



universität
wien

DIPLOMARBEIT

Titel der Diplomarbeit

Comparative metabolic studies in *Arabidopsis thaliana*
via GC-MS and LC-MS

Verfasser

Hannes Dörfler

angestrebter akademischer Grad

Magister der Naturwissenschaften (Mag. rer. nat.)

Wien, 2011

Studienkennzahl lt. Studienblatt:

A 423

Studienrichtung lt. Studienblatt:

Lehramt Chemie

Betreuerin / Betreuer:

Univ.-Prof Dr. Wolfram Weckwerth

Contents

1. Introduction	7
1.1 Metabolomics and its role in the life sciences	7
1.2 Arabidopsis thaliana – the model organism	9
1.3 Primary and secondary metabolism of a plant.....	11
1.4 GC-MS and LC-MS based metabolomics	13
1.4.1 GC-MS and small molecules	13
1.4.2 LC-MS and slightly bigger molecules	17
2. Material and methods	20
2.1 Experimental setup	20
2.1.1 The cold stress experiment.....	20
2.1.2 Growth and harvesting of <i>A. thaliana</i>	20
2.2 The GC-MS approach.....	23
2.2.1 Sample preparation for GC-MS analysis	23
2.2.2 GC-MS analysis – instrumental settings.....	24
2.3 The LC-MS approach	26
2.3.1 Building a plant secondary metabolite database	26
2.3.2 Sample preparation for LC-MS analysis.....	28
2.3.3 LC-MS analysis – instrumental settings	29
3. Results	31
3.1 GC-MS-Measurements	31
3.1.1 Amino acids	31
3.1.2 Sugars.....	33
3.1.3 Polyols.....	34
3.1.4 TCA cycle components.....	35
3.2 LC-MS-Measurements	36
4. Discussion.....	37
4.1 Outlook	39

5. Summary	40
6. Acknowledgements	42
7. Attachment	47
7.1 Tables	47

List of figures

Figure 1	Metabolomics and relative disciplines (8)
Figure 2	<i>Arabidopsis thaliana</i> (9)
Figure 3	Glycolysis and TCA cycle (11)
Figure 4	Primary and secondary metabolism (12)
Figure 5	Gas chromatograph (14)
Figure 6	EI mass spectrum (15)
Figure 7	Library matching (15)
Figure 8	GCxGC-TOF plot (16)
Figure 9	Fields of application of GC and LC (17)
Figure 10	LC columns (18)
Figure 11	Electrospray process (19)
Figure 12	FT-MS (19)
Figure 13	<i>A. thaliana</i> (greenhouse) (21)
Figure 14	<i>A. thaliana</i> (cold stress) (22)
Figure 15	GC derivatisation (24)
Figure 16	GCxGC-MS (25)
Figure 17	Orbitrap-MS (26)
Figure 18	Orbitrap precursor spectrum (27)
Figure 19	Orbitrap MS ² spectrum (27)
Figure 20	Orbitrap MS ⁿ spectrum (28)
Figure 21.x	Orbitrap direct infusion spectrum and recovered spectrum (29,30)
Figure 22.x	Amino acid measurements (31, 32)
Figure 23.x	Sugar measurements (33)
Figure 24.x	Polyol measurements (34)
Figure 25.x	TCA compound measurements (35)
Figure 26	LC data peak annotation (36)
Figure 27	PCA of GC metabolites (37)

List of tables

Table I	Metabolites identified by GC-MS (47)
Table II	Pure substances used for direct infusion (48)

List of abbreviations

ACN	Acetonitrile
CID	Collision-induced dissociation
EI	Electron impact ionisation
ESI	Electrospray ionisation
FA	Formic acid
FT	Fourier transform
GC	Gas chromatography
GCxGC	Two-dimensional gas chromatography
GC-MS	Gas chromatography coupled to mass spectrometry
LC	Liquid chromatography
LC-MS	Liquid chromatography coupled to mass spectrometry
MCW	Methanol / Chloroform / Water
MS	Mass spectrometry
MSTFA	N-methy-N-(trimethylsilyl) trifluoroacetamide
NL	Liquid nitrogen
PCA	Principal component analysis
RI	Retention index
TCA	Tricarboxylic acid
TOF-MS	time-of-flight mass spectrometry
QC	Quality control

1. Introduction

1.1 Metabolomics and its role in the life sciences

In recent years, the term *metabolomics* has emerged to become one of the most important bioanalytical approaches besides *genomics* and *proteomics*. While the aim of the latter two is to elucidate the functions of genes and proteins, respectively, metabolomics is about investigating the entirety of the metabolism of a living system at a current state, the *metabolome*. Especially for “functional genomics” – where the changes in an organism originating from the deactivation of a certain gene are observed, in order to gain information of this gene – metabolomics is an important technology. Due to the diversity of the metabolome, metabolomic techniques need to comprise a broad range of bioanalytical methods.

Metabolomics, as well as *proteomics*, *transcriptomics*, *metabonomics*, *lipidomics*, *peptidomics* etc. are all terms which roughly define a field of application rather than depicting a specific analytical method. Because all those techniques are just being established right now, one will always find slightly different and interchangeable definitions of them in the literature, as well as overlapping meanings in scientific speech. One important factor, however, which sets apart the “omics” techniques from other biochemical approaches, is its global, genome-scale and unbiased nature, meaning that its primary aim is, first and foremost, to measure all the metabolites in a biological sample.

Combined, the “omics” sciences make up the tools needed for a new, large scale approach of investigating biochemical networks, the *systems biology* approach. [1]

The paradigm in *systems biology* is to carry together all the information of a complex biological system, to integrate it, and eventually create a mathematical model apt to predict the conduct of the system under varying factors. [2] This, of course, poses quite an ambitious task, taking in mind that the latest major achievements in biology on a molecular level, such as the sequencing of the human or the *Arabidopsis thaliana* genome, have brought up way more questions than they answered. Still, as lacking our knowledge may be at this time, one cannot deny the tremendous potential these scientific fields could unlock in the scientific

studies of living organisms – the *life sciences* – the long run. Future fields of application of metabolomic techniques include for instance the so called “personalized medicine” [3], as well as crop [4] and biofuel [5] research.

However, due to the immense diversity of the chemical compounds present in the metabolome at any time point, and the dynamic range of them, it is not hard to understand that getting hold of the whole of the metabolome is a non-trivial task. In fact, there exists no method to date to measure all the metabolites in a sample with one single technique or in one sample run. Still, as will be talked about below, the basic tools for the elucidation of the metabolome are available, and methods are permanently being developed and improved.

Also, analytical measurements are always a trade-off between accuracy, complexity, and of course, time. In the field of metabolomics, there exist several approaches to investigate the area of interest to its best extent, depending on which level of information is required. While *metabolomics*, as mentioned, strives to detect all metabolites, *metabolic target analysis* e.g. focuses only on very few specific compounds, or a specific group of them, like some amino acids. [6] [7] Target analyses are very accurate, but they ignore all other compounds of a sample. *Metabolic profiling* describes a larger analytical range than a target analysis. A possible field of application, for instance, would be the elucidation of a specific pathway or the investigation of a group of compounds like amino acids. *Metabolic fingerprinting* on the other side is a way to quickly classify large sample sets, as it is needed for screening methods. Figure 1 depicts the relation of these different metabolic disciplines to each other.

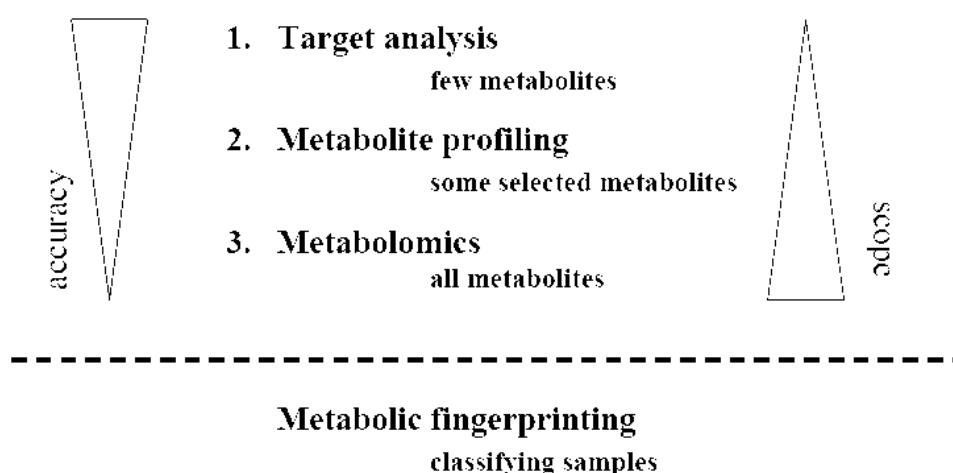


Figure 1: The extent of the interesting metabolites determines the analytical approach as well as the achieved accuracy of the measurements. Image altered according to [8]

Summing up, it is undoubtedly clear that, after the huge progressions in genome and proteome research in the last years, unfolding the metabolome of a biological organism is the next logical step, and finding out the connections between these three major levels of biochemical organisation will create nothing less than a revelation in the field of the life sciences in the upcoming future.

1.2 *Arabidopsis thaliana* – the model organism

Model organisms play an important role in biological research; especially in the era of the already discussed “omics” sciences, where we are dealing with vast numbers of molecules, pathways and unexpected factors, having a reliable, well investigated system, which provides the starting point for further research activities, is an invaluable tool. Bearing in mind that the genetic code is almost universal, newly found genotype – phenotype coherency as well as elucidation of certain pathways in a simple organism often is applicable to other organisms too. Some prominent model organisms in biology are: *Escherichia coli*, as far as prokaryotes are concerned; *Chlamydomonas reinhardtii* for algae, furthermore *Saccharomyces cerevisiae* (yeast), *Caenorhabditis elegans* (a nematode) and *Drosophila melanogaster* (fruit fly), as well as *Mus musculus* (mouse) and *Danio rerio* (zebra fish).

Concerning plants, one could easily say that *Arabidopsis thaliana* might be one of the – if not the – most important model organisms in modern plant research at the present. [9] [10]

Figure 2 shows the plant in its typical habitus.



Figure 2: *Arabidopsis thaliana* on stony underground, flowering

Arabidopsis thaliana is a small, flowering, annual plant of the *Brassicaceae* family. Its relatively short lifecycle of about 5 to 6 weeks is one of the many reasons why it is so extraordinarily well suited for plant researchers. Other advantages of working with *A. thaliana* comprise:

- *A. thaliana* has been sequenced in the year 2000
- It has a relatively small genome: 115,409,949 base pairs of DNA, distributed in 5 chromosomes
- Cultivation is easy and one specimen produces a lot of seeds (about 10.000 per plant)
- Many mutations exist, which are also commercially available
- A lot of transcriptomics, proteomic and metabolomic data already exist from this organism

Since *A. thaliana* is of worldwide scientific interest, a database of genetic and molecular data has been created, where a plethora of information about this important plant is compiled, called The Arabidopsis Information Resource – *TAIR* [11]. Furthermore, Arabidopsis proteome research have established an online platform, where proteomic information from different online resources can be searched: MASC Gator [12], an initiative of the Multinational Arabidopsis Steering Committee *MASC*.

Metabolomic platforms and pathway databases are, too, already available [13] [14], however the problem with metabolomic data is the variety of analytical approaches that are used to retrieve information, which can lead to difficulties when trying to compare data obtained by different instruments. Thus, standardisation and database optimisation are the next steps in order to build a comprehensive approach in the metabolomics area – [15]. For the future, an all – embracing Arabidopsis portal has recently been thought up, which would act as a central, global platform containing all available information, databases, tools, etc. for scientific interest, under a management of an international Arabidopsis informatics consortium – *IAIC*. [16]

1.3 Primary and secondary metabolism of a plant

Plants, just like other living organisms, have to acquire and metabolise nutrients in order to maintain their vital life functions, foremostly survival and propagation. All the anabolic and catabolic processes needed for it can be subsumed under the concept of the *primary metabolism*. The molecules found in the primary metabolism are very similar in all organisms and include compounds like amino acids, sugars, nucleotides, fatty acids, etc., which are part of the basic biochemical processes like the Calvin cycle – where CO₂ fixation takes place and which thus represents the most important way for net primary production on earth – , glycolysis and the TCA (tricarboxylic acid) cycle. Fig. 3 shows two major steps of the central energy metabolism of a cell, and which family of molecules play a part therein.

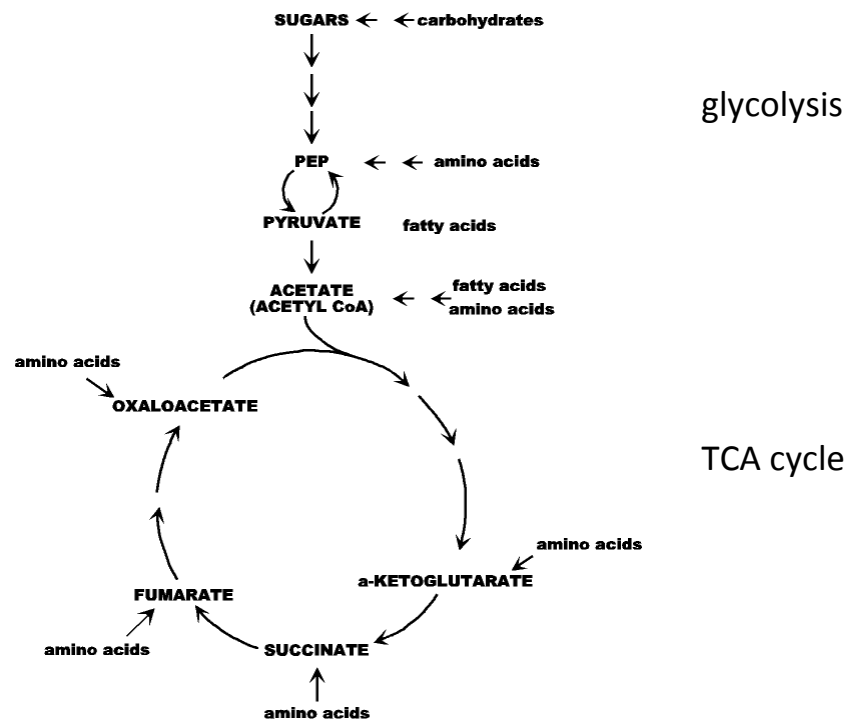


Figure 3: Glycolysis followed by the TCA cycle. These processes not only yield energy for the cell, but also create intermediates needed for other metabolic pathways. Not shown is the oxidative phosphorylation successive to the TCA cycle.

Apart from these metabolic pathways, though, we find another biochemical network in many plants, which is knotted to the primary metabolism, and produces a broad range of bigger molecules with various properties and functions – the so called *secondary metabolism*.

In contrast to the primary metabolism of a plant, which sums up all the central, perpetual biochemical processes, the *secondary metabolism* produces compounds which possess ecochemical functions. Contrary to humans or animals, plants are sessile organisms and thus cannot move away from dangers and unfavourable environmental conditions. Therefore, if they want to survive, they have to be equipped with instruments that help them to overcome disadvantageous situations. Secondary metabolites are a way for plants to weather various ecological strains, like heat, cold, or salt stress, UV light, natural enemies, but also microbes or competing plants. So the denomination “secondary metabolism” does not imply a lower-ranking role of these kind of substances, but rather points to the fact that this metabolism is derived from the primary one. Plant secondary metabolites have accompanied humans throughout the centuries, mostly because of their physiological activity in the human body; may it be as a medicine or illegal drug, a venom or a spice – [17]. Many carry a charge or possess a mesomeric carbon structure with sp^2 -hybridised bonds, thus absorb electromagnetic waves in the range of visible light (400-900 nm) and are used as dyes or food colorings. Figure 4 shows selected secondary metabolites, and their associated primary metabolite or primary metabolite class they originate from.

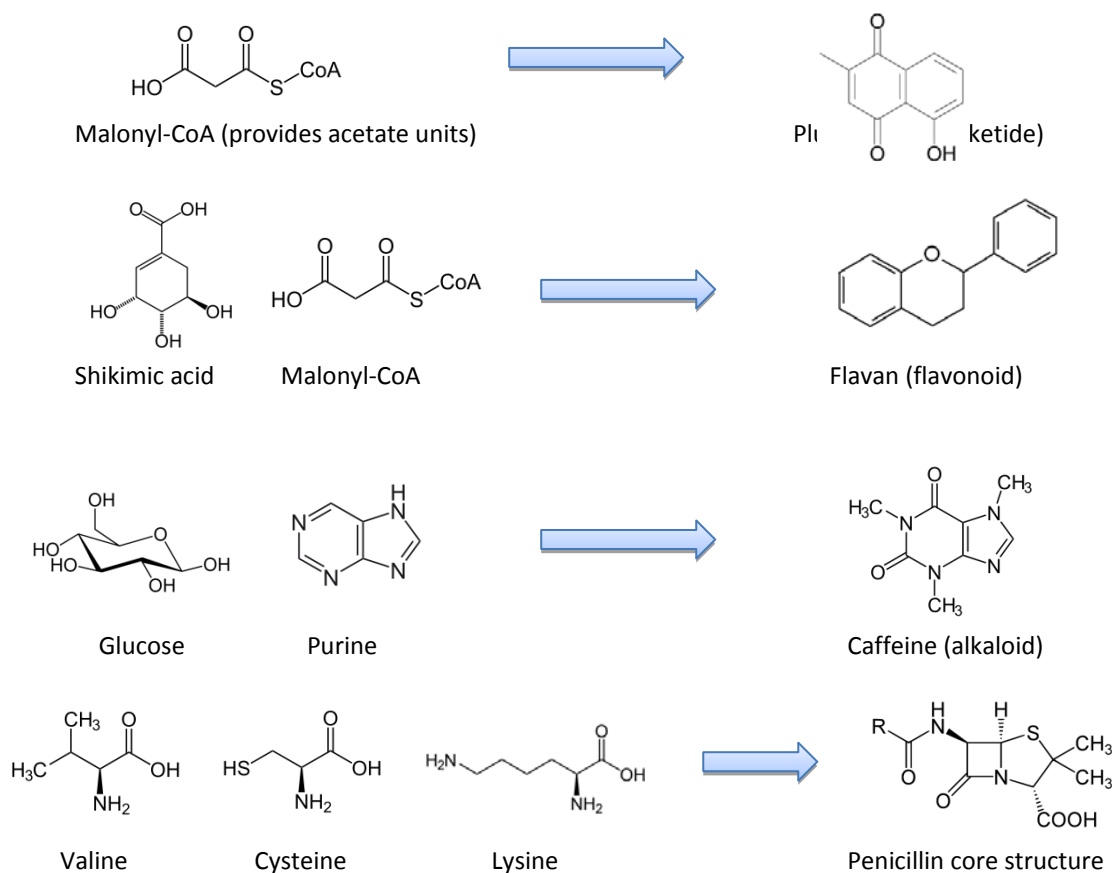


Figure 4: Some plant secondary metabolites and their basic building components. Image created with the help of KEGG pathway database. [18] The actual pathways are of course much more complicated.

Enhanced production of secondary metabolites of pharmacological interest by genetically modified plants (pharming) [19] is one major vision in the field of plant research, which becomes more and more tangible as genomic, proteomic and metabolomic techniques improve.

With regard to the different chemical properties of primary and secondary metabolites, we have to choose from a respectively suitable analytical method when performing measurements with these two types of compounds, as is explained in the upcoming chapter.

1.4 GC-MS and LC-MS based metabolomics

The ultimate goal of metabolomics, to identify and quantitate all the metabolites in a given sample in an unbiased way, is something that cannot be achieved by the means of a single analytical technique – at least not at this day and time.

As stated before, the analytical method has to be orientated towards the chemical properties of the analytes. When dealing with metabolomics, there have emerged several major approaches to date which meet the analyst's need to perform satisfying measurements regarding both sensitivity and selectivity: gas chromatography coupled to mass spectrometry (GC-MS), liquid chromatography coupled to mass spectrometry (LC-MS), and NMR-based metabolomics. [20] [21] [xx]

1.4.1 GC-MS and small molecules

Gas chromatography is an analytical separation technique that has been around for decades and become a standard tool in chemical laboratories for analyses in the field of e.g. drugs, hormones, essential oils, air, seawater, etc. The basic principle of a gas chromatograph is separation of the compounds of a complex mixture by different allocation of each analyte between a stationary and a mobile phase in a column. Stationary phases are either solid – then we talk about *GSC* (gas solid chromatography) – or liquid (which actually stands for a highly viscous polymer), which is the main case when dealing with small organic molecules, and which is then called *GLC* (gas liquid chromatography), or simply GC. The compounds are sent over the column with an inert carrier gas like helium after being evaporated in the so-called *liner*. Figure 5 depicts the principle of a gas chromatographic separation layout.

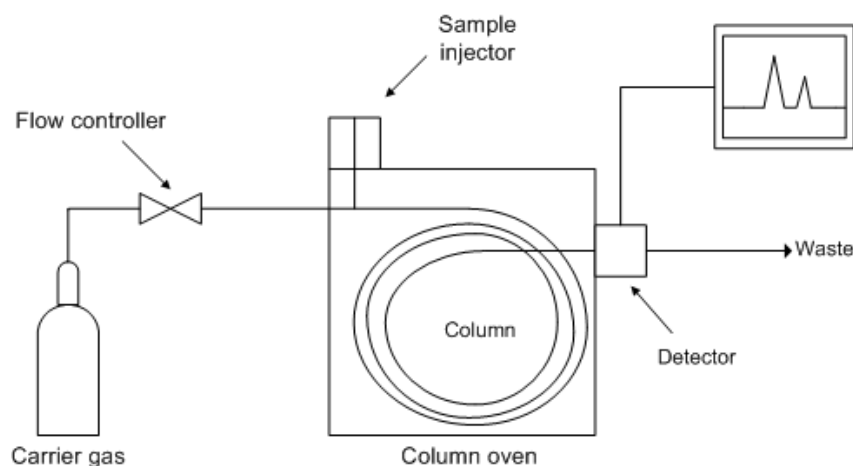


Figure 5: Schematic gas chromatograph. Working principle is explained in the text.

Depending on their allocation and interaction with the column, the compounds of the analysed sample reach the end of the column at different times. The time which one specific molecule needs to pass through the column is called the *retention time* of that molecule. The existence of this retention time is the main reason why GC databases can be built, which provide identification according to the so-called *retention index*: According to Kovats, there is a direct correlation between the number of C-atoms of a homologous n-alkane row and their retention time. Thus, independent from the column or temperature, those alkanes can be assigned values, while the analytes of the mixture elute somewhere between those defined points and are set into relation to the predefined alkanes. This setup provides a good way to separate preferably small molecules. Apolar compounds mostly containing sp^3 -hybridised C and H can be evaporated easily, while for polar analytes, diverse derivatisation agents exist, which can increase their volatility (see also 2.2.1). At the same time, due to volatility reasons, GC is not suited for separation of molecules which exceed a mass of 500 u. After having passed the column, the compounds enter the *detector*. There are several detectors which can be used in combination with GC, like *flame ionisation detectors* (FID) or *thermal conductivity detectors* (TCD), however the technique with the greatest potential for identification, quantification and sensitivity is provided by gas chromatography coupled to *mass spectrometry*, GC-MS.

Mass spectrometry describes the analytical technique where chemical compounds are identified through their mass, and fragments thereof. The background of this method is the finding that molecules tend to break in a characteristic pattern when confronted with particle radiation. In the GC-MS case, the freshly eluted compounds get ionised by a stream of highly energised electrons (70 eV) in a process called *electron impact ionisation* (EI). This is a

destructive ionisation method, resulting in many fragment ions, while the molecular ion peak often is not visible. Figure 6 shows a typical EI spectrum of an organic compound.

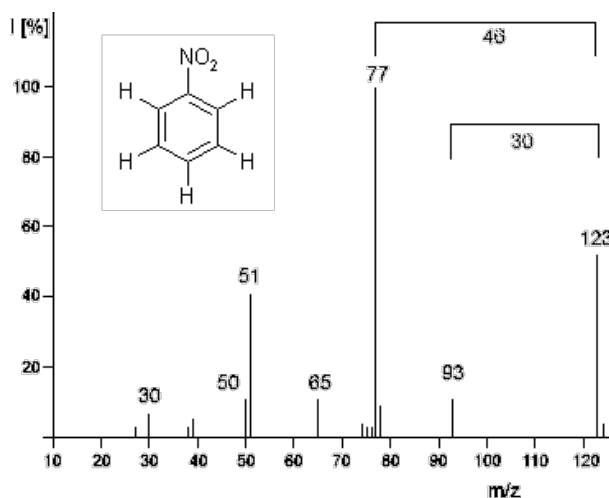


Figure 6:

Electron ionisation mass spectrum of nitrobenzene (M 123.11). The peaks at m/z 93 and 77 correspond to the loss of an NO and NO_2 radical, respectively. The peak at 123 m/z corresponds to the parent ion (actually a radical cation).

The EI mass spectrum, beside the retention time, is the second parameter which allows compound identification by matching of the acquired spectrum against a library. The software will compare the newly generated spectra to the library and suggest fitting compounds using a match factor. Figure 7 depicts the process of library matching featuring an amino acid derivative.

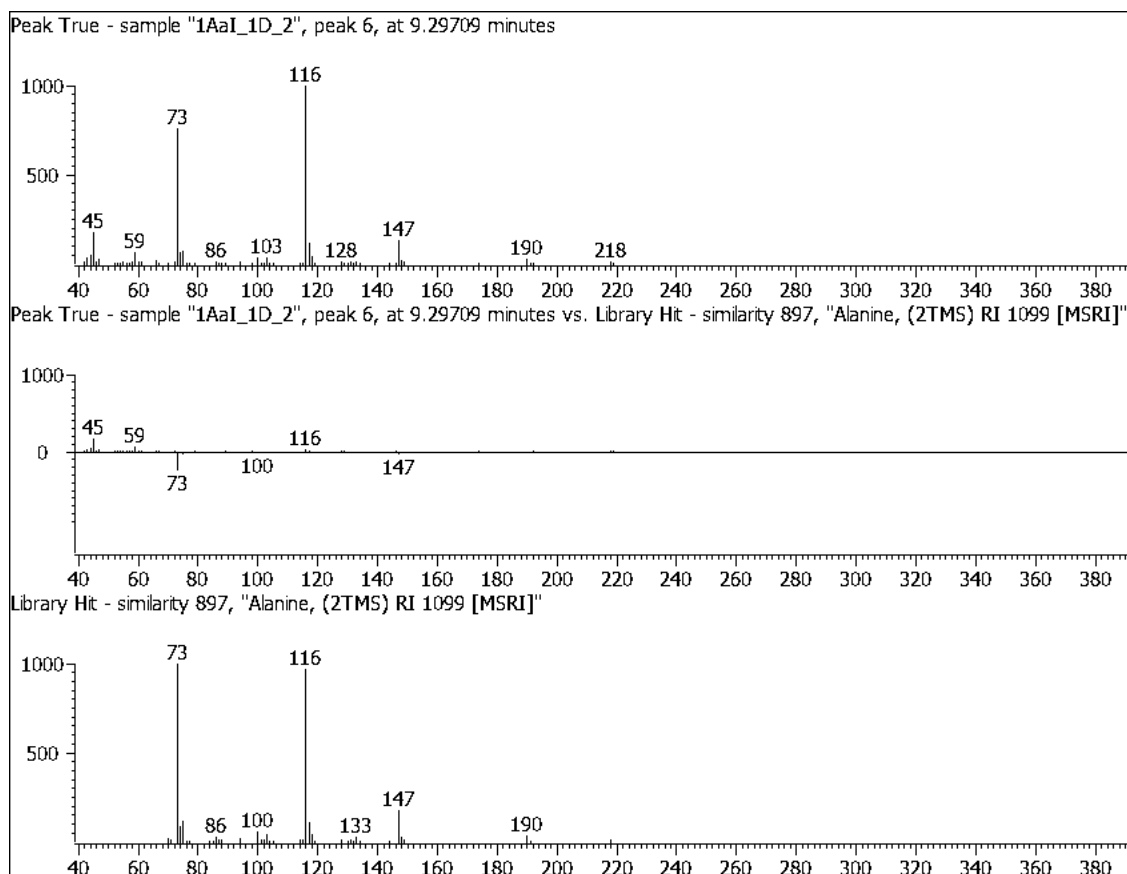


Figure 7: Identifying a substance through library matching. The above spectrum originates from an Arabidopsis chromatogram shortly after the ninth minute. The lower spectrum represents the library entry. Subtracting the lower from the upper spectrum results in the “spectrum” in the middle and a match factor of 897.

This method provides a robust way for identification of many organic compounds, however it falls short of being able to differentiate between isomer forms of molecules, like we find them in sugars or organic acids. For separating and identifying enantiomers, which plays an important role especially in pharmaceutical chemistry, special columns are needed, as well as analytical techniques like NMR.

2-dimensional GC

2D-GC-MS has the potential to yield a broader coverage of the metabolome of a complex sample, due to the presence of a second, moderately polar column directly attached to the first one; this setup allows separation of molecules which co-elute on the first dimension. This results in a 2D plot similar to the one of a gel electrophoresis (Figure 8).

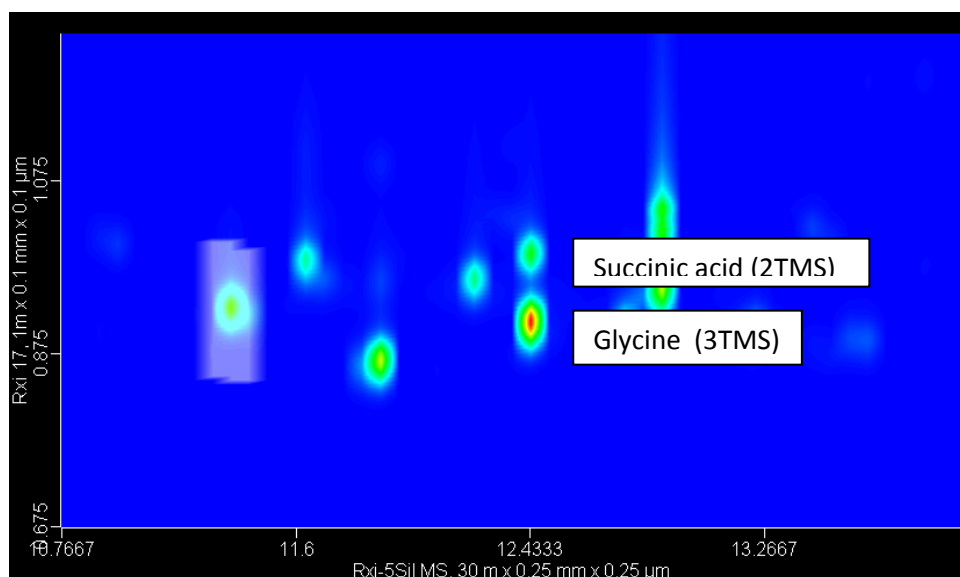


Figure 8: Showcase 2D plot of a selected area in a GCxGC separated metabolite chromatogram. Succinic acid (2TMS) and Glycine (3TMS), for instance, can only be recognized as two discrete peaks owing to the presence of a second column, TMS derivatisation explained in 2.2.1. Data provided by Fragner, 2011.

GCxGC analysis of the Arabidopsis samples is yet to be done – in conjunction with establishing a robust method for further metabolic 2D analyses. To this end, intensive editing of the produced data is needed, as well as having it constantly compared to the 1D data.

Beside the possibility of gaining an extra chromatographic dimension for increasing the number of signals, a still remaining task is to analyse the unpolar fraction left from the phase separation with water after metabolite extraction (see 2.2.1), which should contain all kinds of lipid compounds.

1.4.2 LC-MS and slightly bigger molecules

When advancing in molecular mass of the metabolites, we soon exceed the limits of GC, which lies around 500 u, yet this is where many interesting metabolites with their complicated carbon structure and heteroatoms can be found. These compounds are non-volatile. In order to measure them, we again need to separate them on a column, create charged particles with an ionisation source, and detect them. The compounds are now separated on a liquid chromatography (LC) system with a column of specified chemical properties and using high pressure for the mobile phase, called HPLC (high pressure / performance liquid chromatography). HPLC-MS measurements of metabolites have emerged in the last years and the comparability of the data and database infrastructure is still lacking, but the continuous development of columns which can be operated with high pressure (ultra performance liquid chromatography, UPLC) as well as the availability of high resolution mass spectrometers promises a bright future for LC-MS based metabolomics. Figure 9 shows the application area of chromatography – coupled mass spectrometer systems depending on the molecular mass and polarity of the analyte.

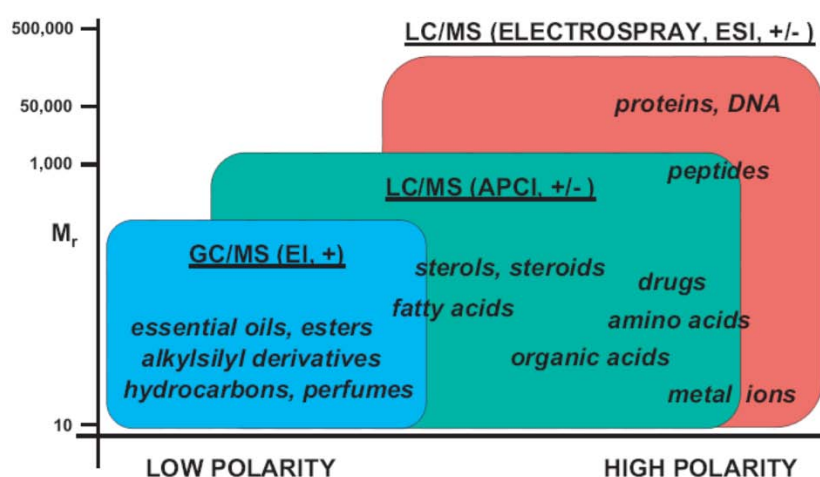


Figure 9: Chemical properties define the choice of chromatography and mass spectrometers. Image altered according to [21]

In LC columns, just as in GC, separation is achieved by having the compounds interact with a stationary phase; the big difference is now, that we are dealing with analytes in liquid samples. Accordingly, the chemistry of the stationary phase determines the elution of the components. In general, there exist 2 major classes of LC columns: *normal-phase HPLC* (NP-HPLC) uses a polar stationary phase (mostly silica gel), while *reversed-phase HPLC* (RP-HPLC) – the most important technique today – utilizes an unpolar stationary phase consisting of alkane-substituted silanes bonded to the Si-OH groups.

The development of RP-LC columns allows the use of watery eluents (which renders the coupling to MS possible in the first place), while NP-LC can only be operated with organic solvents. Since most biological samples arise from an water-based background, RP-HPLC is the method of choice here. In Figure 10, the chemical nature of NP-LC and RP-LC stationary phases can be seen.

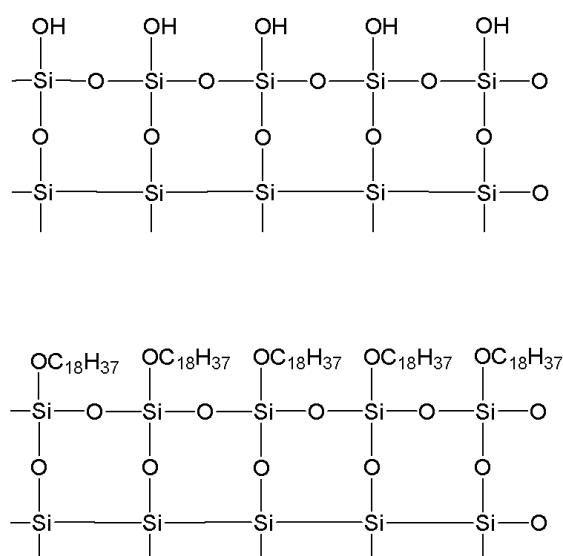


Figure 10: Exemplary stationary phases of normal-phase columns (above) and reversed phase columns.

Common carbon numbers of the alkane derivatives of RP phases are 8 and 18.

Liquid chromatography can be coupled with different types of ionisation techniques, however for rather small and polar molecules, as we are dealing with in metabolomics, the *electrospray ionisation* (ESI) is the dominant procedure for bringing the analytes into a charged state.

The main principle of ESI consists of conveying the eluents from the column with the help of an electric current into an aerosol, which has then evaporate the solvents away from it, ideally leaving only charged analytes behind. Depending on the electric field, obtained ions are either positively ($[M+H]^+$) or negatively ($[M-H]^-$) charged. Figure 11 depicts the electrospray ionisation process.

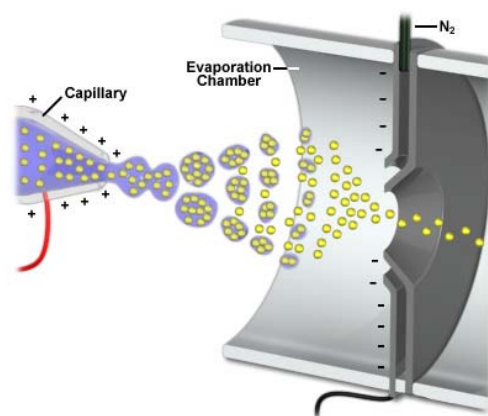


Figure 11: Electrospray process: the needle connected to the column emits the analytes plus the solvent, which then evaporates, leaving charged molecules (ideally analytes, but in fact also matrix components)

For detection, there are various mass spectrometer types suited to be combined with HPLC, mainly triple quad instruments (QqQ), ion traps (IT), time-of-flight applications (TOF), or combined setups like quadrupole – time-of-flight mass spectrometers (Q-TOF). In this project, an Orbitrap mass spectrometer was used. The Orbitrap mass analyser is a combination of a linear ion trap with a specialised electrode, around which charged molecules can circle, effectively describing oscillations, which can be displayed in a frequency spectrum, or, after Fourier transformation, as a mass spectrum (Fig. 12).

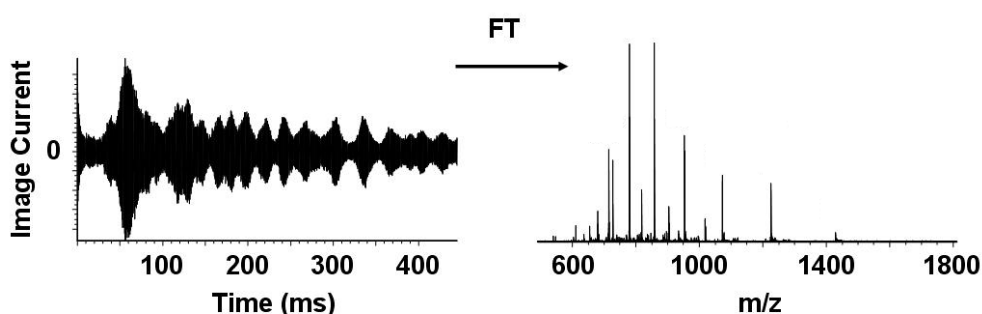


Figure 12: Principle of Orbitrap and FTICR (Fourier transform ion cyclotron resonance) mass spectrometers. Applying Fourier transform on the signals in the frequency domain yields a mass spectrum.

2. Material and methods

2.1 Experimental setup

2.1.1 The cold stress experiment

The stress metabolome of plants is subject to manifold plant physiological investigations [22][23]. As already mentioned, there are multiple ways to create a stressful environment for plants. In this study, the response of *A. thaliana* to cold stress is being observed.

From previous experiments at the Dept. of Molecular Systems Biology it is known that *A. thaliana* undergoes a red coloration during 2 to 3 weeks of cold stress treatment, supposedly due to induction of Flavonoid/Anthocyanin biosynthesis. Thus, besides the changes in the primary metabolism, which are evident [24], we can make the assumption that also the secondary metabolism is affected by the cold stress experiment. However, a combined metabolomic analysis of primary and secondary metabolism during cold stress was not yet performed. Hence, such a cold stress experiment seemed like a convenient way to accompany the setup of a combined platform for both primary and secondary metabolites.

2.1.2 Growth and harvesting of *A. thaliana*

In order to gain access to a sufficient amount of plant material, the aim was to harvest the leaves of approximately 100 *A. thaliana* specimen over a time course of some 3 weeks. From germination to full rosette development the time interval is about 25 days; after the cold stress the plant goes into senescence, which involves metabolic changes undesired for our investigations. An unstressed control group was harvested before the cold stress, the other plants were brought into a climate chamber with a temperature of 4°C.

Shortly after the start of the experiment it became clear that the amount of leaf material produced by the plant was by no way sufficient for metabolomic examination. It was soon realized that the light conditions under which *A. thaliana* had been grown in the glass house had caused the plants to induce flowering – it was June at this time – thus producing less

leaves, and also introducing unwanted metabolic effects. This phenomenon is due to the fact that *A. thaliana* is a facultative long day plant. [25]

Thus, a new batch of *A. thaliana* was being grown, only this time under an “artificial” short day: therefore, the plants were shielded from sunshine during 3 pm to 7 am (16 hours of dark); therefore the light period happened from 7 am to 3 pm (8 hours of light).

Plants were harvested randomly from their places in the rack to minimize the slight differences from illumination. Plant leaves were taken in groups from 3 - 4 plants from 3 biological replicates each. This procedure was repeated 8 times for obtaining enough time points to measure the metabolic changes during the experiment resulting in a total of 9 time points (some plants of the 100 didn't grow as expected or died off).

Figure 13 shows the plants in the glasshouse before undergoing cold stress; figure 14 shows the development of a single plant during the cold stress.



Figure 13: *Arabidopsis thaliana* grown in the greenhouse

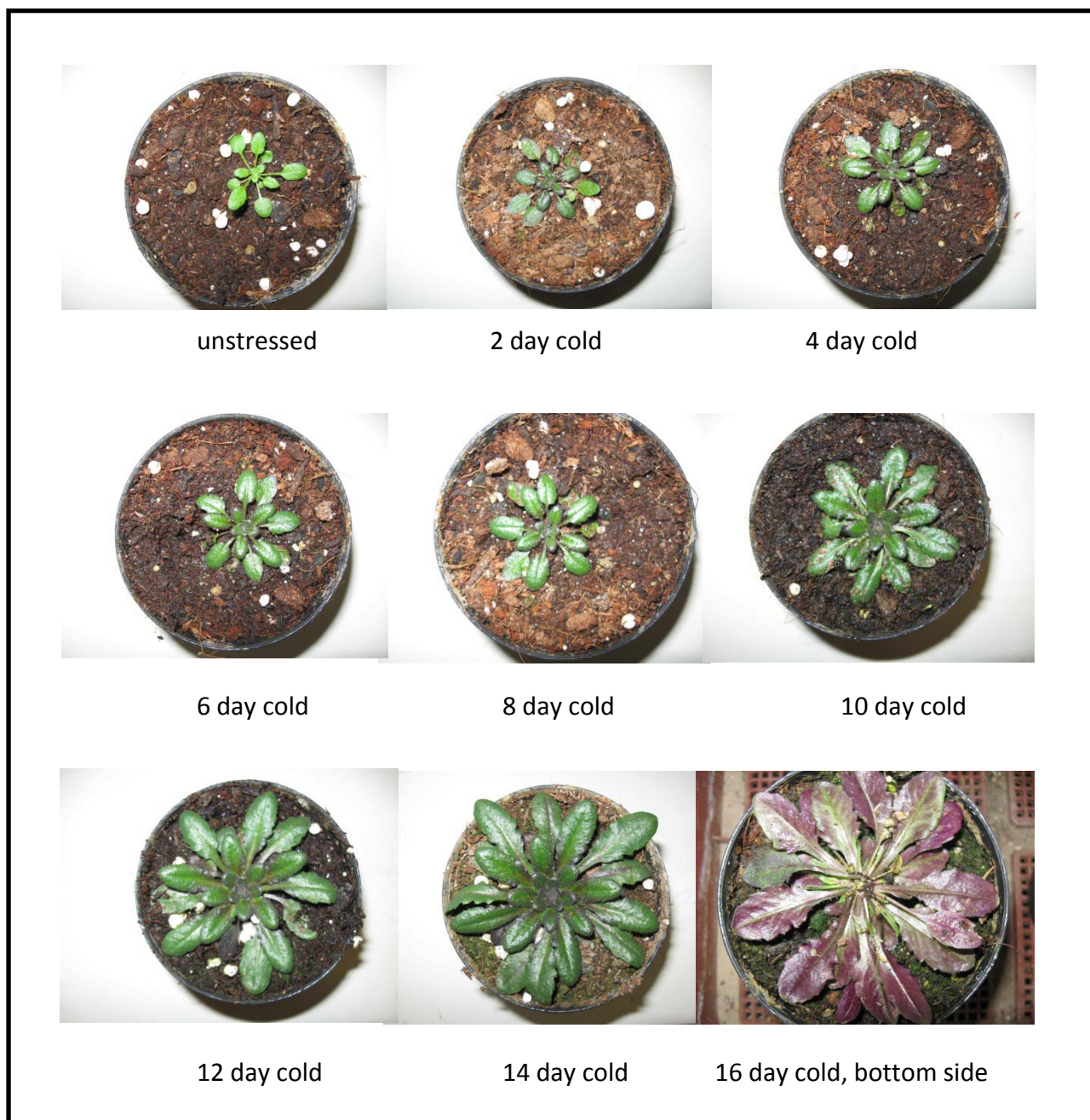


Figure 14: Habitus of short day-grown *A. thaliana* corresponding to the different stages of cold stress

Immediately after cutting the leaves from the plants, they were put in aluminium bags and deep frozen in liquid nitrogen in order to squelch all metabolic processes. Plant material was subsequently stored at -80°C until further treatment.

2.2 The GC-MS approach

2.2.1 Sample preparation for GC-MS analysis

As explained above, the method of choice for measuring the small molecules of the metabolome like sugars and amino acids is gas chromatography coupled to mass spectrometry (GC-MS). In order to make the desired compounds amenable to analysis, extensive sample preparation is needed. In the following a standard metabolite sample treatment is described, which can be applied to other plant material as well, like it is done with metabolomic studies of e.g. *Medicago* and *Populus* in the MoSys department. Sample treatment consists of two big steps: extraction of the polar metabolites and the derivatisation of them.

Extraction process

Deep frozen plant material was grinded using a mortar and pistil under constant adding of LN to prevent thawing and thus undesired reactivation of metabolic processes. Once a fine powder was obtained, about 45 mg of each replicate was conveyed into pre-cooled Eppendorf tubes. For the extraction process, 1 ml of ice cold MCV (5:2:1) extraction mixture was subsequently added into each ET. Tubes were being vortexed for several seconds and incubated on ice for 8 minutes. Hereupon, the samples were centrifuged for 4 minutes at 14.000 g, which should provide proper separation of the metabolic phase from the unwanted macromolecules like proteins and other cell structure components. The MCV phase was afterwards carried over into a new ET. After separating the desired fraction from the remains, 500 µl of milli-q water were added to each tube in order to create an isolated water/methanol phase which should contain most of the polar metabolites. To get rid of the solvent, the samples were put into a speedvac overnight, yielding an ET with a dry pellet consisting of most of the polar metabolites of *A. thaliana* leaves.

Derivatisation process

Two major steps are included in the metabolite derivatisation process: methoximation and silylation. The methoximation is done at first, which is primarily needed for opening ring sugars; this results in two stereoisomers, which are then separated on the GC column. Therefore, a solution of 40 mg methoxyamine hydrochloride in 1 ml pyridine was created, each sample was dissolved in 20 µl, and put on a thermoshaker at 30 °C for 90 minutes.

In the ensuing silylation process, the aim is to increase the volatility of the analytes by introducing trimethylsilyl groups: 80 µl of MSTFA were added into the tubes and incubated on a thermoshaker at 37 °C for at least 30 minutes. Depending on the workflow and if previous samples were measured before, MSTFA vials were spiked with alkane standards in order to create a retention index method for further data processing as explained later. Figure 15 shows the principle of a silylation and methoximation reaction.

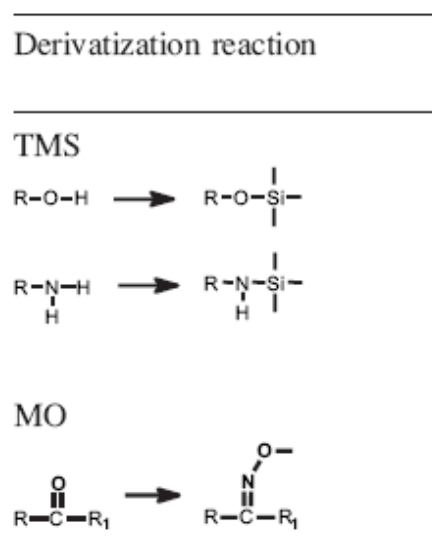


Figure 15: The two derivatisation steps needed in order to render the polar metabolites fit for GC analysis (TMS = trimethylsilylation, MO = methoximation)

After successful silylation, samples were centrifuged at 14.000 g for 2 min. and 50 µl of each sample was conveyed into a GC vial with a micro insert.

2.2.2 GC-MS analysis – instrumental settings

GC measurements were carried out with a LECO Pegasus® 4D GCxGC-TOF mass spectrometer (Figure 16). This device is capable of performing 2D-measurements, meaning that the analytes, after being separated by a column, are cooled down with LN at a specific time and, after being evaporated again, are once again sent onto a column. With this technique, co-elution of similar compounds can be thwarted, and more metabolites are detected in general. During the course of this project, 2D measurements have been done and 2D methods have been developed and compared, however, due to time constraints as well as

the complexity and size of 2D data, the final samples were measured one-dimensional. For establishing reliable and reproducible 2D analyses on the GCxGC mass spectrometer, an in-depth look at the produced data, especially on the peak deconvolution, would be needed.



Figure 16: LECO Pegasus® 4D GCxGC-TOFMS

GC method

The column used was an Agilent HP-5MS (30 m length, 0,25 mm diameter, 0,25 μ m film) with a 4 mm inner diameter tapered liner (deactivated glass wool). Injection was done splitless with 230 °C injection temperature. The initial oven temperature was 70 °C for one minute, followed by a ramp of 9 °C per minute with 310 °C target temperature, being held for 5 minutes, and after that a 20 °C jump to 330 °C, also being held for 5 minutes.

MS method

Length of a run was 35 minutes, after an acquisition delay of 7,5 minutes, with a starting mass of 40 u and an end mass of 600 u. Acquisition rate was 15 spectra / second, detector voltage 1600 V, electron energy – 70 V, ion source temperature 200 °C.

Data processing

Recorded spectra were processed using the LECO Pegasus software, which features spectra deconvolution, automated peak search, library comparison and base line correction. For peak identification, a retention index method had to be created (by running a sole alkane mix, or adding the alkanes into the MSTFA during derivatisation), followed by the arrangement of a reference method, which includes the desired metabolites found in a similar sample (an unstressed standard sample was injected with each batch analysed) through a database [26]. The reference compounds can be found in Table I.

2.3 The LC-MS approach

Contrary to analysing small molecules with GC-MS, metabolomic analysis via LC-MS is a rather novel approach which has emerged in recent years and methods and databases are just now being developed [27] [28] [29]. Thus, there exist not many standardised metabolomic-scale LC-MS protocols or comparable databases yet, at least not in the extent of that of GC-MS systems; mass spectrometric data of LC-separated metabolites varies between instrument types, depending which fragmentation type and collision energy has been chosen to work with. Also, to date there exists no universal retention index method in LC-MS metabolite analyses.

In line with this project, the onset of an in-house plant metabolite database was to be established. The instrument used was a Thermo LTQ Orbitrap XL FT mass spectrometer (Figure 17), which has the potential to deliver the accurate mass and high-resolution measurements.



Figure 17: Thermo Orbitrap® mass spectrometer

2.3.1 Building a plant secondary metabolite database

In order to acquire spectra from molecules, the chemical compounds were presented to the mass spectrometer in a methanolic solution via direct infusion (meaning no LC column was used). Therefore, a batch of about 10 pure substances each in low concentration was injected using a syringe with a flow rate of about 1-2 μl / min. The ensuing electrospray ionisation process has been described in 1.4.1. For protonation, FA (10 %) was added to each set of pure substances (approximately 10 μl FA per ml of solution).

A list of all the injected compounds can be found in Table II.

Subsequently, each compound, whose accurate mass had to be known, was identified and isolated (with an isolation width of 1 m/z) and recorded via the instrument's software using 20 scans for each compound. Fig. 18 shows how the precursor mass of such a typical compound was displayed on the instrument.

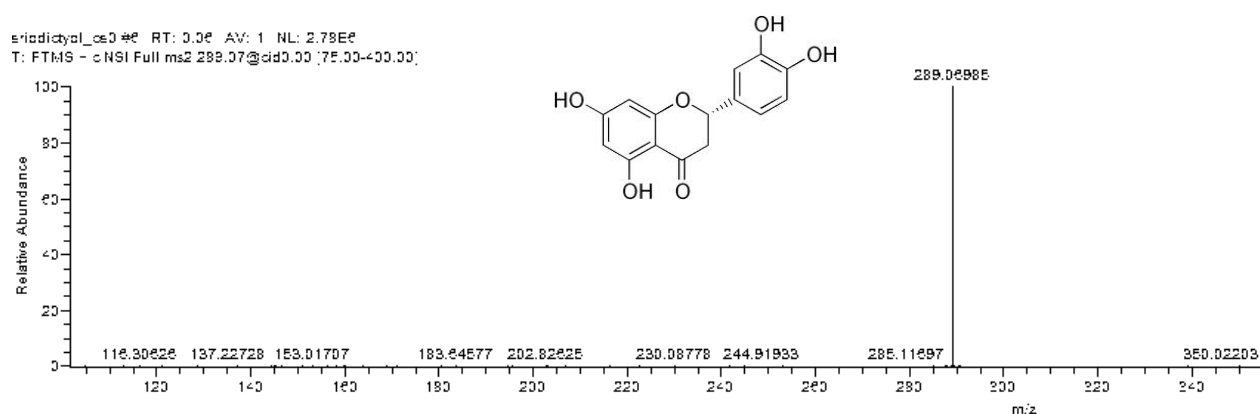


Figure 18: Precursor spectrum of the flavonoid Eriodictyol. The carbon structure is shown in the upper right corner; the dominant peak belongs to the singly protonated form $[M+H]^+$

For obtaining spectral information, the precursor molecule was fragmented, using CID and a CE of 50 eV. The resulting spectra is shown in Figure 19.

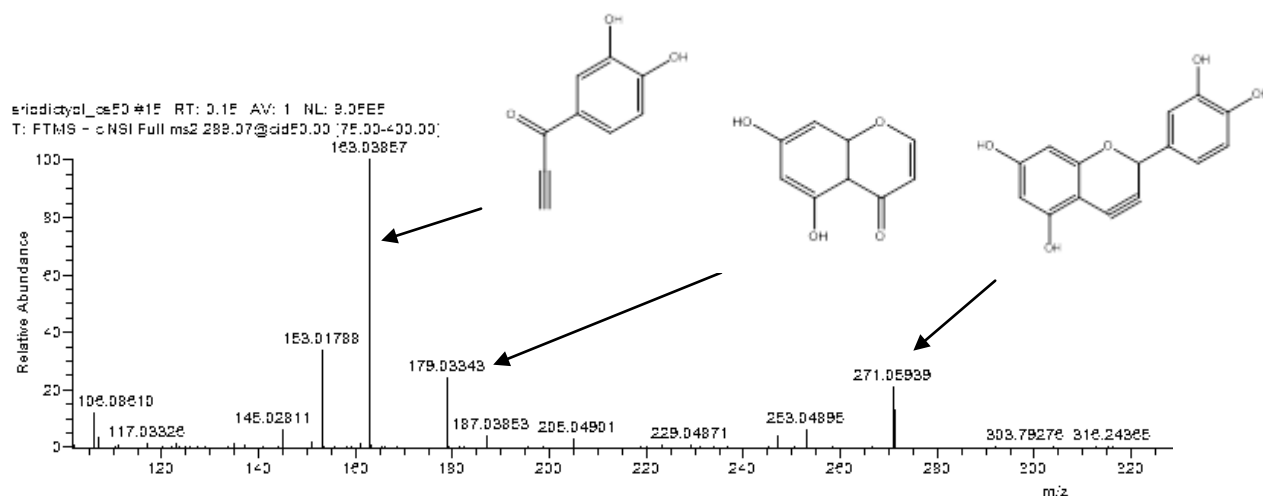


Figure 19: Spectrum of Eriodictyol after applying 50 eV collision energy. The fragments matching the peaks are in accordance with already existing MS² data [30] [31]

The MSⁿ ability of the Orbitrap also allows fragmentation beyond the MS² level, which is useful for further characterising a compound, however at the expense of sensitivity. Figure 20 shows a MS³ spectrum of a selected Eriodictyol fragment.

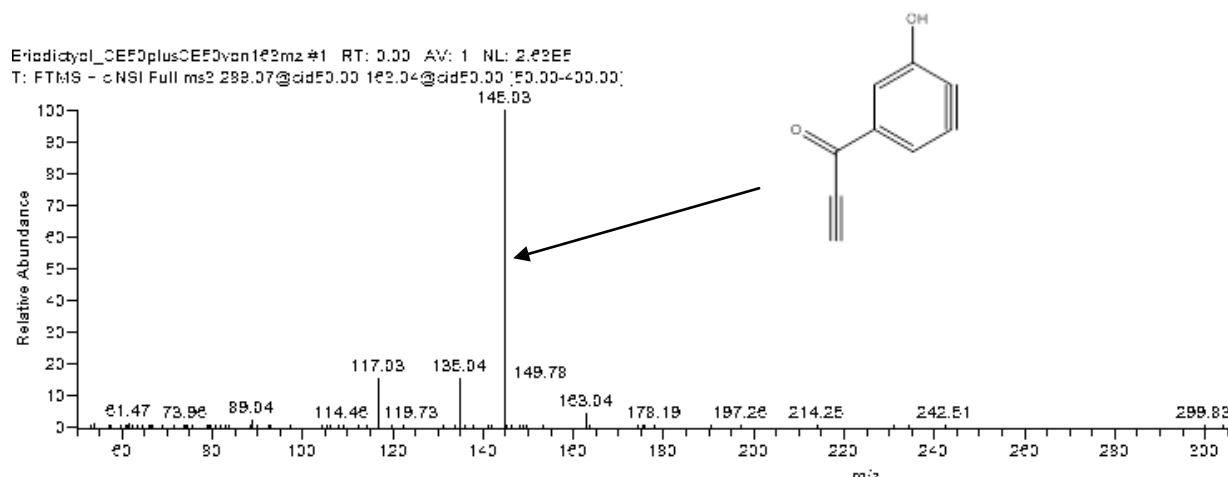


Figure 20: MS³ spectrum of the 163 m/z Eriodictyol fragment, 50 CE. The peak at 145 m/z might result through a loss of a H₂O molecule; a putative compound is given.

2.3.2 Sample preparation for LC-MS analysis

Parallel to recording spectra from pure substances, an attempt was made to find a suitable method for measuring plant metabolites on an LC-Orbitrap setup.

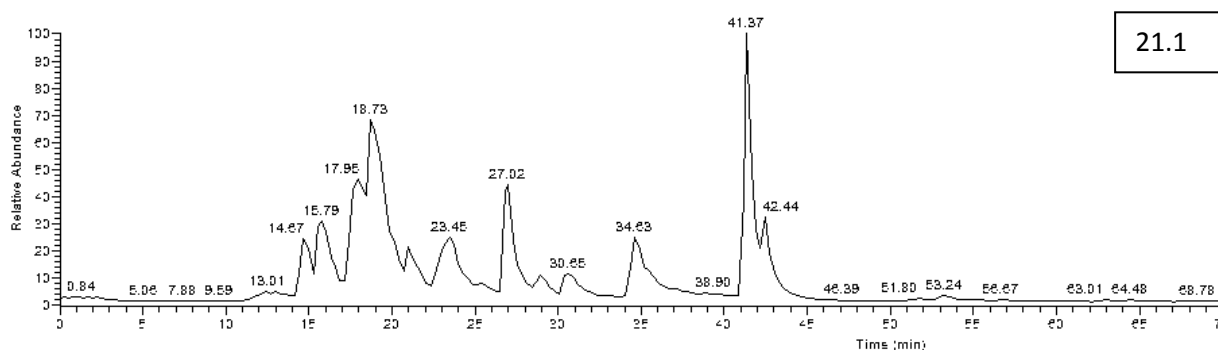
Various extraction tests with Arabidopsis resulted in the development of the following workflow:

- Pulverisation of 50 mg of deep frozen plant material
- Addition of 1 ml of pre-chilled 80/20 MeOH/H₂O extraction solution with 1 µg of internal standard
- Centrifugation at 20.000 g for 15 minutes
- Conveying the supernatant into a new tube and drying in a speedvac
- Dissolving the sample in 100 µl of 50/50 MeOH/H₂O solution
- Centrifugation at 20.000 g for 15 minutes
- Conveying the supernatant and filtering each lot through a stage tip (Empore/Disk C18 diameter 47mm) into a vial with a micro insert for ensuing analysis

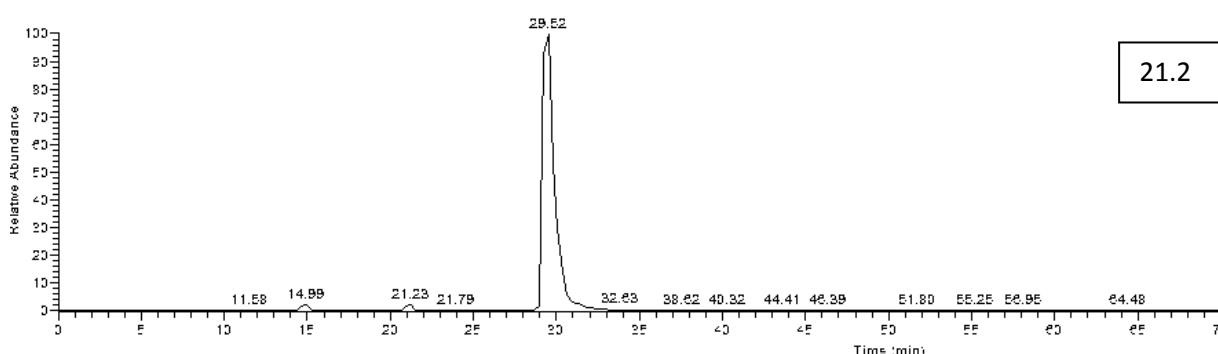
After initial measurements, it was found that the cold stressed *Arabidopsis* samples were so high in fat compounds that they left the ion transfer capillary completely clogged. Therefore, 200 ml of CHCL₃ were added into each vial for removing apolar substances, and after short vortexing, the newly created chloroform phase was separated from the remaining sample.

2.3.3 LC-MS analysis – instrumental settings

The column used was a WATERS HSS T3; mobile phases were H₂O (+FA) (A), and 90% ACN (+FA) (B). The mass spectrometer method was a top 2 method, facilitating a resolution of 60 000 resolution (FS) with a 7500 resolution MS² run. Exclusion list was set to maximum (500), exclusion duration 90, repeat count 1; with a signal threshold of 1000. A quality control mix containing 19 typical plant compounds with different chemical properties was measured several times before actual analyses, to see if they had a reasonable retention time with the chosen column and gradient. Figures 21.1-21.4 show Eriodictyol in a QC mix sample run, and its recovered MS² spectrum in comparison with the one of the directly injected pure substance.



21.1



21.2

Figure 21.1: Chromatogram of the pure substance QC mix

Figure 21.2: Extracted Eriodictyol mass (289.07067) from 20.1. With this gradient and column, retention time is around minute 30

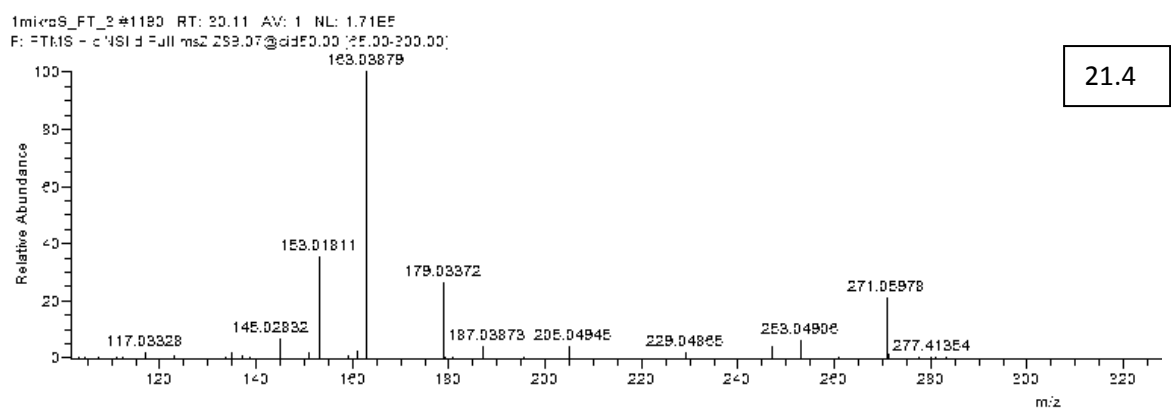
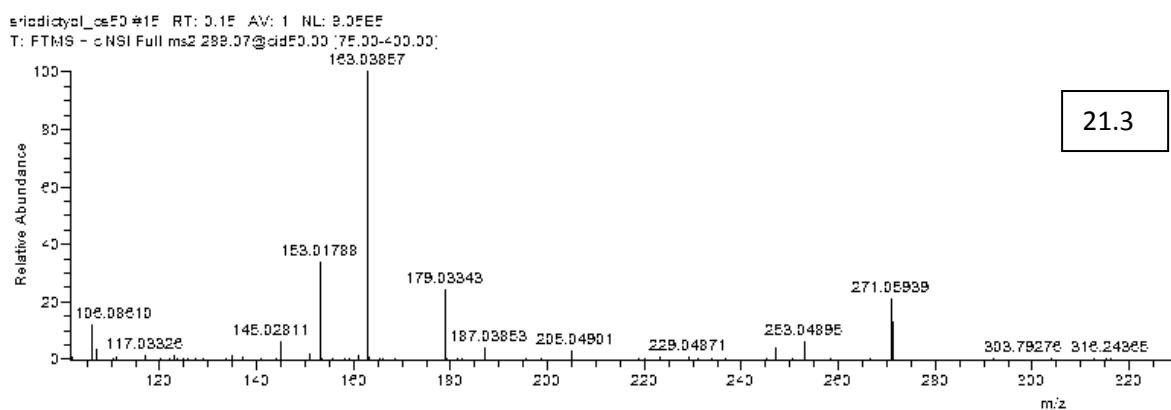


Figure 21.3: MS² spectrum of Eriodictyol as recorded with direct infusion (see also Fig. 18)

Figure 21.4: Recovered MS² spectrum of the m/z 289 peak in the QC mix at minute 30 (highly likely Eriodictyol).

Mass deviation from precursor is 0.104 ppm; -1.50 for the 153m/z fragment, -0.25 for 163m/z, -1.62 for 179m/z, -1.44 for m/z 271.

3. Results

3.1 GC-MS-Measurements

With the NIST and gmd database, 65 metabolites could be identified with a match factor of at least 850, or less if the retention time and the library hits were reliable enough to assign a compound to the peak(s). The results were normalized to fresh weight and an internal standard (C13 sorbitol). It was found that the last 3 time points in the cold stress experiment showed next to no more substantial change to the compound levels once day 14 was reached, but rather brought adulteration to the analyses because of aging of the plant leaves, which was an undesired factor and thus were not accounted for yet.

3.1.1 Amino acids

12 amino acids could be retrieved through the measurements. Fig. 22.1 shows the total and relative amino acid levels throughout the course of the experiment, Fig. 22.2 the total amino acid amount in the samples.

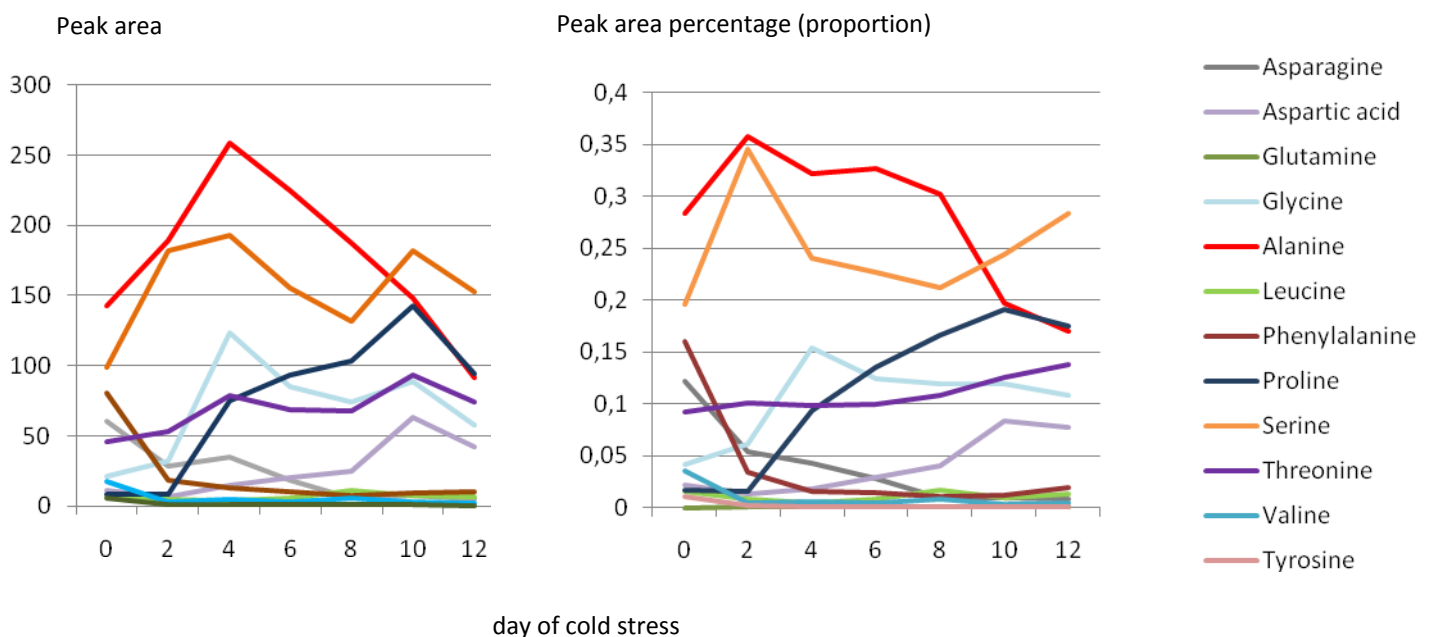


Figure 22.1: Total and relative amino acid levels during the cold stress experiment

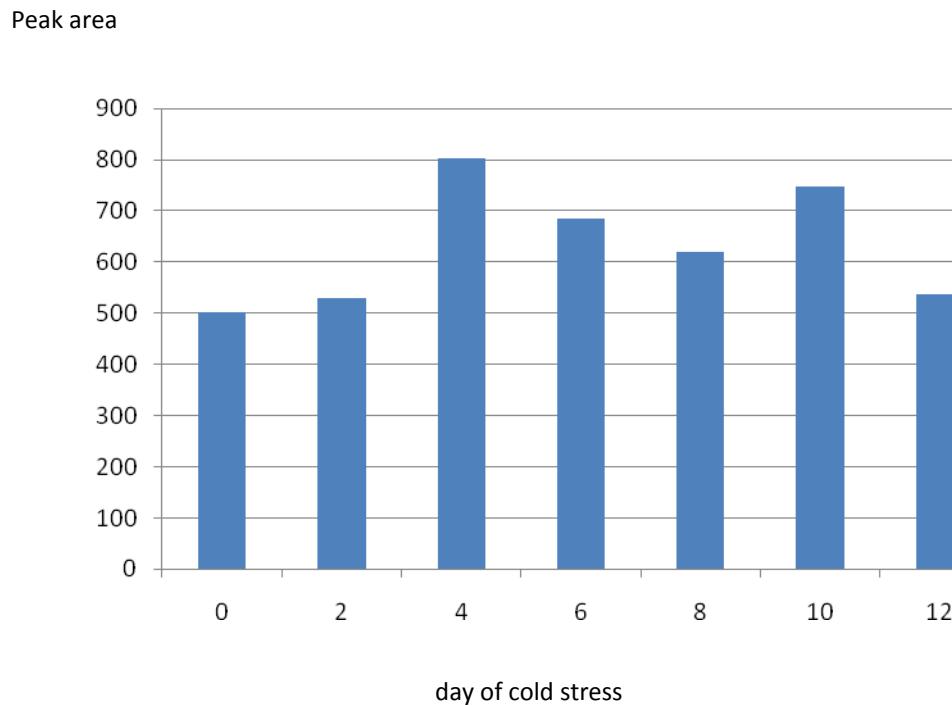


Figure 22.2: Total amino acid amount during the cold stress

It was found that the levels of most of the identified amino acids were strongly changed during the course of the cold stress experiment. The most abundant amino acids, alanine and serine, seem to fall off after being upregulated at the first measured time point, with serine slightly recovering later on. Expected was the upregulation of proline after day 2. Proline is known to be an important stress response factor in plants, acting as an osmoprotectant [23] [32] [33], and its increase in *Arabidopsis* during cold stress has already been reported [24]. An interesting finding is the depletion of the phenylalanine and tyrosine pool, which might be connected to the fact that these two compounds are important starting points in the phenylpropanoid biosynthesis [34], from which later on flavonoids and anthocyanins are resulting, who probably are responsible for the coloration of the plant as shown in Figure 13.

3.1.2 Sugars

8 types of sugars were identified. Fig. 23.1 shows the total and relative sugar levels throughout the course of the experiment, Fig. 23.2 the total sugar amount in the samples.

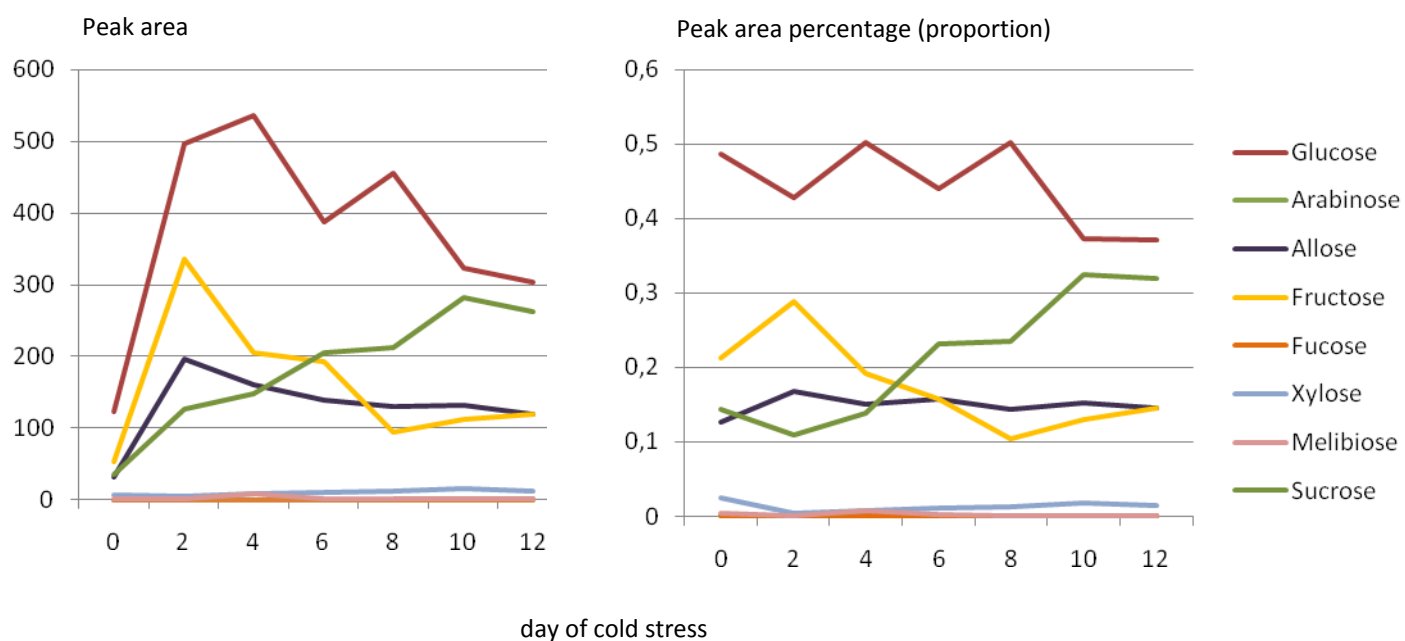


Figure 23.1: Total and relative sugar levels during the cold stress experiment

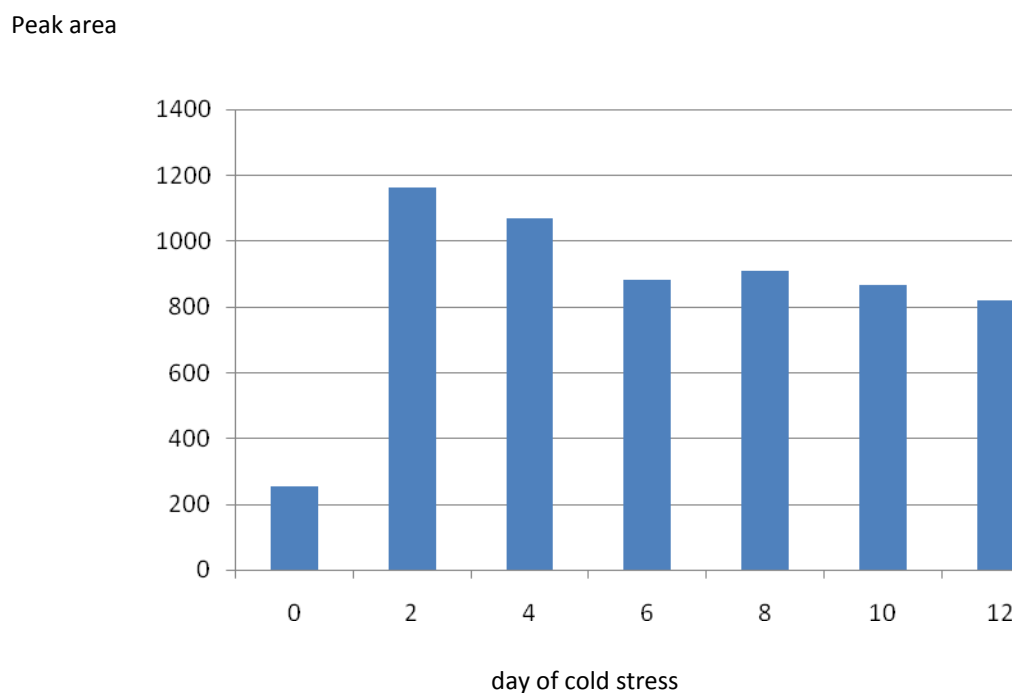


Figure 23.2: Total sugar amount during the cold stress

The most significant changes in the sugar levels were the slow downregulation of glucose and fructose, while sucrose showed a constant upregulation during the course of the cold stress. The total sugar amount decreases slightly after a sudden jump in the first 48 hours.

3.1.3 Polyols

6 types of polyol-like compounds could be identified. Fig. 24.1 shows the total and relative polyols levels throughout the course of the experiment, Fig. 24.2 the total polyol amount in the samples.

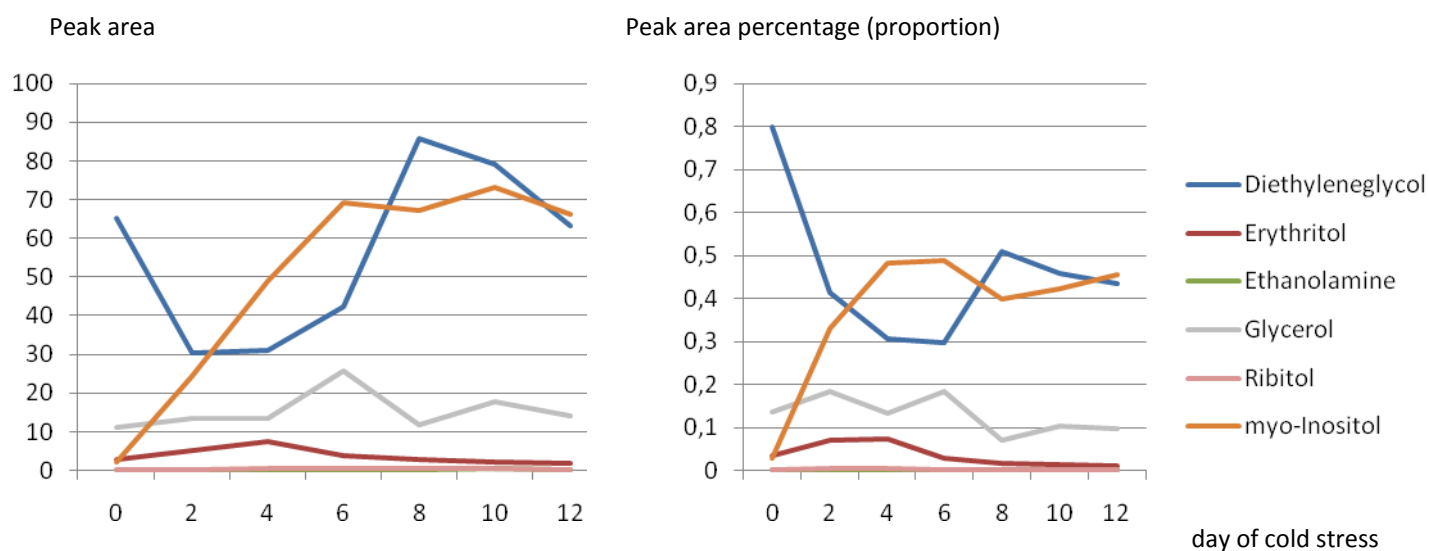


Figure 24.1: Total and relative polyol levels during the cold stress experiment

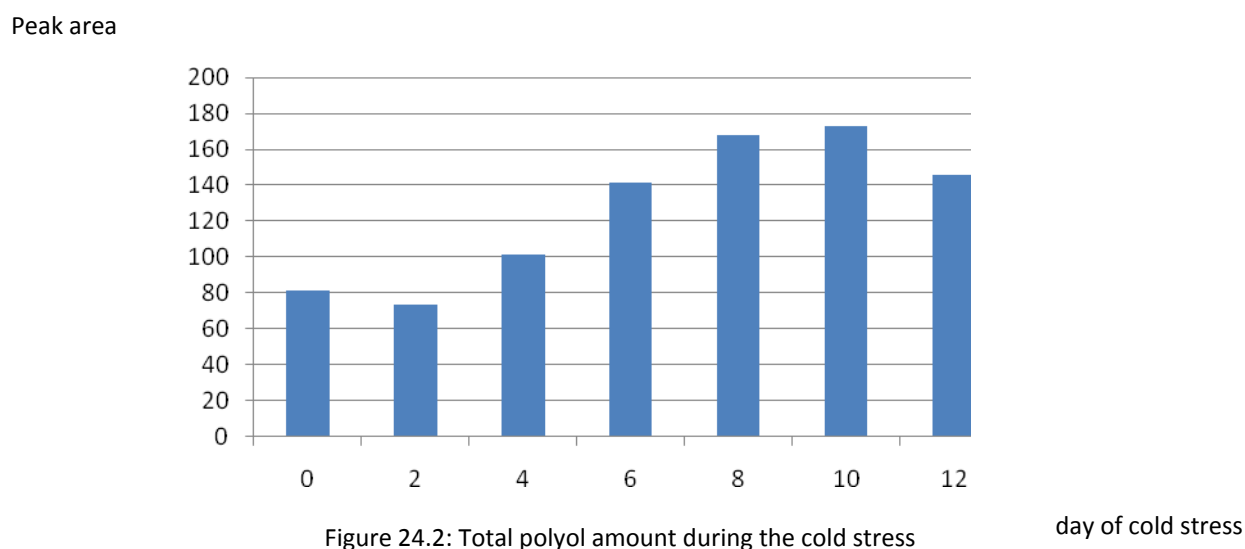


Figure 24.2: Total polyol amount during the cold stress

Figure 23.2 shows that there is a general increase in polyol amount during the cold stress, which is understandable, since sugar alcohols are recognized to act as a osmoprotectants in plants. An interesting finding are the near perfect contrary levels of diethylene glycol and myo-inositol.

3.1.4 TCA cycle components

5 components of the TCA cycle could be identified. Fig. 25.1 shows the total and relative TCA component levels throughout the course of the experiment, Fig. 25.2 the total amount in the samples.

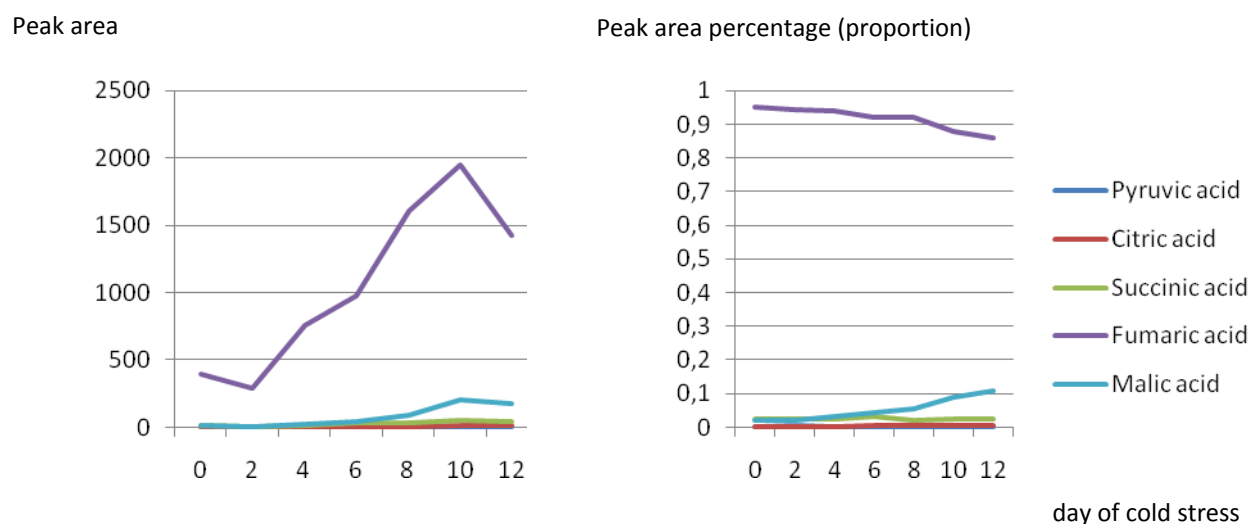


Figure 25.1: Total and relative TCA component levels during the cold stress experiment

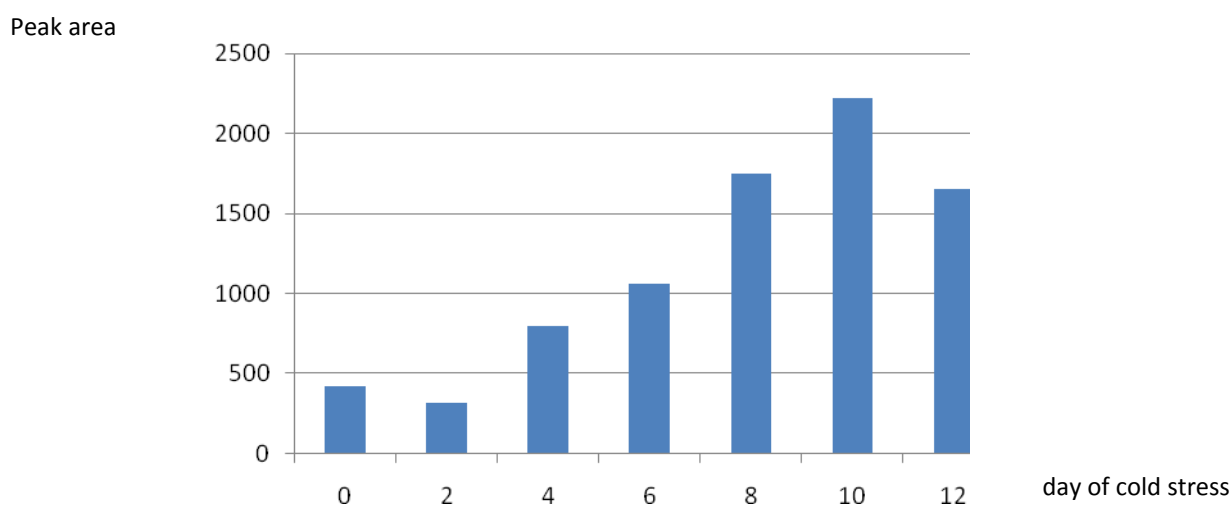


Figure 25.2: Total TCA component amount during the cold stress.

The results in the TCA compound measurements point to a strong amplification of this metabolic pathway; especially fumaric and malic acid are upregulated after the first 48 hours of cold stress.

3.2 LC-MS-Measurements

Databases and software for LC metabolite measurements is just an emerging field right now, thus metabolite identification cannot be done as in GC-MS analyses. However, by obtaining an accurate m/z ratio, preliminary annotations for compounds can be made [36]. Figure 26 shows an LC chromatogram and how annotation of an extracted mass could possibly look like.

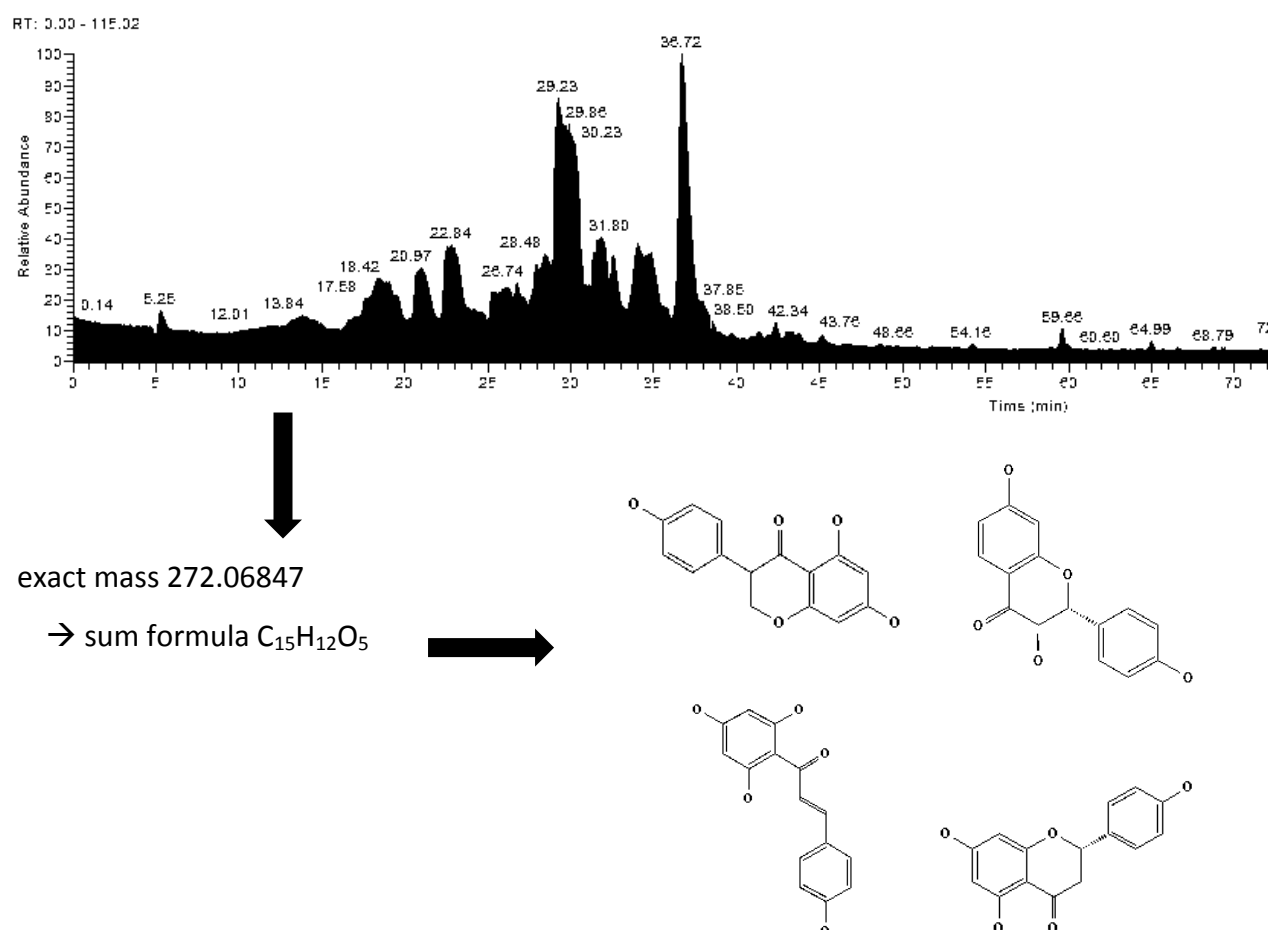


Figure 26: Process of annotation of an exact mass obtained by LC-MS (Orbitrap). When applying C,H and O as constraints of structure search, several compounds are possible (4 are shown). Results yielded by MetFrag [37]

4. Discussion

As shown in 3.1, the cold stress imposed to the *Arabidopsis* plants triggered multiple metabolic changes in the primary metabolism. Figure 27 shows how all the identified metabolites are arranged in a PCA diagram during the cold stress time course.

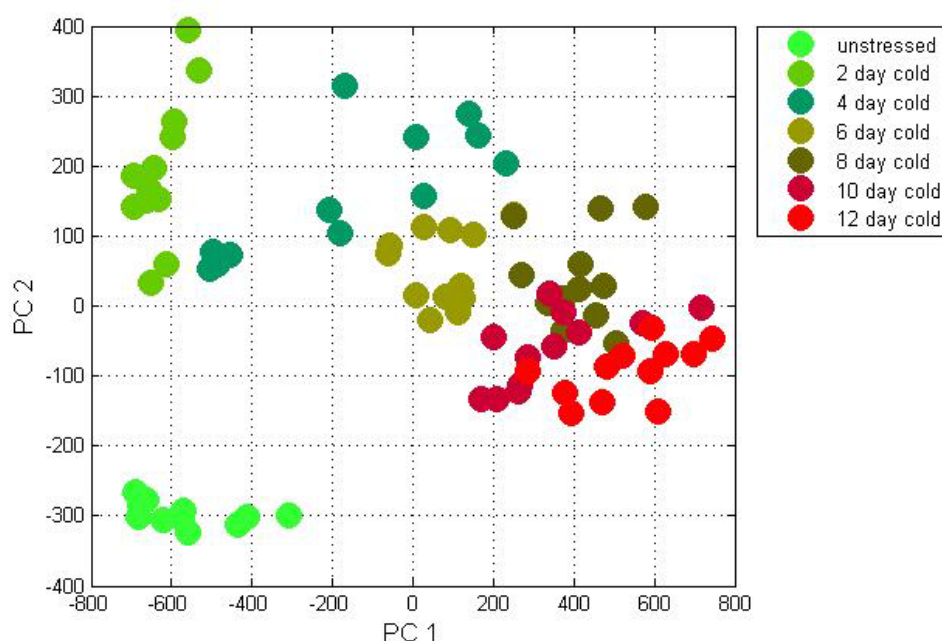


Figure 27: Principal component analysis of the metabolites retrieved by GC measurements. The abscissa (PC 1) explains about 83% of the total variance and acts like a time axis for this experiment.

The distance between the unstressed samples (the bright green dots in the lower left corner) and those of the first cold stress time point is the biggest one in the diagram, indicating that there is a great shift in metabolite levels in the first 48 hours. In fact, it is known that a plethora of metabolic changes in *Arabidopsis* during cold stress happens during the first few hours [24][35]. Thus, early stage metabolite changes could not be resolved in this experiment; however a small time interval between sample drawing might not be well suited for the observation of changes in the secondary metabolism – also desired in this project – which can be assumed to have a delayed reaction judging on the amount of days it took the *Arabidopsis* leaves to become red.

This metabolic analysis is, of course, far from being complete. Not even 100 distinct metabolites could be reliably identified within this experiment; when enhanced data deconvolution is applied to the measurements, the number of potential metabolites is raised into the range of 1000s [38]. Accordingly, highly abundant but unidentified compounds could, or should possibly be included in metabolomic analyses, as the databases are not complete yet and still being improved, and the possibility of finding relations of certain unknowns to already identified compounds and/or pathways is always there.

For absolute quantification (the internal C13 sorbitol standard just corrects the unsteadiness of the instrument when measuring several samples), an external standard quality control mix is yet to be created, which allows quantitating selected and common metabolites (sugars, amino acids, sugar alcohols, etc.). This will be achieved by setting up calibration curves at differing concentrations of certain compounds, ideally yielding a linear concentration / signal plot, from which then the absolute amount of the metabolite can be deduced. When having single analytes measured, a positive side effect of this procedure will be a precise insight of the GC behaviour of several hexose sugars, which are very similar to each other both in retention time and fragment ions, and often cannot be assigned accurately, despite having a high abundance.

Concerning the LC measurements, even though a basic chromatographic separation was successfully achieved by choice of column, gradient and extraction method, there is still a vast area of improvement to be done. Especially the apolar fraction, which had to be isolated from the cold stressed LC samples (see 2.3.2), is supposed to contain interesting compounds with larger carbon structures; thus rendering this phase amenable to LC analysis might be a goal for future method development. Also, the extension of the in-house LC-MS database by various other pure substances, in addition to possibly acquiring a rough retention time when having them elute over a column, is slated for the future.

As far as spectra interpretation goes, a very promising current approach is the so-called precursor ion fingerprinting (PIF), where ions are characterised from their tandem product spectra, in order to elucidate the structure of unknowns [39].

4.1 Outlook

After obtaining data from the GC measurements, as well as establishing a chromatographic and mass spectrometric LC-MS method for metabolite detection, thus achieving a high metabolome coverage, the ultimate goal is to combine both data sets in order to create a relation between both primary and secondary metabolism, ideally highlighting correlations or finding hints for new pathways or compounds [40]. The downregulation of phenylalanine and tyrosine and the upregulation of flavonoids in this study point exactly to this relationship.

Bringing GC-MS and LC-MS data together faces several obstacles, due to the differing nature of the raw data, which stems from the different mass spectrometers from which they are created. Here, the use of Z transform can help. In order to gain a more complete picture of the conducted experiment, and an even higher coverage of the biochemistry of Arabidopsis during cold stress, retrieving supplemental protein data from the corresponding cold stress time points would be an useful approach.

5. Summary

Metabolomics is – besides proteomics and genomics – one of the most important bioanalytical approaches today and pursues nothing less than to elucidate the entity of the metabolism of a living system at a certain point of time – the *metabolome*. Although a rather young technique and under constant development, the first results of metabolic investigations look very promising. Fields of application for metabolomic studies include plant and crop research, as well as medicinal and pharmaceutical problems. Also, regarding the energy issue, metabolomics is going to play a major role for discovering and optimising renewable biofuel sources (“third generation biofuels”).

In this study, the cold stress metabolome of *Arabidopsis thaliana* is being investigated. One part of the metabolome, the so-called primary metabolism - containing the small molecules of the basic vital functions (sugars, amino acids, organic acids, etc.) – is being analysed with a standardised gas chromatography – mass spectrometer setup. The bigger molecules of the secondary metabolism, containing interesting compounds of various ecological functions for the plant (like flavonoids, anthocyanins, lipids etc.), are being measured with a liquid chromatography – mass spectrometer technique, in line with the development of an LC-Orbitrap metabolite detection method as well as an in-house database of plant secondary metabolites.

The GC-MS measurements revealed a plethora of metabolic changes during a two week cold treatment of *A. thaliana*, with various compounds being upregulated over time, whereas others were found to be decreased, or supposedly used up by the secondary metabolism. The LC-MS measurements were attempted to be optimised on a reversed-phase column, with a mass spectrometer method to obtain as much metabolite data as possible while being able to extract exact masses and fragmentation scans as well.

The work done sets the basis for the correlation of both GC-MS and LC-MS data sets, which allows a further elucidation of metabolic relations as well as formulating structure proposals and possibly metabolic pathway prediction.

5. Zusammenfassung

Neben Proteomics und Genomics ist *Metabolomics* heute eine der wichtigsten bioanalytischen Techniken und hat zum Ziel, das *Metabolom* – die Gesamtheit aller Stoffwechselprodukte eines lebenden Systems zu einem bestimmten Zeitpunkt – aufzuklären. Obschon eine junge Technik und in ständiger Entwicklung, versprechen die ersten Daten eine positive Zukunft für metabolische Untersuchungen. Anwendungsgebiete im Bereich Metabolomics sind z.B. Pflanzen- und Kulturpflanzenforschung, sowie medizinische und pharmazeutische Fragestellungen. Ebenso werden metabolische Studien in der Frage um erneuerbare Energien eine entscheidende Rolle spielen (Stichwort: “Biotreibstoffe dritter Generation”)

In diesem Projekt wurde das Kaltstress-Metabolom von *Arabidopsis thaliana* untersucht. Der sogenannte Primärmetabolismus enthält alle kleineren Metaboliten, die für die Aufrechterhaltung der Lebensfunktionen einer Pflanze unabdingbar sind (wie z.B. Zucker, Aminosäuren, organische Säuren, etc.) und wird einer standardisierten GC-MS – Analyse unterworfen. Die größeren Moleküle des Sekundärmetabolismus, welcher ökologisch interessante Verbindungen enthält (Flavonoide, Anthocyane, Fette, etc.) werden mit einer LC-MS – Vorrichtung vermessen, im Zuge einer Etablierung einer Detektionsmethode für Metaboliten an einem LC–Orbitrap – System und der Entwicklung einer internen Datenbank für sekundäre Pflanzenstoffe.

Die GC-MS – Messungen enthüllten eine Reihe von metabolischen Veränderungen während eines zweiwöchigen Kaltstresses an *A. thaliana*, wobei mehrere Verbindungen hochreguliert wurden, während andere verringert vorlagen, oder in den Sekundärmetabolismus einwanderten. Die LC-MS – Messungen wurden für eine Umkehrphasenchromatographie optimiert, wobei versucht wurde, so viele Metabolit-Daten wie möglich zu erhalten bei gleichzeitig hoher Qualität von Massengenauigkeit und Anzahl von Fragmentationen-Scans.

Die Arbeit bildet den Grundstein für die folgende Korrelation von GC-MS und LC-MS – Daten, wodurch eine weitere Aufklärung von metabolischen Relationen möglich werden soll, sowie der Formulierung von Strukturvorschlägen und Aufklärung von Reaktionswegen.

6. Acknowledgements

I want to express my gratitude to everyone who helped me during the work presented in this thesis. I would like to thank Wolfram Weckwerth for the essential guidance and advice and for enabling this project in the first place.

In particular, I thank David Lyon for his instructions concerning LC and Orbitrap handling, and Lena Fragner for support with GC-MS issues. On the bioinformatic side, I want to thank Volker Egelhofer for his great collaboration concerning the metabolite database, as well as Xiaoling Sun for providing data analysis tools.

Thanks to Andreas Schröfl and Thomas Joch for providing me with plant material.

Finally, I want to thank the whole MoSys team for the overall support and a comfortable working atmosphere.

Bibliography

- [1] Weckwerth, W. (2003) Metabolomics in systems biology. *Annu. Rev. Plant Biol.* 54, 669–689
- [2] Ideker, T., Galitski, T., and Hood, L.: *A new approach to decoding life: Systems Biology.*, *Annu. Rev. Genomics Hum. Genet.* 2: 343-372 (2001)
- [3] Van der Greef Jan; Hankemeier Thomas; McBurney Robert N, Metabolomics-based systems biology and personalized medicine: moving towards n = 1 clinical trials? *Pharmacogenomics* 2006;7(7):1087-94.
- [4] Fernie AR, Schauer N: Metabolomics-assisted breeding: a viable option for crop improvement? *Trends Genet.* 2009, 25:39–48.
- [5] Atsumi, S. & Liao, J.C.: Metabolic Engineering for Advanced Biofuels Production from *Escherichia coli* *Curr Opin Biotechnol.* Vol 19, 414-419 (2008)
- [6] Fiehn, O.: Comparative and Functional Genomics. *Comp Funct Genom* 2001; 2: 155–168.
- [7] Goodacre, Royston; Vaidyanathan, Seetharaman; Dunn, Warwick B.; Harrigan, George G.; Kell, Douglas B. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends in Biotechnology.* 2004; 22: 245-252.
- [8] Fiehn, O.: *Plant Mol. Biol.* 48 (155-171), 2002
- [9] Meinke, et al. *Arabidopsis thaliana*: A Model Plant for Genome Analysis. *Science*, 23 October 1998: 662-682.
- [10] Koornneef M, Meinke D (2010) The development of *Arabidopsis* as a model plant. *Plant J.* 61:909-921
- [11] Rhee SY, Beavis W, Berardini TZ, Chen G, Dixon D, Doyle A, Garcia-Hernandez M, Huala E, Lander G, Montoya M, Miller N, Mueller LA, Mundodi S, Reiser L, Tacklind J, Weems DC, Wu Y, Xu I, Yoo D, Yoon J, Zhang P. (2003) The Arabidopsis Information Resource (TAIR): a model organism database providing a centralized, curated gateway to Arabidopsis biology, research materials and community. *Nucl. Acids Res.* 31(1):224-8
- [12] Joshi, H.J., Hirsch-Hoffmann, M., Baerenfaller, K., Gruissem, W., Baginsky, S., Schmidt, R., Schulze, W.X., Sun, Q., van Wijk, K., Egelhofer, V., Wienkoop, S., Weckwerth, W., Bruley, C., Rolland, R., Toyoda, T., Nakagam, H., Jones, A., Briggs, S.P., Castleden, I., Tanz, S., Millar, A.H., and Heazlewood, J.L. (2011). MASCP Gator: An aggregation portal for the visualization of Arabidopsis proteomics data. *Plant Physiol.* 155, 259-270.
- [13] Hiroyuki Ogata, Susumu Goto, Kazushige Sato, Wataru Fujibuchi, Hidemasa Bono, Minoru Kanehisa. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 1999: 29~34
- [14] Ron Caspi, Tomer Altman, Joseph M. Dale, Kate Dreher, Carol A. Fulcher, Fred Gilham, Pallavi Kaipa, Athikkattuvalasu S. Karthikeyan, Anamika Kothari, Markus Krummenacker,

Mario Latendresse, Lukas A. Mueller, Suzanne M. Paley, Liviu Popescu, Anuradha Pujar, Alexander Glennon Shearer, Peifen Zhang, Peter D. Karp. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research*, 2010: 473~479

[15] Oliver Fiehn, Tobias Kind and Dinesh Kumar Barupal. Data Processing, Metabolomic Databases and Pathway Analysis. *Annual Plant Reviews* (2011) 43, 367-406

[16] Weckwerth, W. An International Arabidopsis Informatics Consortium. An international bioinformatics infrastructure to underpin the Arabidopsis community. *Plant Cell*. 2010 Aug;22(8):2530-6.

[17] Rodney Croteau, Tony M. Kutchan, Norman G. Lewis: "Natural Product (Secondary Metabolites)", American Society of Plant Physiologist, 2000.

[18] Okuda,S., Yamada,T., Hamajima,M., Itoh,M., Katayama,T., Bork,P., Goto,S. and Kanehisa,M. (2008) KEGG Atlas mapping for global analysis of metabolic pathways. *Nucleic Acids Res.*, 36, W423–W426.

[19] Abdullah, M.A., A. Rahmah, A.J. Sinskey, and C.K. Rha (2008), Cell Engineering and Molecular Pharming for Biopharmaceuticals, *Open Medicinal Chemistry J.*, 2, 49–61.

[20] Weckwerth, W. and Fiehn, O. (2002) Can we discover novel pathways using metabolomic analysis? , *Current Opinion in Biotechnology* 13: 156-160

[21] Halket, J.M., Waterman, D, Przyborowka, A.M., Patel, R.K.P., Fraser, P,D,, Bramley, P.M. Chemical derivatization and mass spectral libraries in metabolic profiling by GC/MS and LC/MS/MS, *Journal of Experimental Botany* 56(410) (2005) 219.

[xx] NMR-based metabolomics. *Drug Chem Toxicol*. 2002;25:375–382.

[22] Kim, J.K., T. Bamba, K. Harada, E. Fukusaki, and A. Kobayashi. 2007. Time-course metabolic profiling in *Arabidopsis thaliana* cell cultures after salt stress treatment. *J. Exp. Bot.* 58:415-424.

[23] Shulaev V, Cortes D, Miller G, Mittler R (2008) Metabolomics for plant stress response. *Physiologia Plantarum* 132: 199-208. PMID: 18251861

[24] Kaplan, F., Kopka, J., Haskell, D. W., Zhao, W., Schiller, K. C., Gatzke, N., Sung, D. Y. and Guy, C. L. (2004) Exploring the temperature-stress metabolome of Arabidopsis , *Plant Physiology* 136: 4159-4168

[25] Jeong S, Clark SE. Photoperiod regulates flower meristem development in *Arabidopsis thaliana*. *Genetics*. 2005;169:907–915

[26] GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics* 21(8): 1635-1638 (2005)

[27] Kliebenstein, D.J. (2004) Secondary metabolites and plant/environment interactions: a view through *Arabidopsis thaliana* tinged glasses. *Plant Cell and Environment* 27(6):675-684

- [28] von Roepenack-Lahaye, E., Degenkolb, T., Zerjeski, M., Franz, M., Roth, U., Wessjohann, L., Schmidt, J., Schee, D., and Clemens, S. (2004) Profiling of Arabidopsis secondary metabolites by capillary liquid chromatography coupled to electrospray ionization quadrupole time-of-flight mass spectrometry *Plant Physiol* 134(2), 548--559.
- [29] D'Auria, J.C., and Gershenzon, J. (2005). The secondary metabolism of *Arabidopsis thaliana*: Growing like a weed. *Curr. Opin. Plant Biol.* 8, 308–316.
- [30] Tsimogiannis D, Samiotaki M, Panayotou G, Oreopoulou V (2007). Characterization of flavonoid subgroups and hydroxy substitution by HPLC-MS/MS. *Molecules*, 12: 593-606.
- [31] Waridel, P., Wolfender, J.-L., Ndjoko, K., Hobby, K. R., Major, H. and Hostettmann, K. (2001). Evaluation of quadrupole time of flight tandem mass spectrometry and ion-trap multiple-stage mass spectrometry for the differentiation of C-glycosidic flavonoid isomers. *J. Chromatogr. A*, 926, 29-41
- [32] Claussen, W. 2005. Proline as a measure of stress in tomato plants. *Plant Sci.*, 168: 241-248.
- [33] Takagi, H. (2008). *Proline as a stress protectant in yeast: physiological functions, metabolic regulations, and biotechnological applications*. *Appl Microbiol Biotechnol* 81, 211–223
- [34] Vogt T. (2010) Phenylpropanoid biosynthesis. *Mol Plant* 3: 2–20
- [35] Kaplan, F., Kopka, J., Sung, D. Y., Zhao, W., Popp, M., Porat, R. and Guy, C. L. (2007) Transcript and metabolite profiling during cold acclimation of Arabidopsis reveals an intricate relationship of cold-regulated gene expression with modifications in metabolite content , *Plant Journal* 50: 967-981
- [36] Giavalisco P, Kohl K, Hummel J, Seiwert B, Willmitzer L (2009) C-13 Isotope - Labeled Metabolomes Allowing for Improved Compound Annotation and Relative Quantification in Liquid Chromatography-Mass Spectrometry-based Metabolomic Research. *Analytical Chemistry* 81: 6546–6551.
- [37] Sebastian Wolf, Stephan Schmidt, Matthias Müller-Hannemann, Steffen Neumann: In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinformatics* 11: 148 (2010)
- [38] Kell, D.B. (2004) Metabolomics and systems biology: making sense of the soup. *Curr. Opin. Microbiol.* 7, 296–307
- [39] Determination of Ion Structures in Structurally Related Compounds Using Precursor Ion Fingerprinting, *J Am Soc Mass Spectrom.* 2009 March; 20(3): 370-376. Michelle T. Sheldon, Robert Mistrik, Timothy R. Croley
- [40] Metabolomics unravel contrasting effects of biodiversity on the performance of individual plant species. Scherling C, Roscher C, Giavalisco P, Schulze ED, Weckwerth W *PLoS One*. 2010 Sep 7;5(9):e1256

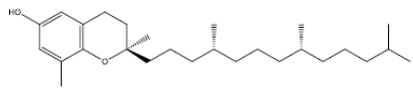
7. Attachment

7.1 Tables

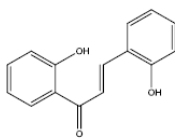
Metabolites identified by GC-MS

*1,3-bisethynylbenzene	*L-Alanine
*2-methyl-Butanedioic acid	*L-Ascorbic acid
*3,4-Dihydroxybenzoic acid	*L-Aspartic acid
*Allose (1MEOX)	*L-Leucine
*Glucose methoxyamine	*L-Phenylalanine
*Arabinose methoxyamine	*L-Proline
*Ascorbic acid	*L-Serine
*Asparagine	*L-Threonic acid
*Aspartic acid	*L-Threonine
*Benzoic acid	*L-Valine
*C13-Sorbitol internal standard	*Lactic acid
*Citric acid	*Maleic acid
*D-Glucopyranose	*Malic acid
*Diethyleneglycol	*Melibiose
*Erythritol	*Nonadecanoic acid
*Ethanolamine	*Oxalic acid
*Fructose methoxyamine	*Phosphoric acid
*Fucose methoxyamine	*Fructose (1MEOX)
*Fumaric acid	*Putrescine
*Gluconic acid	*Pyroglutamic acid
*D-Xylose	*Pyruvic acid
*Allose methoxyamine	*Ribitol
*Glucose, 1,6-anhydro,beta-	*Ribonic acid
*Glutamine, DL-	*Shikimic acid
*Glyceric acid	*Sinapic acid, trans-
*Glycerol	*Spermidine
*Glycine	*Succinic acid
*Glycolic acid	*Sucrose
*Hexadecanoic acid	*Threonic acid-1,4-lactone
*Hexanoic acid	*Tyrosine
*Hexose methoxyamine	*myo-Inositol
*Hydroxylamine	

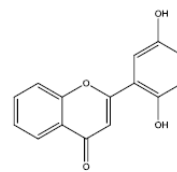
Pure substances used for direct infusion



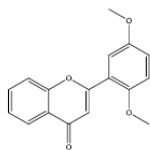
(+)-Delta-Tocopherol



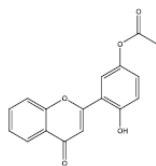
2,2'-diOH-Chalkon



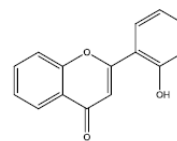
2',5'-diOH-Flavon



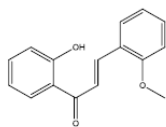
2',5'-diOMe-Flavon



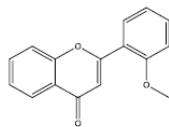
2'-OH-5'-OAc-Flavon



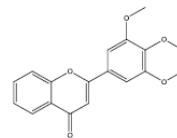
2'-OH-Flavon



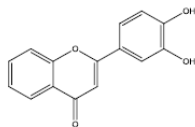
2-OMe-2'-OH-Chalkon



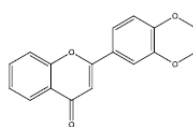
2'-OMe-Flavon



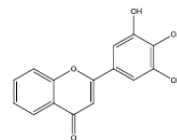
3',4',5'-OMe-Flavon



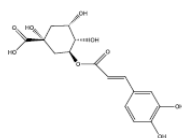
3',4'-diOH-Flavon



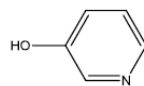
3',4'-diOMe-Flavon



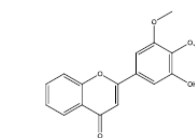
3',5'-diOH-4'-OMe-Flavon



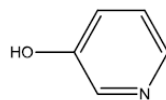
3-Caffeoyl_quinic_acid



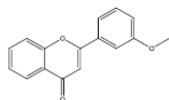
3-Hydroxypyridine



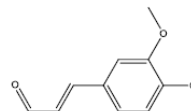
3'-OH-4',5'-diOMe-Flavon



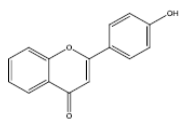
3-OH-pyridine



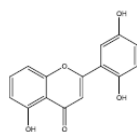
3'-OMe-Flavon



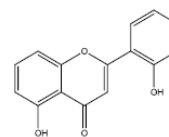
4-hydroxy-3-methoxy-Cinnamaldehyd



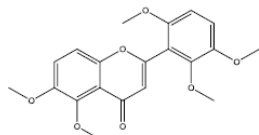
4'-OH-Flavon



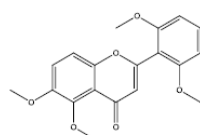
5,2',5'-triOH-Flavon



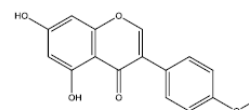
5,2'-diOH-Flavon



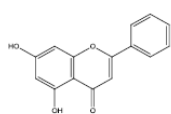
5,6,2',3',6'-OMe-Flavon



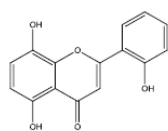
5,6,2',6'-OMe-Flavon



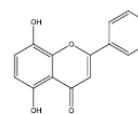
5,7-Dihydroxy-4-Methoxyisoflavone



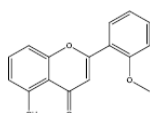
5,7-Dihydroxyflavone



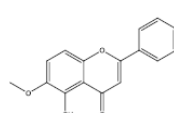
5,8,2'-triOH-Flavon



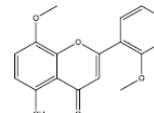
5,8-diOH-Flavon (Primetin)



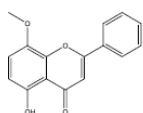
5-OH-2'-OMe-Flavon



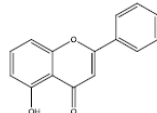
5-OH-6-OMe-Flavon



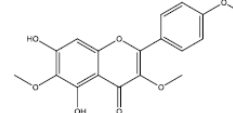
5-OH-8,2'-OMe-Flavon



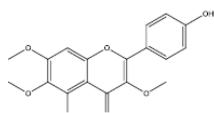
5-OH-8-OMe-Flavon



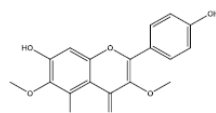
5-OH-Flavon



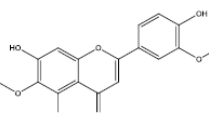
6-OH-Kaempferol-3,6,4'-OMe



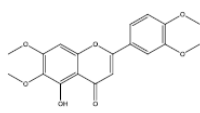
6-OH-Kaempferol-3,6,7-OMe



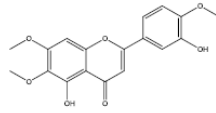
6-OH-Kaempferol-3,6-OMe



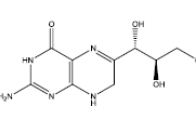
6-OH-Luteolin-6,3'-OMe



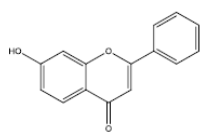
6-OH-Luteolin-6,7,3',4'-OMe



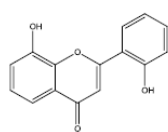
6-OH-Luteolin-6,7,4'-OMe



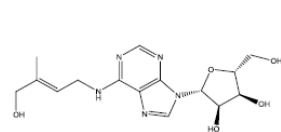
7,8-Dihydroneopterin



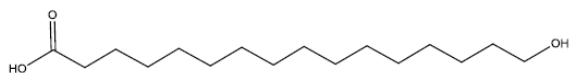
7-OH-Flavon



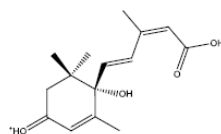
8,2'-diOH-Flavon



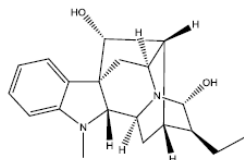
9-beta-ribofuranosyl-trans-Zeatin



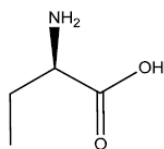
16-hydroxyhexadecanoic acid



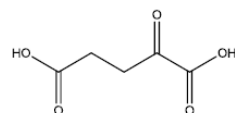
Absciscic acid



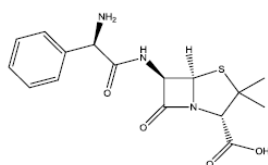
Ajmaline



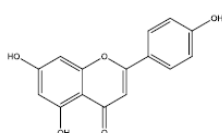
alpha-aminobutyric acid



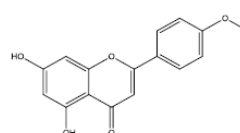
alpha-ketoglutaric acid



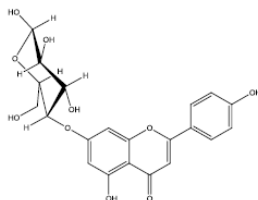
Ampicillin



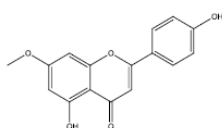
Apigenin



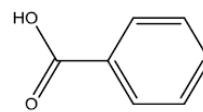
Apigenin-4'-OMe



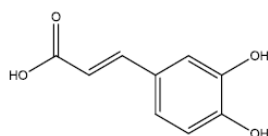
Apigenin-7-glucosid



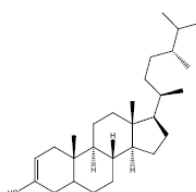
Apigenin-7-OMe



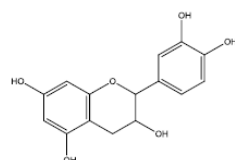
Benzoic acid



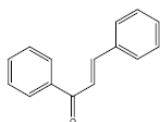
Caffeic acid



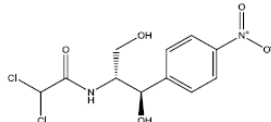
Campesterol



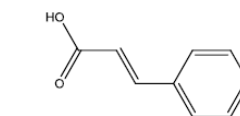
Catechin



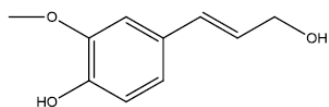
Chalkon



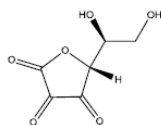
Chloramphenicol



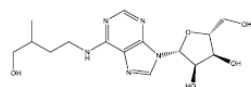
Cinnamic acid



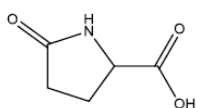
Coniferyl alcohol



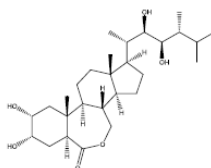
Dehydroascorbic acid



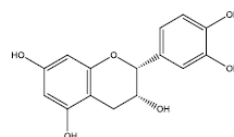
Dihydrozeatine riboside



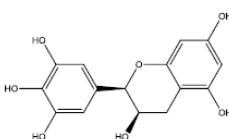
DL-Pyrroglutamic acid



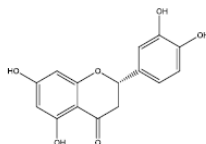
Epibrassinolide



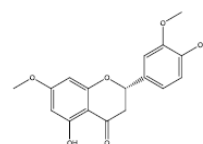
Epicatechin



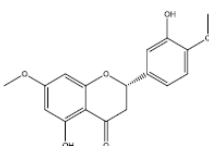
Epigallocatechin



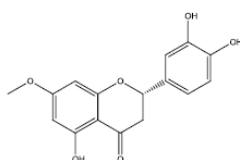
Eriodictyol



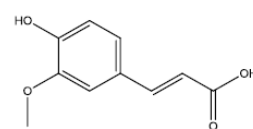
Eriodictyol-7,3'-OMe



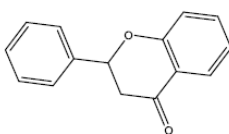
Eriodictyol-7,4'-OMe



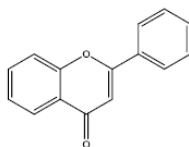
Eriodictyol-7-OMe



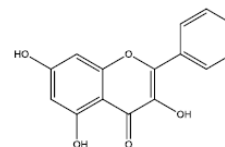
Ferulic acid



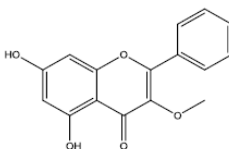
Flavanone



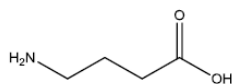
Flavon



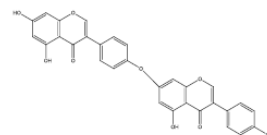
Galangin



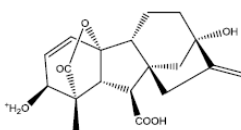
Galangin-3-OMe



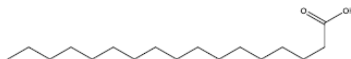
Gamma-aminobutyric acid



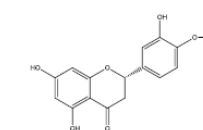
Genistein



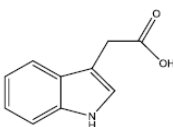
Gibberelic acid



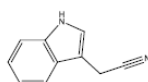
Heptadecanoic acid



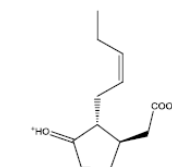
Hesperetin



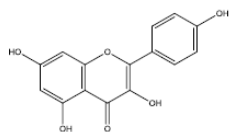
Indole-3-acetic acid



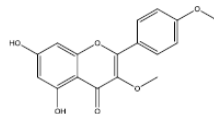
Indole-3-acetonitrile



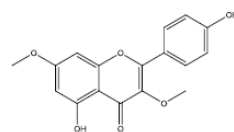
Jasmonic acid



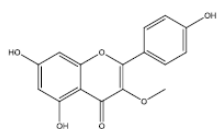
Kaempferol



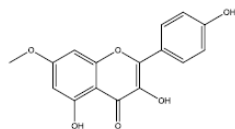
Kaempferol-3,4'-diOMe



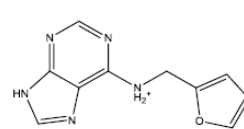
Kaempferol-3,7-diOMe



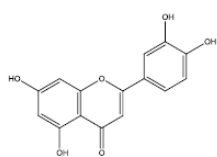
Kaempferol-3-OMe



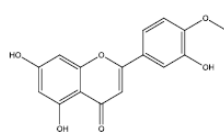
Kaempferol-7-OMe



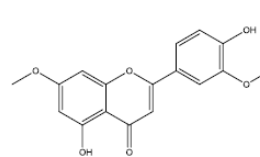
Kinetin



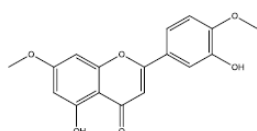
Luteolin



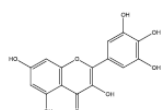
Luteolin-4'-OMe



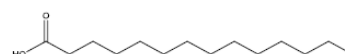
Luteolin-7,3'-OMe



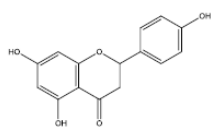
Luteolin-7,4'-OMe



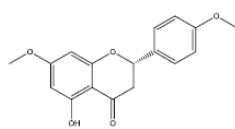
Myricetin



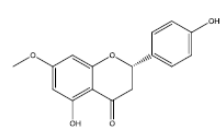
Myristic acid



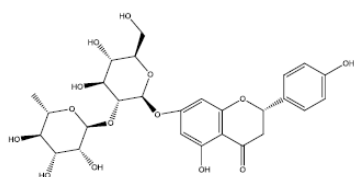
Naringenin



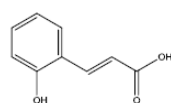
Naringenin-7,4'-OMe



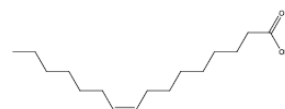
Naringenin-7-OMe



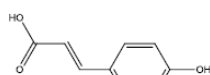
Naringin



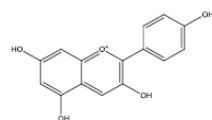
o-Coumaric acid



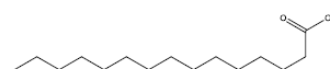
Palmitoleic acid



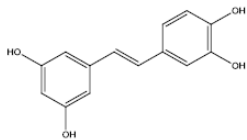
P-coumaric acid



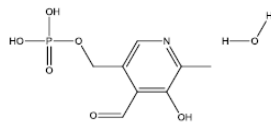
Pelargonidin



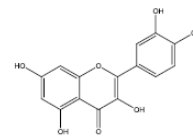
Pentadecanoic acid



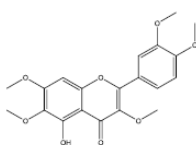
Piceatannol



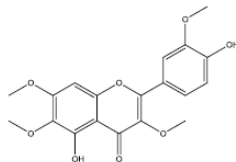
Pyridoxal-5'-phosphate-monohydrate



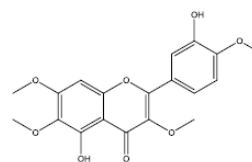
Quercetagenin



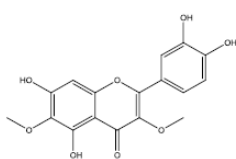
Quercetagenin-3,6,7,3'-OMe



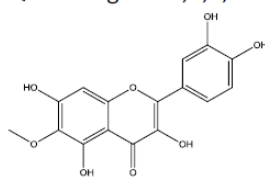
Quercetagenin-3,6,7,3'-OMe



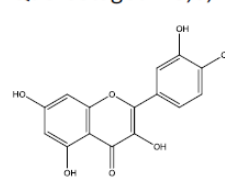
Quercetagenin-3,6,7,4'-OMe



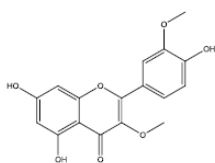
Quercetagenin-3,6-OMe



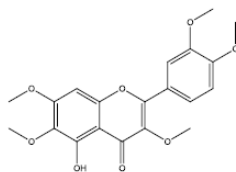
Quercetagenin-6-OMe



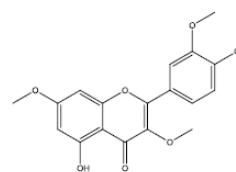
Quercetin



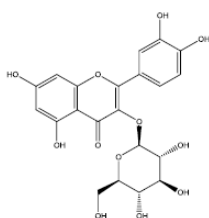
Quercetin-3,3'-OMe



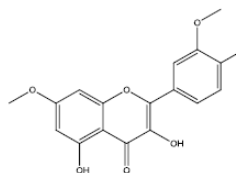
Quercetin-3,6,7,3'-OMe



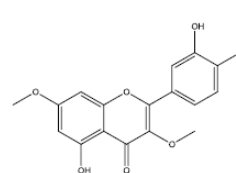
Quercetin-3,7,3'-OMe



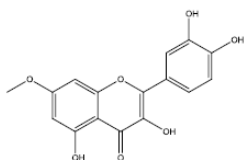
Quercetin-3-beta-D-glucoside



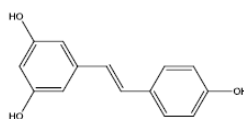
Quercetin-7,3',4'-triOMe



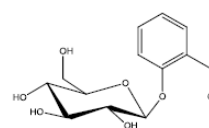
Quercetin-7,3'-diOMe



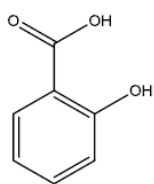
Quercetin-7-OMe



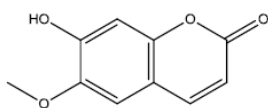
Resveratrol



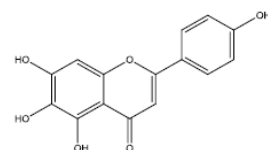
Salicin



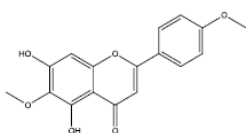
Salicylic acid



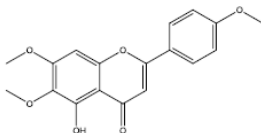
Scopoletin



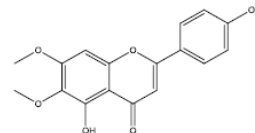
Scutellarein



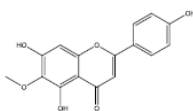
Scutellarein-6,4'-OMe



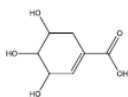
Scutellarein-6,7,4'-OMe



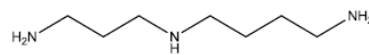
Scutellarein-6,7-OMe



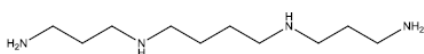
Scutellarein-6-OMe



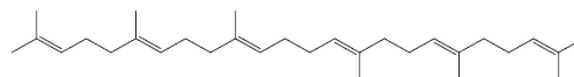
Shikimic acid



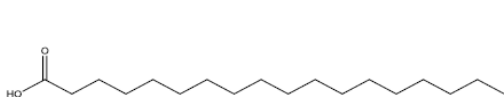
Spermidine



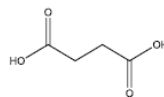
Spermine



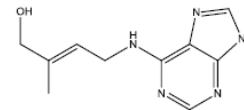
Squalene



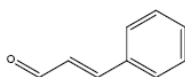
Stearic acid



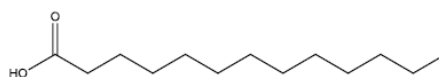
Succinic acid



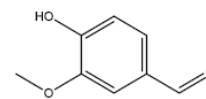
Trans-Zeatin



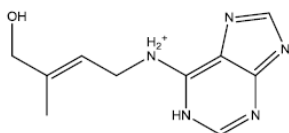
Trans-Cinnamaldehyde



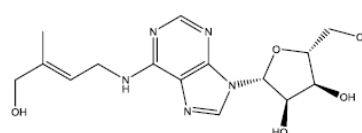
Tridecanoic acid



Vanillin



Zeatin



Zeatin riboside

Hannes Dörfler

Personal details

Date of birth: December 4th 1983

Nationality: Austria

Gender: Male

Education

June 2002	General qualification for university entrance (Matura) at Stiftsgymnasium Melk
2003	Civilian service
Since 2004	Studies of Chemistry and Biology at University of Vienna; projected conclusion in September 2011 with the thesis "Comparative studies in A. thaliana via GC-MS and LC-MS"
2010 - 2011	Coworker at the Department of Molecular Systems Biology under Wolfram Weckwerth; tutor function for "Metabolomics" practical course
2011	Vienna Science day talk: "Establishment of a combined GC/MS and LC/MS platform for metabolomics"