$[M-H]^-$	1	moderate
PG	yes	yes	$[M-H]^-$	2	high
PS	yes	yes	$[M-H]^-$	3	high
LPC	yes	no	$[M+Ac]^-$	2	low
PE-O	no	-	$[M-H]^-$	2	low
SM	no	-	$[M-H]^-$	1	low
PE	no	-	$[M+Ac]^-$	2	low
PA	no	-	$[M-H]^-$	2	very low

Table 4: Chart of level of confidence for respective lipids.

Level of confidence is derived from number of assumed false positives, complexity of the MFQL script, the availability of a standard and the results of the experiments of the pig lavage samples.

4.2 *Proof of principle by analyzing pig lavage*

Here is the list of detected lipids of the pig lavage measurement is shown:

MASS [Th]	NAME	SPECIES				
722,51	PE-O [36:5]	PE-O [16:1 / 20:4]				
750,54	PE-O [38:5]	PE-O [18:1 / 20:4]	PE-O [16:0 / 22:5]	PE-O [16:1 / 22:4]		
778,57	PE-O [40:5]	PE-O [18:1 / 22:4]				
750,54	PE-O [38:5]	PE-O [19:0 / 19:5]				
764,56	PE-O [39:5]	PE-O [19:1 / 20:4]	PE-O [19:0 / 20:5]	PE-O [21:4 / 18:1]	PE-O [17:0 / 22:5]	
790,58	PE-O [41:6]	PE-O [19:1 / 22:5]				PE-O [21:1 / 20:4]
792,59	PE-O [41:5]	PE-O [22:0 / 19:5]	PE-O [19:0 / 22:5]	PE-O [19:1 / 22:4]	PE-O [21:2 / 20:3]	PE-O [19:3 / 22:2]
						PE-O [21:0 / 20:5]
719,49	PG [32:1]	PG [16:0 / 16:1]	PG [18:1 / 14:0]			
721,50	PG [32:0]	PG [16:0 / 16:0]				
745,50	PG [34:2]	PG [16:0 / 18:2]	PG [18:1 / 16:1]			
747,52	PG [34:1]	PG [16:1 / 18:0]	PG [16:0 / 18:1]			
773,53	PG [36:2]	PG [16:0 / 20:2]	PG [18:1 / 18:1]	PG [18:0 / 18:2]		
775,55	PG [36:1]	PG [18:1 / 18:0]				
795,52	PG [38:5]	PG [18:1 / 20:4]				
761,53	PG [35:1]	PG [18:1 / 17:0]				
554,35	LPC [16:0]					
761,58	SM [34:1]					
788,54	PC [32:2]	PC [16:1 / 16:1]				
792,57	PC [32:0]	PC [16:0 / 16:0]	PC [18:0 / 14:0]	PC [17:0 / 15:0]		
820,61	PC [34:0]	PC [18:0 / 16:0]				
840,57	PC [36:4]	PC [20:4 / 16:0]				
842,59	PC [36:3]	PC [18:2 / 18:1]				
844,60	PC [36:2]	PC [18:1 / 18:1]	PC [18:2 / 18:0]			
778,56	PC [31:0]	PC [15:0 / 16:0]				
804,58	PC [33:1]	PC [17:1 / 16:0]				
806,59	PC [33:0]	PC [17:0 / 16:0]				
750,53	PE [32:0]	PE [16:0 / 16:0]	PE [14:0 / 18:0]			
776,54	PE [34:1]	PE [16:0 / 18:1]	PE [18:0 / 16:1]			
802,56	PE [36:2]	PE [18:1 / 18:1]				
792,57	PE [35:0]	PE [15:0 / 20:0]	PE [19:0 / 16:0]	PE [17:0 / 18:0]		
788,54	PS [36:1]	PS [18:1 / 18:0]				
810,53	PS [38:4]	PS [18:0 / 20:4]				
844,60	PS [40:1]	PS [18:1 / 22:0]				

Table 5: List of found lipids using the mentioned settings.

On the left side is the mass in Th. The lipids under 'name' are the possible interpretations for the respective mass, biological possible lipids are marked in green, while lipids containing an oddly numbered fatty acid are marked in red and considered not likely or a false positive. On the right side under 'species' are all isomers which can be supported by MS^2 interpretation. If the species is written in green then all fragments are found with a typical intensity by

manually searching a random MS^2, orange means that the intensity of at least one fragment was low and red means that the spectrum was noisy and noise was interpreted as hit or the base peak was several potencies more powerful than all of the searched fragments

There are a lot of unlikely structures, some of which turn out to be false positives when investigated further. Namely those are PE-Os and PEs which are isobaric to the most abundant PC, the MS^2 noise was interpreted as detection. The reason for this is easily explained. The PCs fatty acids usually show a low abundance in the MS^2 spectra, hence the settings must be chosen to detect low abundant fatty acids in MS^2 spectra. The same settings are then used for other noisy MS^2 spectra, which might yield some false positives. Of course there are solutions, for instance it is possible to create a master scan file with different settings for every lipid class and its MS^2 peculiarities. A different solution is to make the MFQL scripts very specific, for instance the SUCHTHAT section offers many tools; the relation of the intensity of the peaks can be used. An example for how useful the SUCHTHAT is, is shown when looking at the list of PCs. The list of 'oddly numbered fatty acids in the lipid' is its own MFQL script, similar to the screening script detecting only even numbered fatty acids. By comparing the number of evenly numbered fatty acids to the number of oddly numbered fatty acids the reliability of the script can be tested. Hence the PS script is rather reliable, there has never been a hit for oddly numbered fatty acids, while the PE-O script reported several oddly numbered fatty acid lipids. The method seems prone to detect PE-Os not very accurately, but PSs seem to be detected with a high certainty. On the other hand, there is also an option to use the SUCHTHAT section to only detect lipids with evenly numbered fatty acids. This will reduce the false positives but will offer less information about the reliability of the used scripts.

Another way to use the MFQL script is to check for ionization mode. We had no PE standard so we did not know if and how PE would be ionized. There are two possible ions, the first one being an acetate adduct of the PE, the second a deprotonated PE-ion. We used both scripts on the pig lavage measurements. The script for an ionization with acetate found eight PE species, five of them biologically likely. One MS^2 spectrum per species was checked manually and the interpretation of the program seemed plausible, even for the biologically unlikely hits. The deprotonated ionization script found 27 species, however checking the MS^2 revealed a lot of neutral loss peaks of 60 Da or 74 Da, which is typical for ionization with acetate. Hence the PE was most likely not ionized by deprotonation and these hits are not

shown here. Hits of this script were further regarded as false positive, hits of the script using acetate adducts were considered true with a low reliability.

The problem of false positive detection originated from underestimating the differences between the samples, we mainly used amniotic samples for the fine tuning of the system and method establishment. These samples offered little to no lipid peaks. The pig lung lavage was rich in lipids, especially in DPPC (PC [16:0 / 16:0]). The base peak of the DPPC-MS^2 was very intense and all fragments of the highly abundant lipid was found, however the noise was more intense too. The noise was interpreted and several isobaric PE-O [41:5]s were found, which are isobaric to DPPC but not very likely. PE-O [41:5], an isobar of DPPC with a high quantity of questionable hits, was interpreted as seven different lipids. When checking the intensities of the MS^2 spectra, which contains fragments of all suggested lipids, only PE-O [19:0 / 22:5] has a reasonable intensity, while PE-O [19:3 / 22:3] is only present in one MS^2 spectrum with an intensity just above the threshold. The PE [35:0]-hits isobaric to DPPC were checked too and even though some fragments were above the set threshold, considering the high intensity of the base peak of the MS^2 spectrum, were considered false positives.

The simplest solution to avoid false positives is to have a look at the output file of the LipidXplorer search. Checking the intensities of the MS^2 peaks gives hints to which lipids are to be classified as false positives, for instance if a lipid is only found in a single MS^2 spectrum in several runs with a very low intensity, the spectrum should be checked manually. Lipids which are often fragmented in an MS^2 experiment and have a well above threshold intensity have never turned out to be false positives. Further the method can be changed, the threshold of the detection limit can be made more stringent, however the method was optimized for amniotic supernatant samples and the pig lavage measurements were more of a proof of principle. Hence we did not further optimize for pig lavage measurement.

The PC script is one of the most specific ones, not only was it tested on a standard, it also uses three fragments for identification. When looking at the results, it becomes obvious that there are many false positives. For instance, PC [17:0 / 15:0] is an unrealistic isobar of the most abundant lipid PC [16:0 / 16:0]. The fragments of the fatty acids yield a low MS^2 peak intensities. The other false positives featuring oddly numbered fatty acids were checked manually and even though they are considered biologically unlikely, their MS^2 is proper.

A lot of the mentioned false positives can be avoided by setting the requirements for a hit higher (e.g. higher peak intensities, stricter MFQL) however some hits, which are true positives but low abundant like PC [16:1 / 16:1] also have low intensities and MS^2 frequencies and would not be detected with stricter measures.

It is further important to note the difference between the samples. We received pig lavage samples two times. The first lavage samples featured less of the detected lipids than the second lavage samples. Considering this might put the question about false positives into perspective; there obviously is a fluctuation of lipid content of the surfactant.

4.3 Our qualitative measurements vs. literature and fitness for purpose

After measuring the pig lavage samples, we compared our results to literature. All the found publications focused on stating the concentration of the several classes of phospholipids and not on the fatty acid composition of these classes of different mammals. Further the different papers vary in their found lipid composition. All papers state that PC [16:0 / 16:0] makes up most of the lipid composition.[12, 2, 16]

This is also easily confirmed by looking at a random MS1 spectrum of the pig lavage measurements:

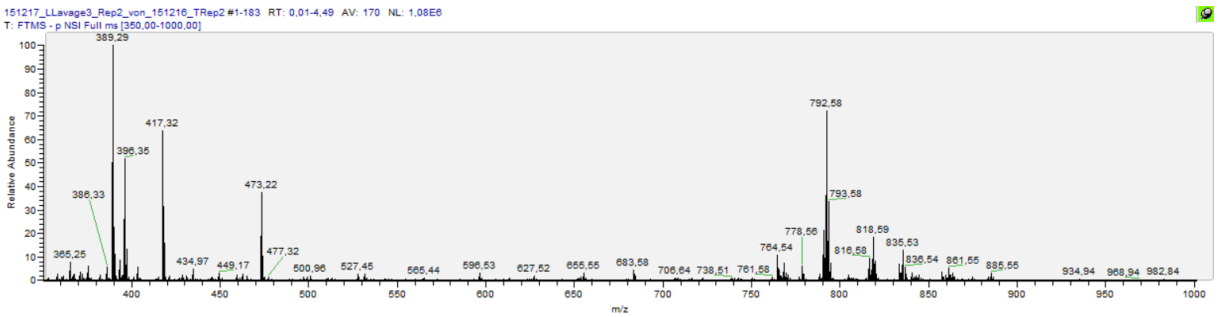

Figure 54: MS1 Spectrum of pig lung lavage sample.

The peak of 792.58 is the acetate adduct of PC [16:0 / 16:0]. This is further supported by several MS^2 spectra.

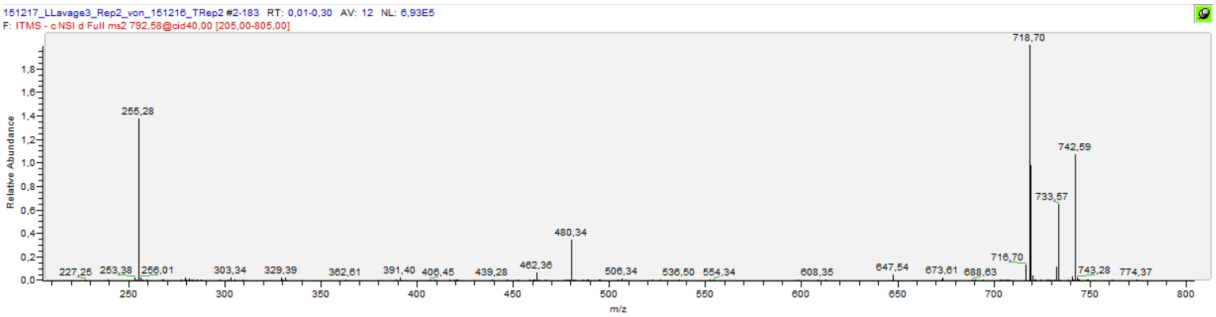

Figure 55: MS² spectrum of 792,58. Several typical peaks are visible.

Said literature described different compositions, with one author just stating that the lung surfactant of most mammals feature similar characteristics. (31). The other sources described pig Broncho alveolar lavage [2] and murine lung surfactant[12].

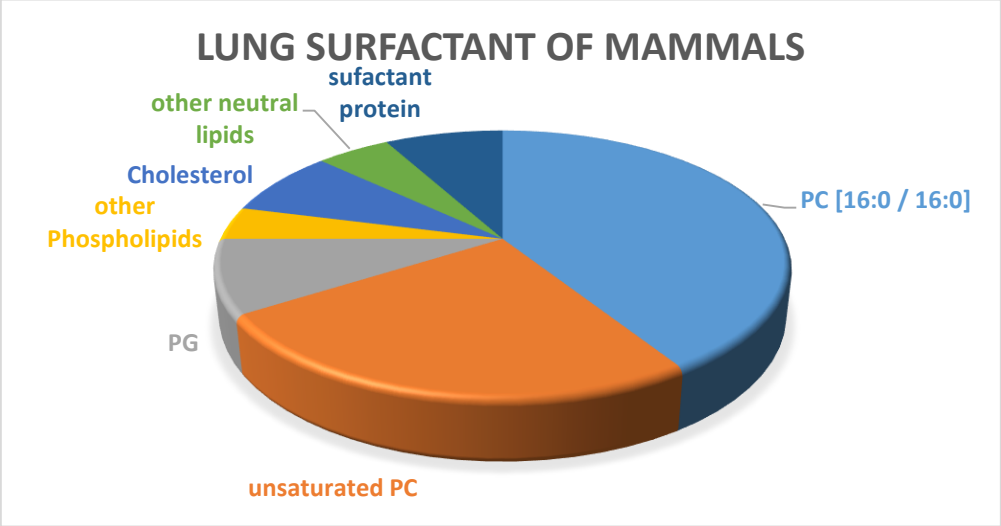

Figure 56: Composition of lung surfactants in mammals. [16]

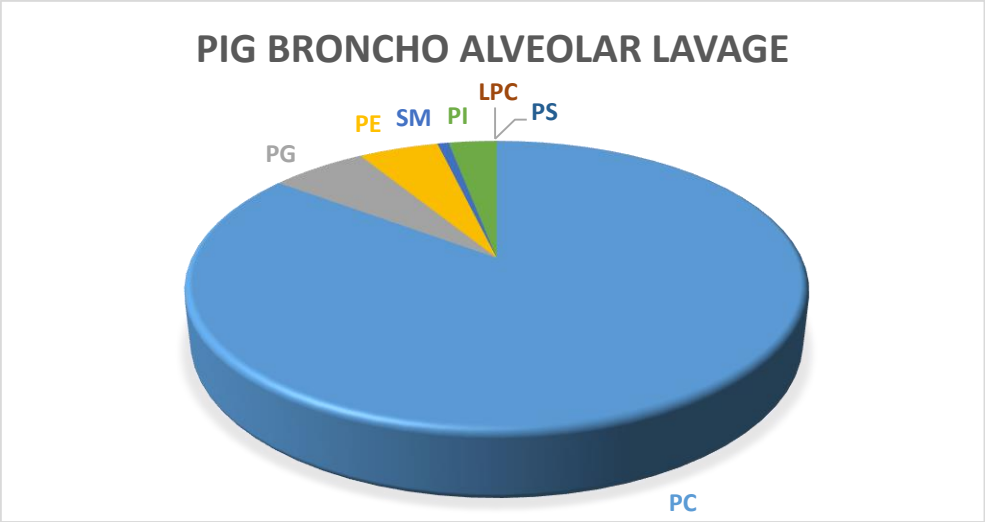

Figure 57: Composition of pig broncho alveolar lavage with LPC and PS only in traces. [2]

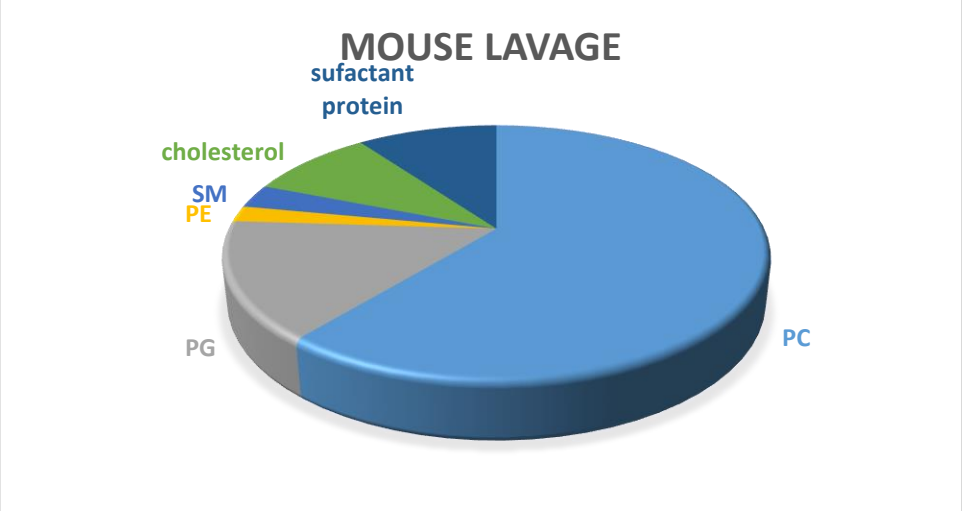

Figure 58: Composition of mouse lung lavage. [12]

As shown here, PC and particularly PC [16:0 / 16:0] are an important feature of lung surfactant. PG and PE are also always shown in the pie charts. When looking at the table of our detected lipids it becomes apparent that many lipids of these three classes were found. This goes along with the data from our sources; if many lipids from these classes are present, we are most likely to detect these. Some lipids, which are not present in all of the pie charts are LPC, SM and PS. These were found using our method, even though they should only be present in traces. [2, 31] Our method also yielded a lot of questionable hits of PE-O, however two PE-O hits seem plausible and have an intense and exemplary MS^2, the other PE-O hits are isobaric and their detecting bases on a single noisy MS^2 spectrum.

One of our sources also feature a table of detectable PCs, we used this table to asses our detection rate:

Source		Our data	
PC Species	pig lung lavage [mol%]	Found	Intensity [*10^-3]
[14:0 / 16:0]	9,2 ± 0,4	manually	
[16:0 / 16:0]	56 ± 1	software	572
[16:0 / 16:1]	7,5 ± 0,4	manually	
[16:0 / 18:1]	10 ± 0,6	manually	
[16:0 / 18:2]	6 ± 0,7	manually	
[18:0 / 18:2]	1,8 ± 0,2	software	12
[16:0 / 20:4]	1,6 ± 0,4	software	29
[16:0 / 22:6]	0,4 ± 0,1	manually	
[18:0 / 20:4]	0,7 ± 0,2	manually	
[18:1 / 18:1]	0,8 ± 0,2	software	12
[18:1 / 18:2]	1,2 ± 0,1	software	13
others	4,6 ± 0,5		

Table 6: Comparison of literature and our findings. [2]

It becomes apparent that all the lipids from the paper were found, however only five of the eleven lipids were found by the software. This is because the data dependent acquisition triggers somewhat randomly and we did not generate enough technical replicates, the manually found lipids were found by looking for their respective masses in three random MS1 experiments and finding their masses in every one. The information about the necessity of an abundant amount of technical replicates became apparent after this measurement series and was also hinted in a paper about a similar shotgun lipidomics approach [18] The LipidXplorer needs MS^2 spectra for detection, which also need to satisfy some criteria. Another reason might be that the first samples of pig lung lavage we received contained less lipids. Also the integrated peak intensity which is calculated by LipidXplorer seems somewhat precise, the most abundant lipid has a higher intensity than the other lipids, however the concentration of the less abundant lipids hardly fits the intensities.

This findings support following conclusions:

- PC [16:0 / 16:0] is the most abundant lipid and the intensities within one class of lipids can be directly compared to estimate their estimated content ratio.
- The data received from utilizing our method for pig lavage samples agrees with the data from literature

- We are able to detect false positive, however there is only one way of distinguishing a false positive and a true positive, if the MS^2 spectrum is proper: the biological possibility. E.g.: Does the majority of the found lipids feature an oddly numbered fatty acid chain? If yes, this detection is prone to false positives.
- Even though the method was not optimized for these samples, but for samples with a lower concentration of analyte and a more complex matrix our results are promising. This means our method is somewhat robust.
- Interpretations relying on questionable peak intensities must be checked manually. Best is to check one MS^2 spectrum for every detected lipid.
- **Our established method is fit for purpose; this means the lipids detected in unknown samples are very likely to be present in our samples.**

4.4 Shotgun lipidomics of Amnion samples

After we set up the method and used it on pig lavage samples, we tried our first analysis of amniotic supernatant samples. We analyzed four different types of amniotic cells, which are distinguished through location on the amnion and through the used medium. The location was either near the placenta or on the 'reflected area' of the placenta, so on the other side of the amniotic sac. We had four donors, which amniotic cells were cultivated three to four days, with a varying amount of time spent in the medium. We measured 3 to 4 technical replicates from one biological replicate of one donor. The cells were usually treated differently, as described the time of incubation varies between three and four days and the time spent in medium varies too. We were able to distinguish media and there is also visible difference between shortly cultivated cells (27h) and longer cultivated cells (7-day cultivation, medium: 45h).

Sadly, the intensity of the measurements had a strong variation, so that interpreting the intensity and calculating the concentration might not be possible. The media used was either a medium called DMEM or SAGM. The used medium affected the lipid-production of the cells and the way of analysis. It also effected the matrix, since the SAGM contained no lipids and the DMEM contained several lipids or lipid isobars namely:

Mass [Th]	Lipid	Intensity [1/1000]
761,6	**SM [34:1]**	50
789,6	**SM [36:1]**	11
871,7	**SM [42:2]**	24
873,7	**SM [42:1]**	7
792,6	**PC [32:0]**	12
818,6	**PC [34:1]**	71
846,6	**PC [36:1]**	52
868,6	**PC [38:4]**	18
832,6	**PC [35:1]**	54
804,6	**PE [36:1]**	66
790,6	**PS [36:0]**	4
not found in amnion samples		
790,6	**PC [32:1]**	4
840,6	**PC [36:4]**	4
870,6	**PC [38:3]**	14
832,6	**PE [38:1]**	60
802,6	**PE [36:2]**	3

Table 7: Lipid isobars found in the DMEM medium.

To determine these lipids, the screening script was used. This means that these lipids are not certified with an MS^2, the point of the experiment was to determine the background of peaks undistinguishable from lipid peaks. Hence it does not matter if these hits are verified with an MS^2 spectrum. The data was acquired by extracting and measuring the medium with our method. We used two biological replicates and ten technical replicates. The peak intensity was averaged and the standard deviation was calculated. If a lipid is found in a sample, but is in the range of the averaged with one standard deviation added it is likely to be part of the medium. These lipids were checked again, if the lipid content was higher in the supernatant of cells which were incubated longer then the cells might still produce this lipid, even though it is also part of the medium.

4.5 *Identifying Lipids in Amniotic cell supernatant*

By using the described method, we found the following lipids:

MASS [Th]	NAME	SPECIES	found in	intensity higher in medium	higher intensity with longer cultivation
609,4	PA [30:5]	PA [14:0 / 16:5]	blank		
673,5	PA [34:1]	PA [16:0 / 18:1]			
649,4	PA [33:6]	PA [16:1 / 17:5]	blank		
		PA [16:5 / 17:1]	blank		
		PA [14:2 / 19:4]	blank		
700,5	PE-O [34:2]	PE-O [16:1 / 18:1]			
722,5	PE-O [36:5]	PE-O [16:1 / 20:4]	blank		
746,5	PE-O [38:7]	PE-O [16:1 / 22:6]			
748,5	PE-O [38:6]	PE-O [18:2 / 20:4]			
750,5	PE-O [38:5]	PE-O [18:1 / 20:4]			
466,3	LPE [17:0]	LPE [17:0]	blank		
697,4	PG [31:5]	PG [15:0 / 16:5]			
761,6	SM [34:1]	SM [34:1]	medium		
789,6	SM [36:1]	SM [36:1]	medium	yes	yes
871,7	SM [42:2]	SM [42:2]	medium		
873,7	SM [42:1]	SM [42:1]	medium	yes	yes
792,6	PC [32:0]	PC [16:0 / 16:0]	medium		
818,6	PC [34:1]	PC [18:1 / 16:0]	medium	yes	yes
846,6	PC [36:1]	PC [18:0 / 18:1]	medium	yes	yes
868,6	PC [38:4]	PC [20:4 / 18:0]	medium		
832,6	PC [35:1]	PC [17:0 / 18:1]	medium	yes	yes
804,6	PE [36:1]	PE [18:1 / 18:0]	medium	yes	NA
788,5	PS [36:1]	PS [18:1 / 18:0]			
790,6	PS [36:0]	PS [18:0 / 18:0]	medium		
846,6	PS [40:0]	PS [20:0 / 20:0]			
820,6	PC [34:0]	PC [17:0 / 17:0]	spike		

Table 8: Identified lipids in the amnion samples.

The first column depicts the mass of the ion of the found lipid, under 'name' is a list of the biologically likely hits, most masses were only interpreted as one species, except for the PA [33:6]. There was only one MS^2 spectrum of said lipid, and the spectrum was very noisy, hence there were a lot of random peaks interpreted as fragments. Other biologically unlikely lipids are LPE [17:0], the precursor was also found in the blank, but the MS^2 spectrum features several fragments expected from the lipid. PG [31:5] was also found, but the MS^2 was noisy and again random peaks were interpreted, it was also found in the blank. The column titled 'found in blank' means that the precursor mass was also found in the blank with comparable intensities to the samples. The mass of PE-O [16:1 / 20:4] was found in the

blank, however the MS^2 spectrum of PE-O [16:1 / 20:4] still features both relevant fragments in acceptable intensity and supports a true positive hit.

The masses of some of the found lipids were also present in the DMEM, but not in the SAGM. Hence we had the duty to determine if the amniotic cells secreted these lipids.

4.5.1 Approach 1: distinctly higher average

Our first approach was to measure the medium several times, to estimate a background. However, the high fluctuation in intensity made it hard to determine the estimated surplus in lipid. The relative standard deviation of the 20 measurements is between 15% and 35%. One standard deviation was added to the average intensity of the lipids in the DMEM and then compared to the average of the amniotic measurements:

	DMEM:		**Amniotic samples**	
	stddev [*10^-3]	**average + 1 stddev** [*10^-3]	**PD - DMEM** [*10^-3]	**RAD - DMEM** [*10^-3]
SM [34:1]	8	57	13	2
SM [36:1]	3	14	<1	-2
SM [42:2]	4	28	3	<1
SM [42:1]	2	9	-6	-6,4
PC [32:0]	4	15	5	<1
PC [34:1]	25	95	-13	-29
PC [36:1]	15	67	-14	-31
PC [38:4]	4	22	4	-3
PC [35:1]	19	72	-48	-53
PE [36:1]	14	80	-53	-56
PS [36:0]	1	7	4	3

Table 9: First approach, many samples and their averages.

When we assume a gauss distribution of the intensity variation, the likelihood that all of green marked lipids are produced by amniotic tissue is at least 68%, because the intensities of these lipids are more intense than the medium average plus at least one standard deviation. The orange colored intensities are slightly more intense than in the medium, however not a whole standard deviation. The PC [32:0] is interesting, because the PD samples are two standard deviations above the average medium sample, however the RAD samples are not even one standard deviation above the average of the medium. This means the PC [32:0] is definitely present in the PD, PS and RAS samples but not necessarily in the

RAD. We further used data, which was measured before to verify the contradicting data of the boldly written lipids.

4.5.2 Approach 2: more lipid with longer cultivation

Another way of approaching this dilemma is, to compare the intensities of lipids, which were produced by cells of the same donor, when being in cultivation for a longer time. The boldly written lipids are covered by this approach. This approach has the advantage of being more comparable since the measurements were done in one queue and there is no comparing of experiments of different queues. The disadvantage is that we only have one donor whose tissue was treated this way and have to assume that this donor is representative.

Here we were looking at the data from cells, which were cultivated for 27h in comparison to cells which were cultivated for seven days, while the medium was changed every 45h. Most of the lipids were also found in this samples and the intensity of the lipid peaks is always higher in the longer cultivated cells.

MASS [Th]	761,58	789,61	871,69	873,71	792,57	818,59	846,63	868,60
NAME	SM [34:1]	SM [36:1]	SM [42:2]	SM [42:1]	PC [32:0]	PC [34:1]	PC [36:1]	PC [38:4]
SPECIES					PC [16:0 / 16:0]	PC [18:1 / 16:0]	PC [18:0 / 18:1]	PC [20:4 / 18:0]
RAD								
PRECURINTENS:160126_23_RAD7_45h_vom160118_1.mzML	185155,3	24889,9	96811,6	14936,3	37154,1	210708,3	40179,5	50145
PRECURINTENS:160126_23_RAD7_45h_vom160118_2.mzML	155497,4	35090,7	120105,4	5784,4	46044,2	358948,2	118393,9	90375,5
PRECURINTENS:160126_23_RAD7_45h_vom160118_3.mzML	111579	24518,2	51446,8	8874,9	31428	313636,3	89129,6	60679,8
(average/1000)_45h	**151**	**28**	**89**	**10**	**38**	**294**	**83**	**67**
std dev	**30**	**5**	**29**	**4**	**6**	**62**	**32**	**17**
PRECURINTENS:160126_23_RAD_1_27h_vom160118_1.mzML	54812,3	12244,8	43419,5	6637	16854,5	128075,5	22254,6	36221,3
PRECURINTENS:160126_23_RAD_1_27h_vom160118_2.mzML	91105,9	16301,5	50232,6	7090,1	18308,7	160761,8	87758,7	35602,9
PRECURINTENS:160126_23_RAD_1_27h_vom160118_3.mzML	82394	11869,6	29517,2	3776,6	18268,3	91417	38893,2	23356,3
(average/1000)_27h	76	13	41	6	18	127	50	32
45h-27h	**75**	**15**	**48**	**4**	**20**	**168**	**33**	**35**
factor	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**

PS								
PRECURINTENS:160126_23_PS7_45h_vom160118_1.mzML	19567,2	0	0	1959,8	0	2085,6	0	0
PRECURINTENS:160126_23_PS7_45h_vom160118_2.mzML	35255,5	0	768,6	3553,2	0	7529,6	196,1	0
PRECURINTENS:160126_23_PS7_45h_vom160118_3.mzML	40168,6	0	690,1	7936,8	552,2	9516,5	269,6	0
(average/1000)_45h	**32**	**0**	**0,5**	**4**	**0,2**	**6**	**0**	**0**
std dev	**9**	**0**	**0**	**3**	**0**	**3**	**0**	**0**
PRECURINTENS:160126_23_PS_2_27h_vom160118_1.mzML	5447,7	0	0	0	0	4966,4	0	0
PRECURINTENS:160126_23_PS_2_27h_vom160118_2.mzML	3076,4	0	0	0	0	4394,2	0	0
PRECURINTENS:160126_23_PS_2_27h_vom160118_3.mzML	4361,4	0	0	0	0	4071,4	0	0
(average/1000)_27h	4	0	0	0	0	4	0	0
45h-27h	**27**	**0**	**0,5**	**4**	**0,2**	**2**	**0,2**	**0**
factor	**7**					**1,4**		

RAS								
PRECURINTENS:160126_23_RAS7_45h_vom160118_1.mzML	803,9	0	0	0	0	2703,4	0	0
PRECURINTENS:160126_23_RAS7_45h_vom160118_2.mzML	731,2	0	0	0	0	922,5	0	0
PRECURINTENS:160126_23_RAS7_45h_vom160118_3.mzML	1882,9	0	0	0	0	2507,5	0	0
(average/1000)_45h	**1**	**0**	**0**	**0**	**0**	**2**	**0**	**0**
std dev	**1**	**0**	**0**	**0**	**0**	**1**	**0**	**0**
PRECURINTENS:160126_23_RAS_1_27h_vom160118_1.mzML	3512,5	0	0	0	1340,4	10850,7	0	0
PRECURINTENS:160126_23_RAS_1_27h_vom160118_2.mzML	7279,1	0	494,2	0	1753,4	17109,1	462,1	1465
PRECURINTENS:160126_23_RAS_1_27h_vom160118_3.mzML	4784,1	0	0	0	0	10700,2	989,9	0
(average/1000)_27h	5	0	0	0	1	13	0	0
45h-27h	**-4**	**0**	**0**	**0**	**-1**	**-11**	**0**	**0**
factor								

PD								
PRECURINTENS:160126_23_PD7_45h_vom160118_1.mzML	228063,7	25709,6	58728,1	6504,8	34775,9	329900,4	38525,4	49970,1
PRECURINTENS:160126_23_PD7_45h_vom160118_2.mzML	107425,5	19602,6	43730,3	11410,7	27963,7	252501,6	13889,9	62220,2
PRECURINTENS:160126_23_PD7_45h_vom160118_3.mzML	118392,5	22315,4	51173	13428,3	29706,1	275335,5	57358,2	42875,3
(average/1000)_45h	**151**	**23**	**51**	**10**	**31**	**286**	**37**	**52**
std dev	**54**	**2**	**6**	**3**	**3**	**32**	**18**	**8**
PRECURINTENS:160126_23_PD_2_27h_vom160118_1.mzML	69905	14328,4	34563,4	6492,8	22680,3	116543,8	31828,5	31751,4
PRECURINTENS:160126_23_PD_2_27h_vom160118_2.mzML	87416,1	20378,7	45182,9	10665,9	28622,3	216368,6	46942,2	40337
PRECURINTENS:160126_23_PD_2_27h_vom160118_3.mzML	80368,9	18810,6	44745,2	9141	26312	139987,2	29779,4	29423,4
(average/1000)_27h	**79**	**18**	**41**	**9**	**26**	**158**	**36**	**34**
45h-27h	**72**	**5**	**10**	**2**	**5**	**128**	**0**	**18**
factor	**2**	**1,3**	**1,2**	**1,2**	**1,2**	**1,8**	**1,0**	**2**

Table 10: Second approach, different cultivation times.

The PD samples show an increase of a factor of roughly two of the intensities for the lipids PC [34:1]. PC [38:4] and SM [34:1]. All other lipids show a slight increase of 1.2, except for PC [36:1] which has a factor of 1. This shows the same as the first approach, where PC [38:4] and SM [34:1] were also more intense and PC [34:1] was only half a standard deviation more intense than in the samples. This means that longer cultivation rises the intensity of all lipids, while all of these lipids (except for SM [42:1]) also feature a higher intensity in the sample measurements than in the background measurements.

The RAD samples show a trend, all of the mentioned lipids mentioned in the column of 'higher intensity with longer cultivation' show an increase of a factor of two. This is in contradiction to the first approach, where only half of signals of the marked lipids are more intense than the average of the medium. However as explained before, both approaches are built on completely different assumptions and feature different advantages and disadvantages. This means the data from the second approach somewhat supports the information about SM [34:1], SM [36:1], SM [42:2] and PC [32:0] being more present in a RAD sample then in the DMEM medium.

The samples cultivated in SAGM were also analyzed like this even though SAGM did not feature any isobaric lipids and was thus background free.

The PS samples feature a more intense lipid peak for SM [34:1] and PC [34:1] with time. All other lipids were detected with a very low intensity in 45h samples, but not at all in 27h samples, except SM [36:1], which was not present.

The RAS samples are void of most lipids, traces of PC [32:0] are present in the 27h samples, but not in the 45h samples. Other lipids have higher intensity in the 27h samples, these lipids are: SM [34:1] and PC [34:1].

It is important to note that for this analysis I often interpreted very low intensities, when using the LipidXplorer I set an intensity of 3000 [] as a threshold, however this threshold only needs to be reached once in all runs of a master scan file. Intensities lower than 3000 [] are obviously just detection of traces, however our analytes were not very concentrated and the background was high, of course the data needs to be manipulated and interpreted critically. However, this is, as I said before a screening approach.

By looking at both of this interpretations, we can conclude that the amniotic tissues do secret the found lipids.

The lipid content is depending on the area of the amniotic sac and also the used medium. Even though it is difficult to compare the cells from different media directly, because of the background of the DMEM, there are a few points which can be said.

- Lipid content is most likely highest in the placenta associated cells in DMEM (PD).
- The reflected area cells in DMEM (RAD) and the placenta associated cells in SAGM (PS) seem to have different lipid compositions, however both have a lower lipid content than PD and a higher lipid concentration than RAS.
- The reflected area tissues in SAGM (RAS) produce some lipids like the other cells however most other lipids are usually only found in traces in RAS samples.

4.6 Comparison of Amnion and pig lavage samples

Even though our first assignment was a direct comparison of the two samples, this venture is admittedly a little like comparing apples and pears. The lavage samples are background free and concentrated, it is easy to measure and interpret them, because of their high concentration and the already established knowledge of its lipid content. The amniotic samples do not only feature a very low concentration of lipids. This is due to the small amounts of cells in around 1cm² of a biopsy compared to the direct lavage, which is produced by the whole lung, which has an absurdly big area. [14: p. 112] However, there are obvious similarities and given the circumstances it can be argued that the amniotic surfactant is similar to the lung surfactant.

First of all, there are only four lipids that are exactly the same: PE-O [16:1 / 20:4], PE-O [18:1 / 20:4], PC [16:0 / 16:0], PS [18:1 / 18:0]. This does not sound like a success, however if only the found classes are compared than the amniotic cell surfactant features mostly the same lipids as the lung lavage. Namely PCs, PE(-O)s, SM and PS with only PGs missing. Further, when only the fatty acids of the lipids are compared, the lavage and the amniotic surfactant feature the same fatty acids, namely saturated and unsaturated 16 and 18 carbon atom chains, 20:4 and seldom 22:6. There are a few other fatty acids which appear, mostly 14 carbon atom chains and oddly numbered fatty acids. Most phospholipids feature two fatty

acids and even though the mentioned fatty acids are few, they can be combined in a number of ways resulting in a lot of similar but obviously different lipids of the same class. This can also be a reason that only four lipids are exactly the same, while there is still a similar lipid class and fatty acid composition.

Hence it can be argued that the two surfactants are quite similar, it has to be kept in mind that the matrix and the analyte content was very different, but not because of biological reasons. Further the natural variance in the surfactant has to be considered. By checking the featured pie charts, it is clear that the phospholipids vary, even though the surfactant has the same function in all mammals. It seems that the surfactant does not have to feature the exact same lipids to function as surfactant. The surfactant-composition is seemingly not a highly preserved mammalian feature, hence a little deviation in lipid-composition might not be a K.O. criterion for a similarity of analyzed surfactants.

5 Conclusion

While working to detect lipids with Thermo's LTQ orbitrap Velos, Advion's Triversa Nanomate and LipidXplorer we learned a lot about practical application of analytical chemistry and also were able to use theoretical knowledge to gain information about our sample. As described in the chapter about the measurements of standards we were confronted with difficult questions, which we had to answer ourselves.

5.1 Overview

We successfully analyzed the phospholipidome of two different samples and also detected trace lipids in the supplied standards. Our semi quantitative approach yielded much information, which when interpreted always leads to similar results. It was a very interesting and fun experience to setup and carry out the whole workflow from extraction to measurement to setting up a platform for automated data interpretation and also to critically check the outcome of said automated interpretation. I enjoyed the trial and error approach of all kind of settings and also the consumption of every paper that helped pushing our method further. It was also interesting to read about how phospholipidomic

experiments were carried out without using a mass spectrometer, the amount of standards and time spent for derivatization is absurdly high compared to the simplicity and data density of a single day of measuring with a high end mass spectrometer. [2]

Sadly, I was not able to conduct a thorough search about surfactants to gain the knowledge to decide if the proven similarity of lung surfactant and amniotic surfactant will suffice, so that the amniotic surfactant could work as surrogate lung surfactant. Further interpretation of our generated data has to be done by biologist and physicians.

5.2 Outlook

Further analysis of the pig lung lavage and amniotic surfactant should include a chromatography-coupled quantitative approach, which checks all masses obtained by our semi quantitative screening method. If all or most of our detected lipids are correct, the same samples can be measured again with our approach, but with less stringent settings to generate more (but less confident hits). While working on our thesis it became apparent that even though the lipidome and especially the phospholipids in lung surfactant samples were well documented, the composition of amniotic fluid is basically not examined. Hence it would be really interesting to setup a method to do quantitative phospholipidomics on all of the samples we already have. As we were able to observe, it is difficult to setup a HPLC method, which is inevitable when doing real quantitative experiments. However, there might be several papers, in which interesting HPLC methods are described. (However some of this methods seem rather forceful and might damage the column). [5] Another interesting experiment would be to use SFC and a triple quad for quantitative experiments. The later setup would complement our employed workflow perfectly, for the technology is orthogonal and the results would thereby be independent of our measurement and thus uninfluenced by any biases or artefacts.

Working on real samples with an unknown analyte was a great experience, it was an incredible feeling to employ all the knowledge I have been gathering during the past six years that I spent in the faculty of chemistry of the University of Vienna. It also feels like I learned a lot about all the aspects of the analysis and I am glad to present somewhat useful results to our sample providers Angela Lemke, Susanne Wolbank and Heinz Redl!

6 Abstract

Lipidomics is a relatively new field of research, the term itself was used first in 2003 and since then has been used to describe experiments, which aim to analyze the lipid content of a sample. Ever since the first omics sciences emerged, namely the human genome project and the human proteome project, bioinformatics has become a crucial part of bio analytics. Many would claim that the bottleneck of modern omics research is the computer science, because of the shear amount of data created by modern analysis methods, especially mass spectrometry. The need for new computer science methods is especially high in lipidomics, because firstly this is one of the newest fields of the omics sciences and secondly because of the chemical heterogeneity of lipids.

In this study we analyzed the phospholipidome of pig lung lavage and excreted lipids from two kinds of amniotic tissue, which were cultivated in different media with a shotgun lipidomics approach, normally used for screening. We did not only set up the extraction method, establish an ESI spray mode and a method for MS^2 analysis, but also installed and used an automated software for lipid identification. The used software is called LipidXplorer is freely available and is not only used by the group which programmed it, but by several other lipidomics researchers. It allows the user to define expected precursors and fragments of these precursors to filter experiments for hits.

After thorough analysis of both sample types, we compared the samples. Because of major differences in the sample matrix a direct comparison was difficult. We decided to show the found similarities and discuss the discovered lipids.

7 Zusammenfassung

Lipidomics ist eine relativ neue wissenschaftliche Disziplin, der Term wurde 2003 das erst mal verwendet und wird seither benützt um Experimente zu beschrieben, die das Ziel haben die Lipid Zusammensetzung einer Probe zu analysieren. Seit den ersten großen Omics-Projekten, wie dem *human genome project* oder dem *human proteome project*, ist Bioinformatik ein wichtiger Teil der Bioanalytik geworden. Viele behaupten, dass die Informatik ein Engpass der modernen Omics-Wissenschaften ist, alleine wegen den riesigen Datenmengen die von modernen Analysemethoden, besonders der Massenspektrometrie, erzeugt werden. Die Notwendigkeit für neue Informatikkonzepte, ist erst recht bei Lipidomics gegeben, weil erstens Lipidomics eines der neusten Omics-Wissenschaften ist und zweitens wegen der hohen Heterogenität der Lipide.

In dieser Arbeit haben wir das ‚Phospholipidome' von Schweinelungenlavage und ausgeschiedenen Lipiden von zwei Arten von amniotischen Gewebe, die in verschiedenen Medien kultiviert wurden, mit einer Shotgun Lipidomics Methode, die normallerweise für Screening Verfahren verwendet wird, untersucht. Wir haben dabei nicht nur die Extraktionsmethode aufgesetzt, eine Methode für ESI und eine Methode für die MS^2 eingeführt, sondern haben auch ein Programm für Lipid Identifikation installiert und verwendet. Die verwendete Software heißt LipidXplorer, ist gratis verfügbar und wird nicht nur von der Gruppe verwendet die sie geschrieben hat, sondern auch bei einigen anderen Lipidomics – Forschern. Sie erlaubt dem User erwartete Precursor und Fragmente zu definieren um MS-Experimente nach Hits zu filtern.

Nach gründlicher Analyse beider Proben – Arten haben wir die Proben verglichen. Da die Matrix der Proben große Unterschiede hat, war ein direkter Vergleich schwierig. Wir haben uns dazu entschieden die Ähnlichkeiten zu zeigen und die entdeckten Lipide zu erörtern.

8 Acknowledgements

Firstly, I want to thank Professor Christopher Gerner, who not only took me in for my bachelor's and master's thesis but also offered me a job as tutor. He provided guidance not only as university education but also for private decisions and views. Working with him and his research group (as a student and a tutor) has always been a lot of fun and very educational. Hence I also want to thank Rupert Mayer for his support and trust, Besnik Muqaku and Ammar Tahir for their advice on mass spectrometry and lipidomics, Andrea Bileck for her advice on creating a scientific document. I further want to thank Dr. Astrid Slany and Dr Samuel Meier, who always had an open door and helped me personally with their anecdotes of working as a bioanalytical chemist and outlook on live as a chemist. Last but not least I want to thank Dr. Mader, whose strict but professional rules for the laboratory, made it possible for all of the master and bachelor students to work together in a single laboratory without misplacing all the equipment and Ass. Prof. Günther Lamprecht, who assisted me personally, while I was still a bachelor's student and later introduced me to the opposite side of teaching, by enabling me to look behind the efforts of a simple practical analytical laboratory course. Further I want to thank Prof. Heinz Redl, Susanne Wolbank and Angela Lemke for the supplied samples.

Secondly, I want to thank Dr. Ronny Herzog, whose software I have been using. He helped me via guidance on his program and sending me papers. Creating and distributing a software for free is very altruistic and should be seen as an example for the scientific community!

Lastly, I want to thank my parents, Gabriele and Mag. Friedrich Prodinger, who supported me while studying in Vienna and always helped me however they could, be it financial support or a book recommendation, they are always there for me. I also want to thank my fiancée, Lisa Fischinger, whose kind heart, sheer endless patience and loving care are encouraging me to give my best and keep me going whenever I think I have reached a dead end.

9 Protocol for setting up LipidXplorer

As I was approached by several people, who were all interested in using LipidXplorer as an automated screening software for their data, I decided to add a short description of how to setup LipidXplorer to my master's thesis. I will only describe the installation in 64bit mod, for all modern PCs use a 64bit system. I will include links to make the installation easier, all Links were used last on 13th June 2016.

The following protocol will be similar to the online tutorial of installing LipidXplorer but more detailed and current: https://wiki.mpi-cbg.de/lipidx/LipidXplorer_Installation

9.1 Installing Python

LipidXplorer is a program written in the Python programming language, hence this language has to be installed in 64bit mod. Please visit:

https://www.python.org/getit/windows/

Download the current version of Python 2.7, it is important to choose the 64 bit mode. At this moment the latest version is: Windows x86-64 MSI installer

After installing Python windows needs to be configured, so that Python can be used. For this a path has to be set. Open the Powershell and type or copy:

```
[Environment]::SetEnvironmentVariable("Path", "$env:Path;C:\Python27", "User")
```

This will set a path, close the Powershell and open it again. Then type ' `Python` '; Now a text should be displayed saying:

```
Python 2.7.11 (v2.7.11:6d1b6a68f775, Dec  5 2015, 20:40:30) [MSC v.1500 64 bit
(AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>>
```

If the highlighted areas vary Python might have been installed in 32bit mod. Make sure to install python in 64 bit mode.

9.2 Adding relevant Packages to Python

Several packages need to be added to python. They are used to process scientific data. Most of these packages must be downloaded and installed using pythons built in feature `pip`.

9.2.1 wxPython2.8 for Python 2.7

Please download exPython2.8 from

https://sourceforge.net/projects/wxpython/files/wxPython/2.8.12.1/wxPython2.8-win64-unicode-2.8.12.1-py27.exe/download?use_mirror=netassist

This file is an executable windows file and can be installed like any familiar windows program.

9.2.2 Numpy for Python 2.7

The following packages are a little more difficult to install, because they have to be installed using python.

Firstly, download `NumPy` from http://www.lfd.uci.edu/~gohlke/pythonlibs/#numpy

It is important to check the numPy Version. The latest working version is

```
numpy-1.10.4+mkl-cp27-cp27m-win_amd64.whl
```

please download this and save it directly in `C:\` so that you can use the prepared code:

```
Python -m pip install C:\numpy-1.10.4+mkl-cp27-cp27m-win_amd64.whl
```

Copy the code with ctrl+c and paste it in the Powershell by right clicking. If you saved the data package somewhere else change the path accordingly. Python will prompt a message saying that the installation was successful.

9.2.3 Lxml for Python 2.7

The next packages which needs to be installed is lxml, which is necessary for python to process html and xml files. Download `lxml-3.4.4-cp27-none-win_amd64.whl` and save it in `C:\`. You can install the package by copying following code

```
Python -m pip install C:\lxml-3.4.4-cp27-none-win_amd64.whl
```

Python will again tell you if the installation was successful.

9.2.4 Downloading and opening LipidXplorer

If both installations were successful, you can now download LipidXplorer with this link:

https://sourceforge.net/projects/lipidxplorer/files/LipidXplorer-1.2.7.zip/download

The latest version at this point is 1.2.7.

Download the zip-file and unpack it. Start LipidXplorer by opening `LipidXplorer.py`. The filename extension `py` means that this is a python code. Now the typical LipidXplorer window should open along with a python window, which shows what the programs works.

9.3 Installing ProteoWizard

ProteoWizard is a toolkit for mass spectroscopy data. This program is needed to convert the *vendor file format* which is `raw` for Thermo products, to a more accessible *open format*. I used `mzML` for my work.

Use this link to download the program in 64bit mod for windows: http://proteowizard.sourceforge.net/downloads.shtml

ProteoWizard can do much more than just convert data, however I only used it for that purpose. To convert data, open the folder in which ProteoWizard was installed and look for a file called 'MSConvertGUI.exe'. This feature can convert data, open it and the graphic user interface will pop up. Change the settings until they look like this:

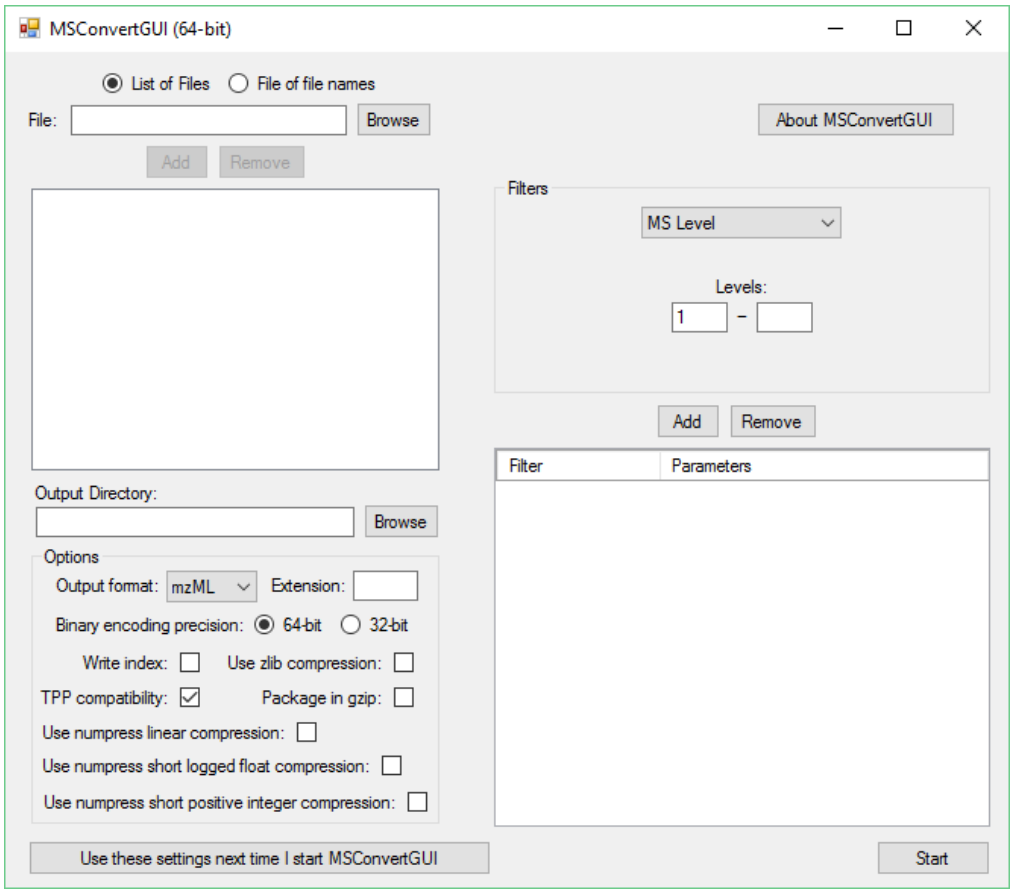

Figure 59: Settings of MSConvertGUI

Now you can choose the files you want to format in the top left corner and press start on the bottom right corner to start the conversion.

9.4 Using LipidXplorer

Using LipidXplorer has two steps, first you create a master scan file with the first two writers. Then you search this master scan file using MFQL-scripts in the third writer.

9.4.1 Creating a master scan using the converted mzML files

The mzML files can be used to create a master scan. For this you need to open LipidXplorer and choose the import source writer.

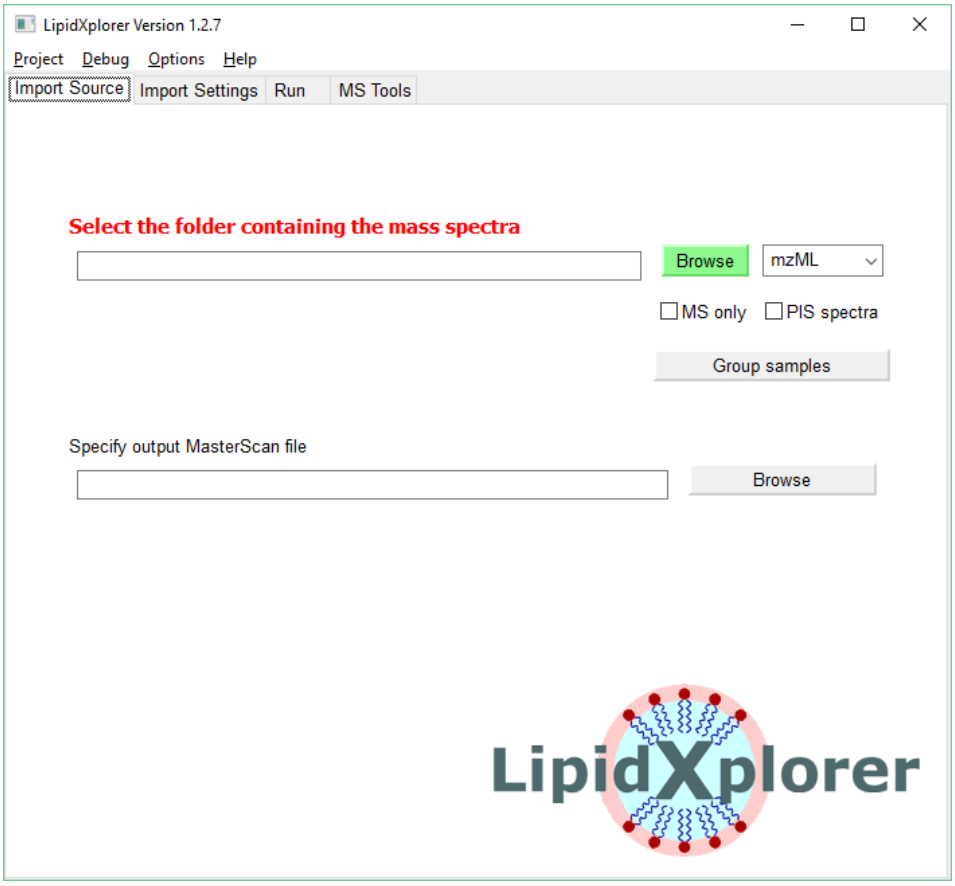

Figure 60: Import Source writer of LipidXplorer

Here you choose the directory of the mzML files you just converted and want to search for lipids. The same directory will be used to store the master scan file (file name extension `sc`), which will be automatically named after the folder holding the mzML files.

On the second writer you need to choose your import settings:

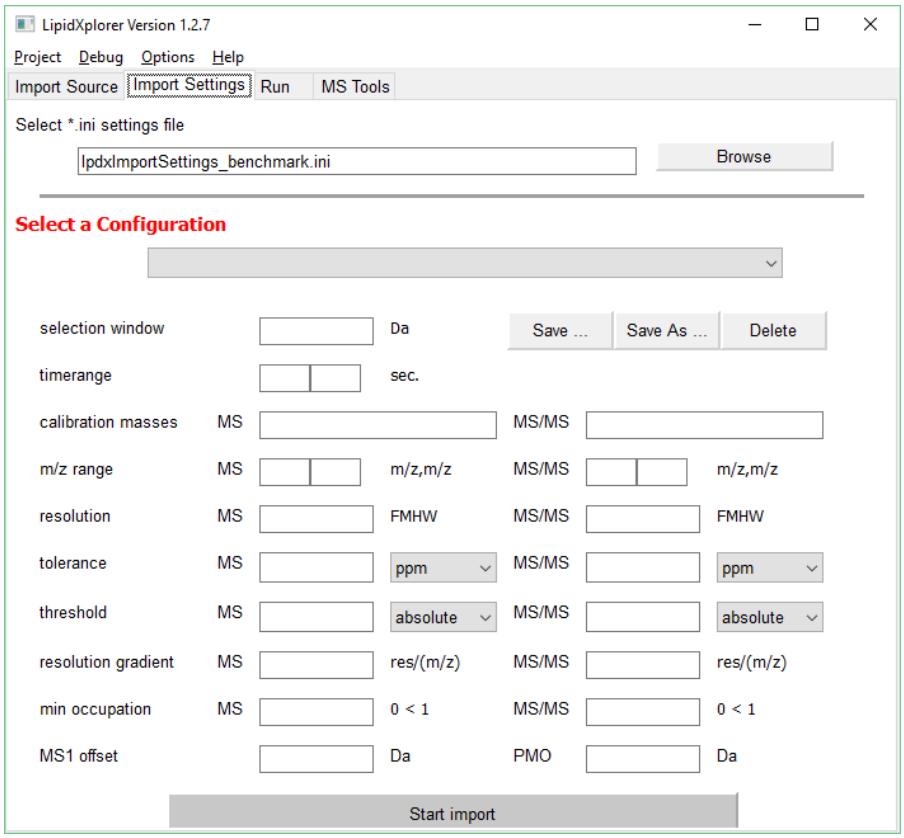

Figure 61: Second writer the Import Settings

Choose the settings according to your mass spectrometer, this is further explained in a previous chapter. By pressing 'start import' you create a master scan out of your converted data.

9.4.2 Searching your new master scan file for Lipids

After creating a master scan file, you obviously want to search it for lipids, to do this you choose the master scan file and an output directory. You also need MFQL-scripts. These are either found in the installation folder of LipidXplorer, or you can use my modified MFQL scripts, which I used for this work. After pressing 'Run LipidXplorer' the program will show the progressing of the search in a Debugging window.

The results will be saved in a csv file, which is saved in the same directory as the master scan. If the search returns no hits, there will not be a file.

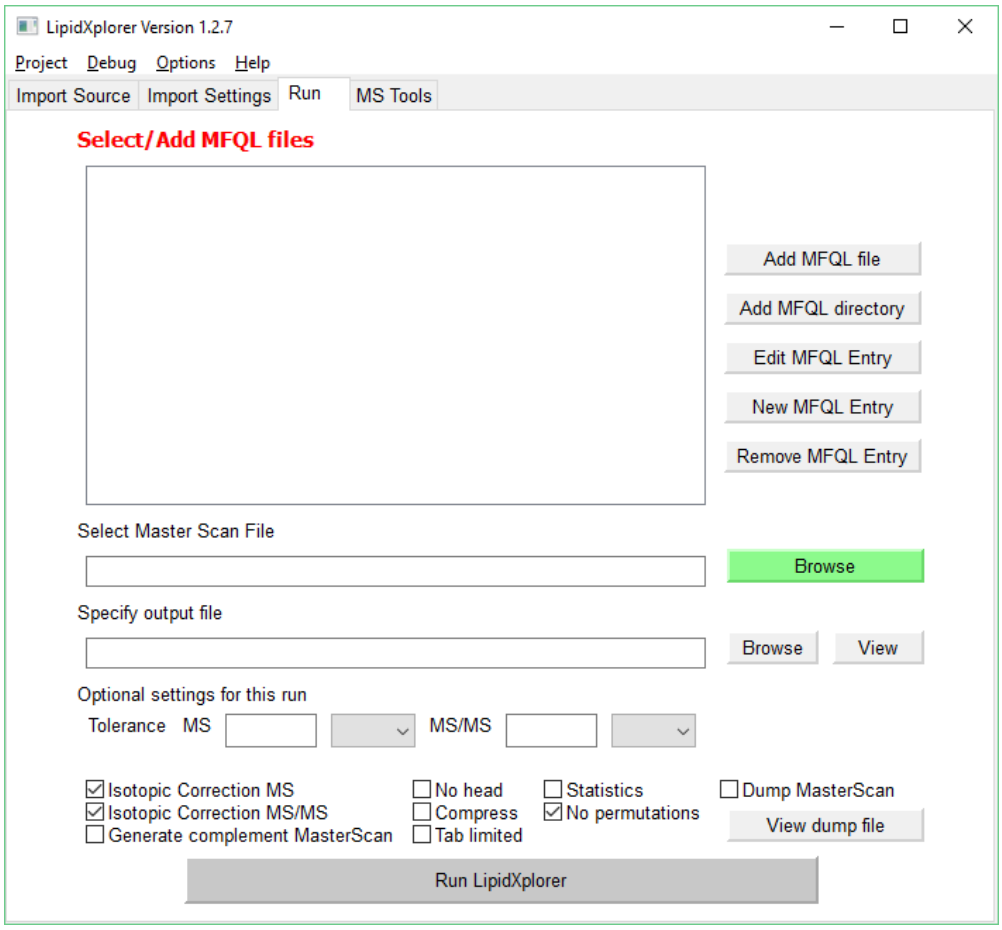

Figure 62: Run writer of LipidXplorer

9.4.3 Used MFQL-scripts

The used MFQL scripts are listed here, most of them are just copies from the files which are present when LipidXplorer is downloaded, some were changed lightly and some were written on my own. They can simply be copied and saved in an editor-like program. It is important to safe the file as with a *mfql* file format extension. Alternatively, they can be downloaded from my personal drobox here, where I also included the screening MFQL files: https://www.dropbox.com/sh/wie9gmc54u8piod/AACA5r5nBSwNp8QjDdGoGxysa?dl=0

9.4.3.1 PA with even numbered fatty acids

```
##########################################################

# Identify PA with checking  the precursor mass and FAS        #

##########################################################



QUERYNAME = PhosphatidicAcidEVEN;

DEFINE PR = 'C[31..47] H[30..120] O[8] P[1]' WITH DBR = (2.5,9.5), CHG = -1;
DEFINE FA1 ='C[12..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FA2 ='C[12..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;



IDENTIFY



       # marking

       PR IN MS1- AND

       FA1 IN MS²- AND

       FA2 IN MS²-



SUCHTHAT


isEven(FA1.chemsc[C]) AND
isEven(FA2.chemsc[C]) AND
FA1.chemsc + FA2.chemsc + 'C3 H6 P1 O4' == PR.chemsc



REPORT

       MASS = "%.4f" % (PR.mass);
       CHEMSC = PR.chemsc;
       ERROR = "%.2fppm" % (PR.errppm);
       NAME = "PA [%d:%d]" % "((PR.chemsc)[C] - 3, (PR.chemsc)[db] - 2.5)";
       SPECIES  =  "PA  [%d:%d  /  %d:%d]"  %  (FA1.chemsc[C],  FA1.chemsc[db]  -  1.5,
FA2.chemsc[C], FA2.chemsc[db] - 1.5);
       PRECURINTENS = PR.intensity;

       FAS = sumIntensity(FA1.intensity, FA2.intensity);;


################ end script #################
```

9.4.3.2 PA with oddly numbered fatty acids

```
##########################################################

# Identify PA with checking  the precursor mass and FAS        #

##########################################################
```

```
QUERYNAME = PhosphatidicAcidODD;

DEFINE PR = 'C[31..47] H[30..120] O[8] P[1]' WITH DBR = (2.5,9.5), CHG = -1;

DEFINE FA1 ='C[12..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;

DEFINE FA2 ='C[12..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;


IDENTIFY


       # marking

       PR IN MS1- AND

       FA1 IN MS²- AND

       FA2 IN MS²-


SUCHTHAT


(isOdd(FA1.chemsc[C]) OR isOdd(FA2.chemsc[C])) AND
FA1.chemsc + FA2.chemsc + 'C3 H6 P1 O4' == PR.chemsc


REPORT

       MASS = "%.4f" % (PR.mass);
       CHEMSC = PR.chemsc;
       ERROR = "%.2fppm" % (PR.errppm);
       NAME = "PA [%d:%d]" % (PR.chemsc[C] - 3, PR.chemsc[db] - 2);
       SPECIES = "PA [%d:%d / %d:%d]" % (FA1.chemsc[C], FA1.chemsc[db] - 1.5,
FA2.chemsc[C], FA2.chemsc[db] - 1.5);
       PRECURINTENS = PR.intensity;
       FAS = sumIntensity(FA1.intensity, FA2.intensity);;


################ end script #################
```

9.4.3.3 PE-O with even numbered fatty acids ionized as acetate adduct

```
############################################################

# Identify PE-O with checking the precursor mass, FAS and FA-O determination
#

############################################################


QUERYNAME = PhosphatidylethanolamineetherEVEN;

DEFINE PR = 'C[33..49] H[50..100] O[7] N[1] P[1]' WITH DBR = (1.5,8.5), CHG = -1;
DEFINE FA1 ='C[12..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FAO ='C[17..27] H[20..80] O[6] N[1] P[1]' WITH DBR = (0.5,6.5), CHG = -1;


IDENTIFY
```

```
        # marking

        PR IN MS1-  AND

        FA1 IN MS²-  AND

        FAO IN MS²-


SUCHTHAT

isEven(FA1.chemsc[C]) AND
isOdd(FAO.chemsc[C]) AND
FA1.chemsc + FAO.chemsc == PR.chemsc + 'O1 H1'
#AND (FA1.intensity + FAO.intensity) > 100


REPORT

        MASS = "%.4f" % (PR.mass);
        CHEMSC = PR.chemsc;
        ERROR = "%.2fppm" % (PR.errppm);
        NAME = "PE-O [%d:%d]" % (PR.chemsc[C] - 5 , PR.chemsc[db] - 1 ) ;
        SPECIES = "PE-O [%d:%d / %d:%d]" % (FAO.chemsc[C] - 5, FAO.chemsc[db] - 0.5,
FA1.chemsc[C], FA1.chemsc[db] - 1.5);
        PRECURINTENS = PR.intensity;
        FAS = FA1.intensity + FAO.intensity;;

#NAME  =  "PE-O  [%d:%d]"  %  (FAO.chemsc[C]  +  FA1.chemsc[C]  ,  FAO.chemsc[db]  +
FA1.chemsc[db] -2) ;
        # NAME added, removed because it is problematic



################ end script #################
```

9.4.3.4 PE-O with oddly numbered fatty acids ionized as acetate adduct

```
############################################################

# Identify PE-O with checking  the precursor mass, FAS and FA-O determination
#

############################################################



QUERYNAME = PhosphatidylethanolamineetherODD;

DEFINE PR = 'C[33..49] H[50..100] O[7] N[1] P[1]' WITH DBR = (1.5,8.5), CHG = -1;
DEFINE FA1 ='C[12..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FAO ='C[17..27] H[20..80] O[6] N[1] P[1]' WITH DBR = (0.5,6.5), CHG = -1;


IDENTIFY



        # marking
        PR IN MS1-  AND
        FA1 IN MS²-  AND
        FAO IN MS²-
```

```
SUCHTHAT

(isOdd(FA1.chemsc[C]) OR isEven(FAO.chemsc[C])) AND
FA1.chemsc + FAO.chemsc == PR.chemsc + 'O1 H1'


REPORT

MASS = "%.4f" % (PR.mass);
       CHEMSC = PR.chemsc;
ERROR = "%.2fppm" % (PR.errppm);
       NAME = "PE-O [%d:%d]" % (PR.chemsc[C] - 5 , PR.chemsc[db] - 1 ) ;
       SPECIES = "PE-O [%d:%d / %d:%d]" % (FAO.chemsc[C] - 5, FAO.chemsc[db] - 0.5,
FA1.chemsc[C], FA1.chemsc[db] - 1.5);
       PRECURINTENS = PR.intensity;
       FAS = FA1.intensity + FAO.intensity;;


################ end script ##################
```

9.4.3.5 LPE ionized as acetate adduct

```
############################################################

# Identify LPE with checking  the precursor mass AND FA  #

############################################################



QUERYNAME = lysoPhosphatidylethanolamineTEST;

DEFINE PR = 'C[19..27] H[30..70] O[7] N[1] P[1]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FA1 ='C[14..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;



IDENTIFY


       PR IN MS1- AND
       FA1 in MS²-


SUCHTHAT


FA1.chemsc + 'C5 H12 O5 N1 P1' == PR.chemsc
#AND isOdd(PR.chemsc[C])
#AND isEven(FA1.chemsc[C])

REPORT

MASS = "%4.4f" % "(PR.mass)";
       CHEMSC = PR.chemsc;
ERROR = "%2.2fppm" % "(PR.errppm)";
       NAME = "LPE [%d:%d]" % "((PR.chemsc)[C] - 5, (PR.chemsc)[db] - 1.5)";
       PRECURINTENS = PR.intensity;
       FAS = FA1.intensity;;



################ end script ##################
```

9.4.3.6 *PG with even numbered fatty acids*

```
############################################################

# Identify PG with checking  the precursor mass and FAS        #

############################################################



QUERYNAME = PhosphatidylglycerolEven;

DEFINE PR = 'C[34..50] H[30..120] O[10] P[1]' WITH DBR = (2.5,8.5), CHG = -1;
DEFINE FA1 ='C[14..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FA2 ='C[14..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;

# Second most intense Peak is PR -FA use this to make you code more specific
# FA as NL and FA as Ketone both at 10^3 too, make specific query

IDENTIFY

       PR IN MS1-  AND
       FA1 in MS²- AND
       FA2 in MS²-



SUCHTHAT
isEven(PR.chemsc[C]) AND
isEven(FA1.chemsc[C]) AND
isEven(FA2.chemsc[C]) AND
FA1.chemsc + FA2.chemsc + 'C6 H12 P1 O6' == PR.chemsc



REPORT
       MASS = "%4.4f" % "(PR.mass)";
       CHEMSC = PR.chemsc;
       ERROR = "%2.2fppm" % "(PR.errppm)";
       NAME = "PG [%d:%d]" % "((PR.chemsc)[C] - 6, (PR.chemsc)[db] - 2.5)";
       SPECIES  =  "PG  [%d:%d  /  %d:%d]"  %  "(FA1.chemsc[C],  FA1.chemsc[db]  -  1.5,
FA2.chemsc[C], FA2.chemsc[db] - 1.5)";
       PRECURINTENS = PR.intensity;
       FAS = sumIntensity(FA1.intensity, FA2.intensity);;


################ end script #################
```

9.4.3.7 *PG with oddly numbered fatty acids*

```
############################################################

# Identify PG with checking  the precursor mass and FAS        #

############################################################



QUERYNAME = PhosphatidylglycerolOdd;

DEFINE PR = 'C[34..50] H[30..120] O[10] P[1]' WITH DBR = (2.5,8.5), CHG = -1;
DEFINE FA1 ='C[14..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FA2 ='C[14..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;



IDENTIFY
```

```
        PR IN MS1-  AND
        FA1 in MS²- AND
        FA2 in MS²-



SUCHTHAT
    (isOdd(FA1.chemsc[C]) OR isOdd(FA2.chemsc[C])) AND
        FA1.chemsc + FA2.chemsc + 'C6 H12 P1 O6' == PR.chemsc



REPORT
        MASS = "%4.4f" % "(PR.mass)";
        CHEMSC = PR.chemsc;
        ERROR = "%2.2fppm" % "(PR.errppm)";
        NAME = "PG [%d:%d]" % "((PR.chemsc)[C] - 6, (PR.chemsc)[db] - 2.5)";
        SPECIES = "PG [%d:%d / %d:%d]" % "(FA1.chemsc[C], FA1.chemsc[db] - 1.5,
FA2.chemsc[C], FA2.chemsc[db] - 1.5)";
        PRECURINTENS = PR.intensity;
        FAS = sumIntensity(FA1.intensity, FA2.intensity);;



################ end script ##################
```

9.4.3.8 LPC ionized as acetate adduct

```
# Find LPCs


QUERYNAME = LysohosphatidylcholineTEST;


DEFINE PR = 'C[10..50] H[20..80] O[9] N[1] P[1]' WITH DBR = (1.5,9.5), CHG =-1;
DEFINE FA1 = 'C[10..26] H[20..60] O[2]' WITH DBR = (0.5,9), CHG=-1;
DEFINE headPC = 'C[3] H[6] O[2]' WITH CHG = 0;


IDENTIFY

PR IN MS1-
AND FA1 IN MS²-
AND headPC IN MS²-



SUCHTHAT

FA1.chemsc + 'C10 H22 O7 N1 P1' == PR.chemsc
#changed suchthat to be more specific



REPORT


MASS = "%4.4f" % "(PR.mass)";
CHEMSC = PR.chemsc;
ERROR = "%2.2fppm" % "(PR.errppm)";
NAME = "LPC [%d:%d]" % "((PR.chemsc)[C] - 10, (PR.chemsc)[db] - 1.5)";
SPECIES = "LPC [%d:%d]" % "((PR.chemsc)[C] - 10, (PR.chemsc)[db] - 1.5)";
PRECURINTENS = PR.intensity;
NLSPIS = headPC.intensity;
FAS = FA1.intensity;
```

```
;

#SPECIES ="FA [%d]" % "(FA1.chemsc[C])";

#NAME = "LPC [%d]" % "FA1.chemsc[C] ";  does not work for some reason: NAME =  "LPC
[%d]" % "PR.chemsc[C] - 10" ;
#stopps to REPORT found LPCs, when PR.chemsc[C] is changed to PR.chemsc[C] - 10
#copied the REPORT - NAME Part from PC.mfql and it works. In the middle of sections
it apparently doesn't like '#'
```

9.4.3.9 PC-O with even numbered fatty acids

This script did not yield any results; this might be because we cannot detect any PC-Os or because the code does not work, or simply because there were no PC-O present.

```
############################################################

# Identify PC-O with checking  the precursor mass, NLS, FAS #

############################################################



QUERYNAME = PhosphatidylcholineetherEVEN;

DEFINE PR = 'C[38..54] H[30..120] O[9] N[1] P[1]' WITH DBR = (1.5,8.5), CHG = -1;
DEFINE headPC = 'C[3] H[6] O[2]' WITH CHG = 0;
DEFINE FA1 ='C[14..29] H[20..70] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FAO ='C[14..29] H[20..80] O[6] N[1] P[1]' WITH DBR = (0.5,6.5), CHG = -1;



IDENTIFY

       # marking
       PR IN MS1- AND
       FA1 in MS²- AND
       FAO in MS²- AND
       headPC in MS²-



SUCHTHAT
       isEven(FA1.chemsc[C]) AND
       isEven(FAO.chemsc[C]) AND
       FA1.chemsc + FAO.chemsc + 'C10 H21 O6 P1 N1' == PR.chemsc


REPORT
       MASS = "%4.4f" % "(PR.mass)";
       CHEMSC = PR.chemsc;
       ERROR = "%2.2fppm" % "(PR.errppm)";
       NAME = "PC-O [%d:%d]" % "((PR.chemsc)[C] - 10, (PR.chemsc)[db])";
       SPECIES = "PC-O [%d:%d / %d:%d]" % "(FAO.chemsc[C], FAO.chemsc[db] - 0.5,
FA1.chemsc[C], FA1.chemsc[db] - 1.5)";
       PRECURINTENS = PR.intensity;
       NLSPIS = headPC.intensity;
       FAS = FA1.intensity + FAO.intensity;;

################ end script #################
```

9.4.3.10 PC-O with oddly numbered fatty acids

```
############################################################
```

```
# Identify PC-O with checking  the precursor mass, NLS, FAS  #

############################################################



QUERYNAME = PhosphatidylcholineetherODD;

DEFINE PR = 'C[38..54] H[30..120] O[9] N[1] P[1]' WITH DBR = (1.5,8.5), CHG = -1;
DEFINE headPC = 'C[3] H[6] O[2]' WITH CHG = 0;
DEFINE FA1 ='C[14..29] H[20..70] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FAO ='C[14..29] H[20..80] O[6] N[1] P[1]' WITH DBR = (0.5,6.5), CHG = -1;



IDENTIFY



       # marking
       PR IN MS1- AND
       FA1 in MS²- AND
       FAO in MS²- AND
       headPC in MS²-


SUCHTHAT

       (isOdd(FA1.chemsc[C]) OR isOdd(FAO.chemsc[C])) AND
       FA1.chemsc + FAO.chemsc + 'C10 H21 O6 P1 N1' == PR.chemsc


REPORT

       MASS = "%4.4f" % "(PR.mass)";
       CHEMSC = PR.chemsc;
       ERROR = "%2.2fppm" % "(PR.errppm)";
       NAME = "PC-O [%d:%d]" % "((PR.chemsc)[C] - 10, (PR.chemsc)[db])";
       SPECIES = "PC-O [%d:%d / %d:%d]" % "(FAO.chemsc[C], FAO.chemsc[db] -0.5 ,
FA1.chemsc[C], FA1.chemsc[db] - 1.5)";
       PRECURINTENS = PR.intensity;
       NLSPIS = headPC.intensity;
       FAS = FA1.intensity + FAO.intensity;;


################ end script #################
```

9.4.3.11 SM

```
############################################################

# Identify SM with checking  the precursor mass, NLS     #

############################################################



QUERYNAME = Sphingomyelin;

DEFINE PR = 'C[39..49] H[30..130] O[8] N[2] P[1]' WITH DBR = (1.5,3.5), CHG = -1;
DEFINE headSMPC = 'C[3] H[6] O[2]' WITH CHG = 0;


IDENTIFY
```

```
        # marking
        PR IN MS1- AND
        headSMPC in MS²-


SUCHTHAT

         isOdd(PR.chemsc[C])


REPORT

        MASS = "%4.4f" % "(PR.mass)";
        CHEMSC = PR.chemsc;
        ERROR = "%2.2fppm" % "(PR.errppm)";
        NAME = "SM [%d:%d]" % "((PR.chemsc)[C] - 7, (PR.chemsc)[db] - 1.5)";
        SPECIES = "SM [%d:%d]" % "((PR.chemsc)[C] - 7, (PR.chemsc)[db] - 1.5)";
        PRECURINTENS = PR.intensity;
        NLSPIS = headSMPC.intensity;;



################ end script #################
```

9.4.3.12 PC with even numbered fatty acids ionized as acetate adduct

```
############################################################

# Identify PC with checking  the precursor mass, NLS, FAS  #

############################################################



QUERYNAME = PhosphatidylcholineEven;
#changing the name of the query disturbs the search, no spacebars and _ allowed!

DEFINE PR = 'C[38..54] H[30..130] O[10] N[1] P[1]' WITH DBR = (2.5,9.5), CHG = -1;
DEFINE headPC = 'C[3] H[6] O[2]' WITH CHG = 0;
DEFINE FA1 ='C[14..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FA2 ='C[14..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;

# when deleting a DEFINE don't forget to delete it from SUCHTHAT and REPORT too!



IDENTIFY

PR IN MS1- AND
FA1 in MS²- AND
FA2 in MS²- AND
headPC in MS²-

#identify can be written like AND FA1 IN MS²-


SUCHTHAT

isEven(PR.chemsc[C]) AND
isEven(FA1.chemsc[C]) AND
isEven(FA2.chemsc[C]) AND
FA1.chemsc + FA2.chemsc + 'C10 H21 P1 O6 N1' == PR.chemsc



REPORT
```

```
        MASS = "%4.4f" % "(PR.mass)";
        CHEMSC = PR.chemsc;
        ERROR = "%2.2fppm" % "(PR.errppm)";
        NAME = "PC [%d:%d]" % "((PR.chemsc)[C] - 10, (PR.chemsc)[db] - 2.5)";
        SPECIES = "PC [%d:%d / %d:%d]" % "(FA1.chemsc[C], FA1.chemsc[db] - 1.5,
FA2.chemsc[C], FA2.chemsc[db] - 1.5)";
        PRECURINTENS = PR.intensity;
        NLSPIS = headPC.intensity;
        FAS = sumIntensity(FA1.intensity, FA2.intensity);;


################ end script #################

#indentation and rows have no effect on the script!!
#deleting some REPORT rows messes up the script. deleting ERROR,MASS, CHEMSC is fine.
```

9.4.3.13 PC with oddly numbered fatty acids ionized as acetate adduct

```
############################################################

# Identify PC with checking  the precursor mass, NLS, FAS  #

############################################################



QUERYNAME = PhosphatidylcholineOdd;


DEFINE PR = 'C[38..54] H[30..130] O[10] N[1] P[1]' WITH DBR = (2.5,9.5), CHG = -1;
DEFINE headPC = 'C[3] H[6] O[2]' WITH CHG = 0;
DEFINE FA1 ='C[14..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FA2 ='C[14..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;

# when deleting a DEFINE don't forget to delete it from SUCHTHAT and REPORT too!



IDENTIFY

PR IN MS1- AND
FA1 in MS²- AND
FA2 in MS²- AND
headPC in MS²-

#identify can be written like AND FA1 IN MS²-


SUCHTHAT

(isOdd(FA1.chemsc[C]) OR isOdd(FA2.chemsc[C])) AND
FA1.chemsc + FA2.chemsc + 'C10 H21 P1 O6 N1' == PR.chemsc



REPORT

        MASS = "%4.4f" % "(PR.mass)";
        CHEMSC = PR.chemsc;
        ERROR = "%2.2fppm" % "(PR.errppm)";
        NAME = "PC [%d:%d]" % "((PR.chemsc)[C] - 10, (PR.chemsc)[db] - 2.5)";
        SPECIES = "PC [%d:%d / %d:%d]" % "(FA1.chemsc[C], FA1.chemsc[db] - 1.5,
FA2.chemsc[C], FA2.chemsc[db] - 1.5)";
        PRECURINTENS = PR.intensity;
        NLSPIS = headPC.intensity;
        FAS = sumIntensity(FA1.intensity, FA2.intensity);;
```

```
################ end script #################

#indentation and rows have no effect on the script!!
#deleting some REPORT rows messes up the script. deleting ERROR,MASS, CHEMSC is fine.
```

9.4.3.14 PE with even numbered fatty acids ionized as acetate adduct

```
############################################################

# Identify PE with checking  the precursor mass, FAS  #

############################################################


QUERYNAME = PhosphatidylethanolamineEVENEAcetat;


DEFINE PRACETAT = 'C[33..49] H[50..100] O[10] N[1] P[1]' WITH DBR = (2.5,9.5), CHG =
-1;
DEFINE FA1 ='C[12..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FA2 ='C[12..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;


IDENTIFY

PRACETAT IN MS1- AND
FA1 IN MS²- AND
FA2 IN MS²-


SUCHTHAT


(isEven(FA1.chemsc[C]) AND isEven(FA2.chemsc[C])) AND
FA1.chemsc + FA2.chemsc + 'C7 H15 O6 N1 P1' == PRACETAT.chemsc


REPORT


       MASS = PRACETAT.mass;
       CHEMSC = PRACETAT.chemsc;
       ERROR = "%.2fppm" % (PRACETAT.errppm);
       NAME = "PE [%d:%d]" % (FA1.chemsc[C] + FA2.chemsc[C], FA1.chemsc[db] +
FA2.chemsc[db] - 2.5);
       SPECIES = "PE [%d:%d / %d:%d]" % (FA1.chemsc[C], FA1.chemsc[db] - 1.5,
FA2.chemsc[C], FA2.chemsc[db] - 1.5);
       PRECURINTENS = PRACETAT.intensity;
       FAS = sumIntensity(FA1.intensity, FA2.intensity);
;


################ end script #################
```

9.4.3.15 PE with oddly numbered fatty acids ionized as acetate adduct

```
############################################################

# Identify PE with checking  the precursor mass, FAS  #
```

```
############################################################


QUERYNAME = PhosphatidylethanolamineODDAcetat;


DEFINE PRACETAT = 'C[33..49] H[50..100] O[10] N[1] P[1]' WITH DBR = (2.5,9.5), CHG =
-1;
DEFINE FA1 ='C[12..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FA2 ='C[12..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;



IDENTIFY

PRACETAT IN MS1- AND
FA1 IN MS²- AND
FA2 IN MS²-


SUCHTHAT


(isOdd(FA1.chemsc[C]) OR isOdd(FA2.chemsc[C])) AND
FA1.chemsc + FA2.chemsc + 'C7 H15 O6 N1 P1' == PRACETAT.chemsc


REPORT


       MASS = PRACETAT.mass;
       CHEMSC = PRACETAT.chemsc;
       ERROR = "%.2fppm" % (PRACETAT.errppm);
       NAME = "PE [%d:%d]" % (FA1.chemsc[C] + FA2.chemsc[C], FA1.chemsc[db] +
FA2.chemsc[db] - 2.5);
       SPECIES = "PE [%d:%d / %d:%d]" % (FA1.chemsc[C], FA1.chemsc[db] - 1.5,
FA2.chemsc[C], FA2.chemsc[db] - 1.5);
       PRECURINTENS = PRACETAT.intensity;
       FAS = sumIntensity(FA1.intensity, FA2.intensity);
;


################ end script #################
```

9.4.3.16 PE with even numbered fatty acids ionized through loss of a proton

In our experiments the PE only ionized through addition of acetic acid, however we also tried this code, which yielded only false positives.

```
############################################################

# Identify PE with checking  the precursor mass, FAS  #

############################################################



QUERYNAME = PhosphatidylethanolamineEVENdeH;


DEFINE PR = 'C[33..49] H[50..100] O[8] N[1] P[1]' WITH DBR = (2.5,9.5), CHG = -1;
DEFINE FA1 ='C[12..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
```

```
DEFINE FA2 ='C[12..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;


IDENTIFY

PR IN MS1- AND
FA1 IN MS²- AND
FA2 IN MS²-


SUCHTHAT


(isEven(FA1.chemsc[C]) AND isEven(FA2.chemsc[C])) AND
FA1.chemsc + FA2.chemsc + 'C5 H11 O4 N1 P1' == PR.chemsc


REPORT


       MASS = PR.mass;
       CHEMSC = PR.chemsc;
    ERROR = "%.2fppm" % (PR.errppm);
       NAME = "PE [%d:%d]" % (FA1.chemsc[C] + FA2.chemsc[C], FA1.chemsc[db] +
FA2.chemsc[db] - 2.5);
       SPECIES = "PE [%d:%d / %d:%d]" % (FA1.chemsc[C], FA1.chemsc[db] - 1.5,
FA2.chemsc[C], FA2.chemsc[db] - 1.5);
       PRECURINTENS = PR.intensity;
       FAS = sumIntensity(FA1.intensity, FA2.intensity);


;

################ end script #################
```

9.4.3.17 PE with oddly numbered fatty acids ionized through loss of a proton

```
############################################################

#    Identify    PE    with    checking        the    precursor    mass,    FAS
#

############################################################


QUERYNAME = PhosphatidylethanolamineODDdeH;


DEFINE PR = 'C[33..49] H[50..100] O[8] N[1] P[1]' WITH DBR = (2.5,9.5), CHG = -1;
DEFINE FA1 ='C[12..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FA2 ='C[12..22] H[20..50] O[2]' WITH DBR = (1.5,7.5), CHG = -1;


IDENTIFY

PR IN MS1- AND
FA1 IN MS²- AND
FA2 IN MS²-


SUCHTHAT
```

```
(isOdd(FA1.chemsc[C]) OR isOdd(FA2.chemsc[C])) AND
FA1.chemsc + FA2.chemsc + 'C5 H11 O4 N1 P1' == PR.chemsc


REPORT


       MASS = PR.mass;
       CHEMSC = PR.chemsc;
       ERROR = "%.2fppm" % (PR.errppm);
       NAME = "PE [%d:%d]" % (FA1.chemsc[C] + FA2.chemsc[C], FA1.chemsc[db] +
FA2.chemsc[db] - 2.5);
       SPECIES = "PE [%d:%d / %d:%d]" % (FA1.chemsc[C], FA1.chemsc[db] - 1.5,
FA2.chemsc[C], FA2.chemsc[db] - 1.5);
       PRECURINTENS = PR.intensity;
       FAS = sumIntensity(FA1.intensity, FA2.intensity);

;


################ end script #################
```

9.4.3.18 PS with even numbered fatty acids

```
QUERYNAME = PhophatidylserineEVEN;

DEFINE PR = 'C[30..50] H[30..120] O[10] N[1] P[1]' WITH DBR = (3.5,10.5), CHG = -1;
DEFINE headPS = 'C[3] H[5] O[2] N[1]' WITH DBR = (0,2) , CHG = 0;
DEFINE FA1 = 'C[12..22] H[23..43] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FA2 = 'C[12..22] H[23..43] O[2]' WITH DBR = (1.5,7.5), CHG = -1;


#most intense peak NL of serine, second most intense is NL of serine and a FA. Use
this to enhance the code
#Achtung auch ein 74NL mit geringerer Int als NL(Serin) sichtbar!
#code entirely written by Florian Prodinger

IDENTIFY

       PR IN MS1- AND
       FA1 IN MS²- AND
       FA2 IN MS²- AND
       headPS IN MS²-


SUCHTHAT

       isEven(PR.chemsc[C]) AND
       FA1.chemsc + FA2.chemsc + 'C6 H11 P1 N1 O6' == PR.chemsc

REPORT

       MASS = "%4.4f" % "(PR.mass)";
       CHEMSC = PR.chemsc;
       ERROR = "%2.2fppm" % "(PR.errppm)";
       NAME = "PS [%d:%d]" % "((PR.chemsc)[C] - 6, (PR.chemsc)[db] - 3.5)";
       SPECIES = "PS [%d:%d / %d:%d]" % "(FA1.chemsc[C], FA1.chemsc[db] - 1.5,
FA2.chemsc[C], FA2.chemsc[db] - 1.5)";
       PRECURINTENS = PR.intensity;
       NLSPIS = headPS.intensity;
       FAS = sumIntensity(FA1.intensity, FA2.intensity);
;
#der code funktioniert und ist fertig!
```

9.4.3.19 PS with oddly numbered fatty acids

```
QUERYNAME = PhophatidylserineODD;

DEFINE PR = 'C[30..50] H[30..120] O[10] N[1] P[1]' WITH DBR = (3.5,10.5), CHG = -1;
DEFINE headPS = 'C[3] H[5] O[2] N[1]' WITH DBR = (0,2) , CHG = 0;
DEFINE FA1 = 'C[12..22] H[23..43] O[2]' WITH DBR = (1.5,7.5), CHG = -1;
DEFINE FA2 = 'C[12..22] H[23..43] O[2]' WITH DBR = (1.5,7.5), CHG = -1;

#code entirely written by Florian Prodinger
#most intense peak NL of serine, second most intense is NL of serine and a FA. Use
this to enhance the code
#Achtung auch ein 74NL mit geringerer Int als NL(Serin) sichtbar!


IDENTIFY

       PR IN MS1- AND
       FA1 IN MS²- AND
       FA2 IN MS²- AND
       headPS IN MS²-


SUCHTHAT

       (isOdd(FA1.chemsc[C]) OR isOdd(FA2.chemsc[C])) AND
       FA1.chemsc + FA2.chemsc + 'C6 H11 P1 N1 O6' == PR.chemsc

REPORT

       MASS = "%4.4f" % "(PR.mass)";
       CHEMSC = PR.chemsc;
       ERROR = "%2.2fppm" % "(PR.errppm)";
       NAME = "PS [%d:%d]" % "((PR.chemsc)[C] - 6, (PR.chemsc)[db] - 3.5)";
       SPECIES = "PS [%d:%d / %d:%d]" % "(FA1.chemsc[C], FA1.chemsc[db] - 1.5,
FA2.chemsc[C], FA2.chemsc[db] - 1.5)";
       PRECURINTENS = PR.intensity;
       NLSPIS = headPS.intensity;
       FAS = sumIntensity(FA1.intensity, FA2.intensity);
;
#der code funktioniert und ist fertig!
```

10 Sources

I tried to contact all owners of used pictures and ask their permission to use their images etc. If any claims copyright infringement should emerge, please contact me!

1.

Alberts, B., Bray, D. & Lewis, J. *Molecular Biology of the Cell*. (Taylor & Francis, 2002).

2.

Bernhard, W. *et al.* Conductive airway surfactant: surface-tension function, biochemical composition, and possible alveolar origin. *American journal of respiratory cell and molecular biology* **17,** 41–50 (1997).

3.

Deutsch, E. W. Mass Spectrometer Output File Format mzML. *Methods Mol Biol* **604,** 319–331 (2010).

4.

Ejsing, C. S. *et al.* Global analysis of the yeast lipidome by quantitative shotgun mass spectrometry. *Proceedings of the National Academy of Sciences* **106,** 2136–2141 (2009).

5.

Gregory, K. E. *et al.* Method development for fecal lipidomics profiling. *Analytical chemistry* **85,** 1114–23 (2013).

6.

Han, X. & Gross, R. W. Global analyses of cellular lipidomes directly from crude extracts of biological samples by ESI mass spectrometry a bridge to lipidomics. *J. Lipid Res.* **44,** 1071–1079 (2003).

7.

Han, X. Lipidomics: Comprehensive Mass Spectrometry of Lipids. (Wiley, 2016).

8.

Herzog, R. *et al.* LipidXplorer: A Software for Consensual Cross-Platform Lipidomics. *PLoS ONE* **7,** e29851 (2012).

9.

Herzog, R. *et al.* A novel informatics concept for high-throughput shotgun lipidomics based on the molecular fragmentation query language. *Genome Biology* **12,** R8–R8 (2011).

10.

Herzog, R., Schwudke, D. & Shevchenko, A. in *Current Protocols in Bioinformatics* (John Wiley & Sons, Inc., 2002).

11.

Kitatani, K., Idkowiak-Baldys, J. & Hannun, Y. A. The sphingolipid salvage pathway in ceramide metabolism and signaling. *Cellular Signalling* **20,** 1010–1018 (2008).

12.

Malcharek, S. Der alveolare Surfactant : Korrelation von Lipidmetabolismus, Funktion und struktureller Organisation /. *ResearchGate*

13.

Matyash, V., Liebisch, G., Kurzchalia, T. V., Shevchenko, A. & Schwudke, D. Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *Journal of lipid research* **49,** 1137–46 (2008).

14.

Mouritsen, O. G. Life - As a Matter of Fat. The Emerging Science of Lipidomics. (Springer, 2004).

15.

Ni, Z., Milic, I. & Fedorova, M. Identification of carbonylated lipids from different phospholipid classes by shotgun and LC-MS lipidomics. *Analytical and Bioanalytical Chemistry* **407,** 5161–5173 (2015).

16.

Parra, E. & Pérez-Gil, J. Composition, structure and mechanical properties define performance of pulmonary surfactant membranes and films. *Chemistry and Physics of Lipids* **185,** 153–175 (2015).

17.

Schuhmann, K. *et al.* Shotgun lipidomics on a LTQ orbitrap mass spectrometer by successive switching between acquisition polarity modes. *Journal of Mass Spectrometry* **47,** 96–104 (2012).

18.

Schwudke, D., Schuhmann, K., Herzog, R., Bornstein, S. R. & Shevchenko, A. Shotgun lipidomics on high resolution mass spectrometers. *Cold Spring Harbor perspectives in biology* **3,** a004614 (2011).

19.

A Guide to Your Genome (October 2007) - guidetoyourgenome07_vs2.pdf. Available at: https://www.genome.gov/pages/education/allaboutthehumangenomeproject/guidetoyourgenome07_vs2.pdf. (Accessed: 23rd June 2016)

20.

KEGG PATHWAY: Fatty acid biosynthesis - Reference pathway. Available at: http://www.genome.jp/kegg-bin/show_pathway?map=map00061&show_description=show. (Accessed: 23rd June 2016)

21.

LIPID MAPS Lipidomics Gateway : Lipidomics Resources. Available at: http://lipidmaps.org/resources/tutorials/lipid_tutorial.html#L. (Accessed: 24th June 2016)

22.

LIPID MAPS Tools : LIPID MAPS Lipidomics Gateway. Available at: http://www.lipidmaps.org/tools/structuredrawing/GP_p_form.php. (Accessed: 24th June 2016)

23.

LipidXplorer MFQL - LipidXplorer. Available at: https://wiki.mpi-cbg.de/lipidx/LipidXplorer_MFQL#Part_2:_The_IDENTIFY_section. (Accessed: 24th June 2016)

24.

LTQ Series Hardware Manual Version A - Man-97055-97072-LTQ-Series-Hardware-Man9705597072-A-EN.pdf.

25.

MetaCyc superpathway of fatty acids biosynthesis (E. coli). Available at: http://www.biocyc.org/META/NEW-IMAGE?type=PATHWAY&object=PWY-6285&detail-level=3. (Accessed: 23rd June 2016)

26.

orbitrap Velos Pro Hardware Manual - orbitrap-Velos-Pro-Hardware.pdf.

27.

The Human Proteome Project. Available at: http://www.thehpp.org/overview.php. (Accessed: 23rd June 2016)

28.

Thermo Fisher :: orbitrap :: orbitrap Velos Pro. Available at: http://planetorbitrap.com/orbitrap-velos-pro#tab:schematic. (Accessed: 21st February 2016)

29.

TRIVERSA IN - TriVersaNanoMateBrochure.pdf.

30.

u:search by UBW – 'lipidomics' filter: Peer-reviewed Journals. (Accessed: 24th June 2016)

31.

A. Dylik & A. Marszałek. Pulmonary surfactant – structure, metabolism and the role in RDS and ARDS treatment. *Archives of Perinatal Medicine* **16(3),** 134–139 (2010).