Mika Koskikallio

Profiling of Cloudberry Cell Cultures by Applying Metabolomics

Metropolia University of Applied Sciences

Bachelor of Laboratory Sciences

Bachelor's Thesis

17 October 2014

Metropolia

Tekijä Otsikko Sivumäärä Aika	Mika Koskikallio Lakkasoluviljelmien profilointi metabolomiikan menetelmin 29 sivua + 6 liitettä 17.10.2014
Tutkinto	Laboratorioanalyytikko
Koulutusohjelma	Laboratorioalan koulutusohjelma
Ohjaajat	FT Tuulikki Seppänen-Laakso FM Miika Kuivikko

Tämä opinnäytetyö tehtiin Teknologian tutkimuskeskukselle VTT:lle. Työn tarkoituksena oli aloittaa laitoksen projekti, jossa sovellettaisiin metabolomiikassa käytettäviä analyyttisiä menetelmiä marjasoluviljelmien koostumuksen selvittämiseen. Päällimmäisinä tavoitteina projektissa oli testata muutamalla eri liuottimella uuttaen marjasolunäytteitä ja kokeilla, soveltuisiko GCxGC-TOFMS -laitteisto tällaisten matriisien analysoimiseen.

Alustavasti työssä oli tarkoitus testata kuutta eri marjalajia, mutta lopulta tutkimus tehtiin lakan soluviljelmillä aikarajoituksien takia. Työn kannalta keskeisimpiä ongelmia olivat sopivan uuttomenetelmän löytämisen lisäksi näytteiden suuresta sokeripitoisuudesta aiheutuvien analyyttisten ongelmien eliminointi.

Työ suoritettiin yritys-erehdys -periaatteella. Näytesarjoja valmistettiin yhteensä seitsemän, joista jokaisen valmistuksessa pyrittiin korjaamaan edellisen virheet ja kehittämään menetelmää entisestään. Menetelmissä muokattiin yksityiskohtia, kuten uuttoliuoksen koostumusta ja -suhteita solukkomääriin verrattuna. Vähitellen kävi ilmi, että eri marjalajeille olisi kehitettävä omakohtaiset uutto-olosuhteet ja esikäsittelyt. Näiden kokeiden perusteella lakka valittiin sopivimmaksi tutkimuskohteeksi.

Tulokset osoittivat, että näytteenvalmistus, esimerkiksi sokereiden poistaminen näytteistä ja kvantitointiin liittyvien parametrien optimoiminen vaatisivat vielä runsaasti kehitystyötä. Tämän työn puitteissa saatiin kuitenkin tärkeää perustietoa marjasoluviljelmien koostumuksesta ja analyyttisistä kehityskohteista.

Keywords	GCxGC-TOFMS, marjasoluviljelmät, metabolomiikka, menetelmän kehitys

Metropolia

Author(s) Title Number of Pages Date	Mika Koskikallio Profiling of Cloudberry Cell Cultures by Applying Metabolomics 29 pages + 6 appendices 17 October 2014
Degree	Bachelor of Science
Degree Programme	Laboratory Sciences
Specialisation option	Chemical Analytics
Instructor(s)	Tuulikki Seppänen-Laakso, Ph.D. Miika Kuivikko, M.Sc.

This thesis project was carried out for VTT Technical Research Center of Finland. The aim of the thesis was to look into possibilities of applying the instrumentation used in metabolomics into profiling berry cell cultures. The main focuses were to test a few general solvent extraction methods on berry cells and evaluating how fitting GC-GC-TOFMS analytics would be when it comes to analyzing such sample matrices.

The project was first set to cover a total of six different berries. After the initial test runs and due to time limits however, it was decided that it would be best to focus merely on cloudberry cell culture samples. There were difficulties in figuring out a proper extraction method since the berry cell culture samples were very fatty and contained high amounts of sugars. The main issue with the extraction was to eliminate the excess of carbohydrates which hindered the profiling of the main compounds of interest.

The work was carried out by preparing successive sample sets in a manner that the preparation of each set would be an improvement over the one before. Such specifics were altered as sample amount and most importantly solvent composition. Set by set it became clear that most culture samples would require more work when it came to figuring out a proper solvent and purification. Cloudberry was chosen as the most simple and useful matrix on the basis of GCxGC-TOFMS chromatograms.

Final results from the time-of-flight detector indicated that several factors such as finding suitable internal standards and establishing sensible dilutions to perform accurate calibrations and have them studied before a reliable method for metabolomics can be developed.

Keywords	GCxGC-TOFMS, berry cell cultures, metabolomics, method development

Metropolia

Contents

Metropolia

Metropolia

Abbreviations

amu	Atomic mass unit
C17:0	Heptadecanoic acid (fatty with 17 carbon atoms and zero double bonds)
CH_2O_2	Formic acid
ECD	Electron capture detector
FID	Flame ionization detector
GC	Gas chromatography
GCxGC	Tandem gas chromatography (two-dimensional)
LC	Liquid chromatography
MeOH	Methanol
MOX	Methoxamine
MS	Mass spectrometer/spectrometry
MSTFA	N-methyl-N-(trimethylsilyl)trifluoroacetamide
PCA	Principal component analysis
TOF-MS	Time-of-flight mass spectrometry/spectrometer
VTT	Technical Research Center of Finland

1 Introduction

Berries are known throughout history as a source of good health. Berry extracts are rich in compounds that provide beneficial physiological effects such as cardioprotection, cancer growth control and blood glucose level control. Behind these effects are compounds such as antioxidants, potential allergens and other such bioactive compounds.

Generally, liquid chromatography (LC) is used when it comes to analyzing these matrices, due to its ability to analyze larger and more polar compounds than most elementary techniques such as gas chromatography (GC). While LC can be very effectively applied to organic analytics, its weakness is weak reproducibility of the chromatogram. GC, on the other hand, provides greater reproducibility. GCxGC-MS is a commonly used gas chromatographic technique in metabolomics and is able to detect and distinguish thousands upon thousands of compounds simultaneously. Coupling of mass spectrometry (MS) to chromatographic techniques is generally desirable due to the sensitivity and specificity that MS offers.

The aim of this thesis project was to establish a start on the research of the plant biotechnology team of VTT into applying the analytics used in metabolomics to analyzing berry cell cultures. The project was meant to map how directly the existing GCxGC-TOFMS methods can be applied to such analysis and how much adjusting there is to be done regarding sample preparation in order to accurately quantify and qualify desired compounds from a berry cell matrix.[1]

2 Metabolomics

Metabolomics is the study of unique chemical traces that metabolic processes leave behind in organisms. Metabolomics aims to profile these traces, known as metabolites. The collection of all metabolites in a biological sample, from a cell to an entire organism, is represented as the metabolome, the entirety of end products of cellular processes in a sample. While other analyses such as mRNA gene expression and proteomic analyses cannot give a precise description of what is happening in a sample, metabolic profiling can provide with an instantaneous snapshot of the physiology of the

sample. This can be applied to signal small yet remarkable changes in a biological system which is important, because these tiniest of fluxes through metabolic pathways can lead to large changes in metabolite concentrations and thus in the organism itself.

The field of metabolomics mainly utilizes mass spectrometric applications in its research. Tandem gas chromatography (GCxGC) coupled with time-of-flight mass spectrometry holds the possibility of identifying thousands of compounds instead of the few dozen that regular mass spectrometry allows. With the improvements in modern analytical capabilities, dedicated bioinformatics and data mining strategies, metabolomics holds enormous potential in explaining the organization of plants and how metabolic systems are simultaneously extremely controlled and highly flexible.[2]

The process of metabolomic analysis focuses on detail and reproducibility. When it comes to carrying out such analysis, there are a number of things to consider ranging from biological trails to conditions of growth of the plant or organism and potential effects of external factors such as potential abiotic or biotic hazards to the organism.

Any slight changes in these factors can alter the metabolome in its entirety. These are monitored over varying periods such as daily cycles or over the development of the entire organism. Thus the capability to record all conditions in detail is key when it comes to understanding what is going on in the organism and unlocking the potential of discovering new, unknown metabolites.[3, p. 1-4]

Changes in the metabolome are an excellent indicator of the status of physiology of an organism. Metabolites in different samples can also vary chemically. This raises the need for applications capable of measuring several metabolites at once. Mass spectrometry is a popular method because of the great sensitivity of the instruments and thus their ability to analyse compounds of various sizes simultaneously. On the other hand the preparation of the samples requires a lot of work and interpretation of the results is somewhat complicated.[4]

3 Sample Matrices

3.1 Cloudberry

Cloudberry (*Rubus chamaemorus*) is a species of raspberry (family *Rosaceae*) that grows on the northern hemisphere in the swampy areas of Alaska, northern Canada, Europe and Asia as far north as the Arctic Circle. The fruit of the berry has an orange-yellow color as is popular in the Scandinavian countries as an ingredient in the making of liqueur, pies and puddings.[5, p. 433]

Cloudberries are rich in citric acid, malic acid, α-tocopherol, anthocyanins, provitamin A carotenoid and β-carotene. The amounts vary by regions based on exposure to sun-light, rainfall and temperature[6, p. 92]. The aroma is accounted for at least 14 volatile compounds such as linalool, but most important benzyl alcohol [7, p. 361; 8, p. 351]

Figure 1 A wild *Rubus chamaemorus.*

3.2 Essential compounds

3.2.1 Fatty acids

In biochemistry, carboxylic acids with long aliphatic chains are generally called fatty acids. These can be either unsaturated or saturated, depending on whether they have one or more double bonds between individual carbon atoms or not respectively. Fatty acids are an important source of fuel in the metabolism due to the large quantities of ATP they yield when metabolized.[9. p. 1335]

Most naturally occurring fatty acids consist of an even number of carbon atoms varying from 4 to 28 atoms. The fatty acid components are generally based on three qualities: chain length, number and position of the possible double bonds and the position of the fatty acid within the glyceride molecule, if present. These characteristics determine the chemical and physical properties of fats and oils consisting of fatty acids.[10, p. 7]

The naming of fatty acids is based on a system depicting the number of carbon atoms and the number of double bonds; For example, the shorthand description according to this system for lauric acid is C-12:0, as it has 12 carbon atoms and zero double bonds.[10, p. 7]

Fatty acids can either be saturated or unsaturated. The degree of saturation depends on how many more hydrogen atoms the molecule has potential to hold. This generally affects the solidity of the fat; Saturated fatty acids are firm and the more unsaturated the fatty acid, the more liquid it appears.[11, p. 161-162]

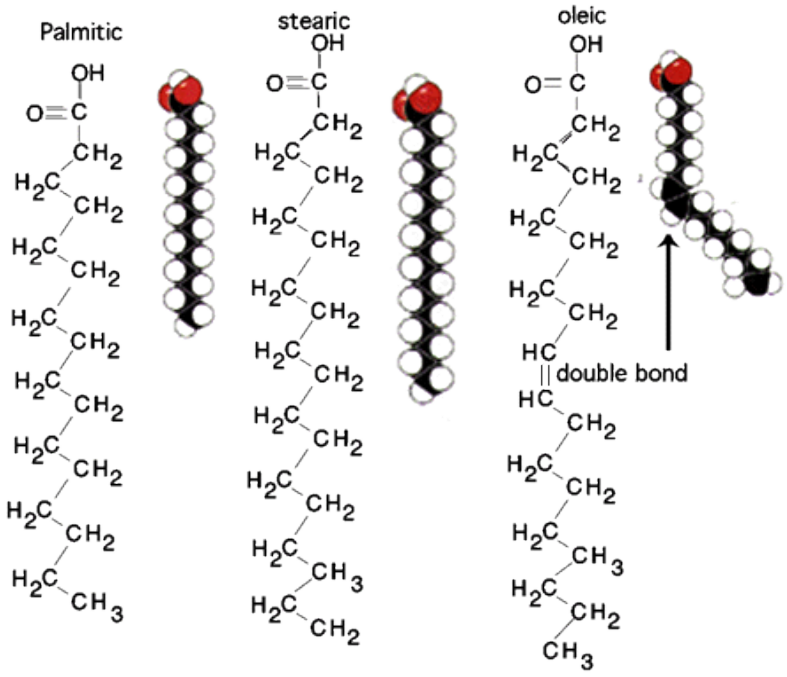


Figure 2 Saturated palmitic and stearic acids and unsaturated oleic acid with its double bond highlighted.

3.2.2 Organic acids

Organic acids are a main group of acidic organic compounds that occur in organisms and products as a result of hydrolysis, biochemical metabolism and microbial activity. The quantitative determination of these compounds is important in food products for technical and nutritional reasons.

Organic acids contain an acid group that varies between kinds of organic acid. The most common types are carboxylic acids with their basic –COOH –group. Other examples of organic compounds that can be acidic include alcohols (-OH) and the relatively strong sulfonic acids ($-SO_2OH$).[12]

Organic acids are generally used as preservatives. Some berries like lingonberry contain high amounts of organic acids, namely benzoic acid, and are thus known to remain edible for long periods of time after harvesting. This is based on the hydrophobicity of benzoic acid. Hydrophobic organic acids can interact with the lipid materials in microbial cell walls and thus disrupt microbial activity. Another decisive factor is the amount of undissociated acid present. The more dissociated the acid is, the less hydrophobic it is. The level of undissociated acid can be predicted from the pKa_1 value of the acid.[12, 13]

3.2.3 Carbohydrates

Carbohydrates are the most abundant type of organic compounds in living organisms. They originate from the endothermic reductive condensation of carbon dioxide utilizing light energy in the process known as photosynthesis. Basically carbon hydrates, carbohydrates are a major source of energy for metabolic systems. The energy-providing carbohydrates are generally known as sugars and starches. In addition to this, carbohydrates also appear as a structural material known as cellulose as well as a component of ATP and one of the three essential components of DNA and RNA.

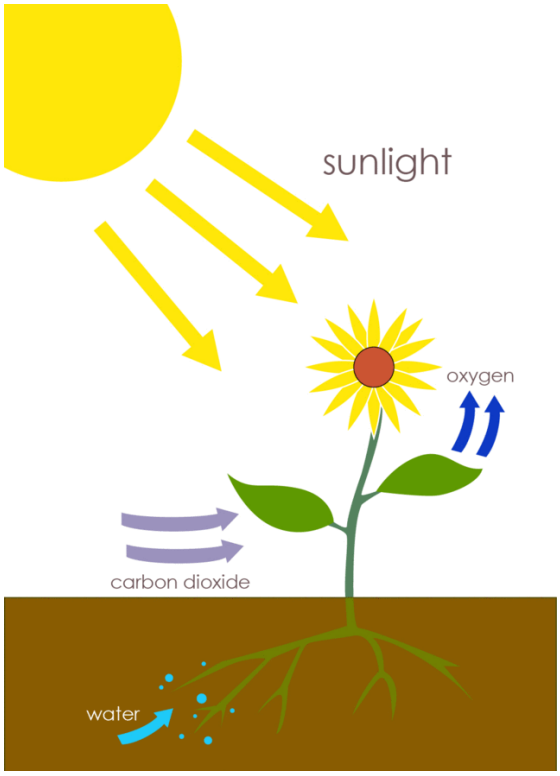


Figure 3 Photosynthesis.

Carbohydrates can be classified as either monosaccharides, oligosaccharides or polysaccharides depending on the amount of monosaccharide units. Monosaccharide is the basic sugar unit, two to ten of which can be included in an oligosaccharide unit. Polysaccharides then again can consist of hundreds of monosaccharides. The presence of hydroxyl groups in carbohydrates make these compounds highly polar allowing them to interact with aqueous environments and to participate in hydrogen bonding within and between chains. Carbohydrate derivatives can contain nitrogen, phosphate and sulfuric groups and they can also combine with lipids and proteins to form glycolipids and glycoproteins respectively.

Polysaccharides form by coining by the reactive carbonyl groups, forming oxygen bonds also known as glycosidic bonds between two saccharides. The reaction involves the remova of an –OH group from one saccharide and the H+ from another. The glycosidic bond then forms with water as byproduct.[14]

Starches are generally a form of stored carbohydrates in plant cells. Its structure is similar to that of glycogen with the difference of much less branching in the chain. Starch is synthetized from glucose-1-phosphate to ADP-glucose utilizing ATP energy.

The ADP-glucose is then added to a growing chain of glucose residues by the starch synthase enzyme creating amylose by liberating ADP. Starch branching enzyme then forms branched amylopectines by adding glucosidic bonds between the chains in the residue. Some of the branches are removed by a debranching enzyme leading to a highly complex synthesis process depending on the isoform of the enzyme.[15, p. 1-2]

3.3 Sterols

Sterols, also known as steroid alcohols, are natural compounds that occur in plants, animals and fungi. Sterols are an important part of the physiology of these organisms, where they work as for example signaling compounds between cells and as parts of the cellular membrane. Sterols are classified based on the type of organism. Phytosterols are sterols of plants and zoosterols are those of animals.

Chemically, plant sterols are a group of complex alcohol molecules that comprise of nine currently identified sterols. These compounds differ from cholesterol by the possession of an extra methyl or ethyl group on the 8-carbon side chain of cholesterol.[16, p. 100]

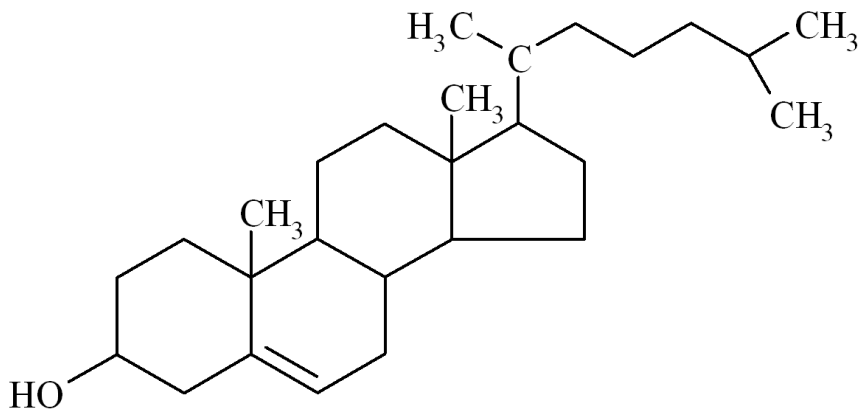


Figure 4 Cholesterol, a familiar zoosterol.

4 Sample Preparation

4.1 Derivatization

In general, samples suitable for gas chromatographic analysis need to be volatile enough to evaporate properly and to be able to withstand high temperatures. Usually ions and compounds whose molar masses exceed 500 can be troublesome to evaporate. In these cases, they can be derivatized by adding functional groups to them that affect their separation or chromatographic behavior.[17, p. 195]

4.1.1 Methylation

Organic acids are usually methylated into methyl esters by replacing a hydrogen atom with a methyl group in a process called methylation. Fatty acids are a common group of compounds to be methylated due to their large size. If the derivate reagent contains atoms such as halogens, another more selective detector such as the electron capture detector (ECD) can be used instead of the FID.[17, p. 195]

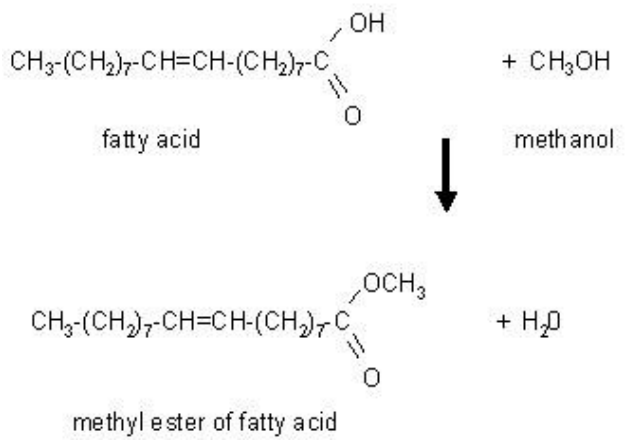


Figure 5 The principle behind the addition of a methyl group to the hydroxy group of a fatty acid.

4.1.2 Acylation

Many compounds such as carboxylic acids, alcohols and amines can be silylated. This process affects the hydroxyl groups in these compounds by replacing the hydrogen atoms with trimethylsilyl groups. This reaction is called asylation and it works by apply-

ing the N-methyl-N-(trimethylsilyl)trifluoroacetamide (MSTFA) reagent on the sample. These reagents are highly reactive and specific focus should be paid into ridding the sample of any impurities. Impurities such as water can produce new compounds in the sample that may cause various disturbances in the analyses.[17, p. 196-197]

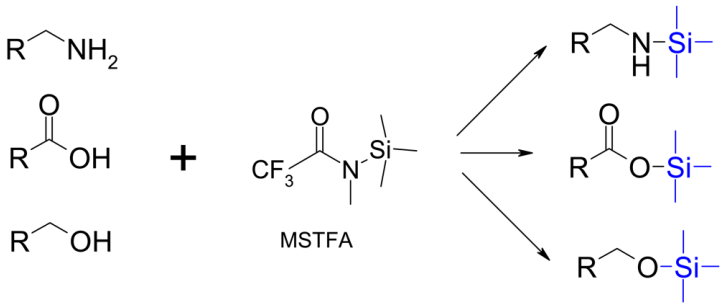


Figure 6 The reactions between MSTFA and three general compounds. On the middle row, a fatty acid is being esterified into a trimethylsilyl derivative.

4.2 Extraction by solvent

Extraction by solvent is a simple method where a solvent is used to acquire compound types of interest from a preferably homogenous sample matrix. This is done by applying a volume of solvent, such as methanol to a sample of dried, homogenized organic mass. This causes for all the components of the sample able to react with the solvent to migrate from the sample to the methanol solution, from which they can be analyzed using desired analytical methods.

In the case of using a nonpolar solvent, it is important for the sample to be dry. This is due to the possibility that residual moisture might isolate compounds of interest from reacting with the solvent. [18, p. 33-36]

4.3 Sonication

Sonication, also known as ultrasonic extraction, is a method used to enhance extraction. In the method, ultrasonic vibration is usually applied to a water bath where the sample container can be placed. The ultrasonic waves then transfer via the water bath to the solvent in the sample and cause the liquid to swirl vigorously due to the introduced energy and thus ensure intimate contact between the solvent and the sample.

Sonication is a simple and inexpensive way to efficiently enhance an extraction process. However it is rarely used alone and usually only applied as an enhancement to other methods.[19, p. 39]

5 Analytical Methods

5.1 One-dimensional gas chromatography

One-dimensional gas chromatography is the basic form of gas chromatography, where the separation process in its entirety happens in a single column. In the column, the sample passes through an either solid or, more commonly, liquid stationary phase in a gaseous mobile phase while the components of the sample travel at varying speeds causing the separation to happen. The time it takes for a compound to separate and pass through the column is called the retention time. After separation, the mobile phase carries the sample components onward to a detector that perceives the components as signals together with a data gathering unit. [17, p. 183]

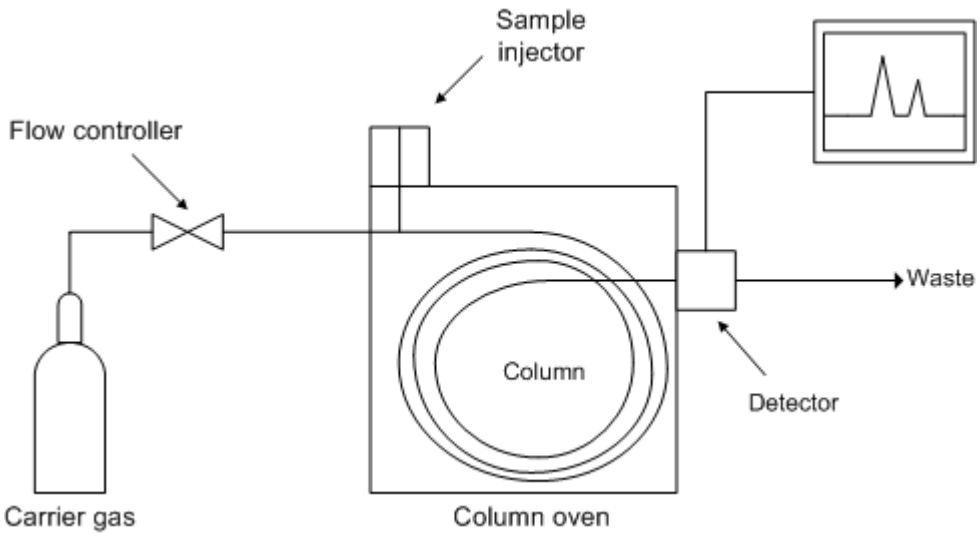


Figure 7 Scheme of a gas chromatographic system.

Before the actual separation process, the sample is injected through a gas-proof septum into a heated injector. Beyond this point the sample continues forward as a gas as the heat in the injector causes it to evaporate and mix in to the mobile phase, also known as the carrier gas.

For carrier gases, it is important for them to be clean. For example Agilent recommend a purity of at least 99,9995% for the helium, nitrogen, hydrogen and argon used as carrier gases in their instruments[1]. Minor impurities may also be cleaned with appropriate adsorbent cleaners. In addition to being clean, the carrier gas has to be inert to prevent any reactions between the carrier gas and the sample or the stationary phase. Out of the primary options, hydrogen is the most efficient option due to its chromatographic properties although safety issues have to be considered. Each carrier gas has its own characteristic optimal linear flow rate and these can be described as the van Deemter plot, which describes plate height as a function of linear flow rate. [17, p. 184-186]

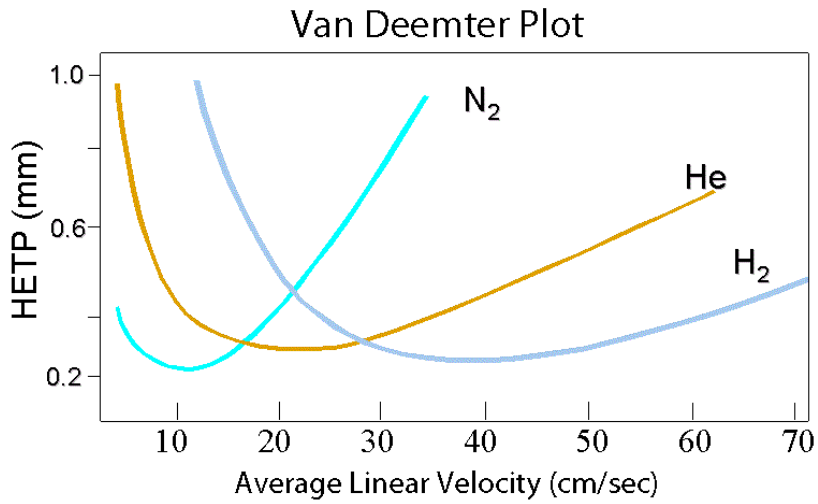


Figure 8 The van Deemter plots for nitrogen, helium and hydrogen described as plate function as a function of linear flow rate.

After mixing in with the carrier gas, the molecules move on into the column. The most common columns used in gas chromatography are 20-30 m long capillary columns.

The most important characteristics of a stationary phase compound is the ability to withstand high temperatures without evaporating and to not react with the carrier gas. Liquid stationary phases are polymers, often polysiloxanes. The structure of polysiloxanes, with the organic groups attached to their silicon atoms, define the abilities of the column itself. These groups can be used to adjust the polarity of the stationary phase and the column, which decides the kind of samples that the gas chromatograph can be used to analyze.

In gas chromatography, the column is located in a column oven, connected by graphite ferrules to the injector and the detector. These ovens are capable of heating up and cooling down fast in order to speed up the separation happening in the column. .[17, p. 190-193]

After departing the column, the components of the sample arrive at the detector that ionizes them. In gas chromatography, the common detector is a flame ionization detector (FID). It is a very widely applicable detector able to detect any compounds that form ions as a result of burning in an oxygen-rich hydrogen flame. Its sensitivity suffers greatly if the molecule contains oxygen, phosphorus, nitrogen, sulfur or halogen atoms. In the FID, the carrier gas and an approximately equal amount of hydrogen burn in an excess of air and the ions are lead to a collector electrode in an electric field and the system measures the electronic current forming. The temperature in the detector has to be high enough to avoid water and other products of burning from condensing inside.

Finally, the results are being interpreted by a computer unit. Most detectors, especially mass spectrometers, utilize a separate program for analyzing the chromatograms. Chromatograms are the final results produced by the detector, where the signals of the electric current are compiled into a graph that displays arriving ions as peaks in a two-dimensional space.[17, p. 193-194]

5.2 Two-dimensional gas chromatography

The basic idea behind two-dimensional gas chromatography (GCxGC) is the linking of two columns together. While the first column is typically the regular nonpolar one, the second one is a shorter and polar one that together allow for analyzing polar components as well as nonpolar ones. The two columns are then connected by a modulator that controls fraction temperatures between the columns.[17, p. 217]

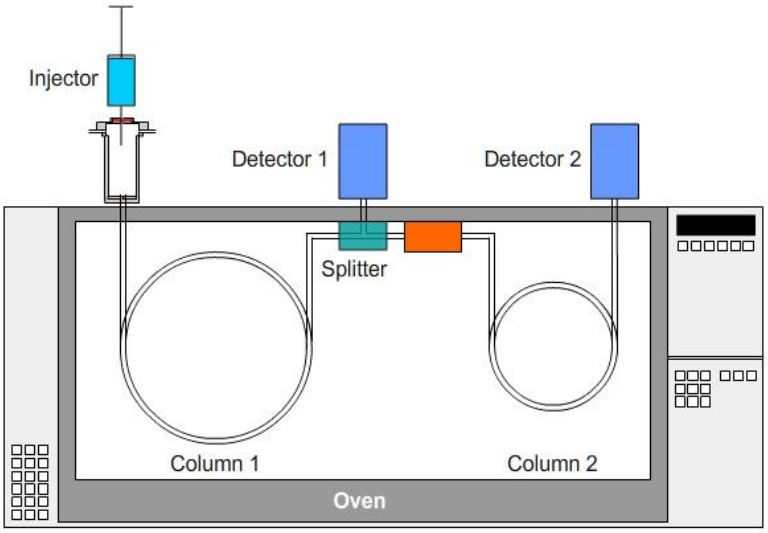


Figure 9 Scheme of a two-dimensional gas chromatographic system.

As in the one-dimensional variant, the molecules run through a septum into the injector, and eventually, the column. Instead of the entire sample arriving at the detector after the first column however, some of the fraction goes through thermal modulation. Cooling of the sample, with e.g. liquid nitrogen, focuses its fractions leading to increased sensitivity. After modulation, the fractions arrive at the second column. This column is usually polar to some extent. This allows for separation of components by polarity. Because this separation mechanism is different to the one in the first column, it is likely that the possible co-eluted components in the first column will further separate in the other. This allows for a highly increased separation efficiency compared to that of regular one-dimensional chromatographic techniques. [20, p. 15]

Due to the shortness of the second column and the narrowness of the eluting peaks, detection has to be significantly faster than in one-dimensional chromatography, all the way down to tens of milliseconds. Two-dimensional separation produces two-dimensional data leading to three-dimensional chromatograms, where the x- and y-axes are represented by the first and second columns respectively. The height of the peaks (z-axis) comes from the response of the detector where a higher peak indicates a higher concentration.[20, p. 15]

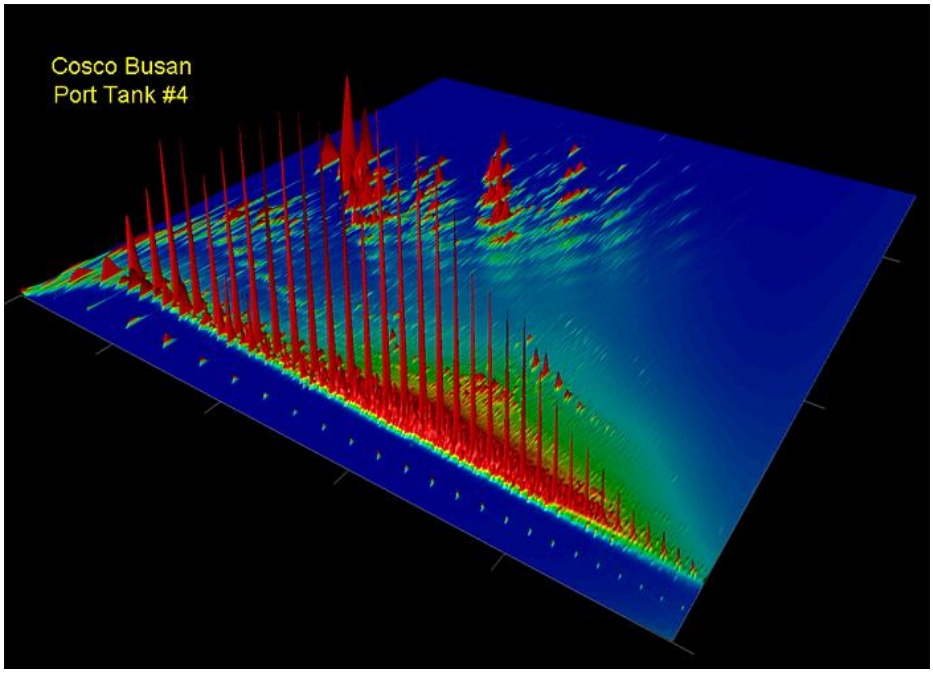


Figure 10 A three-dimensional chromatogram.

5.2.1 Comprehensive GCxGC

As opposed to conventional two-dimensional chromatography that employs a detector after the first column as well and only allows a fraction of the sample to proceed to the second column, in GCxGC the entire sample passes on into the second column. This provides a comprehensive view of all the components in the sample as each co-eluting component of the sample ends up separated.

For a two-dimensional separation to be comprehensive, Schoenmakers et al. [2003] have proposed a set of three criteria to be met:

1. Each component of the sample goes through two different separations.

2. Each component of the sample that passes through both of the two columns eventually reaches the detector.

3. The separated compounds obtained after the first column are maintained.[20, p. 17]

5.3 Time of flight mass spectrometry

Time of flight mass spectrometry (TOF) is a technique that is based on the varying velocities between ions of different masses. In the time of flight detector, ions receive an identical amount of kinetic energy when accelerated at a constant voltage. Then, according to Newton's laws of motion in formula 1, heavier ions fly slower and the detection of a mass is based on the point of arrival at the detector.

$$E = ½mv^2 \qquad (1)$$

The time of flight method achieves great resolutions. The precision of the method can be as accurate as 0,001 amu and the highest measurable m/z values can reach up to approximately 10^6 at speeds of a thousand times faster than regular GC-MS devices. Time of flight mass spectrometric methods are also rather quick and thus fit excellently with chromatographic analysis. Also very small amounts of sample are sufficient in order to be able to obtain spectra.[17, p. 127]

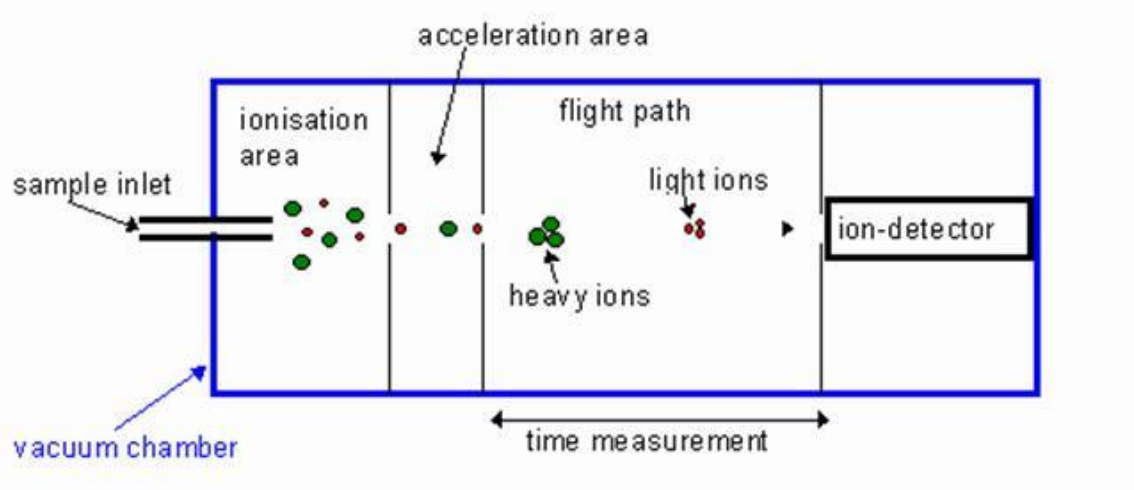


Figure 11 Scheme of a time-of-flight detector.

5.3.1 Retention indices

The Kováts retention index is a concept commonly used in quantitative chromatography as a means to describe an absolute point of retention in isothermal analyses. The basic idea behind the method is to inject a series of n-alkanes in the sample as internal standards, to whose retention times the retention times of the sample compo-

nents are compared to. This creates a quantitation method that allows for accurate repeatability for a sample in varying conditions and even different instruments.

The concept is based on the idea that each component of a sample would have two n-alkanes ($n \cdot C_nH_{2n+2}$) close to their point of retention, one earlier and one later. The retention index for a compound between two n-alkanes would then be a value relative to the distance of its corrected retention time to those of the n-alkanes.

$$RI_x = 100 \cdot \left[n + (N - n) \frac{\log\left(t'_{r(x)}\right) - \log(t'_{r(n)})}{\log\left(t'_{r(N)}\right) - \log\left(t'_{r(n)}\right)} \right] \qquad (2)$$

Where: RI = retention index, n = number of carbon atoms in the smaller n-alkane, N = number of carbon atoms in the larger n-alkane, t'_r = the adjusted retention time. [21, p. 902; 22, p. 42]

6 Statistical methods

These statistical methods are commonly connected to metabolomics research in order to help uncover connections between data sets.

6.1 Principal component analysis

Principal component analysis (PCA) is a statistical method useful for the compression and classification of data. It aims to reduce the dimensions of a data set or a sample by finding a new set of variables that retains the important parts of information of the sample while the set of variables becomes significantly smaller.

Principal component analysis incorporates linearly uncorrelated variables with their characteristic noises, variances and correlations and attempts to project these data in a manner that would maximize their final variance while losing as little information as possible. This leads to new sets of artificial variables, called principal components that reflect the original data as efficiently as possible, and in successful analysis, with much fewer dimensions.[23, p. 1-4]

6.2 T-distribution and t-test

Student's t-distribution is a statistical method based on continuous hyberbolic probability distribution that arises when estimating normally distributed populations and the standard deviation is unknown. While a normal distribution describes a full population, t-distribution is based on samples drawn from that full body of samples and the distribution is thus unique for each sample size. Therefor the larger the sample size, the more the distribution resembles a normal distribution.

A t-test is simply any statistical hypothesis test that follows this concept of distribution if the null hypothesis applies. It can be used to determine how different two sets of data are and whether this difference is significant or not. The test is performed by calculating a t-value that is compared to a limit in the t-distribution that depends on the chosen significance (usually the level of 0.05). A paired test will then approach 1.96 for this level when the sample size increases. The t-value then gains a larger value when the mean of the sample is farther from the null hypothesis and the deviation is small for the chosen sample size.[23, p. 2-3]

7 Experimental

7.1 Materials and equipment

The following instruments were used:

- Balance: Precisa EP520 A with autocalibration
- Ball mill (Retsch MM400)
- Vortex mixer (Vortex Genie 2)
- Freeze dryer (Christ Alpha 1-4 LDplus), with Edwards 5 RV5 vacuum pump

- Ultrasonic bath (Cole Parmer 8890)
- GCxGC-TOFMS: Leco Pegasus 4D with ChromaTOF software
    - 1st column Restek Rxi-5MS (10 m x 0.81 mm x 0.18 µm)
    - 2nd column SGE BPX-50 (1.5 m x 0.1 mm x 0.1 µm)
    - Carrier gas Helium (purity 4,5) at constant 40 psig
    - Method parameters for all analyses are listed in Appendix 1.

The following equipment were used:

- Syringe filters, 13 mm x 0,2 µm (VWR)
- Centrifuge filter tubes, 0.22 µm (Corning Costar Spin-X)

The following solvents and reagents were used:

Table 1 List of solvents and reagents used.

Reagent	Formula	CAS	Manufacturer	Purity
Methanol	CH_3OH	67-56-1	JT Baker	≥99,9%
Formic acid	CH_2O_2	64-18-6	Sigma-Aldrich	99,9%
Heptadecanoic acid C17:0	$C_{17}H_{34}O_2$	506-12-7	Sigma-Aldrich	
Hexane	C_6H_{14}	110-54-3	Sigma-Aldrich	≥99,9%
Ethyl acetate	$C_4H_8O_2$	64-18-6	Sigma-Aldrich	99,9%
Acetone	C_3H_6O	67-64-1	Sigma-Aldrich	99,9%

7.2 Samples

The samples were grown under the following conditions:

Cloudberry, raspberry, bilberry, cranberry, lingonberry cell cultures were obtained from the leaves of the plants and grown for 28 days for first generation and 32 days for arctic raspberry. Second generation samples were all grown for twice as long as the first generation samples.

The second generation samples were suffixed with a number and a light code (letter A-D) higher than 15 (e.g. 15C). The remainder of the samples were of first generation.

Different lighting conditions for each sample are displayed with a letter suffix (A-D) in the sample name.

Light A: LED G2, dominant in 660 nm and 440 nm, 20 W

Light B: AP67SV2, close to regular white light, 20 W

Light C: G13, dominant in 620 nm and 440 nm, 100 W

Light D: AP673L, dominant in 620 nm and 440 nm, also includes some green wavelengths, 100 W

Control: Grown in total darkness

7.3 Assay 1

Assay 1 included a single sample of each of the following berry cell cultures: Lingonberry (*Vaccinium vitis-idaea*), cloudberry (*Rubus chamaemorus*), arctic raspberry (*Rubus arcticus*), raspberry (*Rubus idaeus*), cranberry (*Vaccinium* oxycoccus) and bilberry (*Vaccinium myrtillus*). Berry callus cells were freeze dried and mechanically milled using a ball mill for 1 min at 29 Hz. 20 mg of cells was weighed accurately into test tubes and 1 ml of methanol was added into each. Each tube was agitated in a vortex apparatus for 1 min and sonicated for 20 min. The samples were then syringe-filtered and transferred into sample vials and inserts and analyzed. Analyzing included the derivatization of the samples by Methoxamine (MOX) and MSTFA by the autosampler directly before injection.

7.4 Assay 2

Assay 2 included two samples of each of the six berries. One sample of each was extracted using ethyl acetate and the other using hexane as solvent. No other steps in the process were altered from the method used in assay 1.

7.5 Assay 3

At this stage, the results for cranberry were satisfactory. Assay 3 was a minor experiment based on another thesis that suggested that the components of cloudberry could be extracted efficiently using 70% acetone. In addition to cloudberry, single samples of raspberry and arctic raspberry were prepared using methanol to repeat assay 1. The rest of the procedure was carried out identical to that of the previous assay.

7.6 Assay 4

Assay 4 began the series of applying the extraction method on to parallel samples. Each of the 4 parallels per sample was grown in different lighting conditions. This assay was an experiment of applying centrifuge filtering into the method to replace the slow syringe filtering. Four samples of raspberry and cranberry and eight of arctic raspberry (four white and four red) were extracted using methanol as solvent in volume of 1 ml per sample.

7.7 Assay 5

At this point the results indicated that all of the berry cell cultures excluding cloudberry and cranberry were not fit for the applied extraction method judging from the amount of sugars present in the chromatogram. Assays 4 and 5 again included four parallel samples per berry. Additional samples such as dark-grown specimen of cloudberry and cranberry and control samples for raspberry and both variants of arctic raspberry were added to the sequence. Eight parallels of cloudberry with varying growing times and lighting conditions were analyzed with four red and four white arctic raspberry, four raspberry and four cranberry samples. A 4% formic acid (CH_2O_2) in methanol solution was used as solvent.

7.8 Assay 6

After the promising results for cloudberry in assay 5, the analyses for both arctic raspberry, raspberry and cranberry were redone in hope of better correlation. 18 mg of

sample was weighed and extracted using 800 µl of 4% CH_2O_2 in methanol solution. Also heptadecanoic acid C17:0 was tested as an internal standard in all solutions.

7.9 Assay 7

After assay 6, all samples excluding cloudberry were discarded from the experiment due to excessive amounts of sugars present interrupting the search for desired compounds. The final assay included 3 parallel samples of each cloudberry specimen with different lighting conditions and growing times including the dark sample and zero samples.

8 Results and Discussion

Results from the TOFMS detector can be viewed in Appendix 2.

The TOFMS detector was able to detect 488 individual peaks, out of which 10 components were identified and quantified and additional 40-60 components were roughly identifiable. The general requirements for a good match were a fitting retention index value, a similarity of 80% or higher match to library spectra, presence of the component in most of the samples and concentrations within or very close to calibration area for the component.

The final runs performed on the GCxGC-TOFMS system were performed before calibration of the signals for the components could have been done. This means that several compounds with fitting matches to the library spectra were discarded due to inability to evaluate their concentration because of significant calibration-exceeding peak sizes. Among the discarded samples were both light A samples, not only due to high deviation between the two samples, but also because merely two samples are not enough to provide reliable information in terms of quantification.

Identified components are listed in green in Appendix 2. Components

The detector was unable to detect the presence of the injection standard 4,4'-dibromooctafluorobiphenyl being used, which causes trouble when it comes to approximating the reliability of the results. An unstable correlation of the injection standard would indicate potential detector-related issues. The same happened with the internal standard C17:0.

Among the identified components, desmosterol, ferulic acid, stigmastan-3,5-diene, campesterol and 4-hydroxybutanoic acid were deemed most important while employing reasonable values considering the circumstances. These components were quantified and the results are listed in Appendix 3. Based on these results, the correlations for each component could be determined. These charts are available in Appendix 4. The first two samples in each light setting are first generation samples.

Generally, the correlation charts describe the effects of each light quite well. For desmosterol, light B (AP675V2) displayed clear correlation between growing time, since the slope clearly ascends after the first two (1st generation) samples. The same can be concluded for Light D (AP673L). Light C, however, seems to descend slightly after the first two samples.

For ferulic acid, the ascending phenomenon can be observed for each of the samples For light B, though, the ascent is not as straightforward, because the first two samples and final three should produce similar results with one another and thus the slope should increase more critically as it does for lights B and D.

For stigmastan-3,5-diene, excessive exposure to light B seems to hinder the production of the compound. Stigmastan-3,5-diene being a plant steroid compound… For lights B and D, mild increases in production can be detected.

Campesterol is a problematic component due to the results showing little to no correlation. Also, the enormous amounts of the compound detected in the first two samples for light C indicate that the light source might somehow work as supplement for the presence of campesterol. For lights B and D, the results are more or less inconclusive.

In 4-hydroxybutanoic acid, again, light B seems to trigger a nulling effect due to the significant drop in its levels between the two generations. Lights C and D hardly display any conclusive information.

Additionally, the total fat concentrations were approximated for each sample. This was done by considering the peak areas of each identified fatty acid and adding them together and then that value was compared to the total peak areas of each component in the sample. This resulted in a value for each sample that would provide with a estimation of the effect of each lighting condition to the fatty acid production in the cell culture. Due to the lack of calibration, however, these results cannot be considered anything more than indicative. These values depict descending fatty acid concentrations for all lighting conditions, excluding light C, amongst whose samples the results displayed no signs of change between generations.

8.1 Statistical tests

The t-test and PCA are generally very relevant to metabolomic analysis. However, the excessive scattering of the data indicated that performing a t-test would be pointless. This was due to the fact that the highest p-values among all the compounds of interest were on the scale of 0,002 while the commonly used level of significance is 0,05. These p-values were automatically calculated by the instrument software upon data analyzing. This merely gives another perspective to the fact that the results are not nearly in correlation.

The PCA also hardly proved informative. The results can be viewed in Appendix 6. These results merely indicate that no definitive connection could be established between the data. The first and second factor represent two principal components that explain 43% and 30% of the total variability respectively. Factor 1 is correlated positively with stigmastan-3,5-diene and negatively especially with 4-hydroxybutanoic acid. Factor 2 is mostly related with desmosterol and campesterol. However, with the absence of control values, this analysis is based on absolutely peak areas so it is therefore merely suggestive, but it was able to unite with the fact that the deviation among the data was significant.

9 Conclusion

The main purpose of this study was to determine whether the existing clinical metabolomics method is suitable for analyzing berry cells. The results are clear and speak for themselves; It is not suitable for such samples and a lot of work would need to be done in order to optimize it, especially when taking into account all the different species of berries. However, the results did provide valuable information regarding what went wrong or is missing and what to focus on in the future.

The project is a somewhat extensive one and the main issues are going to be disposing of the excess of carbohydrates and figure out how to manage the triglycerids that form in the derivatization process due to some fatty acids not turning into methyl esters. It is clear that the mere extraction of the samples by solvent either will not yield wanted results or creating a functional method would require very much time. Several other methods came up during the development process, such as experimenting with the Sep Pak column to dispose of the sugars or the soxhlet extraction to determine the total fats. Total costs, however, might then skyrocket due to the amount of berry species and the sheer amount of samples this implies. Additional methods that could be considered are small things such as testing out various split ratios for the GC, or the SPME or head space techniques to determine the potential highly volatile compounds that may escape the sample container during sample preparation.

Another point worth considering is the fact that metabolomic analysis is not especially targeted to specific compounds or compound groups, but instead aims to map the entirety of the metabolome and to gather as much information as possible. In addition to this, several targeted methods could be applied to focus more into specific groups, such as HPLC for organic acids.

When it comes to quantification, detecting of the internal and injection standards becomes crucial. The fact that these compounds were not detected in the samples probably means that the clinical method used is meant for significantly lower concentrations and the standard peaks were simply lost amongst all the larger peaks and was unable to be detected. This could always be changed by increasing the amount of standard, perhaps by some ten-fold for example, but the other side of the issue is the lack of proper calibration. Generally this is done semi-automatically in the control software, but

it is done for selected compounds only because of the sheer amount of manual work the process requires in the software.

Overall, the project proved quite informative. It was able to point out many shortcomings that would need to be fixed for the method to be functional in the analyzing of the examined matrices. It is not, however, only the instrumental method that is lacking, but sample preparation methods will prove an additional problem in the future as well.

10 References

1. Pitt, James J. 2009. Principles and Applications of Liquid Chromatography-Mass Spectrometry in Clinical Biochemistry. *The Clinical Biochemist Reviews* [online] February 2009; 30(1): 19-34. Available through: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2643089/ (Accessed 4.6.2014)

2. Daviss, Bennett. 2005. Growing Pains for Metabolomics. *The Scientist* [e-journal] 25 April 2005 Available through: http://www.the-scientist.com/?articles.view/articleNo/16400/title/Growing-Pains-for-Metabolomics/ (Accessed 4.6.2014)

3. Fiehn, Oliver. 2007. Validated High Quality Automated Metabolome Analysis. In Nikolau, Basil J. and Wurtele, Eve S. ed. 2007. *Concepts in Plant Metabolomics.* The Netherlands: Springer. 1-18.

4. Suvisaari, Jaana and Orešic, Matej: Metabolomiikka psykoositutkimuksessa, Duodecim [e-journal] 21/2013. Available through: http://www.terveysportti.fi/xmedia/duo/duo11303.pdf [in Finnish] (Accessed 4.6.2014)

5. Ensminger, Marion E., Ensminger, Audrey H., 1993: *Foods and Nutrition Encyclopedia, 2nd edition*. U.S.A: CRC Press. 433. Ensminger, Audrey H., 1993.

6. Jaakkola, Mari, Korpelainen, Ville, Hoppula, Kalle and Virtanen, Vesa, 2011: Chemical composition of ripe fruits of *Rubus chamaemorus L*. grown in different habitats, *Journal of food and agriculture* [online] 92(6): 1234-1330. Available through: http://www.ncbi.nlm.nih.gov/pubmed/22083544

7. Angelino, Steven A. G. F. et al., 1991: *Volatile Compounds in Foods and Beverages*. In. Maarse, Henk ed. 1991. U.S.A.: CRC Press. 353-364. Boelens, Mans H., 1991.

8. Fox, Michael E. and Wilcock, Robert J., 1988. Soil, Water and Vegetation Residues of 2,4,5-T applied to a New Zealand Hillside. In. Stanislawek, Sandra D. and Sutherland, Gillian P. ed. 1988. *New Zealand journal of agricultural research* [online] Jan 2012, 31(3). Available through: http://www.tandfonline.com/doi/pdf/10.1080/00288233.1988.10423425 (Accessed 4.6.2014)

9. IUPAC Commission on the Nomenclature of Organic Chemistry and Commission on Physical Organic Chemistry. 1995. *Glossary of Class Names of Organic Compounds and Reactive Intermediates Based on Structure*. IUPAC Recommendations 1995. Prepared for publication by G.P. Moss, P.A.S. Smith, D. Tavernier. Pure Appl. Chem. 67, 1307-1375.

10. O'Brien, Richard D. 2009: *Fats and Oils: Formulating and Processing for Applications*, 3rd edition U.S.A.: CRC Press.

11. Sizer, Francis and Whitney, Ellie. 2014: *Nutrition: Concepts and Controversies*, 13th edition U.S.A.: Cengage Learning.

12. Bartek Ingredients Inc. *Organic Acids as Antimicrobial Agents* [online] Bartek Ingredients Inc. Available through: http://www.bartek.ca/pdfs/Applications/AntimicrobialPreservation/Activity/Organic%20Acids%20as%20Antimicrobial%20Agents.pdf (Accessed 4.6.2014)

13. Heinonen, Marina. 2009. Orgaanisilla hapoilla on monta merkitystä marjoissa ja marjatuotteissa. *Kehittyvä Elintarvike* [e-journal] 3/2009. Available through: http://kehittyvaelintarvike.fi/teemajutut/20-orgaanisilla-hapoilla-on-monta-merkitysta-marjoissa-ja-marjatuotteissa [in Finnish] (Accessed 4.6.2014)

14. King, Michael W. 2013. *Carbohydrates* [online] Available through: http://themedicalbiochemistrypage.org/carbohydrates.php (Accessed 4.6.2014)

15. American Society of Plant Physiologists. 1995: Starch Biosynthesis. *The Plant Cell, Vol 7.* U.K.: American Society of Plant Physiologists. Martin, Cathie and Smith, Alison M. 971-985.

16. Armellini, Fabio et al. 1996.: *Handbook of Lipids in Human Nutrition*, In. Spiller, Gene A. ed. 1996. U.S.A.:CRC Press. Simopoulos, Artemis P. 91-100.

17. Jaarinen, Soili and Niiranen, Jukka. 2008: *Laboratorion analyysitekniikka, 5th-6th edition*. Finland: Edita Publishing Oy and authors.

18. Lampi, Anna-Maija; Piironen, Vieno and Toivo, Jari. 2004.: Analysis of Phytosterols in Foods. *Phytosterols as Functional Food Components and Nutraceuticals*. In Dutta, Paresh C. ed. 2004. U.S.A. Marcel Dekker Inc. Lampi, Anna-Maija; Piironen, Vieno and Toivo, Jari. Ch. 2.

19. De La Guardia, Miguel and Armenta, Sergio. 2011.: Green Analytical Chemistry: Theory and Practice. *Comprehensive Analytical Chemistry, Vol. 57*. In. Barcelo, D. ed. 2011. U.K.: Elsevier B.V.

20. Kallio, Minna. 2008: *Comprehensive Two-Dimensional Gas Chromatography: Instrumental and Methodological Development*. Ph.D. University of Helsinki, Department of Chemistry, Laboratory of Analytical Chemistry.

21. Aboul-Enein, Hassan Y. et al. 2005: *Encyclopedia of Chromatography, 2nd edition, Vol. 2*. In. Cazes, Jack. ed. 2005. U.S.A: CRC Press. Zenkevich, Igor G. 901-907.

22. Kaliszan, Roman. 1997: *Structure and Retention in Chromatography: A Chemometric Approach*. The Netherlands: OPA Amsterdam B.V.

23. Shlens, Johathon. 2005: *A Tutorial on Principal Component Analysis* [online] Available through: http://www.ics.forth.gr/mobile/pca.pdf (Accessed 6.10.2014)

24. Hurst, Simon. 1995: *The Characteristic Function of the Student t Distribution.* Available through: http://maths-intranet.anu.edu.au/research.reports/srr/95/044/SRR95-044-scan.pdf (Accessed 6.10.2014)

10.1 Images

1. Lapland Food. n.d. *Lapland Food – Cloudberry (Rubus Chamaemorus) [fin. lakka]* [image online] Available through: http://www.laplandfood.fi/uploads/images/kuvitus/cloudberry.jpg (Accessed 6.10.2014)

2. Brooklyn College. n.d. *Fatty Acids* [image online] Available through: http://academic.brooklyn.cuny.edu/biology/bio4fv/page/3fatty.gif (Accessed 6.10.2014)

3. Wikipedia. n.d. *Photosynthesis* [image online] Available through: http://upload.wikimedia.org/wikipedia/commons/thumb/d/db/Photosynthesis.gif/553px-Photosynthesis.gif (Accessed 6.10.2014)

4. DNA Learning Center Blog. n.d. *Bad Cholesterol* [image online] Available through: http://blogs.dnalc.org/wp-content/uploads/2012/04/cholesterol.png (Accessed 6.10.2014)

5. NC State University, College of Agriculture and Life Sciences. n.d. *Herbicide Classification and Mode of Action.* [image online] Available through: http://courses.cropsci.ncsu.edu/cs414/cs414_web/CH_6_2005_files/image011.jpg (Accessed 6.10.2014)

6. MDPI AG. 2011: *Alkylation or Silylation for Analysis of Amino and Non-Amino Organic Acids by GC-MS?* [image online] Available through: http://www.mdpi.com/metabolites/metabolites-01-00003/article_deploy/html/images/metabolites-01-00003f1-200.png (Accessed 6.10.2014)

7. Wikipedia. n.d. *Gas Chromatography* [image online] Available through: http://upload.wikimedia.org/wikipedia/commons/8/87/Gas_chromatograph.png (Accessed 6.10.2014)

8. Chromablography. 2011: *Fast(er) GC: How to Decrease Analysis Time using Existing Instrumentation? Part IV: Using Hydrogen as the Carrier Gas* [image online] Available through: http://blog.restek.com/wp-content/uploads/2011/09/Fast-IV-fig.2-300x225.gif (Accessed 6.10.2014)

9. Chromacademy. n.d. *Understanding 2D Gas Chromatography* [image online] Available through: http://www.chromacademy.com/essential-guide/sept_2011/2d_gc_figure_01.jpg (Accessed 6.10.2014)

10. Ocean Leadership. 2012. *Tracking Toxic Chemicals in Oil Spills* [image online] Available through: http://oceanleadership.org/wp-content/uploads/2012/02/Chrom_700_204455.jpeg (Accessed 6.10.2014)

11. Kore Technology. n.d. *Converging Annular Time-of-Flight Mass* Spectrometer [image online] Available through: http://www.kore.co.uk/graphics/MS-200_tof.gif (Accessed 6.10.2014)

1. GCxGC-TOFMS method parameters

Gathered spectrum: 45-700 amu at 100/s

Injection: 1.0 µl split injection at 1:10 at 240°C

Oven temperature program:

1st column 50 °C, 2 min, 7 °C/min → 240 °C, 25 °C/min → 300 °C, 6 min

2nd column 20 °C higher than 1st column for each step

Autosampling: Gerstel MPS autosampler

1. 25 µl MOX, 45°C, 60 min
2. 25 µl MSTFA, 45°C, 60 min

   -The addition of MOX included the following alkanes for determining of retention indices: C11, C15, C17, C21 and C25

3. 50 µl Hexane with injection standard 4,4'-dibromooctafluorobiphenyl

Each sample was derivatized individually and injected immediately upon the completion of the sample.

2. TOFMS results, identification

Selection	ID	RT1	RT2	RTI	N Found	Max similarity	Mean similarity	Similarity std dev	Metabolite name	True name
TRUE	391	1386	1,697	2014,083	6	680	618,5	85,598	Unknown	Sterol
TRUE	96	701	1,314	1321,1	16	962	928,063	21,52	Butane, 1,2,3-tris(trimethylsiloxy)-	1,2,3-Trihydroxybutane
TRUE	138	1031,333	1,815	1620,433	16	820	785,6	37,416	Silane, [1,2,3-benzenetriyltris(oxy)]tris[trimethyl-	1,2,3-Trihydroxybenzene
TRUE	34	2026	3,354	2867,8	16	950	937,25	9,32	*Campesterol, TMS	Campesterol
TRUE	275	1028,143	1,534	1616,943	7	754	724	21,962	Unknown	Amino acid
TRUE	183	710	2,09	1329,36	5	527	505	13,172	Unknown	Carboxylic acid
TRUE	128	1011	1,766	1600	16	917	898,25	20,789	Pentanedioic acid, 2-[(trimethylsilyl)oxy]-, bis(trimethylsilyl) ester	2-Hydroxyglutaric acid
TRUE	452	1786	1,81	2540,4	8	695	649,75	28,991	Unknown	Hydroxy acid
TRUE	429	1611	2,658	2304,3	5	629	609,4	14,398	Unknown	Hydroxy acid
TRUE	102	622,333	1,602	1254,42	15	917	863,333	43,428	Butanoic acid, 4-[(trimethylsilyl)oxy]-, trimethylsilyl ester	4-Hydroxybutanoic acid
TRUE	510	2101	4,583	2971,1	16	798	778,5	9,606	Unknown	Sterol
TRUE	332	1216,667	3,773	1822,96	15	644	631,467	5,914	Unknown	Phenolic acid
TRUE	35	2081	4,25	2943,533	9	822	804,778	9,602	*Desmosterol, TMS	Desmosterol
TRUE	172	626	2,155	1258	10	595	571,9	21,646	Unknown	Carboxylic acid
TRUE	432	1626	2,354	2324,2	5	696	668,4	26,595	Unknown	Hydroxy acid
TRUE	5	716,938	1,981	1335,188	16	938	887,875	50,541	Succinic acid, 2TMS	Succinic acid
TRUE	9	532,875	1,463	1178,188	16	966	868,5	108,204	3-Hydroxybutyric acid, 2TMS	3-Hydroxybutyric acid
TRUE	513	2116	1,19	2986,833	6	934	820,833	72,411	Unknown	Fatty acid
TRUE	400	1418,188	2,588	2051,85	16	738	711,188	31,718	Unknown	Amine
TRUE	417	1547	1,608	2216,14	5	568	533,6	31,722	Unknown	Hydroxy acid
TRUE	36	1463,273	2,718	2104,145	11	897	769,091	102,903	*Ferulic acid, 2TMS	Ferulic acid
TRUE	139	1966	3,316	2786,4	12	823	799,167	12,394	Stigmastan-3,5-diene	Stigmastan-3,5-diene
TRUE	268	1001	1,754	1590	16	750	740,5	11,95	Unknown	Hydroxy acid
TRUE	515	2126,625	4,734	3006,038	8	777	727,5	60,696	Unknown	Alcohol

Table legend

Selection: Compound/functional group identified true/false.

ID: Compound ID in the database/library.

RT 1 and 2: Retention times in columns one and two.

RTI: Retention index.

N Found: Compound identified in N samples.

Similarity: The similarity of the chromatogram, maximum, mean and standard deviation values amongst samples.

Metabolite name: Exact name of the identified compound. (IUPAC name)

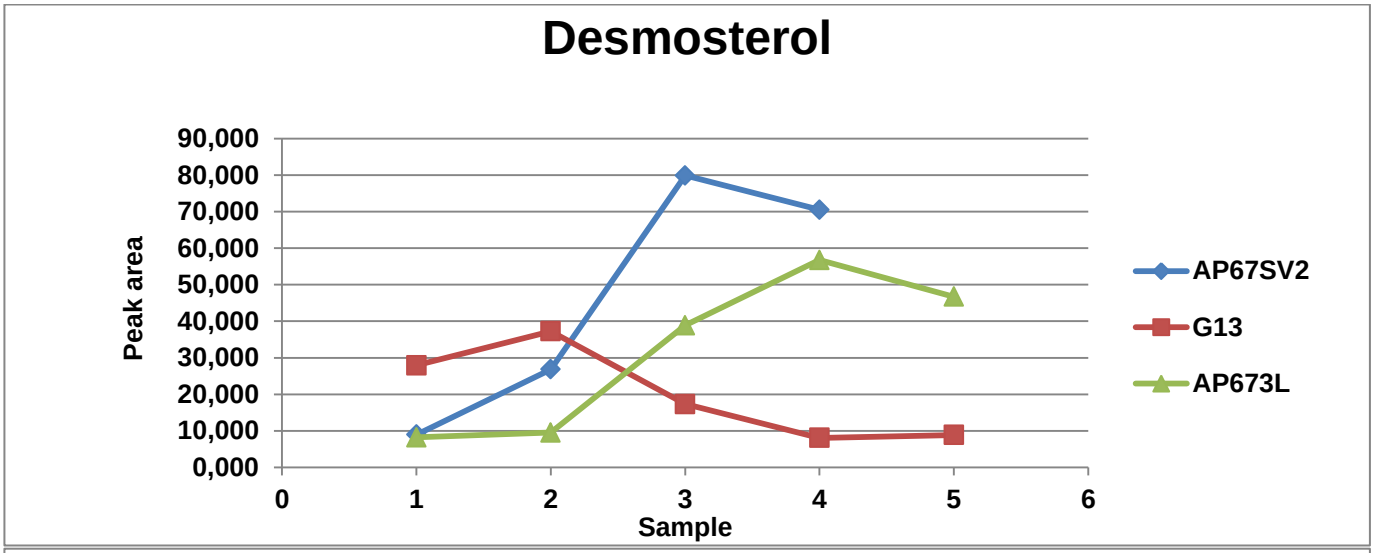
True name: Real name of the identified compound or functional group. (Distinguishes the original component from the derivatized form)

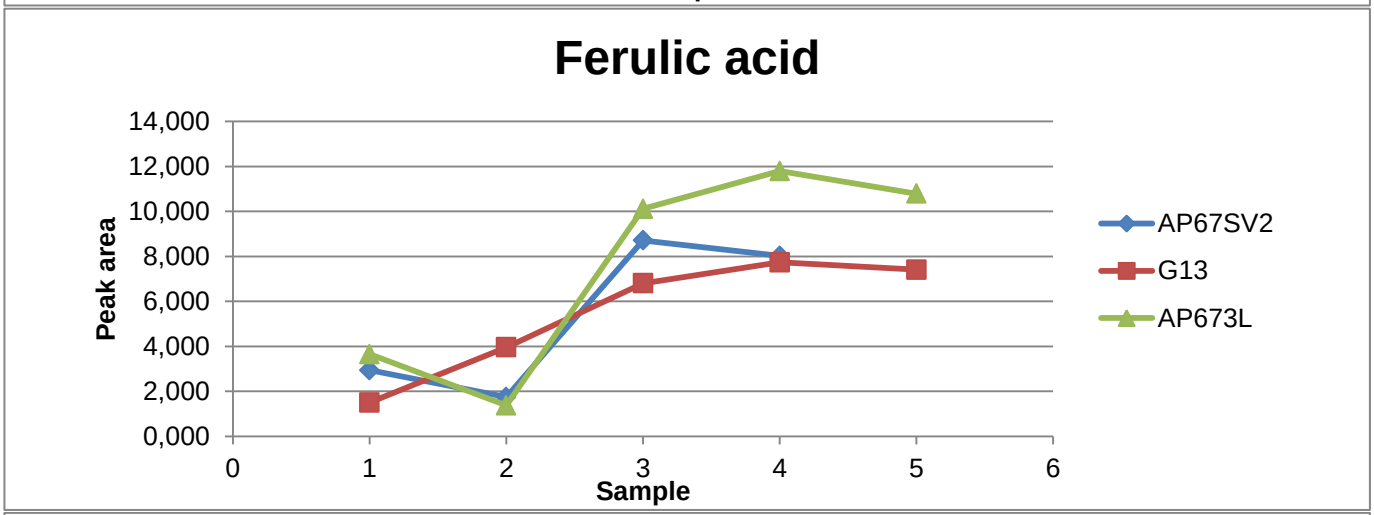
3. TOFMS results, quantification

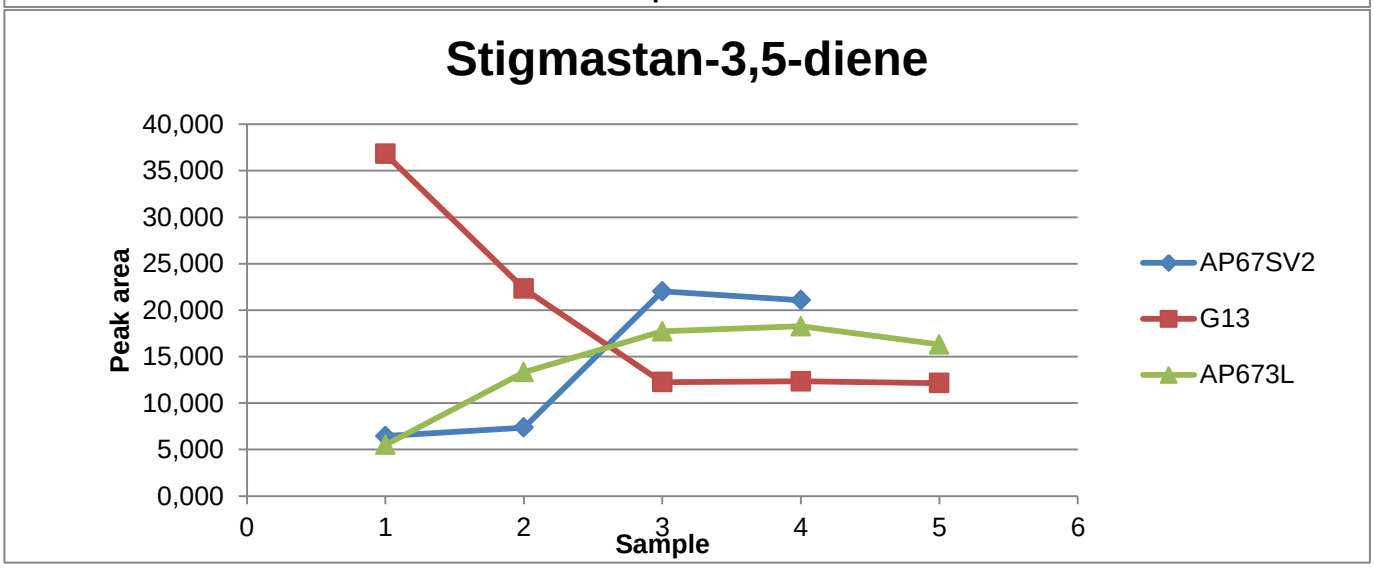
	Sample ID	Desmosterol	Ferulic acid	Stigmastan-3,5-diene	Campesterol	4-Hydroxybutanoic acid
B	1b_Lakka_AP675V2	8,968	2,937	6,449	41,843	67,098
	2c_Lakka_AP675V2_rep2	26,852	1,743	7,367	33,598	70,116
	13c_Lakka_15B_rep1	79,915	8,715	22,029	55,359	8,953
	15c_Lakka_15B_rep3	70,524	8,022	21,073	56,749	30,086
	Mean	46,564	5,354	14,229	46,887	44,063
	Stdev	34,100	3,525	8,471	11,122	29,650
C	31c_Lakka_3C_rep1	27,933	1,502	36,807	~~435,558~~	17,771
	32c_Lakka_3C_rep2	37,273	3,952	22,316	~~357,933~~	16,890
	16c_Lakka_15C_rep1	17,321	6,801	12,250	28,688	26,741
	17c_Lakka_15C_rep2	8,087	7,735	12,349	28,680	21,784
	18c_Lakka_15C_rep3	8,862	7,406	12,148	28,068	28,534
	Mean	11,423	7,314	12,249	28,479	25,686
	Stdev	5,122	0,474	0,100	0,356	3,497
D	2b_Lakka_AP673L	8,215	3,652	5,508	54,719	44,864
	8c_Lakka_AP673L_rep2	9,532	1,369	13,322	31,605	48,645
	19c_Lakka_15D_rep1	38,888	10,114	17,713	41,677	40,387
	20c_Lakka_15D_rep2	56,768	11,794	18,267	42,020	16,994
	21c_Lakka_15D_rep3	46,703	10,789	16,309	33,147	42,055
	Mean	32,021	7,544	14,224	40,633	38,589
	Stdev	22,066	4,703	5,236	9,205	12,471

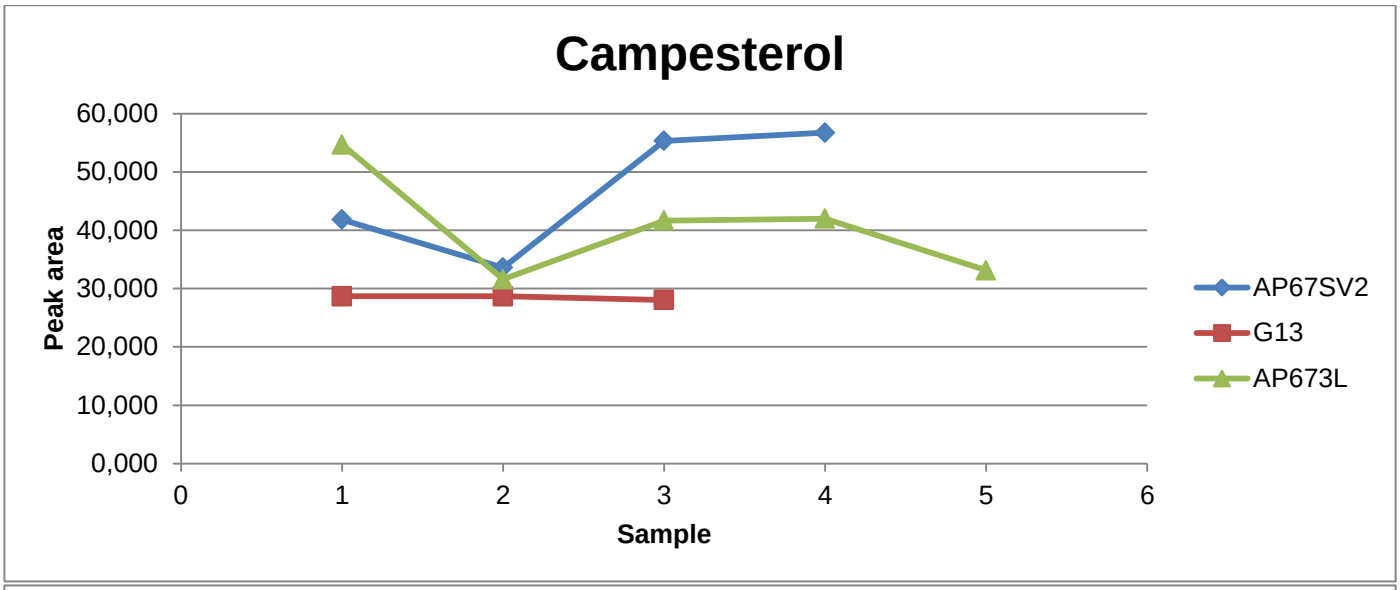
B: AP675V2, **C**: G13, **D**: AP673L

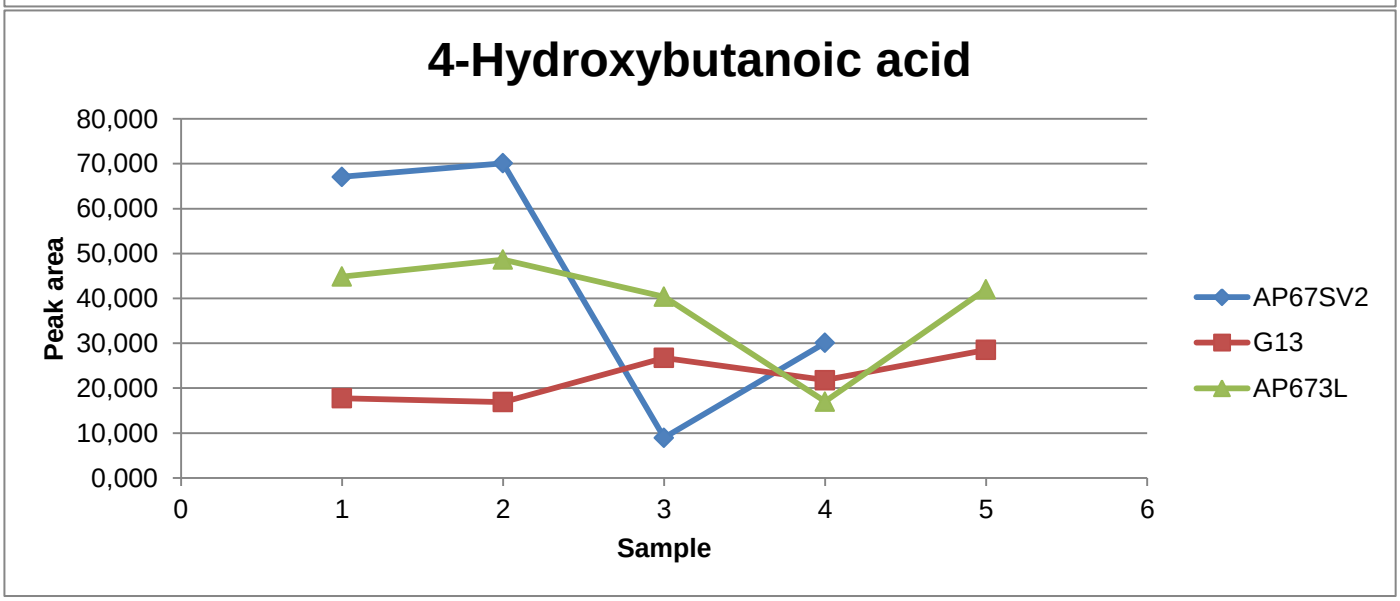
4. Essential compounds, correlation








Campesterol
60,000
50,000
40,000
30,000
20,000
10,000
0,000
Peak area
0
1
2
3
4
5
6
Sample
AP67SV2
G13
AP673L


4-Hydroxybutanoic acid
80,000
70,000
60,000
50,000
40,000
30,000
20,000
10,000
0,000
Peak area
0
1
2
3
4
5
6
Sample
AP67SV2
G13
AP673L

Metropolia

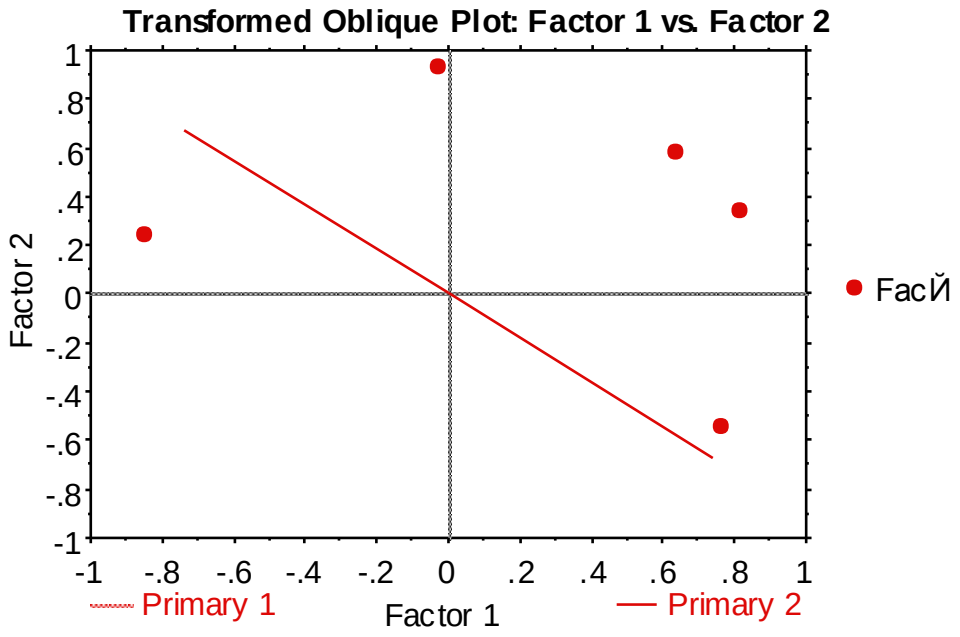
5. Essential compounds, fat concentration

	Sample ID	A(Fats)	A(Total)	**Fat-%**
A	3b_Lakka_G2	67626,524	389653,1	**17,36 %**
	11c_Lakka_15A_rep2	16032,091	394025,9	**4,07 %**
B	1b_Lakka_AP675V2	91417,15	871361,4	**10,5 %**
	2c_Lakka_AP675V2_rep2	39325,65	492128,6	**7,99 %**
	13c_Lakka_15B_rep1	25185,75	606271	**4,15 %**
	15c_Lakka_15B_rep3	22878,47	711188,7	**3,22 %**
C	31c_Lakka_3C_rep1	22350,57	540937,7	**4,72 %**
	32c_Lakka_3C_rep2	18952,23	411437,9	**5,78 %**
	16c_Lakka_15C_rep1	23629,74	524828,4	**4,26 %**
	17c_Lakka_15C_rep2	25525,52	409107,5	**4,63 %**
	18c_Lakka_15C_rep3	23771,38	491112,1	**4,81 %**
D	2b_Lakka_AP673L	82395,15	531111,8	**15,51 %**
	8c_Lakka_AP673L_rep2	36298,72	350448,3	**10,36 %**
	19c_Lakka_15D_rep1	20930,70	433467,1	**4,83 %**
	20c_Lakka_15D_rep2	23274,86	505347,2	**4,61 %**
	21c_Lakka_15D_rep3	16640,44	363337,3	**4,58 %**

A: G2, **B**: AP675V2, **C**: G13, **D**: AP673L

6. Principal component analysis

PCA was performed using StatView 5.0.1 for Macintosh



Left to right: 4-Hydroxybutyric acid, campesterol, stigmastan-3,5-diene, desmosterol, ferulic acid

Orthogonal Transformation Solution-Varimax

	Factor 1	Factor 2
Desmosterol	.423	.775
Ferulic acid	.139	.85
Stigmastan-...	.937	.044
Campesterol	-.602	.719
4-OH-butano...	-.819	-.332