Metabolomics, Proteomics, and Transcriptomics of *Cannabis sativa* L. Trichomes

Zur Erlangung des akademischen Grades eines

Dr. rer. nat.

von der Fakultät Bio- und Chemieingenieurwesen

der Technischen Universität Dortmund

Dissertation

vorgelegt von

Nizar Happyana, M.Sc.

aus

Bandung, Indonesien

Tag der mündlichen Prüfung: 18.12.2014

1. Gutachter: Prof. Dr. Oliver Kayser

2. Gutachter: Prof. Dr. Robert Verpoorte

Dortmund 2014

Acknowledgements

Alhamdulillahirabilallamin.

Foremost, I would like to express my deepest gratitude to my supervisor, Prof. Oliver Kayser. Dear Oliver, I am grateful to have an opportunity to conduct PhD study in your laboratory. It is a pleasure to work with you. Thank you very much for your guidance, patience, and all supports you have given to me during my PhD. I was very happy when you drove me to Veendam for collecting *Cannabis*. I also really enjoyed scientific and non-scientific discussion with you.

I gratefully acknowledge the Directorate General of Higher Education, Ministry of National Education, Indonesia, for providing me the doctoral scholarship. I am deeply indebted to Tjalling Erkelens and Freerk Bruining from Bedrocan BV for supplying the plant material. Thank you very much for your hospitality when I came to Veendam. It was nice to have discussion with you both.

I gratefully acknowledge to Prof. Robert Verpoorte from Leiden University, Prof. Sebastian Engell and Prof. Gabriele Sadowski from TU Dortmund for being my PhD Committee. I am utterly grateful to Prof. Jörn Kalinowski and Oliver Rupp from CEBITEC for helping us in analyzing the cDNA library. I would like to thank Prof. Bernd Schneider and Sara Agnolet from MPI (Chemical Ecology) for cryogenic NMR measurements. Many thanks to Annie Van Dam from Groningen University for the LCMS measurements. I would like to thank Remco Muntendam for fruitful discussions and sharing your research experience with me. I am deeply indebted to Fabian Real for driving me 4 times to Veendam for collecting the plant materials. I would like to thank Prof. Albert Sickmann and Stefan Loroch from ISAS for the collaboration in the proteomic project.

I would like to thank Dr. Armin Quentmeier and Dr. Felix Stehle for all suggestions and comments regarding my research. Many thanks to Kristine Blase for the administrative supports. My special thanks to Jörg, Bettina, and Fabian for their helps during my laboratory works. I would like to thank all the current and former members of the Technical Biochemistry Department. Super thanks to Sayed, Marcello, Beer, Anna, and Parijat. Thank you so much for our friendship and for sure I am going to miss you all. Special thanks to Kathleen, Magda, Friederike, Torsten, Rike, Eva, Bastian, Joanna, Katharina, Alena, and the other students for the good atmosphere in our group.

Most of all, I would like to thank my beloved family for their unconditional love, pray, and supports throughout my entire life. I have dedicated this thesis to you all. *Kanggo Mama, Bapa, A Janjan, Teh Eros, Teh Eneng, A Yanto, Teh Euis, A Heri, Fajar, Risma, Agnes, Oki, Saskia, Marsya, Natya, Keisha, Milan sareng kulawargi sanesna, hatur nuhun pisan kana bantosanna, pidu'ana, dukunganna, sareng nu sanesna*. *Untuk Mama, Papa, dan Cesar, terima kasih banyak atas do'anya, dukungan moralnya, dan bantuan yang lainnya*. Very super special thanks to my dearest wife Prita and my little angel, Tania. Simply, you both make me be a better man.

Table of Contents

Abstract

Cannabis sativa L. trichomes are the main site for synthesizing and storing cannabinoids and other secondary metabolites on this plant. Metabolomics, proteomics, and transcriptomics were applied in this research as analytical approaches in order to study secondary metabolites, proteins, and genes that are synthesized in the trichomes. The function of specific trichomes and their individual parts in the cannabinoid biosynthesis was investigated as well. ^{1}H NMR-based metabolomics has been successfully applied for monitoring the production of metabolites, especially cannabinoids in the *Cannabis* trichomes during the last weeks of flowering period. Proteomics analysis of the *Cannabis* trichomes revealed that many enzymes corresponding in the biosynthesis of secondary metabolites, including cannabinoids, flavonoids, and terpenoids were successfully recorded. This finding supported the function of *Cannabis* trichomes as the main site of secondary metabolite production. Although there is no flavonoid reported from the trichomes, however the identification of enzymes related to its biosynthesis indicated that this compound might be present in this organ. Interestingly, identification of enzymes involved in the biosynthesis of cannabinoids, terpenoids, and flavonoids in the proteomic work has been also confirmed by the detection of their putative transcripts in the cDNA library of *Cannabis* trichomes. Analysis of cannabinoids in the laser-microdissected trichomes of *Cannabis* showed that these compounds were detected not only in the head of capitate-stalked trichomes but also in its stem part. This finding suggest that cannabinoid biosynthesis is not only limited to the expected head cells, but also the stems of *Cannabis* capitate-stalked trichomes might play a role in the cannabinoids biosynthesis.

Zusammenfassung

Cannabis sativa L. Trichome sind der hauptsächliche Ort der Cannabinoid-Biosynthese. Im Rahmen dieser Arbeiten wurden Techniken angewandt und verbessert, um das Metabolom, Transkriptom und Proteom der biosynthetisierenden Zellen der sezernierenden Trichome zu erforschen. Die Biosyntheleistung der drei Trichomtypen wurde mit Hilfe von LC-MS, Laser Dissection Mikroskopie und ^{1}H-NMR Metabolomics untersucht, um die Biosynthese über 8 Wochen qualitativ und quantitativ zu erfassen. Weitere Analysen des Proteom zeigten auf, das Genexpression und Funktion biosynthetischer Proteine für Cannabinoide, Flavonoide und Terpene des ätherischen Öls ist stark abhängig von Alter und Blütenbildung ist. Analyse und Annotierung der Gene ist durch die eigens erstellte cDNA Bank ermöglicht worden. Die Ergebnisse zeigen, das eine Biosynthese in den sezernierenden Kopfzellen sehr stark aber auch in den Stielzellen zu finden ist. Auf Grund der Proteomuntersuchungen im ätherischen Öl ist eine Biotransformation im nicht wässrigen Milieu ausgeschlossen worden. Mit Hilfe von ^{1}H-NMR Metabolomics und statistischer PCA-Analyse konnten Strategien entwickelt warden, um Zuchtlinien von *Cannabis sativa* zu identifizieren und auf Basis des metabolomischen Profil voneinander zu unterscheiden

Chapter 1

Introduction and scope of the thesis

Nizar Happyana

Since ancient time, human beings could not be separated from medicinal plants to fulfill their basic health needs. One of medicinal plants that have long been used for medication is *Cannabis sativa* L. In fact, cultivation method of this plant was recorded in the ancient Chinese literature (H.-L. Li 1974). Although *Cannabis* is classified as a dangerous narcotic in many countries since has a long history as a recreational drug, however this plant possesses the potential to treat a wide range of diseases, such as multiple migraine, asthma, and cancer. Furthermore, Δ^9-tetrahydrocannabinol (THC), the main and most interesting compound of the *Cannabis* has been recognized as a medicine with trade name Marinol®. The Food and Drug Administration of United States (FDA) has approved Marinol® for the treatments of anorexia in AIDS patients, and nausea and vomiting due to cancer chemotherapy. In addition, a mouth spray containing THC and cannabidiol (CBD) under trademark Sativex® is available in the market for the treatment of multiple sclerosis.

More than 450 constituents have been reported in *Cannabis,* including cannabinoids, terpenoids, and flavonoids (ElSohly and Slade 2005), and leads it as one of chemically best studied plants. Cannabinoids are the most studied compounds in this plant as a consequence of their biological activities. So far, at least 100 cannabinoids have been identified in the plant with THC, CBD, CBG, CBC and their acidic forms as the main constituents. In *Cannabis,* 120 terpenoids have been detected, including 61 monoterpenes, 52 sesquiterpenes, 2 triterpenes, one diterpene, and 4 other terpenoid derivatives (ElSohly and Slade 2005). Meanwhile more than 20 flavonoids have been identified from this plant (ElSohly and Slade 2005) representing 7 chemical structures which can be glycosylated, prenylated or methylated (Flores-Sanchez and Verpoorte 2008).

Cannabinoids and other secondary metabolites are mainly synthesized and stored in the trichomes, small epidermal protrusions found almost on the whole surfaces of the plant (Lanyon et al. 1981; Malingre et al. 1975; Petri et al. 1988; Turner et al. 1978). Three types of trichomes are identified in this plant, namely capitate-stalked, capitate-sessile and bulbous trichomes. Additionally, capitate-stalked trichomes contain most cannabinoids compared to other types. *Cannabis* trichomes can be easily isolated since it protrudes from the epidermis of the plant. Thus, proteins, mRNA, and small molecules of the trichomes are accessible to be analyzed. This advantage makes *Cannabis* trichomes become excellent experimental systems for investigating production of secondary metabolites, especially cannabinoids, and studying the biosynthetic pathways that are responsible for synthesize the valuable compounds. Therefore, this thesis focused on the *Cannabis* trichomes. There are 4 main aims of this thesis, namely:

- Analyze the production of metabolites, especially cannabinoids in the *Cannabis* trichomes over flowering period using metabolomics approach.
- Study protein profile and enzymes that are involved in the biosynthesis of secondary metabolites in the *Cannabis* trichomes using proteomics analysis.
- Elucidate biosynthetic pathways of secondary metabolites, especially cannabinoids, terpenoids, and flavonoids in the *Cannabis* trichomes using transcriptomics approach.
- Investigate the function of specific trichomes and their individual parts in the cannabinoid biosynthesis.

The following section of this thesis (**chapter 2**) provides an extensive overview of metabolomics, an analysis approach used in our works. This chapter highlights bio-analytical tools often used in metabolomics of medicinal plant researches, such as NMR and LCMS, with special focus for quality control and metabolic profiling for herbal medicinal products. Moreover we describe the important applications of metabolomics for medicinal phytochemistry by explaining their applications in medicinal plants.

In order to reveal and distinguish metabolite profiles, and study cannabinoids biosynthesis in the leaves, flowers, and trichomes of *Cannabis*, ^{1}H NMR-based metabolomics and real time PCR (RT-PCR) were applied (**chapter 3**). Two different varieties of this plant were analyzed, namely Bedrocan and Bedica which categorized as medicinal-type *Cannabis*. Expression level of THCAS, OLS, and OAC were analyzed in the different organs of the plants with RT-PCR. Beside published cannabinoid genes, we also investigated an unpublished gene from our group that is predicted as CBGA synthase (predicted-CBGAS).

We continued our investigation using ^{1}H NMR-based metabolomics for monitoring metabolite production, especially cannabinoids in the trichomes (**chapter 4**). The investigation was applied over the last weeks of flowering period since based on our previous study that cannabinoids are produced in a high amount in flower *Cannabis* during these weeks. The trichomes from 4 medicinal *Cannabis* varieties: Bedrocan, Bedrobinol, Bediol, and Bedica, were harvested and analyzed over the last 4 weeks of flowering period.

Although many compounds have been reported from *Cannabis* trichomes, however the information of proteome in this tissue is still limited. Because of that reason, a proteomic analysis was carried on trichomes of medicinal *Cannabis* variety Bediol using nano-LC-MS/MS analysis (**chapter 5**). In this work, protein profile of the trichomes was investigated and classified based on their biological

function. Not only house-keeping proteins but also enzymes participating in the biosynthetic pathways of secondary metabolites were described in this section.

To follow up the findings of the proteomic work (**chapter 5**), a transcriptomic analysis of *Cannabis* trichomes was performed (**chapter 6**). A cDNA library of *Cannabis* trichomes variety Bediol was constructed using 454 GS FLX pyrosequencing system. Afterward, the assembled sequences were analyzed using SAMS for gene ontology (GO) annotation and KEGG pathway assignments. Furthermore, RT-PCR was applied to investigate expression level of some genes involved in biosynthesis of secondary metabolites in different organs of this plant.

Analysis of cannabinoids in the laser-microdissected trichomes, particularly in the intact capitate-sessile and capitate-stalked trichomes as well as in the heads and stems of the latter is explained in **chapter 7**. This section is started with determining the distribution and density of various types of *Cannabis* trichomes in the different organs by scanning electron microscopy (SEM). The specific cells of trichomes were separated and collected using laser microdissection (LMD). Afterward, cannabinoids in the dissected samples were analyzed using LCMS and cryogenic NMR.

Finally, in the **chapter 8**, we highlight the findings and integrate the results of our studies explained in the thesis.

Chapter 2

Metabolomics as Bioanalytical Tool for Characterization of Medicinal Plants and their Phytomedical Preparations

Nizar Happyana, Remco Muntendam, Oliver Kayser

Remco Muntendam and Oliver Kayser coordinated and supervised the project and corrected the manuscript

Published as a book chapter in *Pharmaceutical Biotechnology: Drug Discovery and Clinical Applications, 2nd ed. (eds. Kayser, O., Warzecha, H.)*, Wiley-Blackwell, 2012, pp. 527-552.

Abstract

Many centuries humanity is dependent on drugs isolated from medicinal plants for their health care. These plants contain secondary metabolites useful in therapy against various diseases. A Growing human population, emerging of new illnesses, and an increase of resistance to current drugs, have emerged challenges for medicinal plant sciences. Metabolomics is a breakthrough approach to accelerate and streamline the analytical process of medicinal plant researches. Metabolomics allows quick and efficient identification and quantification of the secondary metabolites within plants and is easily coupled to high throughput bioactivity screening. Here we review bio-analytical tools often used in metabolomics of medicinal plant researches, including NMR, MS, GCMS, and LCMS. Besides that, a selection of multivariate data analysis for processing huge data obtained from metabolomics experiments is summarized. Furthermore, the applications of metabolomics in medicinal plants are described as well.

Keywords: Metabolomics, medicinal plant, NMR, LCMS, quality control

1. Introduction

For many centuries, humankind used medicinal plants directly or extracts thereof for their basic healthcare needs. Medicinal plants contain a complex mixture of secondary natural products and show synergistic effects against a broad variety of diseases. According to the World Health Organization (WHO), more than 80 per cent of the world population uses medicinal plant for everyday healthcare (Edwards 2004). Now there are some 50,000 different medicinal plant species used for medication and in Europe alone are around 1,300 medicinal plants commercially used (Vines 2004). In the United Kingdom more than 25 per cent of the population uses medicinal plant regularly (Vines 2004), and in Germany approx. 30% of all over the counter drugs (OTC) drugs are herbal medicinal products (HMPs).

Commercial medicinal plants in the world are mostly harvested from the wild. Not only in so called low income countries that have to face this situation, but also in Europe 90 % of medicinal plants used for extraction are collected from the wild (Vines 2004). The use of HMPs is getting more and more popular and the demographic development in Europe, USA and later in Asia (DESA 2009; United-Nation 2009) will increase significantly the demand in the future. From these backgrounds, plant collection from the wild should be replaced by controlled cultivation to ensure sustainability of HMPs. Moreover, emerging of new illnesses, and an increase of resistance to current drugs, have emerged challenges for medicinal plant sciences as well.

HMPs are made of complex biological matrices. To ensure patient safety and a high level of quality with regard to composition and activity, sophisticated analytical methods have to be developed and applied. Production and quality management is regulated by Good Manufacturing Practice (GMP) regulations and dominantly GC, HPLC, NMR, GCMS and LCMS have been introduced.

Actual developments in analytical chemistry, bioinformatics and computers have speeded up the procedures and a new strategy has been developed to get a huge pile of information from some simple testing, namely metabolomics. It is a breakthrough approach to accelerate and streamline the analytical process of medicinal plant researches. Metabolomics allows quick and efficient identification and quantification of the secondary metabolites within plants and is easily coupled to high throughput bioactivity screening. Moreover metabolomics is well known as the youngest “omics” method. A summary of some definitions related to metabolomics is shown in table 1.

In this chapter we review bioanalytical tools often used in metabolomics of medicinal plant researches, such as NMR, DIMS, LCMS, and GCMS with special focus for quality control and metabolic profiling for herbal medicinal products. We highlight how bioanalytical tools are used and maybe applied in industrial routine work. Furthermore we summarize the important applications of metabolomics for medicinal phytochemistry by explaining their applications in medicinal plants.

Table 1. Some definitions related to metabolomics

Metabolomics	• Holistic, simultaneous and systematic qualitative and quantitative determination of metabolites over time after stimulus
Metabolom	• Dynamic situation • Total number of all low molecular metabolites
Metabolite	• Intermediates and products of metabolism • Primary and secondary natural metabolites (e.g. lipids, sugars, alkaloids, flavonoids) • Low molecular compounds (mw < 1000)
Metabolic profiling	• Analyzes a selected group of compounds or set of metabolites in a specific biochemical pathway • Targeted metabolites analysis
Metabolic fingerprinting	• Global screening approach • Classify samples based on metabolite patterns or "fingerprints" • Detect discriminating metabolites without identifying all of the compounds present

2. Bioanalytical tools

The main purpose of metabolomics is to analyze all metabolites both qualitatively and quantitatively in medicinal plant samples. Basically metabolomics analysis consists of three steps, namely sample preparation including extraction process; metabolite measurements using bioanalytical tools, such as NMR, LCMS, GCMS; and mining the raw complex data using chemometric software. An illustration of the steps required in a metabolomic analysis of medicinal plants is explained in figure 1.

2.1 Sample preparation

Sample preparation is considered as one of the most important steps in metabolomics analysis. The procedure must be standardized; starting from harvesting, extraction, storage and applying validated analytical parameters. Standardization and validation procedures are essential to receive reproducible results over time.

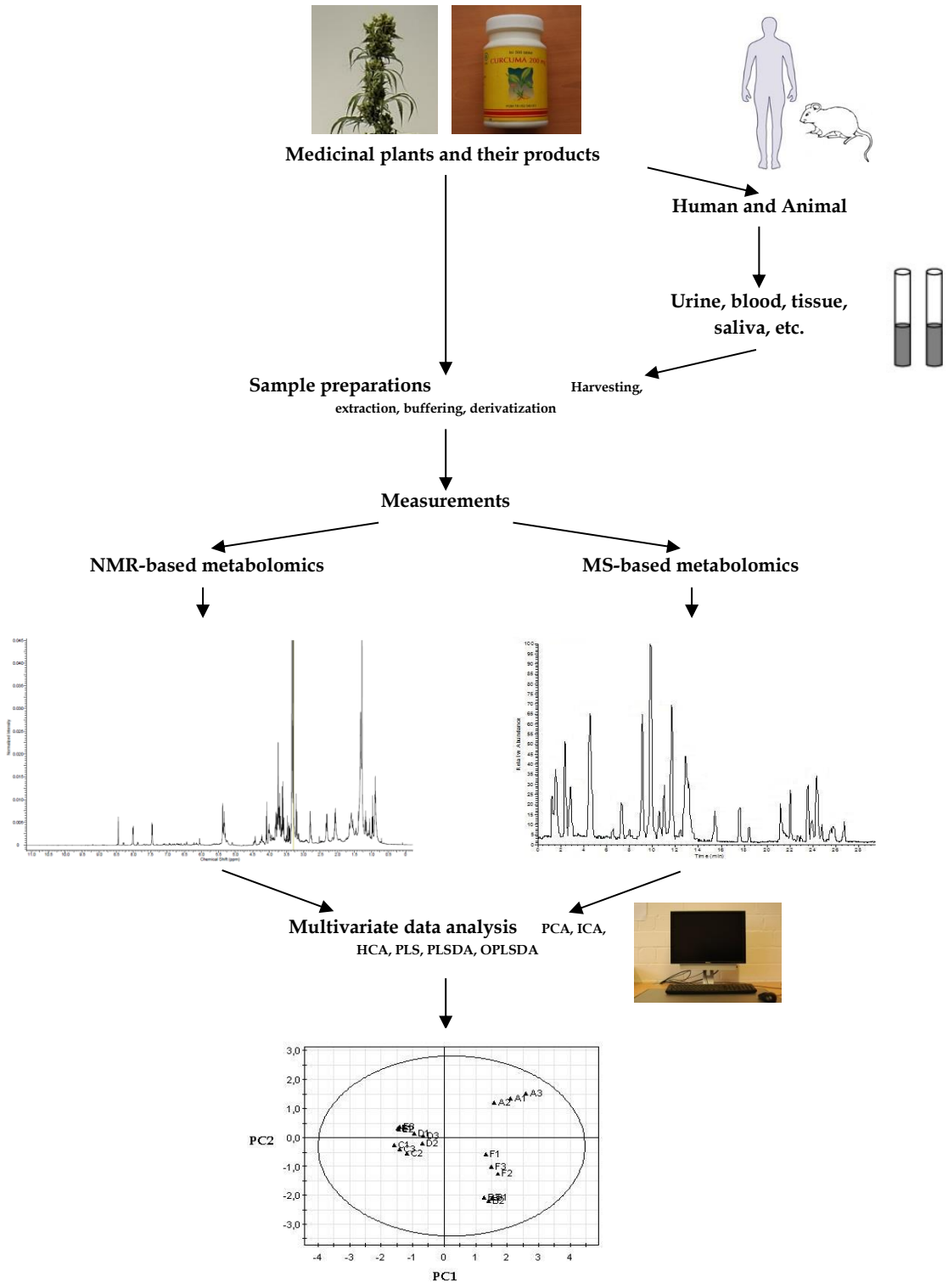


Figure 1. Workflow of metabolomics analysis in medicinal plant researches

In harvesting, it is critical that metabolism process in plants should be stopped Subsequently. Damaging by cutting plants could release active enzymes (e.g. glycosidases) that can change the metabolite profile significantly. Typical degradation reactions after harvesting are oxidation, hydrolysis, and decarboxylation. To prevent those, harvesting should be done

rapidly and the harvested material should be dried to stop further metabolism. Alternatively lab samples can be frozen Subsequently (-20˚C or -80 ˚C) or submerged in a liquid nitrogen tank and followed by mechanical disruption to release metabolites from the cells (Y. H. Choi et al. 2008). However, for commercial preparation this technique is difficult to be implemented.

First step of extraction of medicinal plants can be classified as a solid-liquid extraction. It means metabolites from a plant sample as the solid phase to transfer into the liquid extraction medium. Generally used solvents for extraction can be divided into four classes: polar solvents (e.g., methanol, ethanol, and water), medium-polar solvents (e.g., chloroform, dichloromethane, diethyl ether), non-polar solvents (e.g., n-hexane), and combined solvents thereof (Y. H. Choi et al. 2008). Each solvent will give different profiles of metabolites extracted. Metabolites that have high polarity mostly are extracted in polar solvents, while those that have low polarity mostly are extracted in medium-polar solvents. We have to consider that for commercial HMPs ethanol with and without water is the standard solvent for extraction. It seems that by one unique solvent metabolomic analysis is getting more simple, but different water:ethanol ratios and different extraction processes varying in time and temperature give a high diversity of extracts which are reflected by a high number of different HMPs on the market.

The goal of metabolomics is to identify and quantify metabolites in the biological samples consisting of a complex spectrum of different natural product classes. Currently there is no single solvent that can be used to extract all compounds directly. However, solvents which can dissolve most diverse group of compounds must be chosen, but due to discussed limitations several extractions with different solvents can be conducted to have a total view of the metabolites (H. K. Kim and Verpoorte 2010). In the case of metabolic profiling studies with the purpose to quantify a selected group of metabolites in medicinal plants, selection of solvents should be based on the physico-chemical properties of target metabolites to be analyzed.

For NMR-based metabolomics, polar solvents such as methanol and combined solvents such as water:methanol are often used for the extractions. A two-phase solvent system, composed of a mixture of chloroform, methanol and water (2:1:1, v/v), also has been used successfully for the extractions in NMR-based metabolomics (H. K. Choi et al. 2004; Y. H. Choi et al. 2004a; Y. H. Choi et al. 2004b). Moreover, the use of deuterated NMR solvents for the

extraction also has been reported (Y. H. Choi et al. 2006; Hendrawati et al. 2006; Yang et al. 2006). This method avoids the need to evaporate the original extraction solvent and to redissolve the sample in the NMR solvents (H. K. Kim and Verpoorte 2010). In LCMS-based metabolomics samples must be dissolved in solvents preferable similar to the eluent of the HPLC system. Regarding GCMS-based metabolomics, compounds must be volatile compounds to be measured, thus for non-volatile compounds derivatisation before measurements is needed.

2.2 NMR

NMR spectroscopy is based on magnetic nuclei resonance in a strong magnetic field to determine physical and chemical properties of molecules. NMR spectroscopy basically consists of a magnet, radio-frequency (rf) transmitter or oscillator, and a suitable rf detector (J. D. Roberts 1959). If an organic compound or an extract is placed in a magnetic field, interactions between NMR active nuclei, such as ^{1}H and ^{13}C, and electromagnetic radiation will produce resonance signals to be collected by the detector. Resonance frequency of NMR active nuclei is dependent on the chemical environment. Different chemical environment will give different resonances, thus each compound will possess specific NMR spectra.

NMR is widely used as bioanalytical tools for the analysis of organic molecules and considered as one of the most promising metabolomic tool (Dunn et al. 2005). It is well known as a powerful technique for elucidation of compound structures, including stereochemistry in details. NMR is non-destructive and can be used for structural analysis of metabolites in crude extracts, cell suspensions, intact tissues or whole plants (Fan and Lane 2008; Ratcliffe et al. 2001). Moreover NMR allows the exploration of metabolic pathways, leading to qualitative information on the link between labelled precursors and their products and quantitative information on metabolic fluxes (Bacher et al. 1998; Krishnan et al. 2005; Kruger et al. 2003; J. K. M. Roberts 2000).

For NMR the ^{1}H NMR is the most popular technique used for qualitative and quantitative metabolomics analysis. It is very fast in its measurement, typically less than 5 minutes for one measurement (depending on the concentrations and the resonance frequency), facilitates high-throughput analysis and mostly has a simple sample preparation. Moreover, quantitative analysis using ^{1}H NMR has no need for a calibration curve because the molar concentration of a compound is directly represented by the intensity of a proton signal (Y. H. Choi et al. 2008).

By adding an internal standard to the sample, we can compare proton signals of the internal standard with those of the sample and thus we can quantify compounds in the sample. For ^{1}H NMR, the concentration threshold for a routine detection of a metabolite in an extract using a modern high field spectrometer is probably 10 μM, corresponding to a quantity of 5 nmol in the typical sample volume of 500 μl (Krishnan et al. 2005). Furthermore, metabolomics based on ^{1}H NMR approaches is highly reproducible, it means that NMR-metabolomics data are valid for ever, as long as the same extraction procedures and the same NMR-solvents are used (Verpoorte et al. 2008).

Although the sensitivity of ^{1}H NMR is low, this weakness can be solved by various approaches. Today, NMR microprobes are available, which can be used for measuring low quantity samples. With this probe, the low concentrated sample is diluted with a small volume of NMR solvent, thus it will make the sample more concentrated and give spectra with improved quality. High-resolution NMR is also available. It has high sensitivity and can give a substantial improvement in the detection of the signals, thus it produces high quality NMR spectra as well. The last approach is cryogenic NMR probes. This probe is small (3 or 5 mm in diameter) but has a capability for improving sensitivity and reducing noises with cooling the receiver coil and preamplifiers to cryogenic temperatures. It is a powerful probe that can be used for measuring low samples as has been demonstrated by Schneider et al. (Schneider and Holscher 2007).

Another problem of using ^{1}H NMR in metabolomics studies for medicinal plants is signal overlap that can obstruct the identification and quantification of metabolites. Nevertheless 2D NMR techniques can be used to solve this problem. These techniques give better signal resolution and reduce signal overlap by distribution of the resonances in a second dimension. Moreover 2D NMR has all advantages of ^{1}H NMR but it consumes longer recording time. ^{1}H J-resolved NMR (JRES) is one of 2D NMR techniques that mostly used in metabolomics studies. JRES has capability to split the effects of chemical shift and J-coupling into two independent dimensions. The use of JRES in metabolomics is reviewed in detail by Ludwig et al. (Ludwig and Viant 2010). Besides JRES, other 2D NMR techniques such as HSQC (heteronuclear single quantum coherence), COSY (correlation spectroscopy), TOCSY (total correlation spectroscopy), HMBC (heteronuclear multiple quantum coherence) and NOESY (nuclear overhauser effect spectroscopy) have been applied in plant metabolomics studies (Agurell and Nilsson 1972; Ratcliffe et al. 2001; Ratcliffe and Shachar-Hill 2005).

2.3 Mass Spectrometry (MS)

MS is an analytical instrument measuring the mass-to-charge (*m/z*) ratio of ions. MS instruments consist of three main parts, namely an ionization chamber where the molecules are being ionized, a mass analyzer which separates ions according to their *m/z* by applying electromagnetic fields, and a detector to record *m/z*. The common techniques used as ionization source are electron ionization (EI), chemical ionization (CI), electrospray ionization (ESI) and atmospheric pressure chemical ionization (APCI). Meanwhile, mass analyzers such as single quadrupole, triple quadrupole, ion trap and time-of-flight (TOF) analyzer are usually applied in metabolomics analysis.

Besides NMR, MS is well known as an analytical tool for metabolomic analysis, particularly in metabolic profiling and metabolic fingerprinting. In metabolomics, MS separates metabolites based on *m/z* ratios of their ions. Furthermore MS also provides both sensitive detection and metabolite identification through mass spectrum interpretation and comparison or molecular formula determination *via* accurate mass measurements (Dunn et al. 2005). Moreover MS is considered as the most sensitive method for metabolomics analysis (Y. H. Choi et al. 2008), because it can identify ionized components at very low quantities. A major disadvantage of using MS is its reproducibility. MS measurements are dependent on the type of MS, operating parameters and matrix effect on ionization, making it difficult to produce similar results across laboratories (Y. H. Choi et al. 2008).

2.3.1 Direct injection mass spectrometry (DIMS)

A DIM analyses metabolites with injection of a sample directly into the ionization chamber without prior chromatographic separation and provides fast and high throughput measurements. DIMS mostly use ESI and APCI as ionization sources. Both are known as soft ionizations which provide minimal fragmentation of molecular ions and a less complex mass spectrum. Therefore, with interpreting molecular ions, metabolites can be identified without chromatographic separations (Schroder 1996).

TOF is usually used as a mass analyzer in DIMS method. TOF separates ions based on the time of ions to reach the detector. All ions in TOF are accelerated by an electric field that makes ions have same kinetic energy. Thus the velocity of ions just depends on their *m/z* and heavier ions will reach the detector later compared to lighter ions. Depending on the flight-tube geometry and instrument tuning, TOF-MS instruments provide mass of 6,000-17,000

with mass accuracy in the range of 3-5 amu (Dettmer et al. 2007). In addition, to improve performance work of TOF-MS, a quadrupole has been attached as a scanning device or a mass filter (Scholz et al. 2004). The quadrupole has four rods with high voltage to create a quadrupole field to select ions according to their *m/z* (only *m/z* within a certain range can pass the rods). Therefore the quadrupole enhances capability of TOF-MS in separation of metabolites.

Different metabolites with same molecular weights cannot be separated with the previous techniques. However, the problem can be addressed by using tandem MS/MS. This technique provides great selectivity through the specific fragmentation of each metabolite. After ions (precursors) pass through the first MS, they are activated by collision with an inert gas such as nitrogen or argon to produce ion fragments. The newly created fragment ions can subsequently be analyzed by the second MS. MS/MS is usually coupled with a quadrupole as the scanning device and TOF as the mass analyzer. The quadrupole-TOF-MS/MS instrument can clarify fragmentation process by distinguishing *m/z* of precursor and fragment ions. Therefore it allows high interpretation of spectra (Dettmer et al. 2007). Fourier transform ion cyclotron resonance mass spectrometry (FTICR-MS) is a powerful MS. It is provided with high resolution to separate and distinguish very complex mixtures and has high mass accuracy allowing calculation of elemental compositions to aid in structural differentiation and characterization (Sumner et al. 2003). Nevertheless it cannot be used to separate structural isomers. Furthermore FTICR-MS is an expensive instrument, thus the utilization is not widespread.

2.3.2 Gas chromatography mass spectrometry (GCMS)

GCMS is well known as one of the popular techniques for global metabolic profiling (J. Allwood, William. et al. 2008). It provides fast analysis at relatively cheaper costs compared to other mass spectrometry techniques, while remaining the ability to specific metabolite detection and quantification. Furthermore GCMS can be used to separate large number of compounds in a single measurement and identify unknown metabolites as well.

GCMS is a combined technique of gas chromatography (GC) and mass spectrometry (MS) to analyze different metabolites within a measurement. In this technique, the GC separates the metabolites while MS functions as metabolite detection tool. The GC is equipped with a capillary column as the stationary phase, a carrier gas as a mobile phase (He, N_2, H_2) and a

sample injector. Electron impact (EI), known as hard ionization, is mostly used for ionization in GCMS-based metabolomics. In EI, electrons are produced through thermionic emission interacting with the molecules in the gas phase to form molecular ions and fragment ions. TOF provides fast metabolite detection and can be used as mass detection, meanwhile single quadrupole and ion trap also has been used as mass analyzer in GCMS.

Metabolites identification in GCMS analysis is conducted by comparing retention time or retention index of metabolites sample with retention time or retention index of pure reference metabolites or spectral library database (C. Wagner et al. 2003). Quantitative metabolomics using GCMS requires making calibration curves of each metabolite, because sensitivity of GCMS varies for all metabolites (Y. H. Choi et al. 2008).

GCMS only can be used for analysing volatile metabolites with thermal stability. Derivatization can be used to measure none volatile and polar metabolites and increases volatility, thermal stability and reduce the polarities of the functional groups of metabolites. Silylation is a derivatization procedure mostly applied by replacing active hydrogen with trimethylsilyl (TMS) group. Silylation reagent is moisture sensitive and reacts heavily with water and decreasing the efficiency. All samples should be fully dry and solvents should be as pure as possible, to avoid inefficient derivatization. Alternatively, alkylation and acylation can be used for derivatization of functional groups, such as –COOH, -OH, -NH, and –SH.

2.3.3 Liquid chromatography mass spectrometry (LCMS)

LCMS is the other widely used technique for targeted or non-targeted metabolomics analysis and offers high selectivity and good sensitivity analysis. Great number of metabolites can be separated with LCMS, since overlapping peaks of different metabolites on the chromatogram can be identified as separate compounds in the mass analyzer of LCMS. However LCMS has disadvantages in reproducibility of separations that can be caused by matrix effects of the complex sample extract and different parameters in LCMS systems (Y. H. Choi et al. 2008).
LCMS instruments consist of two main parts, namely liquid chromatography for metabolites separation and MS for detection. ESI is commonly used for ionization of molecules in LCMS, although APCI also has been used to ionize more difficult metabolites. In ESI, the sample in a suitable solvent at atmospheric pressure is ionised by application of a high electric charge to the sample needle (J. W. Allwood and Goodacre 2010). Moreover ESI can cover broad metabolites, since it operates ionization in negative and positive modes. LCMS commonly

applies TOF and single quadrupole as mass analyzer and a combination of both has been used as well. In LCMS analysis each metabolite has a different sensitivity, therefore calibration curves of each metabolite are definitely needed for quantitative analysis.

Reversed phase column is mostly employed in metabolomics analysis using LCMS, since it is easy to use and can be applied for separation of majority metabolites. The characteristics of LC column are such as column internal diameter and packing particle size of column influences the level of LCMS resolution. Improving chromatographic resolution can be achieved by reducing diameter particle of column packing material as has been applied in ultrahigh pressure liquid chromatography mass spectrometry (UPLCMS) system (Plumb et al. 2006). This technique can reduce ionization suppression significantly and decrease co-elution of metabolites. High polar metabolites are mostly difficult separate by reversed phase columns. Alternatively, hydrophilic interaction chromatography (HILIC) columns can be used for separation of high polar metabolites. Moreover, HILICMS also has been used for analysis of highly polar compounds in *Curcubita maxima* leaves (Tolstikov and Fiehn 2002).

2.3 Data processing

The data obtained from medicinal plant metabolomics experiments are numerous and complex, thus very difficult to be interpreted by eye. However, after development of computer technologies and emerging of analytical software, the large amounts of data sets can be automatically visualized and interpreted. The common method for data processing in metabolomics is multivariate data analysis. It is a chemometric method that can visualize large number of compounds resulted from metabolomics experiments and data mine information about the relationships between levels of different metabolites (Jansen et al. 2010). To provide appropriate data for multivariate data analysis, the spectra from metabolomics experiments must be extracted.

Today, three methods have been developed for the raw data extraction, namely binning, peakpicking and deconvolution (Scalbert et al. 2009). Binning or bucketing is the most used method for data extraction in metabolomics (Spraul et al. 1994). By this method, spectra are subdivided into several regions, called “bins” or “buckets” and the total area within each bucket is used as a representation of the original spectra (J. L. Izquierdo-García et al. 2011). In NMR-based plant metabolomics, binning can reduce 16k data points to 250 data points (Schripsema 2010) while in MS-based plant metabolomics the number of bins is usually below 2000 at a bin size of 1 amu (Dettmer et al. 2007). However, binning in crowded

spectra has the potential for significant loss of information, for example by including peaks belonging to multiple compounds within a single bin (Staab et al. 2010). Peakpicking method as an alternative of bucketing basically consists of peak finding, baseline subtraction and alignment steps (Scalbert et al. 2009). An investigation has been conducted to compare peakpicking with bucketing on the data extraction, and the results showed that the peakpicking approach was more interpretable than the bucketing (Forshed et al. 2005). In deconvolution, defined as targeted profiling, the data is integrated by an algorithm from pure compound spectra and interrogated for identifying and quantifying the metabolites in the mixture (Weljie et al. 2006). In NMR-based metabolomics, this method provides NMR signal vectors and quantitative metabolite data (Scalbert et al. 2009), whereas in MS-based metabolomics, deconvolution reduces complexity of the data (Dettmer et al. 2007).

Normalization is the next step after raw data extraction. In NMR-based metabolomics, integral normalization is a standard method for normalizing and required to control possible variations in sample concentrations and variable sample dilutions (J. L. Izquierdo-García et al. 2011). Meanwhile in MS-based metabolomics, normalization coupled with transformation is introduced to minimize the impact of variability of high-intensity peaks (Dettmer et al. 2007).

After normalization, the extracted data are further analyzed by multivariate data analysis. Basically multivariate data analysis can be distinguished into two general types, namely unsupervised and supervised approaches. Unsupervised approaches are often known as clustering techniques, not using independent variables, provide a simplified description of the data with describing general information, and visualize the relationship between the dependent variables (Jansen et al. 2010). Moreover unsupervised approaches are powerful methods for sample classification. Whereas supervised approaches describe the subset of variation in the dependent variables and do not describe the variation irrelevant to the experimental question, thus provide more simplified compared to unsupervised approaches (Jansen et al. 2010). Supervised approaches are appropriate for metabolomics analysis that aims to discover characteristic compounds and the sample identity is often known (Dettmer et al. 2007).

One of the most popular of multivariate data analysis for unsupervised approaches is principal component analysis (PCA) (Jackson 1991). It is basically a data reduction technique and represents multivariate data in a low dimensional space. Furthermore PCA has capabilities for finding relationships and variances in the data, making a model of how chemical system

behave, and separating an underlying systematic data from noise (Wold et al. 1987). Figure 2 describes a graphical overview of the matrices and vectors used in PCA**.** In PCA, extracted data are represented by a set of new variables known as principal components (PC). Similarities and variances of samples according to metabolomic data are shown in a score plot and influence of each metabolite signal is visualized by the loading plot of PCA (Y. H. Choi et al. 2008).

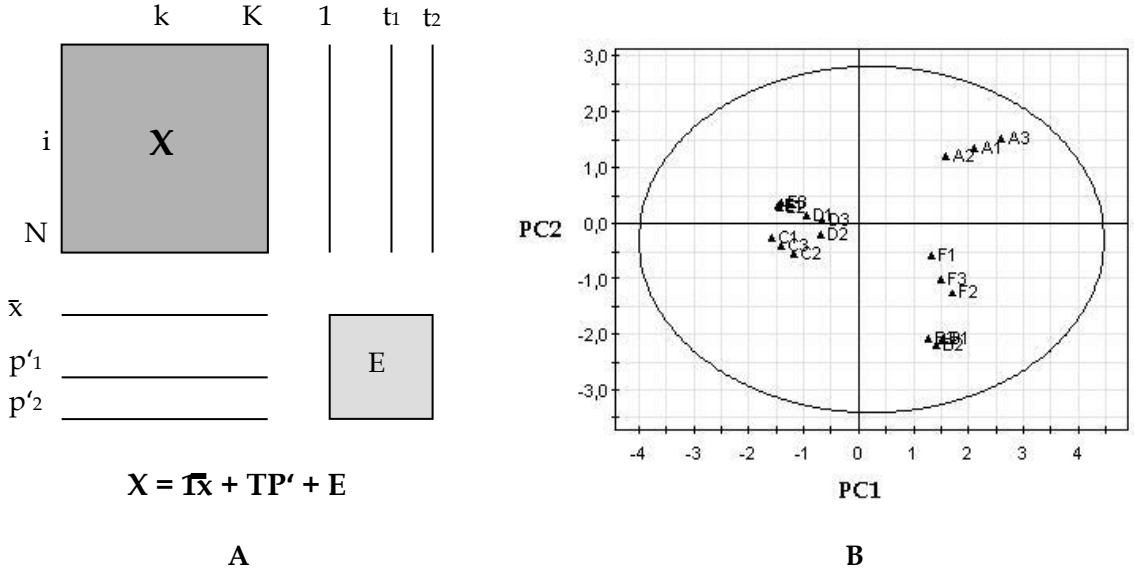


Figure 2. A is a graphic of a data matrix X with its first two principal components. Index is used for objects (rows) and index k for variables (columns). There are N objects and K variables. The matrix E contains the residuals, the part of data not "explained" by the PC models (Wold et al. 1987). B is an example of score plot.

The other common methods have been used for unsupervised approaches are independent component analysis (ICA) and hierarchical cluster analysis (HCA). ICA apparently is an improvement of PCA, because in the beginning step it needs PCA for reduction of the high dimension of the dataset and the quality of ICA is determined by the number of principal components as well (Scholz et al. 2004). In ICA, new set of components known as independent components (ICs) are calculated to detect more meaningful components, and different ICs represent different non-overlapping information (Scholz et al. 2004). ICA has shown a good result in metabolic fingerprinting for a small number of high-dimensional samples when PCA failed to do it (Scholz et al. 2004). HCA classifies samples in a data set based on their similarity. It creates a hierarchy of clusters that commonly visualized in a tree

structure called a dendogram where the root consists of a single cluster containing all observations and the leaves correspond to individual observations. Therefore HCA draws an easy description of the similarities of the samples within data sets (Sumner et al. 2003).

Partial least squares (PLS) regression is one of the common methods for supervised approaches. PLS combines features from PCA and multiple regressions. It can be used for discrimination with creating a linear regression model by projecting the predicted and the observable variables to a new dimension. PLS has been improved to several variants that also can be used for supervised approaches, such as partial least squares discriminant analysis (PLSDA), and orthogonal partial least squares discriminant analysis (OPLSDA). PLSDA can discriminate samples in a data set with identification of variables that significantly show relevant variations in the data set. OPLSDA is an improvement of PLSDA and removes irrelevant information and aligns the projections precisely to the aspect of interest, thus gives better interpretation than PLSDA. This technique discriminates between two or more groups in a data set (Bylesjo et al. 2006) in which the regression model is calculated between the multivariate data and a response variable that only contains class information (Westerhuis et al. 2010).

3. Metabolomics applications in medicinal plants

Metabolomics represents the phenotype on a metabolic level and is considered as the most informative technique for understanding biology systems (Sumner et al. 2003). Moreover it is a beneficial technique that quickly provides required information in the study of medicinal plants. In addition, metabolomics has been applied in various research fields of medicinal plants, such as classification of medicinal plants, characterization of plant cell cultures and transgenic medicinal plants, quality control, and proof for efficacy of medicinal plants from urine or blood samples.

3.1 Discrimination for classification of medicinal plants

Many factors influence the profiles of metabolites biosynthesised of medicinal plants. Different environmental conditions like soil, fertilizers, climate, pest control and insects for plant cultivation can create a high diversity of biochemical composition. Factors affecting metabolite production in medicinal plants are described in figure 3. Because of its commercial utility, some medicinal plants such as *Rhodiola rosea*, a folk medicine in Scandinavia and Russia (Ioset et al. 2011), are now cultivated at many locations which have different

geographies. These cultivations show variations in the metabolite profile effected by geographic and soil conditions. Furthermore morphological diversity represents varieties of the plants also leads to variations in profile of metabolites. As consequence, the medicinal value and pharmacological activity of an herbal medicinal product strongly depends on a desired metabolite composition, and there is a high interest by the pharmaceutical company to have cultivation and production process as much as possible under control. A smart, reliable and cost efficient analytical method is of high interest for the in process control to characterize and classify medicinal plants constantly.

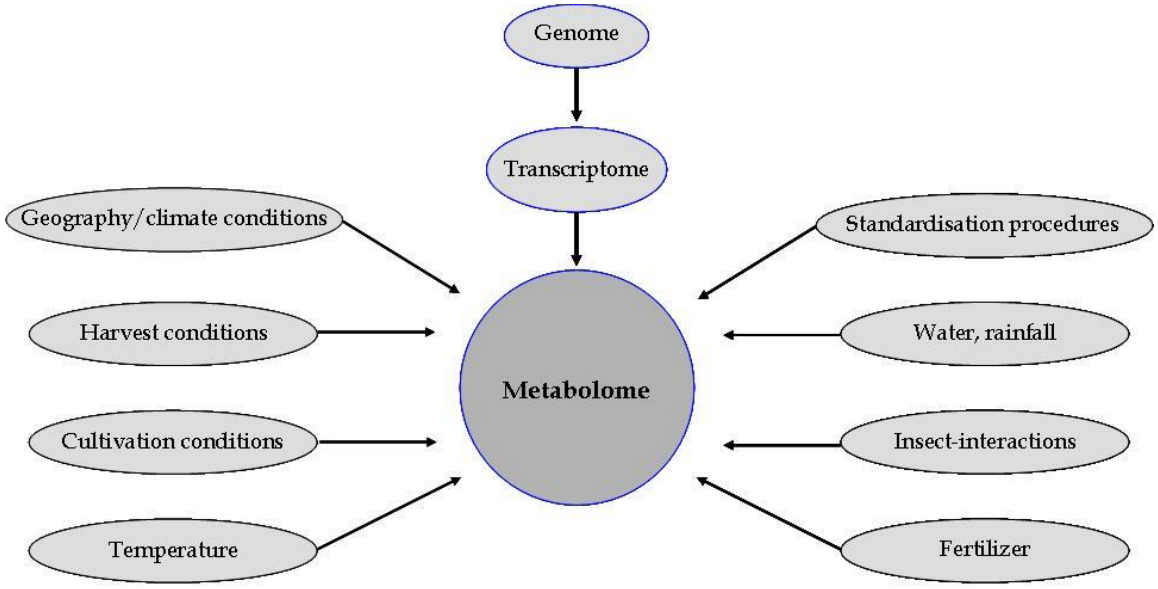


Figure 3. Factors affecting the metabolome of a medicinal plant

Metabolomics meet these requirements for metabolite analysis in herbal medicinal products, as it covers the complete spectrum of all natural products affected by cultivation conditions (temperature, rain and soil conditions) and location. All relevant data for internal use or for declaration at the authorities can be systematically compared and stored. As an example, using ^{1}H NMR coupled with PCA, an unsupervised metabolic profiling has been conducted to discriminate 9 rhizomes of *R. rosea* collected from Switzerland, Russia, and Finland (Ioset et al. 2011). According to PCA models, volatile compounds characterising the essential oil spectra have been identified as the discriminating factor of rhizomes of *R. rosea* (Ioset et al. 2011). Similarly the approach addressed to rhizomes and roots of *Notopterygium incisum* Ting ex H.T. chang (Qianghuo), which is known in Traditional Chinese Medicine (TCM). Comprehensive GCxGC and GCxGCMS combined with PCA analysis have been applied to

analyze 15 samples of volatile oils of Qianghuo from Sichuan, Ningxia, Inner Mongolia, Hunan and Gansu. The results indicated that monoterpenes and oxygenated sesquiterpenes were responsible for the differentiation of Qianghuo (Qiu et al. 2007).

Metabolomics also has been used to discriminate medicinal plants which were cultivated under the same environmental conditions but from different varieties. Metabolomic analysis using ^{1}H NMR coupled with PCA has been applied to discriminate 12 *Cannabis sativa* cultivars cultivated under the same standardised conditions. In the chloroform fraction THCA and CBDA were identified as important metabolites to discriminate cultivars from each other, whereas in the water fraction, sucrose, glucose, asparagine, and glutamic acid were identified as major discriminating metabolites (Y. H. Choi et al. 2004a). Discrimination of 11 varieties of *C. sativa* has also been conducted by quantification of the concentration of monoterpenoids, sesquiterpenoids, and cannabinoids using GC coupled with PCA (Fischedick et al. 2010).

Besides discrimination of medicinal plants from same species, metabolomics can also be applied to distinguish closely related species with similar metabolic profile. Three species of the genus *Panax*, namely *P. ginseng*, *P. notoginseng* and *P. japonicus*, have different pharmacological activities and highly similar morphology making it hard to differentiate between them. By using UPLCMS combined with PCA, the three species of *Panax* have been discriminated (G. X. Xie et al. 2008). PCA showed that chikusetsusaponin IVa, ginsenoside R0, ginsenoside RC, ginsenoside Rb1, ginsenoside Rb2 and ginsenoside Rg2 were the critical markers for discrimination of each species (G. X. Xie et al. 2008). The same problem, similarity in morphology, has been addressed for root from *Astragalus membranceus* Fisch. and *Hedysarum polybotrys* Hand. The identification of nine active compounds using HPLC-DAD, LCMS coupled with HCA showed significant differences in the metabolome and allowed exact and definitive identification of both roots (J. Zhao et al. 2008).

In another case, the discrimination of *Ephedra sinica*, *Ephedra intermedia* and *Ephedra distachya* var. *Distachya*, known as old Chinese medicinal herbs, has been conducted using ^{1}H NMR coupled with PCA (H. K. Kim et al. 2005). Benzoic acid analogues in the water fraction and ephedrine-type alkaloids in the organic fraction were identified to be the discriminating chemical markers in *Ephedra* species (H. K. Kim et al. 2005). The method that combines ^{1}H NMR with PCA has also been applied to discriminate the three most common species of *Strychnos* with different pharmacological activities. Brucine, loganin, fatty acids,

icajine and sungucine were identified as the key compounds for discrimination of *S. nux-vomica*, *S. ignatii* and *S. icaja* (Frederich et al. 2004).

3.2 Characterization of metabolites in plant cultures and transgenic medicinal plants

Plant cultures may offer an attractive solution for solving the limitations of extracting active metabolites from natural medicinal plants. Some attempts have been conducted with introducing elicitors or molecular biological approaches to the cultures for improving metabolite production and studying metabolite biosynthesis. Moreover metabolomics as a powerful method has been used as a promising method for identification of metabolites in medicinal plant cultures.

Metabolomics has been applied for monitoring elicitation process in medicinal plant cell cultures. By using ^{1}H NMR coupled with PCA, a metabolomic profiling approach has been conducted for identifying effects of elicitation of silver nitrate in cell suspension cultures of *Vitis vinifera*. It was found that the metabolite profile in an elicited suspension culture is different from the control and that lactate, alanine, acetic acid, choline, fructose, glucose and sucrose were the key compounds for differentiation of elicited cultures from the control (H. K. Choi and Yoon 2007).

Moreover metabolomics has also been used for studying effects of elicitation in metabolite biosynthetic pathway of cell cultures made from medicinal plants. A metabolomic profile using ^{1}H NMR combined with PCA has been taken for monitoring elicitation effects of yeast extract and yeast extract-methyl jasmonate in suspension cultures of *Silybum marianum*. The results revealed that elicitation influenced the phenylpropanoid pathway (Sanchez-Sampedro et al. 2007). Furthermore elicitation of yeast extract and methyl jasmonate caused a metabolic reprogramming affecting amino acid and carbohydrate metabolism, whereas yeast extract alone affected threonine and valine metabolism (Sanchez-Sampedro et al. 2007). GCMS coupled with PCA has been conducted to investigate the response of *Medicago trunctula* cell cultures towards elicitations of methyl jasmonate, yeast extract and ultraviolet. The results indicated that methyl jasmonate increased production of β-amyrin and saponins, while yeast extract induced accumulation of shikimic acid, an early precursor in the prenylpropanoid pathway, and ultraviolet gave insignificant effects (Broeckling et al. 2005). However not all elicitations can stimulate targeted metabolite production such as elicitation in cell suspension

cultures of *C. sativa* L which did not fulfil the expectation of a stimulation of cannabinoids production (Flores-Sanchez et al. 2009).

Metabolomics is also a powerful tool for identifying metabolites in hairy root cultures. A metabolomic profiling approach has been applied to discriminate two hairy root clones of *Psoralea corylifolia* L. and untransformed control. GCMS and LCMS analysis showed that the isoflavanoid formononetin and its glycoside were present only in the hairy root clones and suggested that clones are dissimilar in their secondary metabolism (Abhyankar et al. 2005). Beside that, metabolomics has also been used to validate a cryopreservation protocol for suspension cell cultures *Tabernaemontana divaricata* (Suhartono et al. 2005). The validation was based on metabolic profiling of fresh control and cryopreserved *T. divaricata.* Comparing obtained data clearly indicated that the level of the main alkaloid precursor tryptamine did not change and differences in the metabolome was only detectable on the level of several amino acids, carbohydrates, and fumaric acid (Suhartono et al. 2005).

In medicinal plant biotechnology, metabolomics has not been applied in plant cell cultures only, but also for transgenic plants. Metabolomics has been used to identify effects of different genetic modification approaches in metabolite production. For example, two lines of the transgenic *Artemisia annua* L. where amorpha-4, 11-diene synthase (ADS) was over-expressed and suppressed were analyzed by metabolic profiling approach using GCxGCMS combined with PLSDA. According to the data analysis, over-expression of the ADS influenced whole terpenoid metabolic pathway especially artemisinin biosynthesis, whereas suppression of ADS showed insignificant effects in artemisinin biosynthesis (Ma et al. 2009). Beside ADS, effects of over-expression of farnesyl diphosphate synthase at five developmental stages of *A. annua* L. have been analyzed by metabolomic profiling. The results showed that clear differences during all flowering stages (Ma et al. 2008).

3.3 Quality control of medicinal plant

Many factors can cause variations in the metabolite profile of an HMP, such as harvesting treatment, storage process, extraction technique, preparation method, and packaging (Fig. 3). Alterations of the metabolite profile change the pharmacological activities of the products. Moreover, these lead to different quality grades, alters efficacy of the products and also gives ambiguous results in clinical trials using medicinal plant products. In some cases, several additional components were illegally mixed into HMPs like aminophylline and prednisone

acetate, as reported for TCM products (Q. Liu et al. 2007). Beside that, composition of metabolites in the HMPs is very complex and difficult to be analyzed. Therefore, good analytical methods are required for guarantying high quality level of medicinal plant products. Furthermore metabolomics with different approaches has been used as a smart analytical method for quality control of medicinal plant products, since it offers efficient quantitative and qualitative analysis of metabolites comprehensively.

Commercial medicinal plant products mostly contain complex mixtures of compounds. Besides that, the correlation between composition of compounds in the mixtures and pharmacological action is commonly unclear. Moreover many pharmaceutical companies sell equal herbal products and claiming the same pharmacological activity and efficacy. A closer look clearly shows that extraction process are different, extraction media are not the same, sometimes the species used are different and the formulation (tablet, liquids, and sachets) cannot be compared. Typical examples for this phytochemical dilemma are *Gingko biloba*, *Echinacea purpurea*, St. John´s wort HMPs and many more. However, metabolomics has been successfully applied for monitoring complex compound mixtures and discriminating commercial herbal products. A simple method using ^{1}H NMR coupled with PCA has been used to characterize 10 samples of commercial St. John's wort from some suppliers. This method not only discriminate the samples from different suppliers but also show differences of samples from the same suppliers (Rasmussen et al. 2006). Moreover 14 commercially available *Tanacentrum parthenium* L. samples have also been discriminated using ^{1}H NMR combined with PCA (Bailey et al. 2002). Therefore all of these have shown that ^{1}H NMR coupled with PCA was an effective method in quality control of medicinal plant products.

Metabolomics in combination with ^{1}H NMR and PCA has also been used to investigate commercial capsules which claim to be from *A. annua* and contain high artemisinin content, an active compound for malaria medication. The results have clearly shown that the capsules contain low concentration of artemisinin and are indeed not *A. annua* but *A. afra* (Van der Kooy et al. 2008). Therefore it was shown that this method also provides a simple method for investigating fake commercial medicinal plants.

In some cases, metabolomics combining ^{1}H NMR and PCA is not sufficient for identification of each compound in very complex mixtures of medicinal plant products. However, ^{1}H J-resolved NMR (JRES) coupled with PCA offers an alternative solution to solve this problem. It has been demonstrated in quality control of commercial ginseng preparations. This method

can discriminate the preparations based on accumulation of ginsenosides, alanine, arginine, choline, fumaric acid, inositol and sucrose (Yang et al. 2006).

Improvement of multivariate data analysis in metabolomics such as combination of PCA and PLSDA has been applied for quality control of medicinal plant products. Combination of PCA and PLSDA in data processing can reduce dimensionality of multivariate data with preserving most data information, maximize separation between classes and minimize distance between intragroup clustering (Xiang et al. 2011). This technique has been demonstrated in quality control of *Curcuma*. By GCMS coupled with of PCA and PLSDA, the key compounds for discrimination of *C. phaeocaulis, C. kwangsiensis* and *C. wenyujin* has been identified, namely curzerenone, germacrone, curdione and epicurzerenone (Xiang et al. 2011). Furthermore, this combination has also been used to assess the quality control of *Pericarpium Citri reticulatae* and *Pericarpium citri reticulatae viride* (Ioset et al. 2011). Beside combination of PCA and PLSDA, orthogonal projection (OP) technique has also been proposed for the quality control. This technique was demonstrated in metabolic fingerprinting of *Houttuynia cordata* Thunb. and its results were compared with those of other techniques, such as PCA. According to this study, OP has shown better results in discrimination and identification of fingerprints than the others and thus it might be a promising tool in evaluating and discriminating the quality of medicinal plants products in the future (Liang et al. 2006).

Beside the improvement of data processing, combination of LCMS and GCMS analysis can also be another alternative technique in the quality control. This technique coupled with HCA has been used to identify 10 samples of *Caulophyllum robustum*. According to HCA, this technique can discriminate all samples and can also distinguish unacceptable samples for quality control (Y. P. Li et al. 2007). As a consequence, this technique offers a meaningful technique for quality control of medicinal plant products.

3.4 Identification of medicinal plant bioactivity

Identification of the components that are responsible for bioactivity of medicinal plants can be conducted using metabolomics coupled with bioactivity assays. This approach has been demonstrated for identification of the components of xiaoyaosan that is responsible for anti-depression. Xiaoyaosan is a product of TCM which composed from several medicinal plants (Zhou et al. 2011). Different fractions were administrated to stressed rats and subsequently

the urine was measured by NMR and the data was mined using PLSDA. The results showed that the metabolite profile of the rat treated with petroleum ether fraction was more similar to that of the positive control. The authors conclude that the fraction containing mostly lipophilic compounds had the highest anti-depression activity (Zhou et al. 2011).

Metabolomics also has been used to investigate therapeutic and synergistic effects of three major ingredients (TSG) of Compound Danshen Formula (CDF), namely tanshinone IIA (T), salvianolic acid B (S) and ginsenoside Rb1 (G). CDF is a formula of TCM derived from the herbs of *Radix salviae miltiorrhizae* (Danshen in Chinese), *Radix notoginseng* (Sanqi) and *Borneolum syntheticum* (Bingpian) (Lu et al. 2011) and used to improve coronary and cerebral circulation (P. Li et al. 2007). Myocardial ischemia rats were administrated with T, S, G, TSG and CDF. Afterward the blood plasma was analyzed by LCMS and PLSDA and revealed that the activity TSG was closer to activity of CDF whereas the activities T, S and G were low. Therefore it was confirmed that T, S and G have a synergistic effect when they are used together (Lu et al. 2011).

In other case, metabolomics also can be used to predict biological activity. This method has been demonstrated in prediction of anti plasmodial activity of *A. annua* extracts. Some extracts of *A. annua* obtained from different locations have been discriminated by ^{1}H NMR combined with PCA and it was found that artemisinin is the main discriminate factor for clustering. Afterward PLS and PLSDA were used to create the models for prediction of anti-plasmodial activity. The results showed that the predictive activities were very similar with experimental activities (Bailey et al. 2004).

3.5 Study efficacy of medicinal plants

Efficacy and toxicity (see following 3.6) of HMPs have to be assessed by in vivo studies (phase I) and later by clinical studies (phase II and III). In vivo studies are carried out in animal models such as rats, later clinical studies are carried out in humans and samples from blood and urine are sampled for further analytics. An illustration of hierarchy levels for quality assurance in HMPs is explained in figure 4. Besides of approving clinical efficacy, bioequivalence and phytoequivalence of HMPs are important questions. In contrast to phytoequivalence defining bioequivalence is rather simple as a term to describe biological and pharmacological equivalence of two proprietary preparations of an HMP. In principal, phytoequivalence can be considered as a physiological equivalence between the effects of two

herbal extracts, the efficacy of one of which has been clinically demonstrated, but according to drug regulating authorities this definition does not include the condition of two phytochemical identical extracts.

To achieve a phytoequivalent extracts both preparations must have qualitatively and quantitatively the same metabolic profile. Due to different extraction procedures which are mostly disclosed by the company as marketing authorization holder, as generic company it is nearly impossible to get two completely equal HMPs based on the original production process. To solve this dispute it would be in the interest of the patient and consumer to define phytoequivalence according to existing regulation in the biotech field for so called "biosimilars". Here, for recombinant therapeutic proteins the basic amino acid structure must be identical to the innovator product, but other decorations like glycolisation may differ if both have the same or comparable biological effect. To follow up this idea, HMPs are phytoequivalent, if main pharmacological relevant compounds or biomarkers are present and both have similar pharmacological and pharmacokinetic profiles. To meet the criteria of similar phytochemical profiles for pharmacological active components, metabolomic is an important tool to identify relevant compounds in complex mixture.

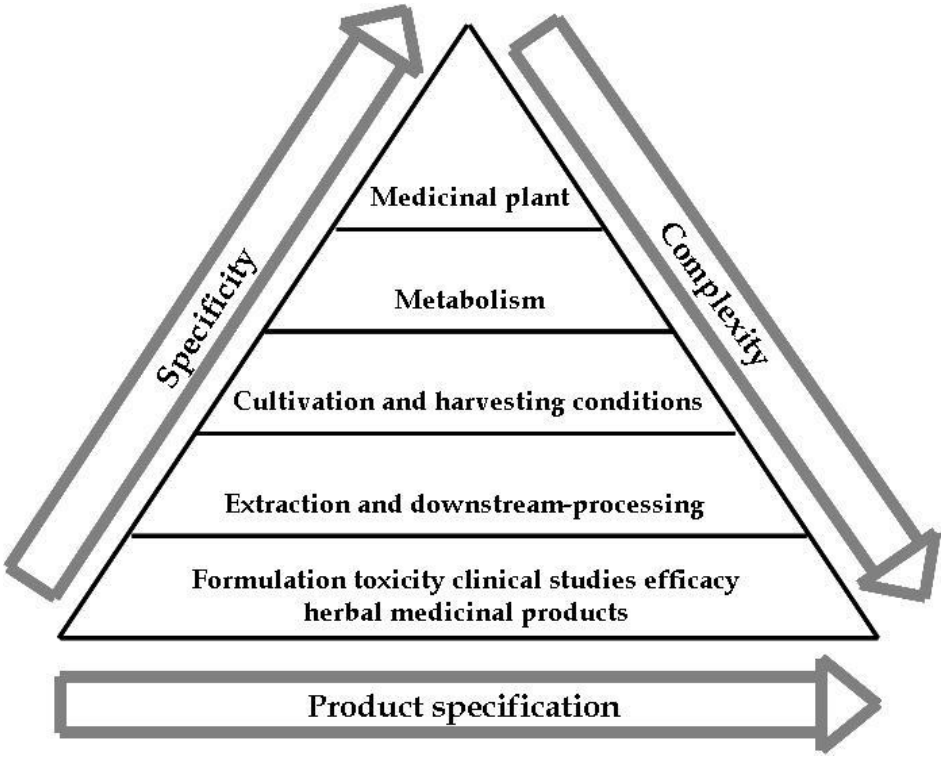


Figure 4. Hierarchy levels for quality assurance in HMPs

UPLCMS and PCA coupled analysis has been used to study efficacy of *Epimedium brevicornum* Maxim, a traditional Chinese medicine, in kidney deficiency syndromes. A high dose of hydrocortisone was injected to rats to induce pathological condition similar to kidney deficiency syndrome. Afterward, these rats were treated with the extract of *E. brevicornum* Maxim and urine and blood was analyzed. Based on PCA, the metabolite profiles of treated rats were shifted closer to that of pre-hydrocortisone administration. Therefore, the efficacy of *E. brevicornum* for kidney deficiency syndromes was proven (F. M. Li et al. 2007). Beside that, efficacy of anti-depressive HMP of xiaoyaosan has been investigated using GCMS and three multivariate data analysis techniques, PCA, PLSDA and HCA. Thirteen metabolites relating to anti-depression were identified in urine sample of rats. Moreover the multivariate data analysis revealed that the metabolite profile of the rats treated with high dose of xiaoyaosan was much closer to the control and gave clear evidence to the efficacy of xiaoyaosan (Dai et al. 2010).

In other case, ^{1}H NMR and PCA have been conducted to proof anti-ageing efficacy of the total flavones in *Epimedium* (TFE), considered as the major active components of *Epimedium koreanum Naka*. Data showed 26 characteristic resonances related to the age of the rats and identified successfully 10 of them, namely as creatinine, lactate, alanine, acetate, acetone, succinate, allantonin, methylamine, dimethylamine and trimethylamine-*N*-oxide (B. Wu et al. 2008). Furthermore, according to the PCA, the metabolite profile of 24-month-old rats treated with TFE was similar with that of 18-month-old rats without TFE. It thus showed the anti-ageing efficacy of TFE (B. Wu et al. 2008). ^{1}H NMR coupled with PCA has also been used to study efficacy of *Ginkgo biloba*. Stressed rats were treated with *Ginkgo*, afterward their urine were analyzed with ^{1}H NMR. Results showed the metabolite profile of the treated rats has changed close to that of the control rats and it suggested that *Ginkgo* responsible for that action (Wang et al. 2005).

3.6 Investigation of medicinal plant toxicity

Toxicity is one of the obstacles in the use of medicinal plants can be investigated with metabolomics as well. This method is based on the fact that organ damage could be associated with the corresponding changes in metabolite profiles. This has been documented in a metabolomics study of the toxicity of Hei-Shun-Pian, a traditional Chinese medicine derived from root of *Aconitum Carmichaelii* Debx. and known for its analgesic, antipyretic, anti-rheumatoid arthritis and anti-inflammation effects (L. Li et al. 2008). Urine samples and

plasma of the rats that had been administered orally with the decoction of Hei-Shun-Pian were measured with ^{1}H NMR and data analyzed by PCA. In the early stage of the dosing period, urinary excretion of trimethylamine-*N*-oxide and taurine which is closely related to the cardiovascular system (Militante and Lombardini 2001) were decreased, while the concentration of citrate, 2-oxoglutarate, succinate and hippurate were increased. These results revealed that the extract of Hei-Shun-Pian has cardiac toxicity and caused disturbances on heart and liver.

Besides of HMPs, metabolomics has also been applied to investigate toxicity of single compounds. A metabolomic approach has been conducted to study toxicity of aristolochic acid, widely present in *Aristolochia*, *Bragantia* and *Asarum* species (Ni et al. 2007). Urine samples form treated rats and control rats were analyzed in combination with GCMS and LCMS and data mined by PCA and OPLSDA. The results revealed that metabolite profiles of treated rats were different from that of the control rats. Furthermore the concentrations of homocysteine and serine, accepted for diagnosis of kidney morbidity, were lower than those of the control. It thus showed that aristolochic acid caused disruption on kidney (Ni et al. 2007).

Toxicity of aristolochic acid has also been studied in comparison with the urinary profiles from rats treated with aristolochic acid and known toxicants such as sodium chromate ($NaCrO_4$), and mercury II chloride ($HgCl_2$) as controls at various time intervals. The results also revealed that aristolochic acid induced renal toxicity (X. Y. Zhang et al. 2006). From this communication we learn that the combination of metabolomics and histopathological experiment gives a new perspective over the potential of theses new bioanalytical tools (LCMS coupled with PCA and linear discriminate analysis) (Chen et al. 2006). In a more recent case on aristolochic acid toxicity, a metabolomic approach using LCMSMS coupled with PCA has been used to investigate renal toxicity but in addition the authors also identified potential biomarkers related to kidney disease, namely kynurenic acid and hippuric acid (Chan and Cai 2008).

4. Conclusions

Many factors influence the profiles of metabolites biosynthesised by medicinal plants. Different environmental conditions like soil, fertilizers, climate, pest control and insects for plant cultivation can create a high and not constant diversity of biochemical composition.

However metabolomics has been proven as powerful method to discriminate and classify same medicinal plants cultivated at different geographies, different varieties plants, and different plants but have same morphology by identifying the key compounds for discriminating. Moreover metabolomics has been used for monitoring elicitation process in medicinal plant cell cultures and for identifying effects of different genetic modification approaches in metabolite production of transgenic plants.

Harvesting treatment, storage conditions, extraction technique, preparation method, and packaging affect the quality of medicinal plant products significantly. Therefore, good analytical methods are required for guarantying high quality level from production to the pharmacy shelf. However metabolomics with different approaches has been successfully applied as a smart analytical method for quality control of medicinal plant products, since it offers efficient quantitative and qualitative analysis of metabolites comprehensively. On the other hand, metabolomics has been coupled with bioactivity assays to identify components responsible for bioactivity, investigate synergistic effects, and even to predict bioactivity of a medicinal plant product. Moreover metabolomics has also been applied to study efficacy of medicinal plant products with measuring urine, blood or other biofluids of addressed object (human or rats) and comparing it with the control. Furthermore the fact that organ damage could be associated with the corresponding changes in metabolite profiles leads to application of metabolomics in toxicity investigation of medicinal plant products.

All of these have shown that emergence of metabolomics has opened new opportunities to answer challenges in medicinal plant fields. Metabolomics has been proven as a breakthrough method to accelerate and streamline the analytical process of medicinal plant researches by allowing quick, efficient identification and quantification of the metabolites within the samples.

Chapter 3

^{1}H NMR-Based Metabolomics Differentiation and Real Time PCR Analysis of Medicinal *Cannabis* Organs

Nizar Happyana, Oliver Kayser

Oliver Kayser coordinated and supervised the project.

Abstract

Cannabis sativa L. is an annual plant that has been cultivated and used as a source of medicine, fiber, oil, and food, since prehistoric time. In this report, ^{1}H NMR-based metabolomics along with RT-PCR analysis was performed to study metabolite profiles and cannabinoids biosynthesis in different organs of *Cannabis*. Two medicinal varieties, Bedrocan and Bedica, were analyzed. Principal component analysis (PCA) models have been successfully applied to classify metabolite profiles of *Cannabis* samples. Δ^{9}-tetrahydrocannabinolic acid [(THCA) (1)], cannabichromenic acid [CBCA (2)], cannabigerolic acid [CBGA (3)], and Δ^{9}-tetrahydrocannabinol [THC (4)] were identified contributing on the classification in the chloroform extracts. Among those, THCA (1) was the most important discriminant compound. The 13 identified compounds in the water extracts were detected contributing in the classification of metabolite profiles of water samples. Furthermore, the absence of asparagine signals in the ^{1}H NMR spectra of the leaves indicated this compound as a leading differentiate compound in the water extracts. RT-PCR analysis in the medicinal plants revealed that the highest gene expression level of THCA synthase (THCAS) was in the leaves of the plants. Predicted-CBGA synthase (predicted-CBGAS) gene was expressed in the same level in all samples. Meanwhile, genes of olivetolic acid synthase (OAC) and olivetol synthase (OLS) were less expressed in the leaves than in the other organs of medicinal Cannabis. Combination of ^{1}H NMR-based metabolomics and RT-PCR revealed expression level of cannabinoid genes and metabolite profiles, especially cannabinoids, in trichomes, flowers, and leaves of the plants. Therefore this alliance provided more complete study of cannabinoids production.

Keywords: *Cannabis sativa* L., metabolomics, trichomes, cannabinoids, ^{1}H NMR, PCA, THCA

1. Introduction

Cannabis (*Cannabaceae*) is an important genus of flowering plants that has been cultivated and used by humankind as a source of medicine, fiber, oil, and food since prehistoric time (Chopra 1969; Fleming and Clarke 2008). Ancient Chinese literature described cultivation methods of *Cannabis* that had been used for thousand years (H.-L. Li 1974). *Cannabis* is an annual dioecious herb (Flemming et al. 2007; Pacifico et al. 2006) and native from Central and Eastern Asia (Candolle 1886; de Barge 1860). Now the plant can be found worldwide from the equator to about 60 ºN latitude, and throughout much of the southern hemisphere (Hillig 2005).

Cannabinoids are well known as the most responsible compounds for the biological activities of this plant. Cannabinoids are a unique group of terpenophenolics possessing alkylresorcinol and monoterpene moieties in their molecular structure. More than 100 cannabinoids have been identified from this plant (Happyana et al. 2013). Because of their psychoactivity, Δ^9-tetrahydrocannabinol [THC (4)] and cannabidiol (CBD) are the most studied and interesting compounds of the class. Compared to other organs of *Cannabis*, trichomes contain a highest amount of cannabinoids (Petri et al. 1988; Turner et al. 1978).

Based on cannabinoid composition, generally this plant is classified into medicinal-type and fiber-type *Cannabis*. Biomarker cannabinoids used for this classification are Δ^9-tetrahydrocannabinolic acid [THCA (1) and or its neutral form, THC (4)], and cannabidiolic acid [CBDA and or its neutral form, CBD] (Grlic 1968, 1962; Grlic and Andrec 1961; C. E. Turner et al. 1980). Medicinal *Cannabis* contains high THCA (1), while the fiber type has a high amount of CBDA. THCA (1) as the biomarker was confirmed also by a report of ^{1}H NMR-based metabolomics of *Cannabis* cultivars (Y. H. Choi et al. 2004a). The samples used in this report were flowers of each cultivar and a leave sample from one of cultivars. However, as far as our knowledge, metabolomics analysis to differentiate metabolite profiles of trichomes, flowers, and leaves of *Cannabis* had not reported yet.

The biosynthesis of THCA (1) is started with the formation of two cannabinoid precursors, namely geranyldiphosphate (GPP), originating predominantly from the non-mevalonate pathway (MEP) (Fellermeier et al. 2001), and olivetolic acid, derived from condensation of malonyl-CoA and Hexaonyl-CoA by olivetolic acid cyclase (OAC) and olivetol synthase (OLS) (Gagne et al. 2012). Both precursors are condensed to form cannabigerolic acid (CBGA (3)) by an enzyme predicted to be a representative of the geranyltransferase group

(Fellermeier and Zenk 1998). CBGA (3) is subsequently transformed to THCA (1), by THCA synthase (THCAS) (Taura et al. 1995).

In this study, we report ^{1}H NMR-based metabolomics of different organs of *Cannabis* coupled with a real time PCR (RT-PCR) analysis. The purpose of this study is to investigate and differentiate metabolite profiles, and study cannabinoids biosynthesis in the leaves, flowers, and trichomes of *Cannabis*. Two different varieties of this plant were analyzed, namely Bedrocan and Bedica that are categorized as medicinal-type *Cannabis*. Expression level of THCAS, OLS, and OAC were analyzed in the different organs of the plants with RT-PCR. Beside published cannabinoid genes, we also investigated an unpublished gene from our group that predicted as CBGA synthase (predicted-CBGAS). This work indicated that the combination of both methods provides a complete study of cannabinoids production in the different plant organs.

2. Material and methods

2.1 Chemicals and reagents

Analytical grade chloroform, methanol, $CDCl_3$ (99.8%), D_2O (99.8%), tetramethylsilane (TMS), trimethylsilane propionic acid sodium salt (TSP), and NaOD were purchased from Carl Roth GmbH (Karlsruhe, Germany). Anthracene was obtained from Sigma-Aldrich Chemie GmbH (Munchen, Germany). Reference compounds of cannabinoids, Δ^9-tetrahydrocannabinol [Δ^9-THC], cannabidiol [CBD], Δ^9-tetrahydrocannabinolic acid [THCA], cannabidiolic acid [CBDA], and cannabigerolic acid [CBGA], were purchased from THC Pharm GmbH (Frankfurt, Germany).

2.2 Plant material

Two medicinal *Cannabis* varieties, Bedrocan and Bedica were supplied by Bedrocan BV (the Netherlands). The plants were grown indoors under standardized conditions. They were initially generated from cuttings of standardized mother plants and cultivated under controlled, long day-light conditions (18 h/day). After the vegetative growth phase, the flowering stage was induced under a shorter (12 h/day) light regime. Plant specimens were assigned voucher numbers (Bedrocan: N8.16.07.2012; Bedica: C8.16.07.2012) and deposited at Department of Technical Biochemistry, Technical University of Dortmund. All plant

handling and experimental procedures were carried out under the license No. 4584989 issued by the Federal Institute for Drugs and Medical Devices (BfArM), Germany.

2.3 Trichomes isolation

Trichomes were isolated according to Yerger et al. (Yerger et al. 1992) with slightly modification. Fresh flowers of *Cannabis* were put in the liquid nitrogen. Floral leaves and stigma on the flowers were removed using a forceps with occasionally the flowers were put back into the liquid nitrogen. Afterward, a 5-10 gram of the flower was transferred into a 50 ml falcon tube and placed into liquid nitrogen. The tube containing the flower was removed from the liquid nitrogen tank and approximately 2 to 3 cm^3 of finely powdered dry ice (prepared by wrapping a piece in clean paper towels and crushing with a pestle) was added to the tube. Subsequently, the tube was loosely capped and vortexed at maximum speed for approximately 1 minute and then the flower in the tube was removed. For obtaining the trichomes, the contents of the tube were sieved through a nylon net filter with 140 μm pore diameter (Merck Millipore, Germany) into a 500-mL beaker glass that surrounded by dry ice. The trichomes in the beaker were transferred into a 2 ml frozen microcentrifuge tube with spatula and placed in the liquid nitrogen.

2.4 Extraction

One hundred milligrams of dried powder of *Cannabis* materials (trichomes, flowers, and leaves) was transferred into a test tube. Two milliliters of 50% aqueous methanol and 2 mL chloroform were added to the tube followed by vortextion for 1 minute and sonication for 1 min. The sample was shaken at 200 rpm for 1 hour at 30 ºC. The water and chloroform phases were separated with pipetting and filtering into new centrifuge tubes. The chloroform fraction was dried in a rotary vacuum evaporator at temperature 30 ºC and pressure 31 mbar. The water fraction was dried in a freeze dryer machine. Totally 54 samples of chloroform fractions and 54 samples of water fractions that obtained from different organs of Bedrocan, and Bedica (trichomes, flowers, and leaves) were prepared.

2.5 NMR measurements

Dry chloroform extracts were dissolved in deuterated chloroform containing tetramethylsilane (TMS) as an internal standard and anthracene as a quantitative internal standard. ^{1}H NMR spectra of chloroform extracts were recorded using a Varian Inova 500 spectrometer operating

at 499.78 Hz. The spectra were measured with following parameters: acquisition time = 4.09 s, relaxation delay = 5 s, pulse width = 4.75 μs, Free Induction Decay (FID) data points = 90122, spectral width = 11001.10 Hz, and number of scans = 64. Potassium dihydrogen phosphate was added to deuterium oxide as a buffering agent. The pH of the deuterium oxide was adjusted to 6.0 using a 1 N sodium deuteroxide solution. Afterward, water extracts were dissolved in the deuterium oxide containing trimethylsilane propionic acid sodium salt (TSP, 0.01%, w/v) as an internal standard. ^{1}H NMR spectra of water extracts were recorded using a Bruker Avance DRX 500 spectrometer (Bruker BioSpin GmbH, Rheinstetten, Germany) operating at 500.13 MHz. The spectra were measured with a presaturation pulse program with following parameters: acquisition time = 2.72 s, relaxation delay = 2 s, FID data points = 32768, spectral width = 12019. 23 Hz, number of scans = 128. Processing of FIDs was performed with line broadening set to 1.0 Hz using ACD/Labs 12.0 software. All ^{1}H NMR spectra were manually phased and baseline corrected.

2.6 Data analysis and quantification

For the $CHCl_3$ extracts, ^{1}H NMR spectra were scaled to TMS and then reduced to integrated regions of equal width (0.02 ppm) corresponding to the region δ 0.50-13.00 ppm. The region δ 7.24-7.27 ppm was excluded from the analysis because of the residual signal of $CHCl_3$. The region of antrachene peaks, namely δ 7.44-7.48 ppm; δ 7.96-8.06 ppm; δ 8.42-8.48 ppm, were removed as well. ^{1}H NMR spectra were scaled to TSP for the water extract and reduced to integrated regions of equal width (0.02 ppm) from δ 0.50-10.00 ppm. The solvent regions were excluded from the analysis. Bucketing was performed by ACD/Labs software with scaling on total intensity. Principal component analysis (PCA) was performed with SIMCA-P software (demo version. 13.0, Umetrics, Umeå, Sweden). Scaling of PCA was based on the Pareto method. Metabolites that have individual signals were quantified. For the identified cannabinoids, quantitation was performed by calculating the relative ratio of the peak area of selected proton signals of the target cannabinoids to the singlet peak area of anthracene.

2.7 RNA isolation and real time PCR

Fresh plant materials were ground to a fine powder using a pestle and a mortar under a cold condition (with adding liquid nitrogen). Afterward, around 90 mg of fresh fine powder of plant materials was prepared for RNA isolation using RNeasy^{TM} Plant Mini Kit (Qiagen N.V., The Netherlands). The isolation procedure followed manufacturer's instructions.

Concentration of isolated RNA was determined with a Qubit® RNA Assay Kit (Life Technologies, California, USA). Equivalent quantities of RNA isolated from leaves, flowers, and *Cannabis* trichomes were reverse transcribed using iScript™ cDNA synthesis kit (BIO-RAD Laboratories, California, USA) according to the manufacturer's instructions. Quantitative real-time PCR experiments were performed with Applied Biosystems StepOnePlus™ Real-Time PCR Systems (Applied Biosystem, Foster City, Canada) and SYBR® Green PCR Master Mix (Applied Biosystem). These experiments were carried out using comparative CT method (ΔΔCT) (Livak and Schmittgen 2001). S18 housekeeping gene was used as an endogenous control gene. Gene sequences of THCAS (Sirikantaramas et al. 2004), OLS (Taura et al. 2009), and OAC (Gagne et al. 2012) were used as the templates for designing primers using Primers Express Software v.3.0.1 (Applied Biosystem). The list of primers could be seen in the table 1. Real-time PCR was run with following conditions: pre-incubation at 95 ºC for 10 min, followed by 40 cycles of 95 ºC for 15 s and 60 ºC for 1 min, and then melt curve stage at 95 ºC for 15 s, 60 ºC for 1 min, and 95 ºC for 15 min.

Table 1. Primers of cannabinoids genes for RT-PCR analysis.

Gene	**Forward Primer**	**Reverse Primer**
18S	GAGAAACGGCTACCACATCCA	CCGTGTCAGGATTGGGTAATTT
THCAS	AAGTTGGCTTGCAGATTCGAA	TGTAGGACATACCCTCAGCATCA
Predicted-CBGAS	TTGGGAAGGCATGTTGGAA	AATCCACAAGCGCATGAAGTAA
OLS	GGGCTGCTGCGGTGATT	TATCGGCCTTTCCCCAACT
OAC	TGTTGGATTTGGAGATGTCTATCG	TTCGTGGTGTGTAGTCAAAAATGA

3. Results and Discussions

The common method for processing a large set of data produced from metabolomics experiments is multivariate data analysis. It is a chemometric method that visualizes a large number of compounds resulted from metabolomics experiments and reveals information about the relationships between levels of different metabolites (Jansen et al. 2010). One of the most used methods of multivariate data analysis is principal component analysis (PCA). It is an unsupervised approach, often known as a clustering technique, reduces the dimensionality of a given data set by producing new linear combinations of the original variables. PCA provides a simplified description of the data with describing general information, visualizes the relationship between the dependent variables and is a powerful method for sample

classification (Jansen et al. 2010; Happyana et al. 2012b). In this study, PCA has been applied as a multivariate data analysis for processing huge data sets obtained from the experiments.

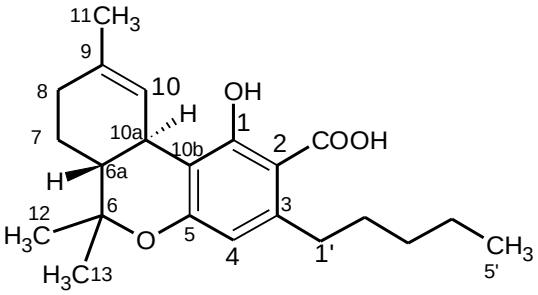
In the first metabolomics experiment, we performed PCA analysis in order to classify metabolite profiles in the chloroform extracts of trichomes, flowers, and leaves of the varieties Bedica and Bedrocan. As shown in the figure 2A, the PCA model of the chloroform extracts of all samples provided a well preview of classification of metabolite profiles in all samples. The eight-component model explained 98.8% of the variance (R2), with the first three components explaining 93.4%, and 96.3% of predictive ability (Q2). Among the principal components (PCs), the combination of PC1 (81.3%) and PC2 (6.4%) gives well separated clusters in the score plot as seen in the figure 2A. Examination on the score plot revealed that cluster position of Bedrocan leaves was close to the position of Bedica leaves which indicated that both had similar metabolite profiles. The same case was also found on the other same organs of both varieties. Moreover, the cluster distance between same organs of Bedrocan and Bedica was closer compared to the distance within organs of each variety. For example, cluster distance between Bedrocan and Bedica flowers was closer than the distance between Bedrocan flowers and Bedrocan trichomes. Therefore, it confirmed that the differences of metabolite profiles within organs of Bedrocan or Bedica was bigger compared to the differences between same organs of Bedrocan and Bedica.

For identifying responsible compounds for the classifications of chloroform extracts of all samples, the loading plot of PC1 was investigated. The loading plot showed some signals giving important contributions to the classification, namely the chemical shifts at δ 6.40, 6.25, 3.23, 1.69, 1.45, 1.12, and 12.20 (see figure 2B). Examination of ^{1}H NMR spectra of chloroform extracts revealed that the signals belong to THCA (1), namely H-10, H-4, H-10a, H-11, H-12, H-13, and H-hydroxyl of carboxylic acid, respectively. Further investigation showed that other signals contributed also to the loading plot of PC1, namely δ 0.90, 1.33, 1.33, 1.55, 2.72, and 2.92. Based on examination of ^{1}H NMR spectra of chloroform extracts, the signals belong to pentyl signals of THCA (1) which assigned as H-5', H-4', H-3', H-2', H-1a', H-1b', respectively. Therefore, it proved that THCA (1) was a responsible compound for the classification.

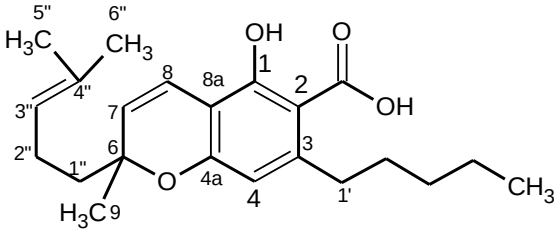
Furthermore, this result has been confirmed by quantification of THCA (1) in the chloroform extracts using ^{1}H NMR method. The quantification showed that concentrations of THCA (1) varied on the different organs of Bedrocan and Bedica as seen in table 2. The highest

concentration of THCA (1) was found in the trichomes of Bedrocan, while the lowest was discovered in the leaves of Bedrocan. Besides that, the similarity of THCA (1) contents of Bedrocan leaves (4.58±0.33 mg/g) and Bedica leaves (4.81±0.47 mg/g) lead to both clusters being present close each other on the score plot.

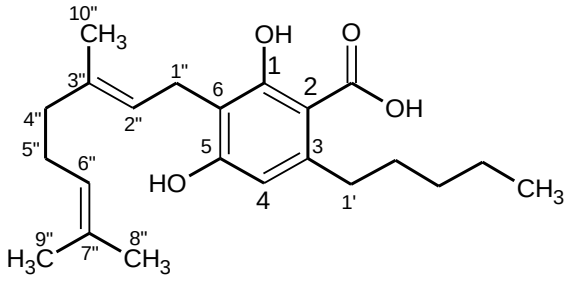
In addition to THCA (1), some signals belonging to CBCA (2), CBGA (3), and THC (4) gave contributions also to the classification of metabolite profiles in the chloroform extracts. The signals at δ 6.74, 6.13, and 2.10 assigned as H-8 of CBCA (2), H-2 of THC (4), and H-2" of CBGA (3), respectively, affected the loading plot of PC1. Quantification analysis showed that concentrations of CBCA (2) and CBGA (3) only could be determined in the trichomes of Bedrocan and Bedica as seen in the table 2. Meanwhile, the amounts of THC (4) only could be quantified in the trichomes and flowers of Bedrocan, and Bedica trichomes. These results indicated that the contents of the three compounds varied in the samples and confirmed that its impact on the classification.



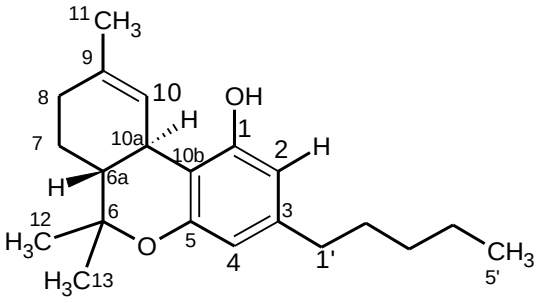
Tetrahydrocannabinolic acid, THCA (1)



Cannabichromenic acid, CBCA (2)



Cannabigerolic acid, CBGA (3)



Tetrahydrocannabinol, THC (4)

Figure 1. Structures of identified cannabinoids

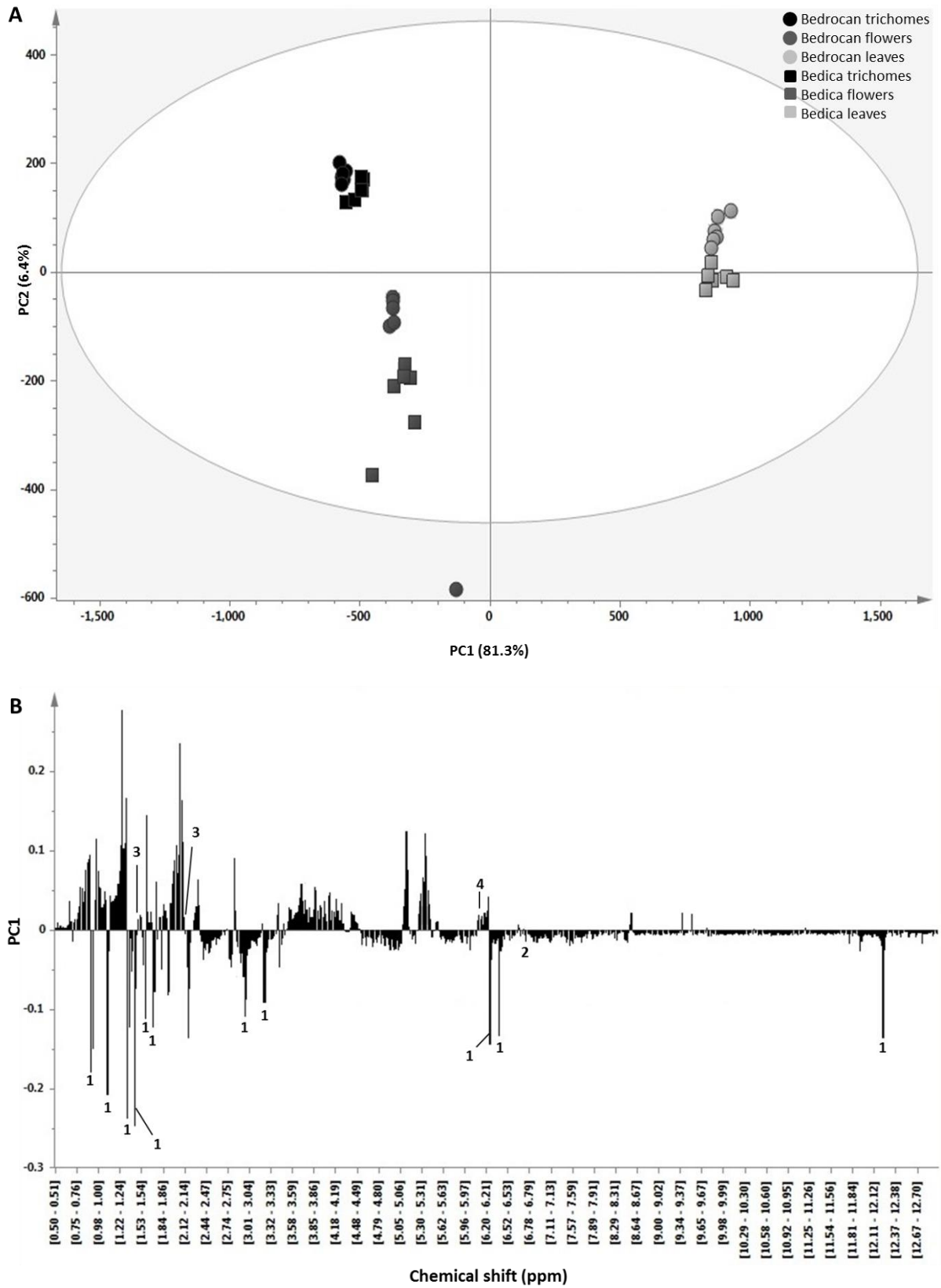


Figure 2A-B. Score plot (A) and loading plot (B) of chloroform extracts of all samples

Table 2. List of identified compounds. +: compound is detected in the samples; -: compound is not detected in the samples.

Compounds	Bedrocan (mg/g±SD)*			Bedica (mg/g±SD)*		
	Trichomes	Flowers	Leaves	Trichomes	Flowers	Leaves
CBCA	3,58±0,39	+	-	2,93±0,41	+	-
CBDA	-	-	-	-	-	-
CBD	-	-	-	-	-	-
CBGA	15,46±0,65	-	-	15,73±0,65	-	-
THCA	204,53±7,81	83,96±6,28	4,58±0,33	155,06±11,32	62,42±6,22	4,81±0,47
THC	3,32±0,38	1,36±0,11	-	2,13±0,41	+	-
α-Glucose	+	+	+	+	+	+
Alanine	+	+	+	+	+	+
Asparagine	+	+	-	+	+	-
Choline	+	+	+	+	+	+
Formic acid	+	-	-	+	-	+
Fructose	+	+	+	+	+	+
Glutamic acid	+	+	+	+	+	+
Glutamine	+	+	+	+	+	+
Inositol	+	+	+	+	+	+
Proline	+	+	+	+	+	+
Sucrose	+	+	+	+	+	+
Threonine	+	+	+	+	+	+
Valine	+	+	-	+	+	+

PCA was applied also in the water extracts of all samples in order to study the metabolite profiles in these fractions. This resulted in a model with 7 PCs accounting for 94.8% of total variance (R2) with the first three components explaining 81.5%, and 85.4 % of predictive ability (Q2). Combining PC1 (51.3%) and PC2 (20.2%) in the score plot provided separated clusters of the samples as showed in the figure 4A. The cluster position of Bedrocan flowers was close to the position of Bedica flowers indicating similar metabolite profiles. Same results were obtained from analyzing metabolite profiles of trichomes and leaves of both varieties on the score plot. Further examination on the score plot of water extracts revealed that the cluster distance between Bedrocan flowers and Bedica flowers was closer than the distance between Bedrocan flowers and Bedrocan trichomes. This result confirmed that the similarity rate of metabolite profiles between same organs of Bedrocan and Bedica was higher than the similarity within organs of each variety.

The loading plot of the first PC was investigated in order to detect responsible compounds for the classification of metabolite profiles in the water extracts. The investigation revealed that the signals between δ 0.8-4.0 ppm of the ^{1}H NMR contributed to the differentiation as shown

in figure 4B. Signals at δ 1.00, 1.33, 1.48, 2.07 and 2.38, 2.14 and 2.46, 2.35, 2.87 and 2.96, and 3.25 affected the PC1 values. Based on ^{1}H NMR analysis of water extracts, signals were assigned as H-3 of valine, H-5 of threonine, H-3 of alanine, H-4 and H-3 of glutamine, H-4 and H-3 of glutamic acid, H-3 of proline, H-3b and H-3a of asparagine, and H-5 of inositol, respectively as shown in the table 3. While, in the δ 4.0-6.0 ppm, signals at δ 4.06, 4.22, and 5.23 corresponded to H-1 of fructose, H-1' of sucrose, and H-1 of α-glucose, respectively, contributed in the loading plot of PC1 as well. Besides amino acids and sugar compounds that mentioned before, the signals of choline at δ 3.21, and formic acid at δ 8.46 gave also a correlation in the loading plot.

Table 3. Characteristic chemical shifts (signals) of ^{1}H NMR spectra of identified compounds. *: the signal is used for quantification.

Compounds		Chemical shift (ppm) and coupling constants (Hz)
Chloroform fraction	**CBCA**	δ 5.49 (H-7, d, *J* = 10.1), δ 6.23 (H-4, s), δ 6.74* (H-8, d, *J* = 10.1)
	CBGA	δ 1.53 (H-9", s), δ 2.10* (H-2", t, *J* = 7.9)
	THCA	δ 0.90 (H-5', m), δ 1.12 (H-13, s), δ 1.33 (H-3', m), δ 1.33 (H-4', m), δ 1.45 (H-12, s), δ 1.55 (H-2', m), δ 1.69 (H-11, s), δ 2.79 (H-1a', m)and 2.92 (H-1b', m), δ 3.23 (H-10a, brd, *J* = 10.9), δ 6.26 (H-4, s), and δ 6.39* (H-10, s) , δ 12.20 (COOH, s)
	THC	δ 6.13* (H-2, brs), δ 6.28 (H-4, s), δ 6.30 (H-10, brs)
Water methanol fraction	**α-Glucose**	δ 5.24 (H-1, d, *J* = 3.8)
	Alanine	δ 1.48 (H-3, d, *J* = 7.2)
	Asparagine	δ 2.87 (H-3b, dd, *J* = 16.9, 7.6), δ 2.96 (H-3a, dd, *J* = 16.9, 4.3), δ 4.01 (H2, dd, *J* = 16.9, 4.3)
	Choline	δ 3.20 (CH_3, s)
	Formic acid	δ 8.46 (s)
	Fructose	δ 4.06 (H-1, d, *J* = 3.5)
	Glutamic acid	δ 2.14 (H-4, m), δ 2.46 (H-3, m)
	Glutamine	δ 2.07 (H-4, m), δ 2.38 (H-3, m)
	Inositol	δ 3.25 (H-5, t, *J* = 9.3), δ 3.49 (H-1, dd, *J* = 9.9, 2.9), δ 3.61 (H-4 , t, *J* = 9.3)
	Proline	δ 2.35 (H-3, m)
	Sucrose	δ 4.22 (H-1' of fructose moiety, d, *J* = 8.6), δ 5.42 (H-1, d, *J* = 3.8)
	Threonine	δ 1.33 (H-5, d, *J* = 6.6)
	Valine	δ 1.00 (H-3, d, *J* = 6.8), δ 1.05 (H-4, d, *J* = 6.8)

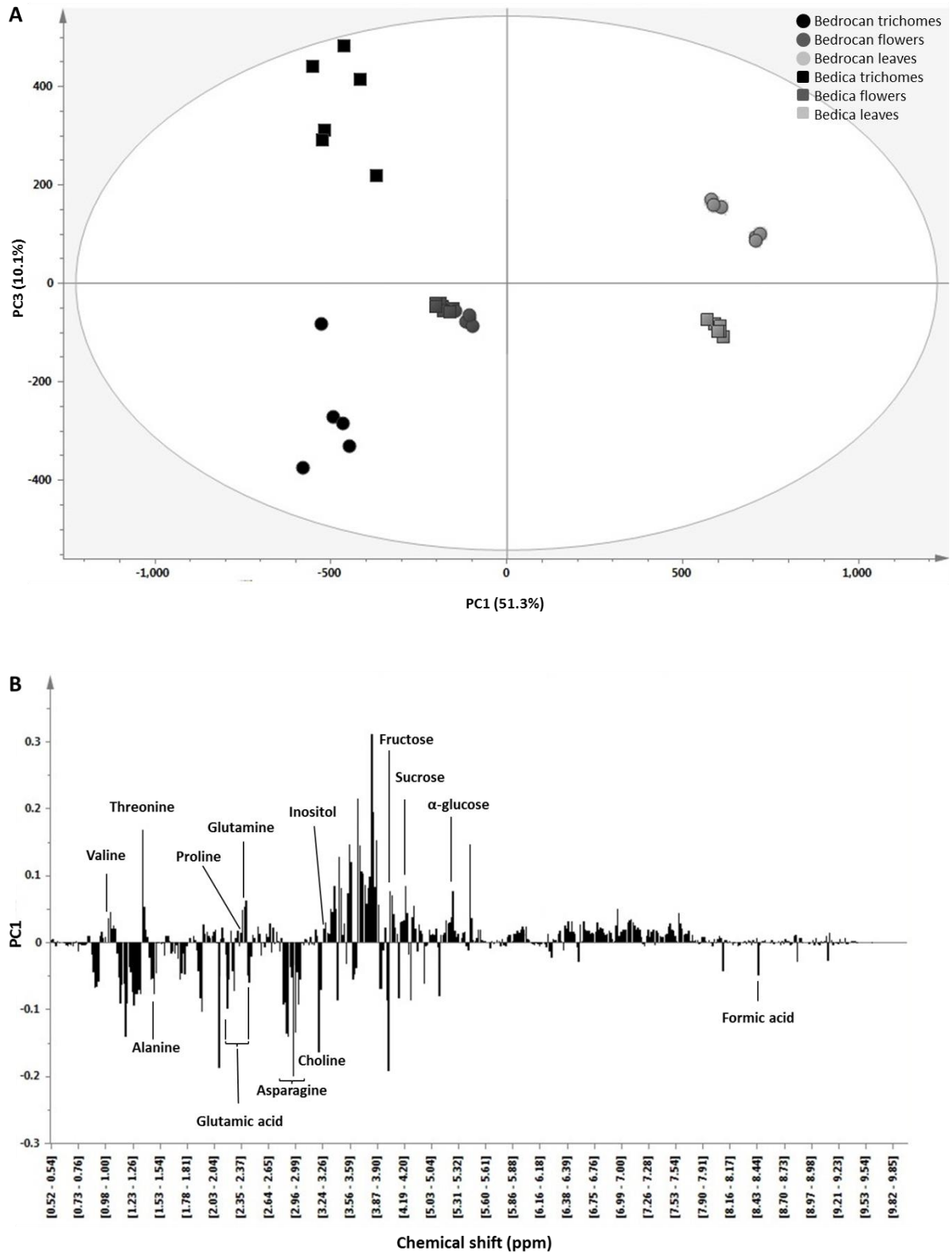


Figure 4. Score plot (A) and loading plot (B) of water extracts of all samples

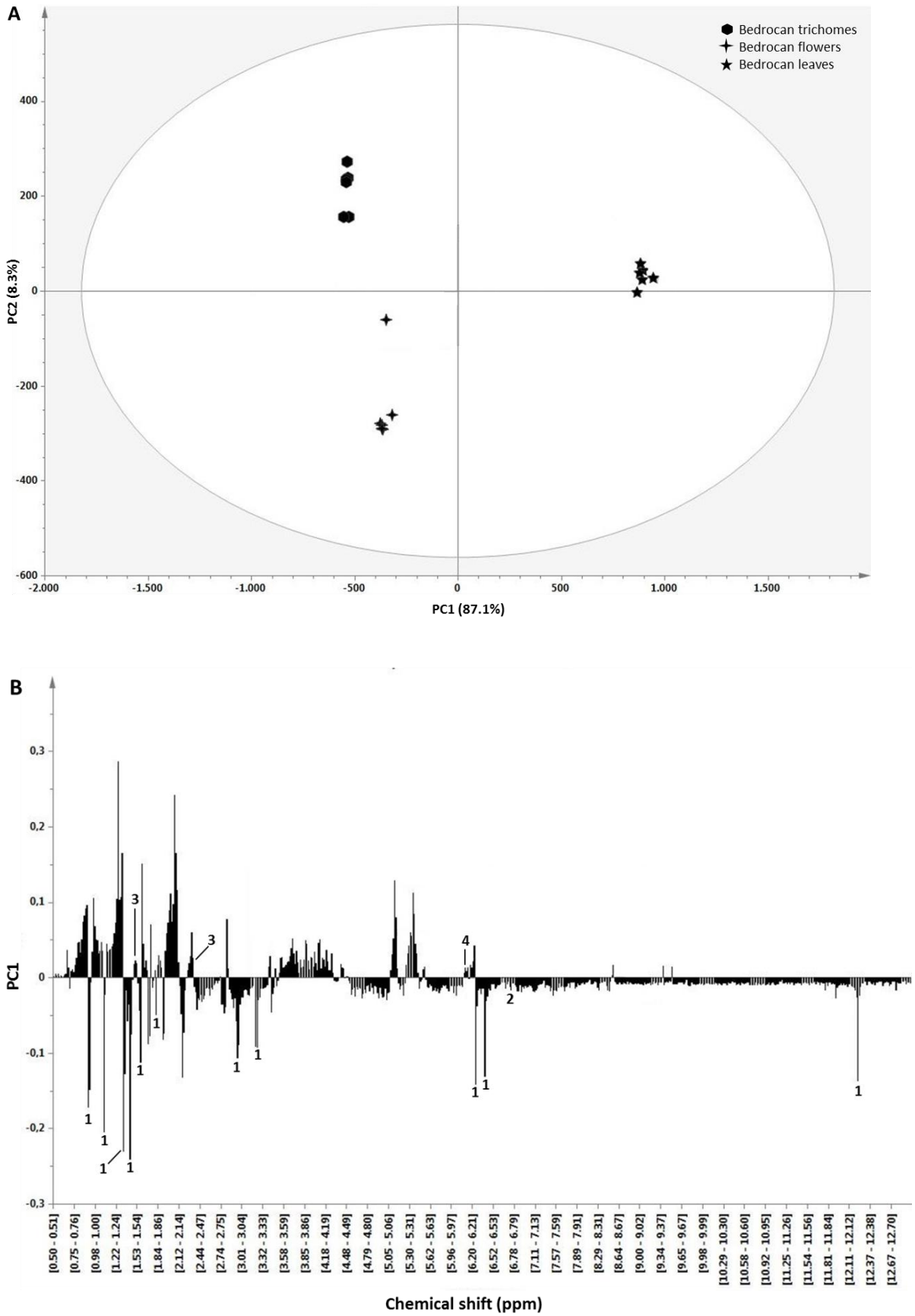


Figure 5. Score plot (A) and loading plot (B) of chloroform extracts of Bedrocan samples

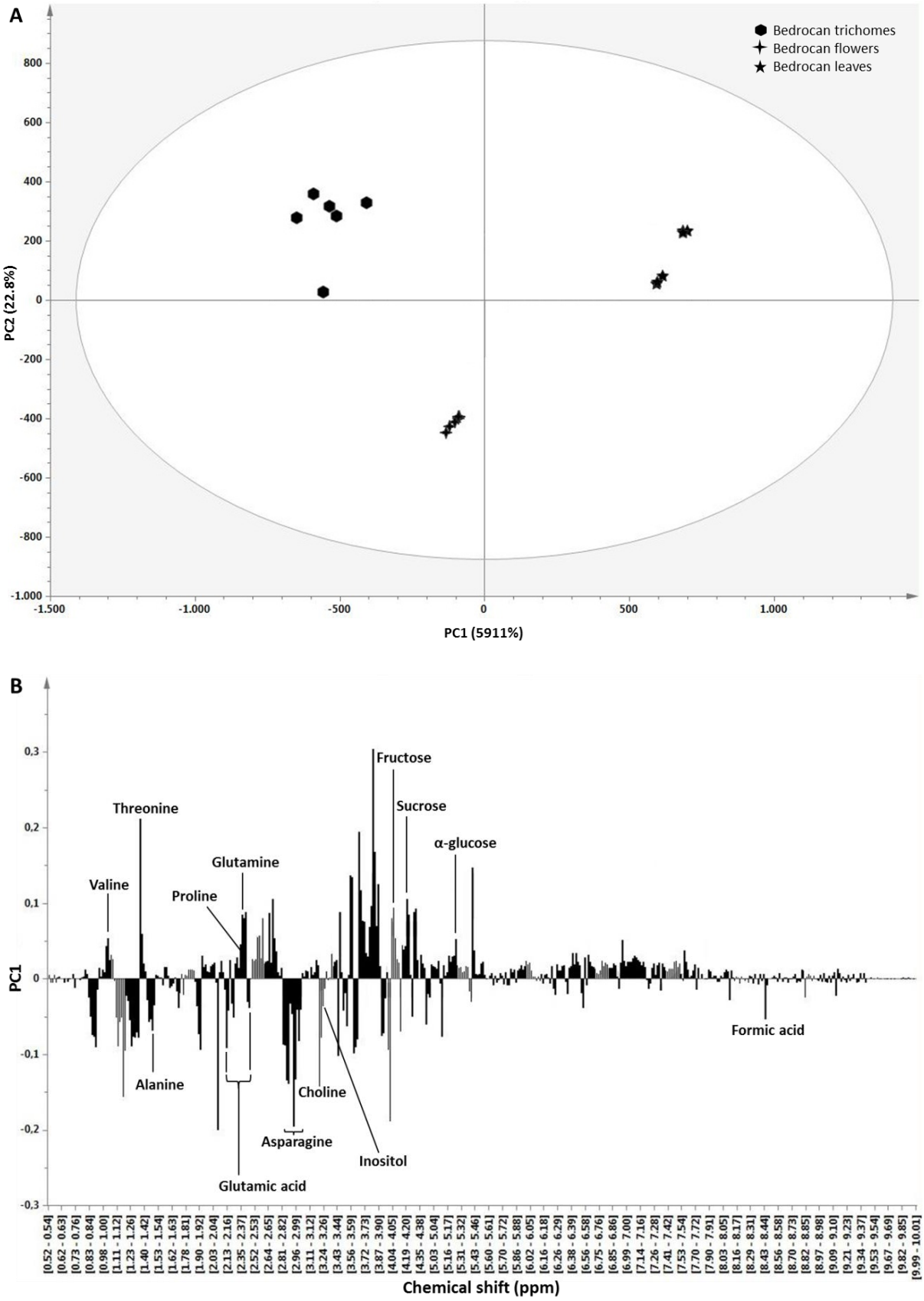


Figure 6. Score plot (A) and loading plot (B) of water extracts of Bedrocan samples

In an up following experiment, metabolomics analysis was applied to differentiate metabolites profiles in trichomes, flowers, and leaves of the variety Bedrocan. PCA was performed on the chloroform extracts of Bedrocan samples. A four component model revealed 99.0% of the variance (R2) and 97.2% of predictive ability (Q2). The combination of PC1 (87.1%) and PC2 (8.2%) showed clearly well-separated clusters based on their organs as showed in the figure 5A. Cluster position of Bedrocan flower was closer to the trichomes cluster than its leaves cluster. It confirmed that the metabolite profile of Bedrocan flowers had more similarity with its trichomes profile compared to the leaves profile. The loading plot of PC1 was used to find the metabolites that are responsible for the separation (figure 5B). As a result, signals of THCA (1) were found as an important compound for the discrimination since its signals contributed greatly in the PC1 loading plot. This was in accordance with the results for quantification of THCA (1) in the Bedrocan samples as seen in the table 2. Besides THCA (1), the signals of CBCA (2), CBGA (3) and THC (4) gave a contribution to PC1 as well.

PCA model of the water extracts also separated metabolite profiles of Bedrocan samples according to their organs. This model possessed 91.7% of total variance (R2) and 85.5% of predictive ability (Q2) using 3 PCs. The score plot that constructed by PC1 (59.1%) and PC2 (22.8%) revealed three well-separated clusters of Bedrocan samples as shown in figure 6A. Investigation on the loading plot of the first PC revealed that some amino acids and sugars contributed in the separation significantly as shown in the figure 6B. Signals of valine, threonine, alanine, glutamine, glutamic acid, proline, and asparagine greatly affected the loading plot of PC1. Furthermore the characteristic signals of sugar compounds, namely sucrose, fructose, and α-glucose were detected contributing in the loading plot. Beside of mentioned primary compounds, the differentiation was due to the signal at δ 3.21 corresponded to choline. Among the discriminant compounds, asparagine was the most important compound in this differentiation since its absence in the ^{1}H NMR spectra of Bedrocan leaves as shown in the figure 7.

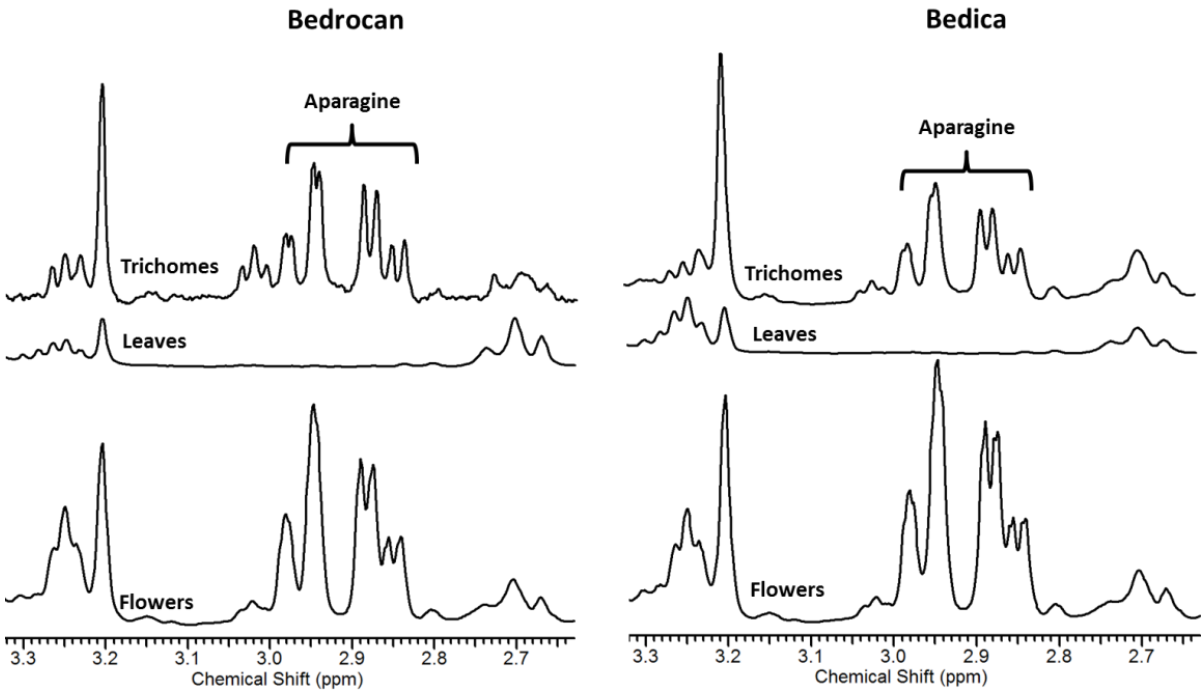


Figure 7. ^{1}H NMR Spectra of water extracts of Bedrocan and Bedica samples. Chemical shifts of asparagine in both leaves could not be detected.

Analyzing of metabolites profiles using metabolomics approach was also performed in the Bedica samples. PCA analysis was applied in the chloroform and water extracts. PCA model of the first extracts showed 98.9 % of the variance (R2) using six PCs with the first three components explaining 94.6%, and 94.8% of predictive ability (Q2). The score plot of PC1 (81.4%) and PC2 (9.5%) were applied successfully to separate samples based on the plant organs as shown in the figure 8A. Based on investigation of PC1 loading plot, THCA (1) was the most important discriminant compound of Bedica samples as it signals affected greatly PC1 values (see figure 8B). Further analysis of PC1 loading plot revealed that CBCA (2), CBGA (3), and THC (4) contributed in this differentiation as well. Only using 2 PCs, PCA model of the water extracts explained 91.4% of total variance (R2) and 84.9% of predictive ability (Q2). Well separated clusters of the samples were shown in the score plot of PC1 (61.2%) and PC2 (24.5%) as documented in figure 9A. The responsible compounds in this separation were detected in the loading plot of PC1 (see figure 9B). Based on this loading plot, valine, threonine, alanine, glutamic acid, asparagine, sucrose, fructose, α-glucose, choline, and formic acid were found as responsible compounds in the differentiation of Bedica samples. The absence of asparagine signals in the ^{1}H NMR spectra of Bedica trichomes lead it as the most important discriminant compound.

Asparagine is an amino acid amide with a molecular mass of 132.12 and an isoelectric of 5.41 (Lea et al. 2007). Some reports investigated accumulation of asparagine in the plants under stress conditions. Asparagine accumulations were reported on *Pennisetum glaucum* under highly dehydrated conditions (Kusaka et al. 2005) and wild *Hordeum* species in salt stress environments (Garthwaite et al. 2005). Furthermore, increasing asparagine amounts were also recorded under mineral deficiencies, such as in soybean and tobacco deprived of phosphorus (Rufty et al. 1993; Rufty et al. 1990), and in the shoots of sulphur-deficient wheat varieties (F. J. Zhao et al. 1996) and Italian rye grass (Mortensen et al. 1993). Moreover, accumulations of asparagine also happened on tomato leaves that infected by *Pseudomonas syringe* (Pe´rez-Garcı´a et al. 1998), and *Theobroma cocoa* which infected with *Crinipellis perniciosa* (Scarpari et al. 2005). Thus these reports indicated that asparagine is involved in the defense and protection system of plants. In other hand, trichomes perform important functions for plants, such as: protection against destructive insects and herbivores (Romero et al. 2008; Van Dam and Hare 1998; Simmons and Gurr 2005); reduction of leaf temperature and increase of light reflectance (G. J. Wagner 1991; G. J. Wagner et al. 2004); prevention of water loss and minimization of leaf abrasion (Gonzalez et al. 2008). Therefore we suggested that the accumulation of asparagine in *Cannabis* trichomes correlated with the role of trichomes in plant defense.

Quantitative RT-PCR experiments were performed in order to analyze expression level of cannabinoids genes, in the different organs of Bedrocan and Bedica. The specific gene primers of THCA synthase (THCAS), predicted-CBGA synthase (predicted-CBGAS), olivetol synthase (OLS), and olivetolic acid cyclase (OAC) involved in cannabinoids biosynthesis were constructed. In Bedrocan samples, the highest expression level of THCAS gene surprisingly was found in the leaves as documented in the figure 12. It was around 3 times higher in the leaves than in the trichomes and the flowers. Different case was obtained on the levels of OAC and OLS in Bedrocan samples. The quantity of OAC gene in the leaves was almost 8 times and 5 times lower than in the flowers and the trichomes, respectively. While, the level of OLS gene in the trichomes was around 5 times higher than in the leaves and slightly higher than in the flowers. The last analyzed transcript, predicted-CBGAS gene, was expressed almost in the same level in all Bedrocan organs.

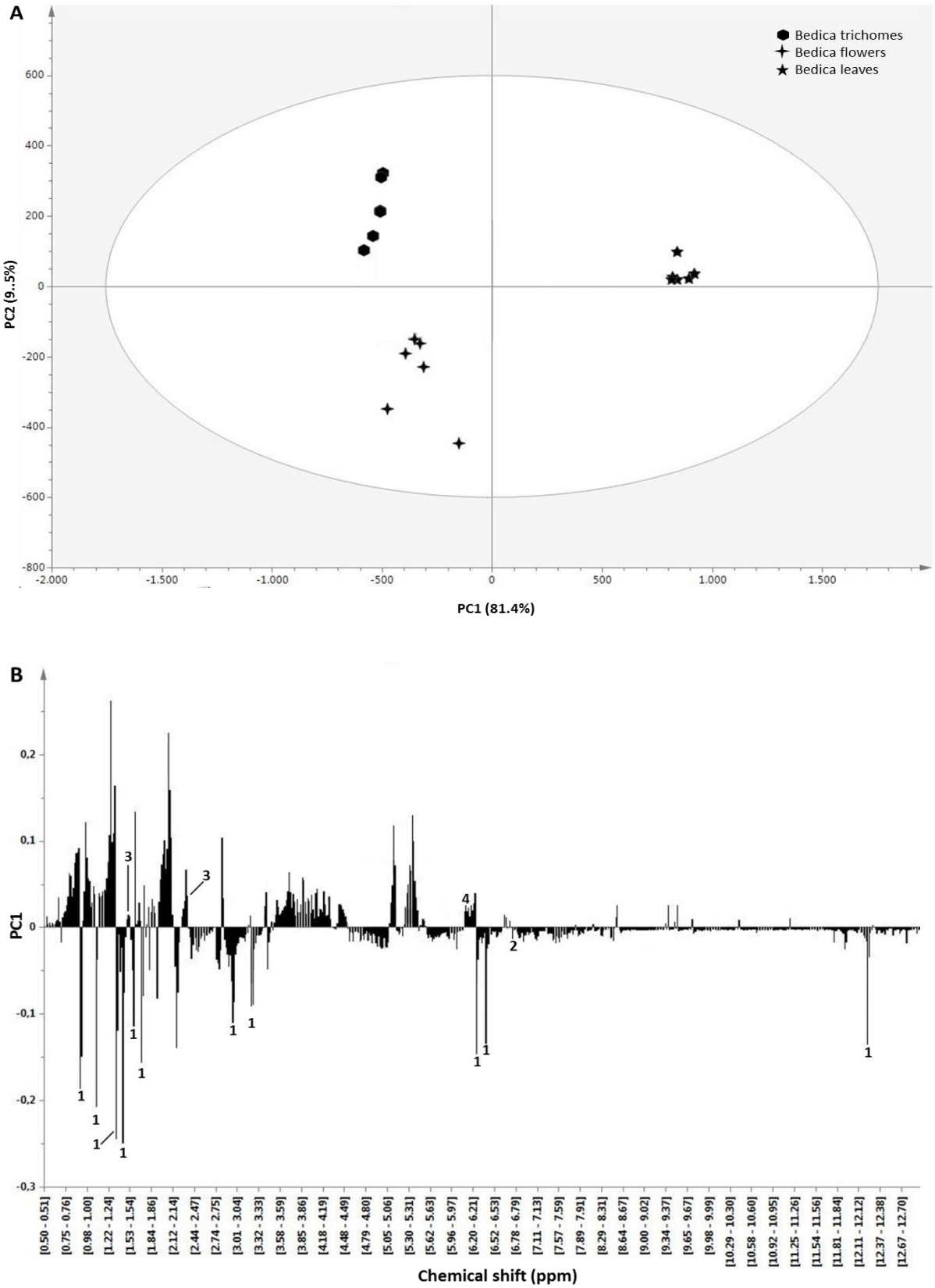


Figure 8. Score plot (A) and loading plot (B) of chloroform extracts of Bedica samples

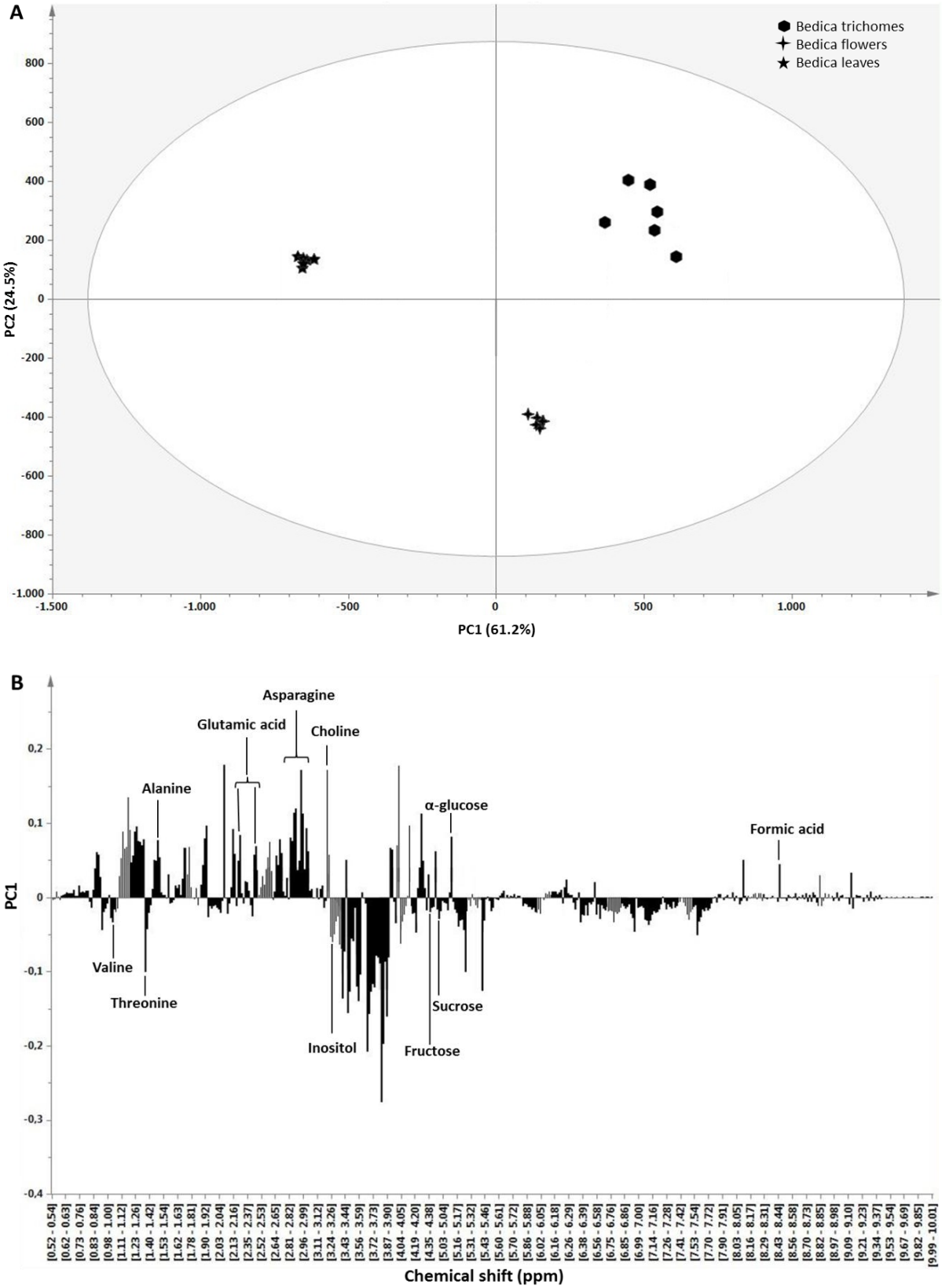


Figure 9. Score plot (A) and loading plot (B) of water extracts of Bedica samples

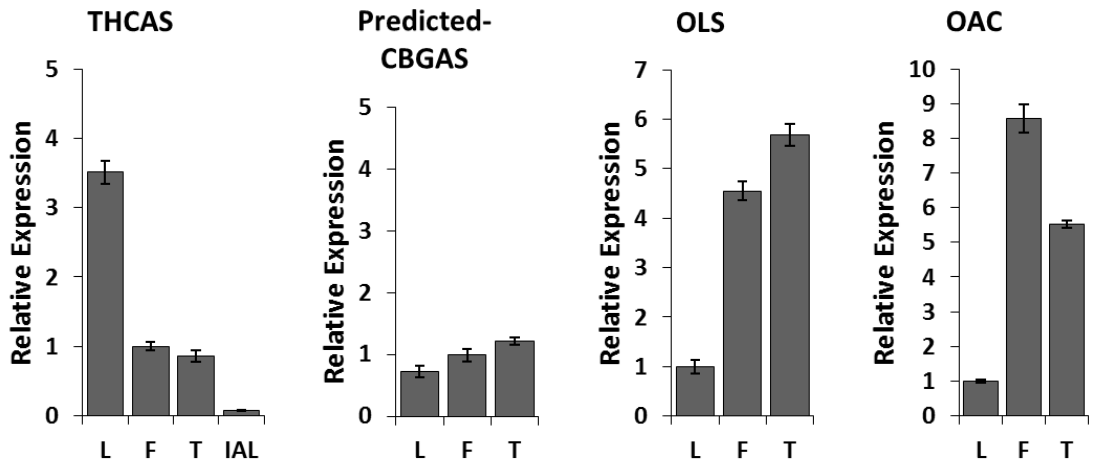


Figure 12. Expression level of cannabinoids genes in the different organs of Bedrocan variety. L: Leave; F: Flowers; T: Trichomes; IAL: inflorescence-associated leaves.

In other medicinal type, Bedica, the level of THCAS gene in the leaves was 6 times higher than in the trichomes and the flowers samples (see figure 13). Contrary results were found in the analysis of other transcripts. The amount of OAC in Bedica leaves was 10 fold and 6 fold lower than in the trichomes and the flowers, respectively. Meanwhile, the amount of OLS gene in the leaves was around 17 times and 8 times lower than in the trichomes and flowers, respectively. However, expression levels ofpredicted-CBGAS gene in the samples did not differ significantly.

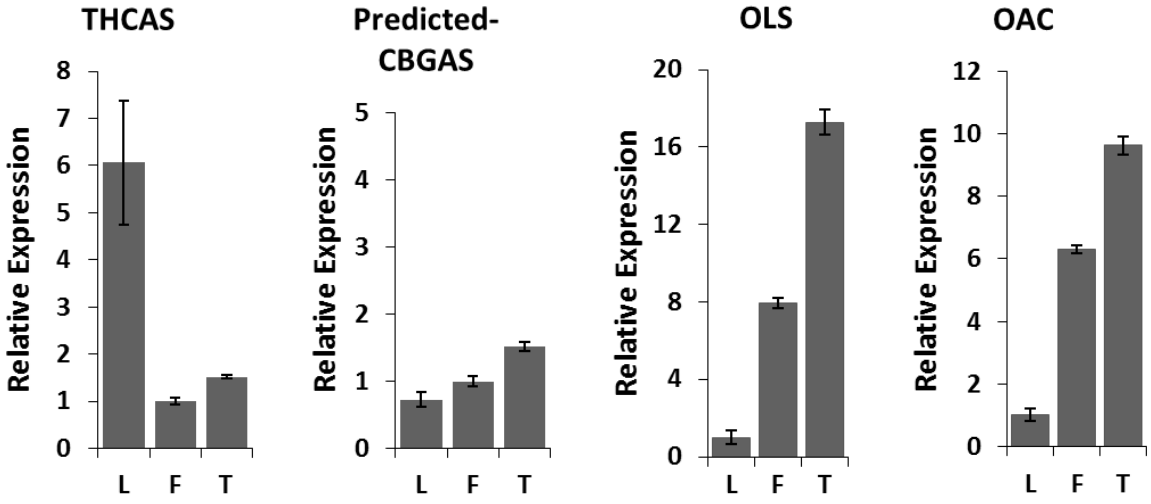


Figure 13. Expression level of cannabinoids genes in the different organs of Bedica variety. L: Leave; F: Flowers; T: Trichomes.

Discovering the highest level of THCAS in leaves of Bedrocan and Bedica was surprise because previous report mentioned that the THCAS was more highly expressed in trichomes than in leaves (Marks et al. 2009a). However, the leaves used in this report were young inflorescence-associated leaves. For confirming this result with our finding, THCAS level in young inflorescence-associated leaves of Bedrocan was analyzed. We found that the THCAS amount in this organ was 11 times lower than in the trichomes (see figure 12), thus the finding was accordance with the previous report (Marks et al. 2009a).

Quantitative analysis showed that THCA (1) was found in a higher concentration in the trichomes than in the leaves as seen in the table 2. However, different results were discovered from analyzing expression level of THCAS, the gene that is responsible for formation of THCA (1), by RT-PCR analysis. This gene was less expressed in the trichomes than in the leaves. These results probably might be a suggestion that there is no positive correlation between the abundance of THCA (1) with its expression level in the trichomes. Previous report revealed that OLS and OAC cooperate in the condensation of malonyl-CoA and Hexaonyl-CoA to form olivetolic acid, a starting compound in THCA (1) biosynthesis (Gagne et al. 2012). RT-PCR analysis showed that both were expressed higher in the trichomes than in the leaves. Thus, an alternative opinion might be proposed, that is the high expression levels of OLS and OAC in the trichomes trigger the increasing of production of olivetolic acid and then lead to the abundance of THCA (1) in the trichomes.

Species classification in genus *Cannabis* is still debatable since different reports support the concepts for a mono or poly species genus. The concept for one species, *C. sativa* L., first was proposed by Linnaeus (Linnaeus 1753) and then supported by Small at al. (Small E. and A. 1976). Lamarck considered that morphology of *Cannabis* plants from India is different from the common plants in Europe, thus he added a new species, *C. indica* Lam. into the genus (de Lamarck 1785). Other botanists divided *Cannabis* into three species, *C. sativa, C. indica* and *C. ruderalis* (Anderson 1974, 1980; Schultes et al. 1974). Furthermore McPartland et. al proposed the fourth species, *C. afghanica* (J. M. McPartland et al. 2000). Based on their morphology, Bedrocan used in this work are assigned as *C. sativa* L., while the other medicinal variety, Bedica, is belonging to *C. indica*. However, the results of metabolomics analysis of different organs of Bedrocan and Bedica revealed that both plants possessed similarity in the term of metabolite profiles and the discriminant compounds, either in the chloroform extracts or water extracts. Furthermore, analysis of RT-PCR indicated that both

had a similar pattern in the term of expression level of cannabinoids genes as described before. Therefore the results of this work seem supporting the concept for one species, namely *C. sativa* L.

4. Conclusions

In this paper, ^{1}H NMR-based metabolomics differentiation together with RT-PCR analysis was performed to analyze metabolite profiles and cannabinoid biosynthesis of different organs of 2 medicinal *Cannabis* varieties, Bedrocan and Bedica. Combination of both methods revealed expression level of cannabinoid genes, metabolite profiles, and discriminant compounds, in trichomes, flowers, and leaves of the plants. Metabolomics analysis revealed that the differences of metabolite profiles within organs of Bedrocan and Bedica were bigger than the differences of same organs between both varieties. THCA (1), CBCA (2), CBGA (3), THC (4), and 15 identified compounds in the water extracts were detected contributing in the classification of metabolite profiles of *Cannabis* samples. Furthermore, THCA (1) was the most important discriminant compounds in the chloroform extracts, while the absence of asparagine signals in the ^{1}H NMR spectra of Bedrocan and Bedica leaves indicated this compound as the leading differentiate compound in the water extracts. We suggest that the high gene expression levels of OLS and OAC in the trichomes of identified plants increase the production of olivetolic acid and then lead to the abundance of THCA (1).

Chapter 4

Monitoring and Differentiation of Metabolites of Medicinal *Cannabis* Trichomes during Flowering Period using ^{1}H NMR-based Metabolomics

Nizar Happyana, Oliver Kayser

Oliver Kayser coordinated and supervised the project and corrected the manuscript

Abstract

Cannabis sativa L. trichomes are glandular structures predominantly responsible for the biosynthesis of cannabinoids-the biologically active constituents of the plant. This study reports on the findings of ^{1}H NMR-based metabolomics of trichomes of 4 medicinal *Cannabis* varieties, Bediol, Bedica, Bedrobinol, and Bedrocan, during the last 4 weeks of the plant flowering period. Analysis of the ^{1}H NMR spectra revealed the presence of 6 cannabinoids (Δ^{9}-tetrahydrocannabinolic acid [THCA (1)], cannabidiolic acid [CBDA (2)], cannabichromenic acid [CBCA (3)], cannabigerolic acid [CBGA (4)], Δ^{9}-tetrahydrocannabinol [THC (5)], and cannabidiol [CBD (6)]) in the chloroform extracts, and 20 non-cannabinoid compounds, including sugars, amino acids, and other acidic constituents, in the aqueous extracts. Subsequent Partial Least Squares Discriminant Analysis (PLSDA) afforded identification of distinct metabolite profiles within trichome varieties. THCA (1) and CBDA (2) constituted the vital differential components of the organic preparations, while asparagine, choline, fructose, and glucose proved their water-extracted counterparts. Application of PLSDA metabolomic models enabled further classification of trichomes of every investigated *Cannabis* variety according to the time of harvest, with THCA (1) identified as the chloroform-extracted discriminant and threonine, asparagine, and fructose, as the water-soluble differential metabolites. The obtained results demonstrated that ^{1}H NMR-based metabolomics have been successfully applied for discriminating metabolite profiles of *Cannabis* trichomes. To the best of our knowledge, this is the first report on time-dependent monitoring of biosynthetic processes in *Cannabis* trichomes during flowering period by means of ^{1}H NMR-based metabolomics.

Keywords: *Cannabis sativa* L., metabolomics, trichomes, cannabinoids, ^{1}H NMR, PLSDA

1. Introduction

Trichomes are small protrusions of epidermal origin found on the surfaces of leaves and other organs of many plants (Schilmiller and others 2008). The morphology of thrichome structures is greatly diverse, with over 300 types reported in various tissues and species (Wagner 1991). Trichomes perform important biological functions, such as: protection against destructive insects and herbivores (Romero and others 2008; Simmons and Gurr 2005; Van Dam and Hare 1998); pollinator attraction (Moyano and others 2003); reduction of leaf temperature and increase of light reflectance (Wagner 1991; Wagner and others 2004); and loss of water prevention and minimization of leaf abrasion (Gonzalez and others 2008).

Although trichomes vary in size, shape, number of cells, and morphology, they are classified into two general categories: non-glandular and glandular trichomes (Schilmiller and others 2008). Non-glandular trichomes are simple outgrowths that do not have glands, and can be found, for example, in *Arabidopsis thaliana* as uni-cellular structures consisting of a stalk and several branches. These trichomes have been studied extensively, providing a model to address general questions in cell and developmental biology (Day and others 2005; Hulskamp 2000; Hulskamp and Kirik 2000; Larkin and others 2003; Marks and others 2009; Schwab and others 2000). Glandular trichomes, on the other hand, arise from a series of anticlinal and periclinal divisions to form supporting auxiliary cells and glands (Wagner 1991). They have a unique capability to synthesize, store, and sometimes secrete large amounts and varied types of chemicals. Various classes of terpenoids and phenylpropanoids (e.g. flavonoids) are commonly produced in glandular trichomes (Wagner and others 2004). Many of these compounds have significant commercial value as pharmaceuticals, fragrances, natural pesticides, and food additives (Duke and others 2000; Wagner 1991; Wagner and others 2004). Therefore, glandular trichomes have been described as chemical factories for the production of high value secondary metabolites of diverse classes (Schilmiller and others 2008; Wagner and others 2004).

One of the valuable compound classes produced in glandular trichomes are cannabinoids (Lanyon and others 1981; Petri and others 1988; Turner and others 1978). These natural products are a unique group of terpenophenolics, possessing alkylresorcinol and monoterpene moieties in their molecular structure, and found only in *Cannabis* spp., a prospecting medicinal plant. Because of their biological activities, Δ^9-tetrahydrocannabinol [THC (5)] and

cannabidiol [CBD (6)] are the most studied compounds within their class. THC, a decarboxylated form of tetrahydrocannabinolic acid [THCA (1)], is a well-established psychoactive component of *Cannabis* spp. (Gaoni and Mechoulam 1964), activating mainly cannabinoid receptors CB1 and CB2 (Matsuda and others 1990; Munro and others 1993). It has been investigated for treatment of nausea and vomiting (Sallan and others 1975), spasticity (Brenneisen and others 1996; Maurer and others 1990), Tourette syndrome (Muller-Vahl and others 2002; Muller-Vahl and others 1999), neuropathic pain (Weber and others 2009), multiple sclerosis (Petro and Ellenberger 1981), pruritus (Neff and others 2002), asthma (Tashkin and others 1977), glaucoma (Merritt and others 1980), migraine (Volfe and others 1985), and appetite stimulation (Beal and others 1995; Beal and others 1997). However, THC is not the only active cannabinoid. CBD, a decarboxylated form of cannabidiolic acid [CBDA (2)], was reported a potent anti-oxidative and anti-inflammatory agent providing neuron protection against acute and chronic degeneration (Hampson and others 1998; Lastres-Becker and others 2005).. Moreover, Mechoulam *et al*. reviewed anticonvulsive, anti-anxiety, antipsychotic, anti-nausea, and antirheumatic properties of CBD (Mechoulam and others 2002).

Cannabis spp. trichomes can be sub-divided into three types, capitate-stalked, capitate-sessile, and bulbous trichomes. Capitate-stalked trichomes, predominantly found on flower surfaces, contain most cannabinoids and generally consist of two parts, the gland (head) and the stem. Our previous work showed that not only the heads but also the stems of capitate-stalked glands play a role in the biosynthesis of cannabinoids (Happyana and others 2012a). Capitate-sessile trichomes, bearing resemblance to hairs, are present on the surfaces of flowers, stems, and leaves. Bulbous trichomes, shaped like balloons, are the smallest of the three types. They are generally made up of two, but never more than four cells (Dayanandan and Kaufman 1976).

As an annual plant, *C. sativa* has two growth periods: vegetative and flowering. In the former, the plant is growing rapidly and producing only low amounts of the relevant metabolites, whereas the flowering stage is the period of accelerated cannabinoid biosynthesis. According to our previous report, cannabinoid production increases with the time of flowering, and reaches the highest level during the last few weeks of the stage (Muntendam and others 2012). On *Cannabis*, flowers contain higher amounts of cannabinoids as compared to the other parts of the plant due to high density of trichomes, especially of the capitate-stalked type. However,

the details of trichome-specific cannabinoid production in *C. sativa* flowers during blooming are still unclear, since, to the best of our knowledge, no systematic investigation, pertaining especially to the late flowering period, has been reported.

Metabolomics, as a recent innovative analytical approach, is broadly used to carry out systematic metabolic investigations. It represents a holistic strategy for qualitative assessment and quantitative measurement of metabolites in the defined samples (species, organ, tissue, etc.) at specific times, and under certain conditions (Happyana and others 2012b). The method has been applied in medicinal plant research, e. g. for the classification and discrimination of *Rhodiola rosea* (Ioset and others 2011), investigation of elicitation effects on suspension cultures of *Silybum marianum* (Sanchez-Sampedro and others 2007), quality control of St. John's wort (Rasmussen and others 2006), bioactivity studies of Xiaoyaosan, a traditional Chinese medicine (TCM) product (Zhou and others 2011), and efficacy investigation of *Epimedium brevicornum* (Li and others 2007). Proton nuclear magnetic resonance spectroscopy (^{1}H NMR) is widely used as an analytical technique in metabolomics. It is reproducible, rapid, simple (as regards sample preparation), non-destructive, and provides primarily quantitative results.

In this study, we report on *C. sativa* trichome metabolomics in order to monitor metabolite profiles-especially cannabinoids, and determine the differential compounds during the flowering period. Our metabolomic investigation was carried out using the ^{1}H NMR technique together with multivariate data analysis. The trichomes of 4 medicinal *C. sativa* varieties: Bedrocan, Bedrobinol, Bediol, and Bedica, were collected and analyzed over the last 4 weeks of the flowering period.

2. Materials and methods

2.1 Chemicals and reagents

First grade chloroform, methanol, deuterated chloroform (99.8%), deuterium oxide (99.8%), tetramethylsilane (TMS), trimethylsilane propionic acid sodium salt (TSP), and sodium deuteroxide were purchased from Carl Roth GmbH (Karlsruhe, Germany). Anthracene was obtained from Sigma-Aldrich Chemie GmbH (München, Germany). Reference compounds of cannabinoids, Δ^9-tetrahydrocannabinol [THC], cannabidiol [CBD], Δ^9-tetrahydrocannabinolic acid [THCA], cannabidiolic acid [CBDA], and cannabigerolic acid [CBGA (4)], were

purchased from THC Pharm GmbH (Frankfurt, Germany) and tested for purity and identity by NMR and LC-MS prior to use.

2.2 Plant material

4 medicinal *C. sativa* varieties, Bedrocan, Bedrobinol, Bediol, and Bedica were supplied by Bedrocan BV (the Netherlands). The plants were grown indoors under standardized conditions as explained in the previous report (Happyana and others 2013). Briefly, they were initially generated from cuttings of standardized mother plants and cultivated under controlled, long day-light conditions (18 h/day). After the vegetative growth phase, the flowering stage was induced under a shorter (12 h/day) light regime for 8 weeks. The trichomes were isolated and analyzed at week 5 to week 8 of the flowering period. Plant specimens were assigned voucher numbers (Bedrocan: N5.26.06.2012, N6.02.07.2012, N7.09.07.2012, N8.16.07.2012; Bedrobinol: B5.26.06.2012, B6.02.07.2012, B7.09.07.2012, B8.16.07.2012; Bediol: D5.26.06.2012, D6.02.07.2012, D7.09.07.2012, D8.16.07.2012; Bedica: C5.26.06.2012, C6.02.07.2012, C7.09.07.2012, C8.16.07.2012) and deposited at the Technical Biochemistry Department, TU Dortmund. All plant handling and experimental procedures were carried out under the license No. 4584989, issued by the Federal Institute for Drugs and Medical Devices (BfArM), Germany.

2.3 Trichome isolation

Trichomes were isolated according to Yerger *et al*. (Yerger and others 1992) with slight modifications. Fresh flowers of *C. sativa* were put into liquid nitrogen. Floral leaves and stigma were removed using forceps, with occasional re-submerging of the flowers in liquid nitrogen. Afterwards, a 5-10 g flower-sample was transferred into a 50 mL falcon tube and placed in liquid nitrogen. The tube was then removed from the liquid nitrogen tank and approximately 2 to 3 cm^3 of finely powdered dry ice (prepared by wrapping a piece in clean paper towels and crushing with a pestle) were added. Subsequently, the tube was loosely capped and vortexed at maximum speed for approximately 1 min, and then the flowers were removed. To obtain the trichomes, the content of each tube was sieved through a nylon net filter with the pore diameter of 140 μm (Merck Millipore, Germany) into a 500 mL glass beaker surrounded by dry ice. The trichomes were subsequently transferred into 2 mL frozen microcentrifuge tubes with a spatula; the samples were placed in liquid nitrogen

2.4 Extraction

200 mg of fresh *C. sativa* trichomes were transferred into a centrifuge tube. 2 mL of 50% aqueous methanol and 2 mL of chloroform were added to the tube followed by vortexing for 1 min and sonication for 1 min. Subsequently, the sample was shaken at 200 rpm for 1 h at 30 ºC. The water and chloroform phases were separated by pipetting and filtering into new centrifuge tubes. The chloroform fraction was dried in a rotary vacuum evaporator at 30 ºC under pressure of 31 mbar. The water fraction was dried in a freeze dryer. In total, 96 of chloroform fraction samples and 96 of water fraction samples were prepared.

2.5 NMR measurements

^{1}H NMR spectra of the samples were recorded using a Bruker Avance DRX 500 spectrometer (Bruker BioSpin GmbH, Rheinstetten, Germany) operating at 500.13 MHz. Dry chloroform extracts were dissolved in deuterated chloroform containing tetramethylsilane (TMS) as an internal standard and anthracene (1 mg/ sample) as a quantitative internal standard. ^{1}H NMR spectra of chloroform extracts were recorded with following parameters: acquisition time = 5.23 s, relaxation delay = 5 s, pulse width = 3 µs, Free Induction Decay (FID) data points = 64 K, spectral width = 12531.32 Hz, and number of scans = 128. Potassium dihydrogen phosphate was added to deuterium oxide as a buffering agent. The pH of the deuterium oxide was adjusted to 6.0 using a 1 N sodium deuteroxide solution. Afterward, water extracts were dissolved in the deuterium oxide containing trimethylsilane propionic acid sodium salt (TSP, 0.01%, w/v) as an internal standard. ^{1}H NMR spectra of water extracts were recorded using a presaturation pulse program with following parameters: acquisition time = 2.72 s, relaxation delay = 2 s, FID data points = 32 K, spectral width = 12019. 23 Hz, number of scans = 128. Processing of FIDs was performed with line broadening set to 1.0 Hz using ACD/Labs 12.0 software. All ^{1}H NMR spectra were manually phased and baseline corrected.

2.6 Data analysis and quantification

For the chloroform extracts, ^{1}H NMR spectra were scaled to TMS and then reduced to integrated regions of equal width (0.02 ppm), corresponding to the region of δ 0.50 - 13.00 ppm. The region δ 7.24 - 7.27 ppm was excluded from the analysis because of the residual signal of chloroform. The region of anthracene peaks, namely δ 7.44 - 7.48 ppm, δ 7.96 - 8.06 ppm, and δ 8.42 - 8.48 ppm, was removed as well. ^{1}H NMR spectra of the water extracts were scaled to TSP and reduced to integrated regions of equal width (0.02 ppm) within δ 0.50 -

10.00 ppm. The regions of δ 3.28 - 3.37 ppm and δ 4.50 - 5.00 ppm were removed from the analysis due to the residual signals of methanol and water. Bucketing was performed by ACD/Labs 12.0 software with scaling on total intensity. Partial Least Squares Discriminant Analysis (PLSDA) was performed with SIMCA-P software (v. 13.0, Umetrics, Umeå, Sweden). PLSDA scaling was based on the Pareto method. Metabolites giving individual signals were quantified. For the identified cannabinoids, quantification was performed by the relative ratio of the peak area of selected proton signals of the target cannabinoids to the singlet peak area of anthracene. The relative quantification of the metabolites identified in water extracts was performed by measuring the ^{1}H NMR signal area of corresponding signals and comparing them to the TSP signal. Analysis of variance (ANOVA) for relevant ^{1}H NMR signals was performed using Microsoft Excel 2010.

2.7 RNA isolation and real time PCR

Fresh *Cannabis* trichomes were ground to a fine powder using a pestle and a mortar under a cold condition (with adding liquid nitrogen). Afterward, around 90 mg of fresh fine powder of plant materials was prepared for RNA isolation using RNeasy™ Plant Mini Kit (Qiagen N.V., The Netherlands). The isolation procedure followed manufacturer's instructions. Concentration of isolated RNA was determined with a Qubit® RNA Assay Kit (Life Technologies, California, USA). Equivalent quantities of RNA isolated from leaves, flowers, and *Cannabis* trichomes were reverse transcribed using iScript™ cDNA synthesis kit (BIO-RAD Laboratories, California, USA) according to the manufacturer's instructions. Quantitative real-time PCR experiments were performed with Applied Biosystems StepOnePlus™ Real-Time PCR Systems (Applied Biosystem, Foster City, Canada) and SYBR® Green PCR Master Mix (Applied Biosystem). These experiments were carried out using comparative CT method (ΔΔCT) (Livak and Schmittgen 2001). S18 housekeeping gene was used as an endogenous control gene. Gene sequences of THCAS (Sirikantaramas et al. 2004), OLS (Taura et al. 2009), and OAC (Gagne et al. 2012) were used as the templates for designing primers using Primers Express Software v.3.0.1 (Applied Biosystem). The list of primers could be seen in the table 1. Real-time PCR was run with following conditions: pre-incubation at 95 ºC for 10 min, followed by 40 cycles of 95 ºC for 15 s and 60 ºC for 1 min, and then melt curve stage at 95 ºC for 15 s, 60 ºC for 1 min, and 95 ºC for 15 min.

Table 1. Primers of cannabinoids genes for RT-PCR analysis.

Gene	Forward Primer	Reverse Primer
18S	GAGAAACGGCTACCACATCCA	CCGTGTCAGGATTGGGTAATTT
THCAS	AAGTTGGCTTGCAGATTCGAA	TGTAGGACATACCCTCAGCATCA
Predicted-CBGAS	TTGGGAAGGCATGTTGGAA	AATCCACAAGCGCATGAAGTAA
OLS	GGGCTGCTGCGGTGATT	TATCGGCCTTTCCCCAACT
OAC	TGTTGGATTTGGAGATGTCTATCG	TTCGTGGTGTGTAGTCAAAAATGA

3. Results and discussion

3.1 Metabolite identification

In order to identify metabolites in the investigated *C. sativa* trichomes, all samples were extracted and grouped into two fractions, namely chloroform and water extracts. Metabolites specific to trichomes of all tested varieties (Bedrocan, Bedrobinol, Bediol, and Bedica) were determined by detecting their fingerprint signals in the ^{1}H NMR spectra and confirmed with reference spectra and literature data. Visual inspection of the obtained spectra showed that cannabinoids were most abundant in the chloroform extracts of all trichome samples. It should be noted that the ^{1}H NMR spectra characteristic of Bediol clearly differed from those obtained for the other varieties (Fig. 1).

For identification of cannabinoids, the ^{1}H NMR spectra of all chloroform extracts were investigated. 4 acidic and 2 neutral cannabinoids were successfully detected and confirmed (Table 1). The signals for all protons of THCA (1) were clearly distinguishable in all ^{1}H NMR spectra, as shown in the table 1, thus, pointing to it being the predominant component of all trichome samples. Other cannabinoids identified in all trichome samples were CBCA (3) and THC (5). The signals of H-4, H-7, and H-8, CBCA (3) were detected at δ 6.23 (s), δ 5.49 (d, J = 10.1), and δ 6.74 (d, J = 10.1), respectively; they were further confirmed based on comparison with published data (Lee and Wang 2005). Signals at δ 6.13 (brs), δ 6.28 (s), and δ 6.30 (brs) corresponded to the signals of H-2, H-4, and H-10, of THC (5), respectively. However, the signal areas of CBCA (3) and THC (5) in all ^{1}H NMR spectra were small, indicating low concentration values of both compounds in the trichomes. The signals of CBGA (4), an important precursor of other cannabinoids, were detected only in the ^{1}H NMR spectra of Bedrocan and Bedica trichomes at δ 1.53 (H-9”, s), and δ 2.10 (H-4”, t, J= 7.9). The fingerprint signals of CBDA (2) were detected only in Bediol trichomes as shown in the Fig. 2

and table 1. Moreover CBD (6), as a neutral form of CBDA (2), was identified solely in Bediol trichomes as well. The signals of H-2, H-12, and H-13*trans* of CBD (6) were detected at δ 6.18 (brs), δ 1.63 (d, J = 10.1), and δ 4.66 (brs). Identification of CBDA (2) and CBD (6) were in accordance with the profile of Bediol plant which contents THCA (1) and CBDA (2). All assigned signals of thus identified cannabinoids were further confirmed by analysis of 2D NMR spectra, including COSY, HSQC, and HMBC, and compared with the signals of reference cannabinoids. The chemical structures of the identified cannabinoids are depicted in figure 3.

Further visual inspection of the obtained ^{1}H NMR spectra indicated that water extracts of all samples contained amino acids and carbohydrates (see supplementary data 1, Figure 6). The fingerprint signals of alanine, asparagine, glutamine, glutamic acid, glycine, leucine, proline, threonine, and valine were detected within the δ 0.8 - 4.0 ppm region of the spectra, while in the δ 4.0 - 6.0 ppm region, the anomeric signals of α-glucose, β-glucose, β-mannose, fructose, and sucrose were recorded. Moreover, a sugar alcohol, inositol was found in all samples. Other organic compounds characteristic of the investigated aqueous extracts included: acetic acid, formic acid, fumaric acid, succinic acid, and choline. All the distinguishing signals of the identified water-soluble compounds were confirmed with literature (Abreu and others 2011; Ali and others 2011; Broyart and others 2010; Choi and others 2004; Kirk and others 2012; Liu and others 2010; Nuringtyas and others 2012; Zhi and others 2012) and can be found in table 2.

3.2 Quantitative analysis

To quantitatively evaluate cannabinoid production in trichomes of the investigated medicinal *C. sativa* varieties over the last 4 weeks of the flowering period, the obtained ^{1}H NMR data were further processed according to a technical report (Hazekamp and others 2004). The proton signals recorded for the chloroform extracts in the range of δ 2.0 - 7.0 ppm were selected for quantification to ascertain optimal discrimination of resonance patterns indicating best selectivity. Anthracene was used as an internal standard, since it has a simple ^{1}H NMR spectrum consisting of a singlet (δ 8.43) and two quartets (δ 8.01, δ 7.48), and does not overlap with the signals characteristic of cannabinoids. The quantification was conducted by calculating the relative ratio of the peak area of selected proton signals of the target cannabinoids to the singlet peak of anthracene.

In Bedica trichomes, 4 cannabinoids, including THCA (1), CBCA (3), CBGA (4), and THC (5), were quantified (Figs. 4A-B). The amount of THCA (1) increased from week 5 till week 7, and then decreased in week 8. However, the concentration of its neutral derivative, THC (5) increased steadily during the entire monitoring period. In contrast, a continuous decline in the concentration of CBCA (5) and CBGA (4) was observed from week 5 till week 8 of the flowering period.

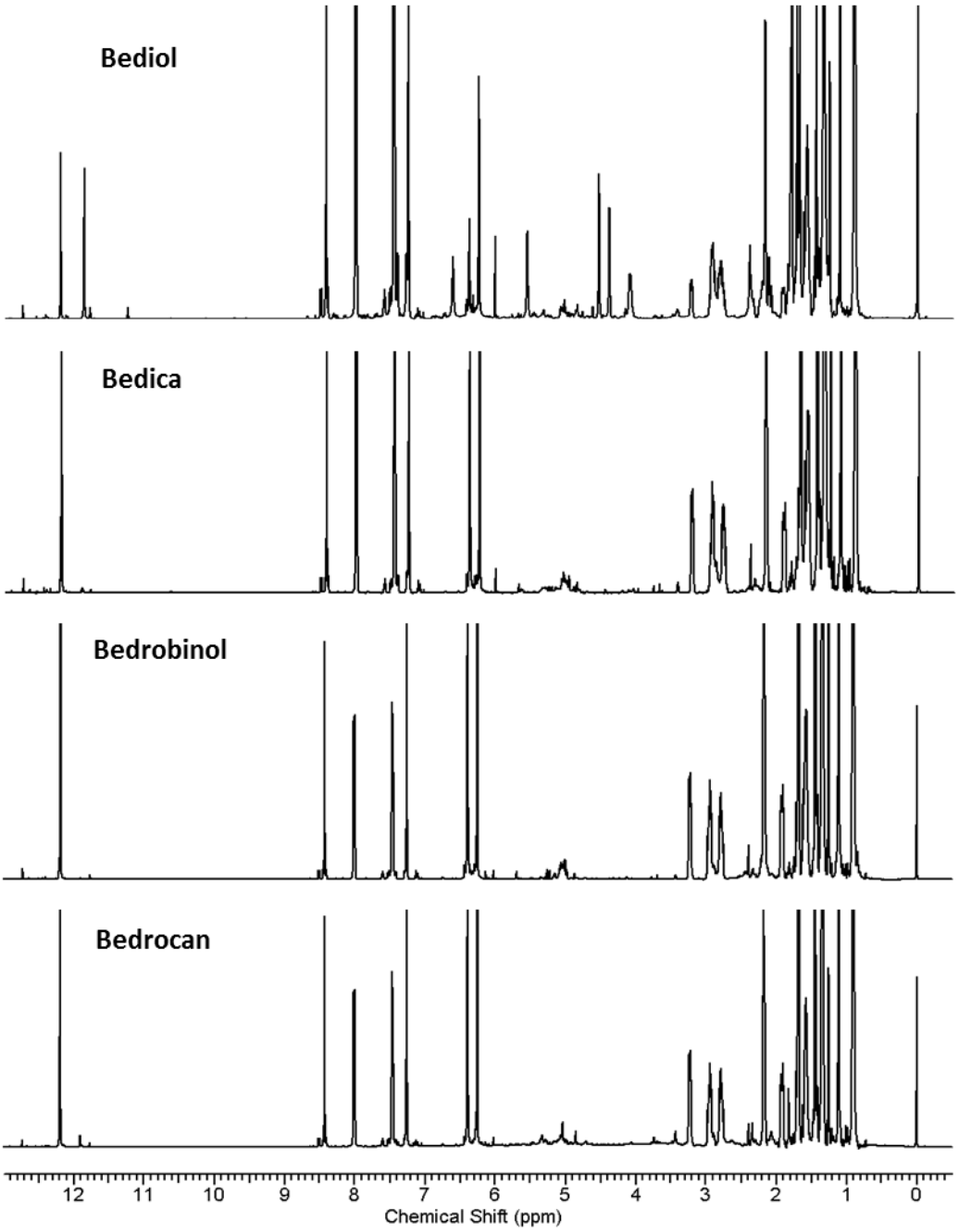


Figure 1. ^{1}H NMR spectra of the chloroform extracts of *C. sativa* trichomes.

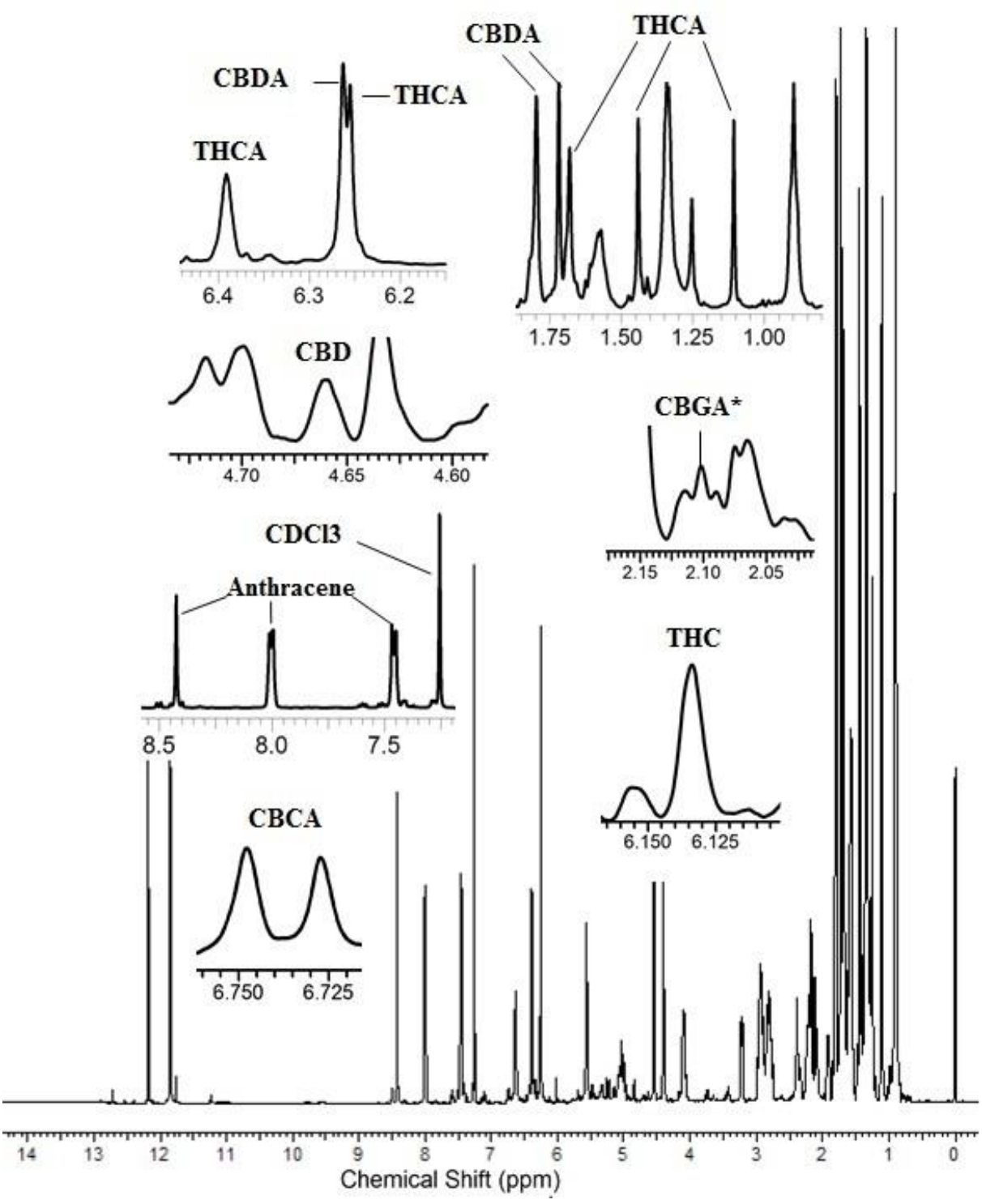
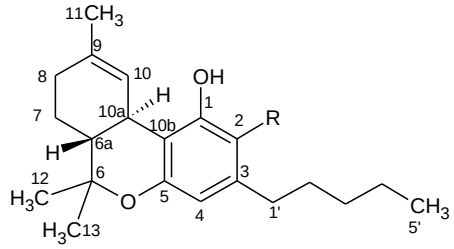
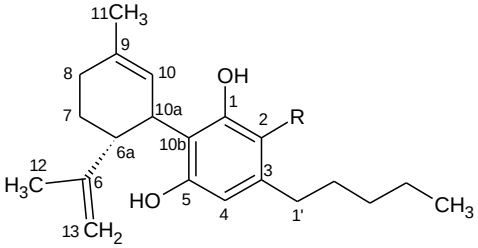


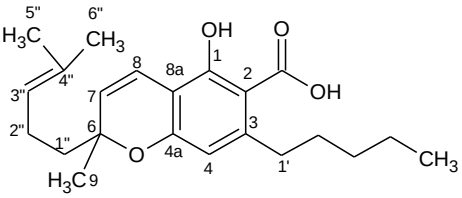
Figure 2. Characteristics signals of cannabinoids identified in the ^{1}H NMR spectra of Bediol trichome chloroform extracts. *: CBGA signal was collected from the ^{1}H NMR spectra of Bedrocan trichome extracts.



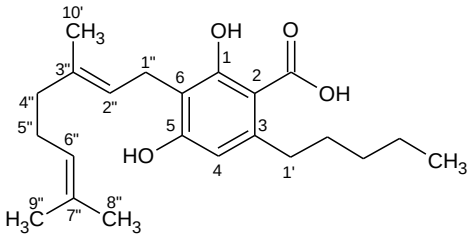
R = COOH, Tetrahydrocannabinolic acid, THCA (1)
R = H, Tetrahydrocannabinol, THC (5)



R = COOH, Cannabidiolic acid, CBDA (2)
R = H, Cannabidiol, CBD (6)



Cannabichromenic acid, CBCA (3)



Cannabigerolic acid, CBGA (4)

Figure 3. Chemical structures of identified cannabinoids.

THCA (1), CBDA (2), CBCA (3), THC (5), and CBD (6) were quantified in the Bediol trichome samples, as shown in figure 4A-B. The concentrations of THCA (1) and CBDA (2) increased during the monitoring time. The same was observed for CBD (6). Production of CBCA (3) increased from week 5 till week 6 and then it was steady in the following weeks. From week 5 till week 7, the production of THC (5) rose steadily, but decreased in the last week of the flowering period.

The concentrations of THCA (1), CBCA (3), and THC (5) were successfully determined in Bedrobinol trichomes (figure 4A-B). THCA (1) production was steady in week 5 and week 6, and then increased at week 7 and week 8. The concentration values recorded for CBCA (3) did not differ significantly during the monitoring time, while the concentration of THC (5) increased during the last 4 weeks of the flowering period.

THCA (1), CBDA (2), CBCA (3), THC (5), and CBD (6) were quantified in the Bediol trichome samples, as shown in Fig. 4A-B. The concentrations of THCA (1) and CBDA (2) increased during the monitoring time. The same was observed for CBD (6). Production of CBCA (3) increased from week 5 till week 6 and then it was steady in the following weeks. From week 5 till week 7, the production of THC (5) rose steadily, but decreased in the last week of the flowering period.

The concentrations of THCA (1), CBCA (3), and THC (5) were successfully determined in Bedrobinol trichomes (Fig. 4A-B). THCA (1) production was steady in week 5 and week 6, and then increased at week 7 and week 8. The concentration values recorded for CBCA (3) did not differ significantly during the monitoring time, while the concentration of THC (5) increased during the last 4 weeks of the flowering period.

4 cannabinoids, THCA (1), CBCA (3), CBGA (4), and THC (5), were quantified in the trichome samples of the final among the investigated *C. sativa* variety, Bedrocan (Fig. 4A-B). The production of THCA (1) increased from week 5 to week 7 and decreased significantly in week 8. The same pattern was observed for the concentrations of CBGA (4) and THC (5). However the production of CBCA (3) did not differ significantly during the flowering period.

Quantification analysis showed that THCA (1) in all trichomes samples was recorded in a high concentration, for example its concentration in Bedrocan trichomes at week 7 was 100 mg/g of fresh trichomes weight. Therefore it confirmed that this cannabinoid is a major compound in *Cannabis* trichomes. This result is in accordance with previous reports (Petri and others 1988; Turner and others 1978). Based on quantification analysis, CBDA (2) was only found in a high amount in Bediol trichomes and established as another major compound on these samples. Furthermore, the concentrations of CBDA (2) were slightly more pronounced than those of THCA (1) as shown in Fig. 4A. As another quantified compound in all trichomes samples, CBCA (3) was noted in low amounts as described in the Fig. 4B. Thus it confirmed CBCA (3) as a minor cannabinoid produced in the trichomes.

Acidic cannabinoids, such as THCA (1) and CBDA (2), are decarboxylated into neutral cannabinoids in the living plant. Furthermore the decarboxylation rate can be affected by harvesting, sample preparation, and storage (Lewis and Turner 1978). Cannabinol (CBN) is an artifact of THC (4) and can be an indicator of the degradation process of cannabinoids (Flemming and others 2007). We were concerned if our experimental methods such as trichomes isolation, extraction, and other sample preparation steps would enforce decarboxylation of the acids. Nevertheless, according to the results of quantification analysis, the concentrations of THC (5) and CBD (6) were lower than those of their acidic counterparts as seen in the Fig 4A-B. Moreover, CBN as the indicator of degradation were not identified. Therefore, we conclude that the methods do not influence the decarboxylation of cannabinoids.

Cannabinoid biosynthesis is started with formation of olivetolic acid from condensation of malonyl-CoA and hexanoyl-CoA by olivetolic acid cyclase and olivetol synthase (Gagne and others 2012), and formation of geranyldiphospahte (GPP), predominantly originating from non-mevalonate pathway (Fellermeier and others 2001). Olivetolic acid and GPP are condensed to form CBGA (4) by an enzyme predicted to be a representative of the geranyltransferase group (Fellermeier and Zenk 1998). CBGA (3) is subsequently transformed to THCA (1), CBDA (2) and CBCA (3) by THCA synthase (Taura and others 1995), CBDA synthase (Taura and others 1996), and CBCA synthase (Morimoto and others 1998), respectively. This cannabinoid biosynthesis shows that CBGA (4) is the key precursor of THCA (1), CBDA (2), and CBCA (3). Besides that it seems logic to assume that increasing of CBGA (4) concentration will increase the production of its final products, especially THCA (1). However, in this research, CBGA (4) could be only detected and quantified in Bedrocan and Bedica trichomes. Thus correlation of THCA (1) production and the amount of its precursor only was studied in both varieties. In Bedrocan trichomes, concentrations of THCA (1) and CBGA (4) increased from week 5 till week 7 and then decreased at week 8 which seems indicating a positive correlation. Nevertheless different case found in Bedica samples, CBGA (4) amount tended to decline during monitoring time although THCA production increased from week 5 to week 7 and decreased at week 8. These results probably showed an unknown complex regulation of cannabinoid production in the trichomes. Besides that, it indicated that further investigations are still needed to shed more light on cannabinoid production in *Cannabis* trichomes during flowering period.

Relative metabolite quantification utilizing the obtained ^{1}H NMR data was carried out for the water extracts of all trichome samples, with trimethylsilane propionic acid (TSP) as an internal standard. Due to extensive signal overlapping, the quantification of all identified compounds was impossible. Nevertheless, alanine, asparagine, glutamine, glutamic acid, threonine, valine, α-glucose, fructose, sucrose, inositol, choline, and formic acid were successfully quantified in all samples investigated within the monitoring period. The recorded production patterns of all the quantified water-soluble compounds are shown in Fig. 4C-D (supplementary data1).

3.3 Partial Least Squares Discriminant Analysis (PLSDA)

Multivariate analysis (MVA) methods were applied for in-depth profiling of metabolite production processes in the trichomes of diverse *C. sativa* varieties over the last 4 weeks of

their flowering period. Since unsupervised multivariate method (principal component analysis) could not provide enough separations (data not shown), PLSDA (a supervised approach) was selected as the principal investigative procedure. It uses a discrete class matrix and is based on the partial least squares (PLS) model, in which the dependent variable is chosen to represent class membership (Westerhuis and others 2008).

Each type of extracts of all trichome samples were divided into 4 classes based on plant variety and analyzed with PLSDA. The percent of the response variation explained by PLSDA model (R^2X and R^2Y), and the percent of the response variation predicted by the model according to cross validation (Q^2) were calculated. The models were validated with the permutation test applying 300 permutations. Based on the obtained results (supplementary data 2), all Q^2 values of the permuted Y vectors were lower than the original ones and the regressions of Q^2 lines intersected the y-axis at points below zero, thus confirming the validity of the applied PLSDA models.

The analytical procedure was initiated by metabolomic differentiation, in order to reveal the distinguishing trichome metabolites characteristic of the varieties Bedica, Bediol, Bedrobinol, and Bedrocan. As shown in Fig. 5A1, PLSDA model of the chloroform extracts of all the investigated samples discriminated trichomes based on the variety of their plant of origin. This model possessed 6 PLS components with R^2X = 97.6%, R^2Y = 96.1%, and Q^2 = 94.8%. The explained variation values (R^2X and R^2Y) and predictive ability (Q^2) in this model are high, indicating that it is a good model. The combination of components PLS1 (89.0%) and PLS2 (2.4%) resulted in clear score plot definition of well separated clusters, corresponding to the trichomes harvested from the distinctive *C. sativa* varieties. The position of the Bediol-cluster on the score plot was significantly divergent from those of the other clusters, indicating the unique metabolic profile of the trichomes of this variety. Subsequent investigation of the PLS1 loading plot indicated that CBDA (2) was the crucial factor, separating Bediol trichomes from the trichomes of other varieties (Fig. 5A2). The signals at δ 6.26, δ 5.57, δ 4.55 and δ 4.40, δ 1.80, δ 1.72, and δ 11.85, assigned as H-4, H-10, H-13 (*trans* and *cis*), H-11, H-12, and H-hydroxyl of carboxylic moiety of CBDA (2), were the PLS-affecting ones. THCA (1) was another important discriminant compound, as it proved to be the discriminant between the varieties of Bedica, Bedrobinol, and Bedrocan. The chemical shifts of THCA (1) at δ 6.40 (H-10), δ 6.25 (H-4), δ 3.23 (H-10a), δ 1.69 (H-11), δ 1.45 (H-12), δ 1.12 (H-13), and δ 12.20 (COOH) were identified as discrimination-contributing signals (Fig. 5A2).

Table 2. Characteristic ^{1}H NMR chemical shifts of the identified compounds. NQ, not quantified; **+,** detected; **-,** not detected; *, signal used for quantification.

Compounds		Chemical shift (ppm) and coupling constants (Hz)	Bedrocan	Bedrobinol	Bediol	Bedica
Chloroform fraction	CBCA	δ 5.49 (H-7, d, *J* = 10.1), δ 6.23 (H-4, s), δ 6.74* (H-8, d, *J* = 10.1)	+	+	+	+
	CBDA	δ 0.90 (H-5', m), δ 1.33 (H-3', m), δ 1.33 (H-4', m), δ 1.55 (H-2', m), δ 1.72 (H-12, s), δ 1.80 (H-11, s), δ 2.39 (H-6a, m), δ 2.79 (H-1a', m)and 2.92 (H-1b', m), δ 4.10 (H-10a, s), δ 4.40 (H-13cis, s), δ 4,55 (H-13trans, s), δ 5.57* (H-10, s), δ 6.26 (H-4, s), δ 11.85 (COOH, s)	-	-	+	-
	CBD	δ 1.63 (H-12, s), δ 4.66* (H-13trans, brs), δ 6.18 (H-2, brs)	-	-	+	-
	CBGA	δ 1.53 (H-9", s), δ 2.10* (H-4", t, *J* = 7.9)	+	-	-	+
	THCA	δ 0.90 (H-5', m), δ 1.12 (H-13, s), δ 1.33 (H-3', m), δ 1.33 (H-4', m), δ 1.45 (H-12, s), δ 1.55 (H-2', m), δ 1.69 (H-11, s), δ 2.79 (H-1a', m)and 2.92 (H-1b', m), δ 3.23 (H-10a, brd, *J* = 10.9), δ 6.25 (H-4, s), and δ 6.39* (H-10, s) , δ 12.20 (COOH, s)	+	+	+	+
	THC	δ 6.13* (H-2, brs), δ 6.28 (H-4, s), δ 6.30 (H-10, brs)	+	+	+	+
Water fraction	α-Glucose	δ 5.23* (H-1, d, *J* = 3.8)	+	+	+	+
	β-Glucose	δ 4.64 (H-1, d, *J* = 7.9)	+, NQ	+, NQ	+, NQ	+, NQ
	β-Mannose	δ 4.99 (H-1, d, *J* = 7.9)	+, NQ	+, NQ	+, NQ	+, NQ
	Acetic acid	δ 1.91 (s)	+, NQ	+, NQ	+, NQ	+, NQ
	Alanine	δ 1.48* (H-3, d, *J* = 7.2)	+	+	+	+
	Asparagine	δ 2.87* (H-3b, dd, *J* = 16.9, 7.6), δ 2.96 (H-3a, dd, *J* = 16.9, 4.3), δ 4.01 (H2, dd, *J* = 16.9, 4.3)	+	+	+	+
	Choline	δ 3.21* (CH_3, s)	+	+	+	+
	Formic acid	δ 8.46* (s)	+	+	+	+
	Fructose	δ 4.06* (H-1, d, *J* = 3.5)	+	+	+	+
	Fumaric acid	δ 6.59 (s)	+, NQ	+, NQ	+, NQ	+, NQ
	Glutamic acid	δ 2.14 (H-4, m), δ 2.46* (H-3, m)	+	+	+	+
	Glutamine	δ 2.07* (H-4, m), δ 2.38 (H-3, m)	+	+	+	+
	Glysine	δ 3.58 (H-2, s)	+, NQ	+, NQ	+, NQ	+, NQ
	Inositol	δ 3.25 (H-5, t, *J* = 9.3), δ 3.49 (H-1, dd, *J* = 9.9, 2.9), δ 3.61* (H-4 , t, *J* = 9.3)	+	+	+	+
	Leucine	δ 0.95 (H-5, d, *J* = 6.6), 0.97 (H-6, d, *J* = 6.6)	+, NQ	+, NQ	+, NQ	+, NQ
	Proline	δ 2.35 (H-3, m), δ 4.06 (H-2, m)	+, NQ	+, NQ	+, NQ	+, NQ
	Succinic acid	δ 2.45 (s)	+, NQ	+, NQ	+, NQ	+, NQ
	Sucrose	δ 4.22* (H-1' of fructose moiety, d, *J* = 8.6), δ 5.42 (H-1, d, *J* = 3.8)	+	+	+	+
	Threonine	δ 1.33* (H-5, d, *J* = 6.6)	+	+	+	+
	Valine	δ 1.00 (H-3, d, *J* = 6.8), δ 1.05* (H-4, d, *J* = 6.8)	+	+	+	+

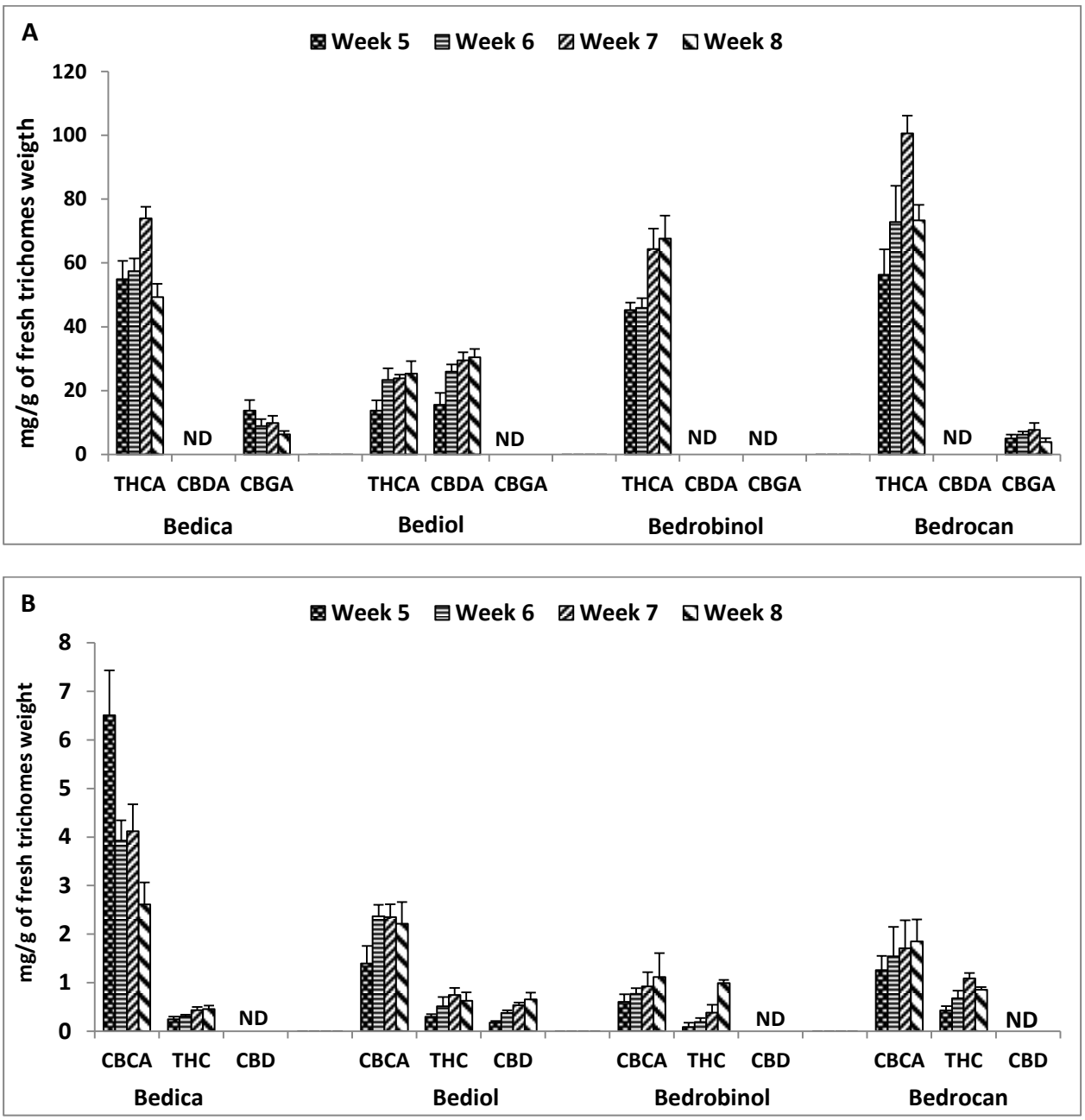
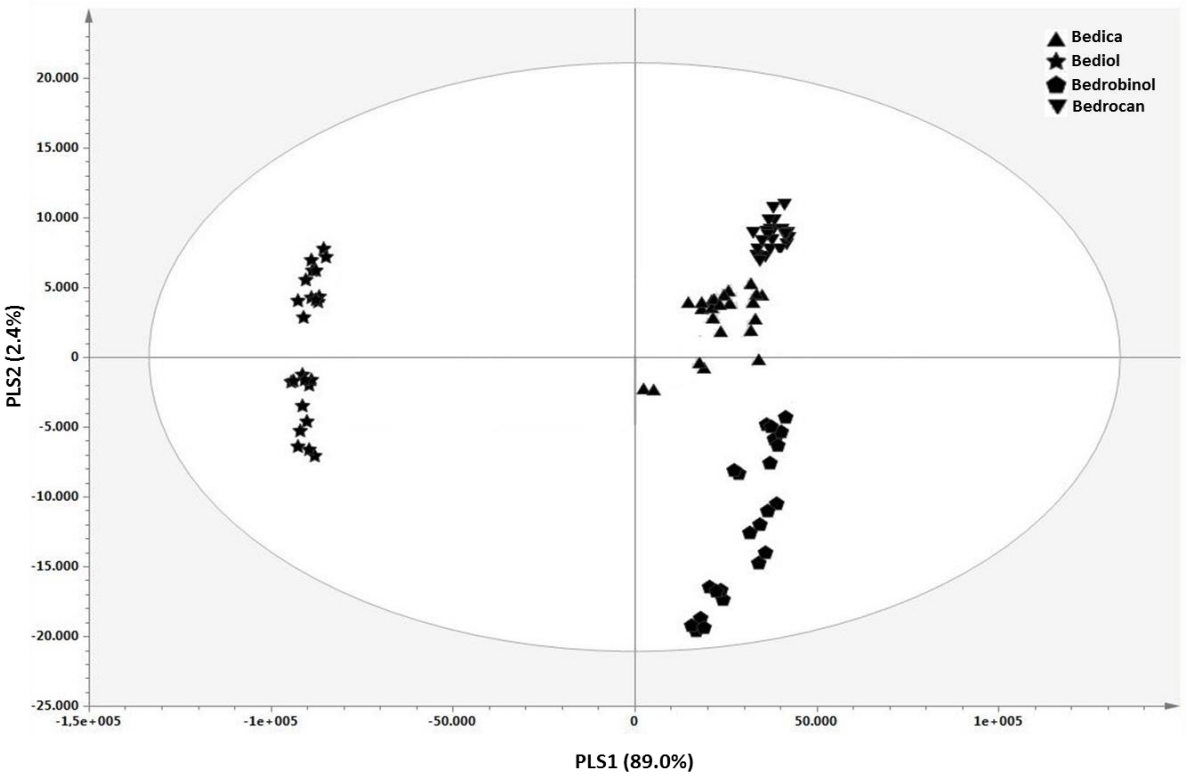


Figure 4A-B. Quantification of cannabinoids in the chloroform trichome extracts of 4 *C. sativa* varieties during the last 4 weeks of the flowering period; ND: compound is not detected.

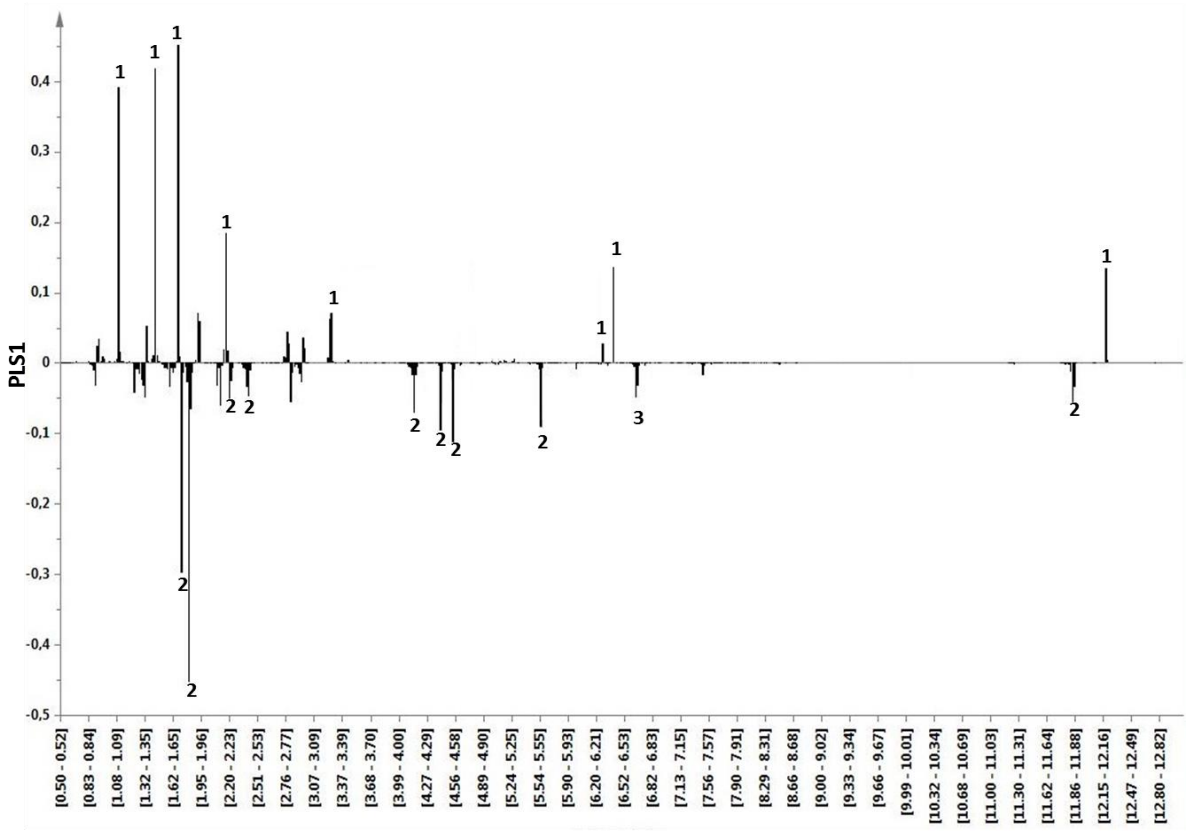
PLSDA of the trichome water extracts resulted in a model comprising 7 PLS components with R^2X = 69.9%, R^2Y = 95.4%, and Q^2 = 92.3%. The score plot of the proposed model showed a separation between the extract components of Bedica, Bediol, Bedrobinol, and Bedrocan trichomes (Fig. 5B1). The subsequent investigation of the loading plot of the first PLS component (PLS1) revealed that the ^{1}H NMR signals recorded for all the quantified water-soluble compounds fostered the differentiation between the *C. sativa* varieties (Fig. 5B2). Furthermore, the signals of α-glucose, fructose, and asparagine at δ 5.23 (H-1), δ 4.06 (H-1), and δ 2.96 (H-3a); 2.87 (H-3b), respectively, were more contributing in the discrimination compared to the signals of the other quantified compounds. According to the score plot between PLS1 and PLS2, the metabolites in the water extracts of Bedica trichomes had more similarity with those of the variety Bedrocan. This result was in accordance with metabolites of both trichomes in the chloroform extracts showing similar profiles.

The following metabolomic analysis involved specific monitoring of metabolite production in the Bedica trichomes during the last 4 weeks of the flowering period. A 4 component PLSDA model of the investigated organic extracts (supplementary data 3) indicated a variance explanation value of 75.3% (R^2X); 85.7% (*R2Y*), with the first 3 accounting for 55.2% of R^2X, while the estimated predictive ability (Q^2) value was 67.3%. The combination of PLS1 (29.7%) and PLS2 (14.0%) resulted in clear definition of 4 well-separated clusters corresponding to the 4 trichome harvesting time-points, as shown in Fig. 5C1. The loading plot of PLS1 was used to identify metabolites responsible for the observed separation (Fig. 5C2). In accord with the cannabinoid quantification results obtained for the Bedica trichomes, THCA (1) once again proved to be the main discriminant. Other PLS1 contributing signals were characteristic of CBGA (4), δ 2.10 (H-2') and THC (5), δ 6.30 (H-10). PLSDA of the water extracts of Bedica trichomes resulted in a model comprising 8 PLS components (supplementary data 3), accounting for 92.0% and 98.5 % of total variations (R^2X and R^2Y, respectively) and with 98.5% of predictive ability (Q^2). The obtained score plot further confirmed harvest time-dependent discrimination of trichome metabolite profiles (Fig. 5D1). The investigation of the PLS1 (37.7%) loading plot (Fig. 5D2) revealed the discriminants among the following amino acids: threonine, asparagine, and glutamic acid, characterized by the respective contributing signals: δ 1.33 (H-5), δ 2.96 (H-3a); δ 2.87 (H-3b), and δ 2.14 (H-4); δ 2.46 (H-3), and sugars: sucrose, δ 5.42 (H-1) and fructose, δ 4.06 (H-1). Beside the afforested primary compounds, formic acid, with a characteristic shift recorded at δ 8.46, proved to be a differentiating water-soluble secondary metabolite.

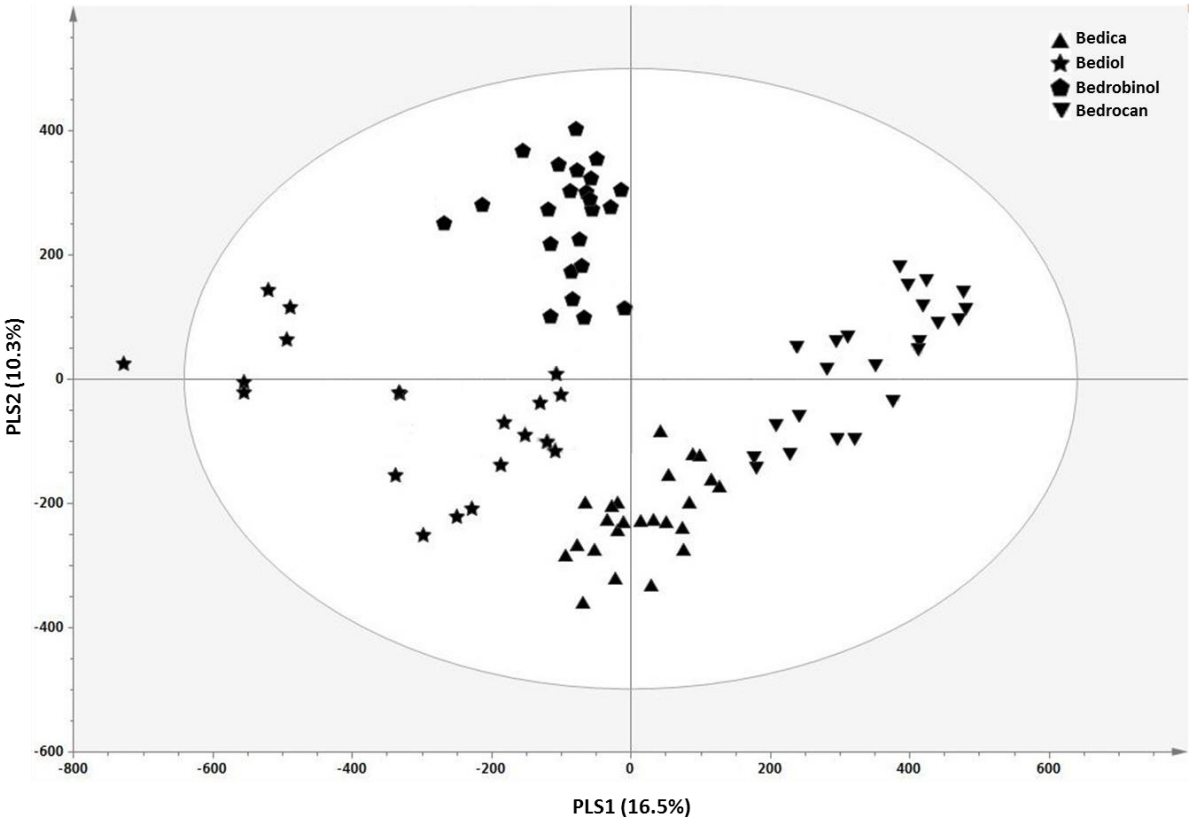
A1. Score plot of chloroform extracts of *Cannabis* trichomes



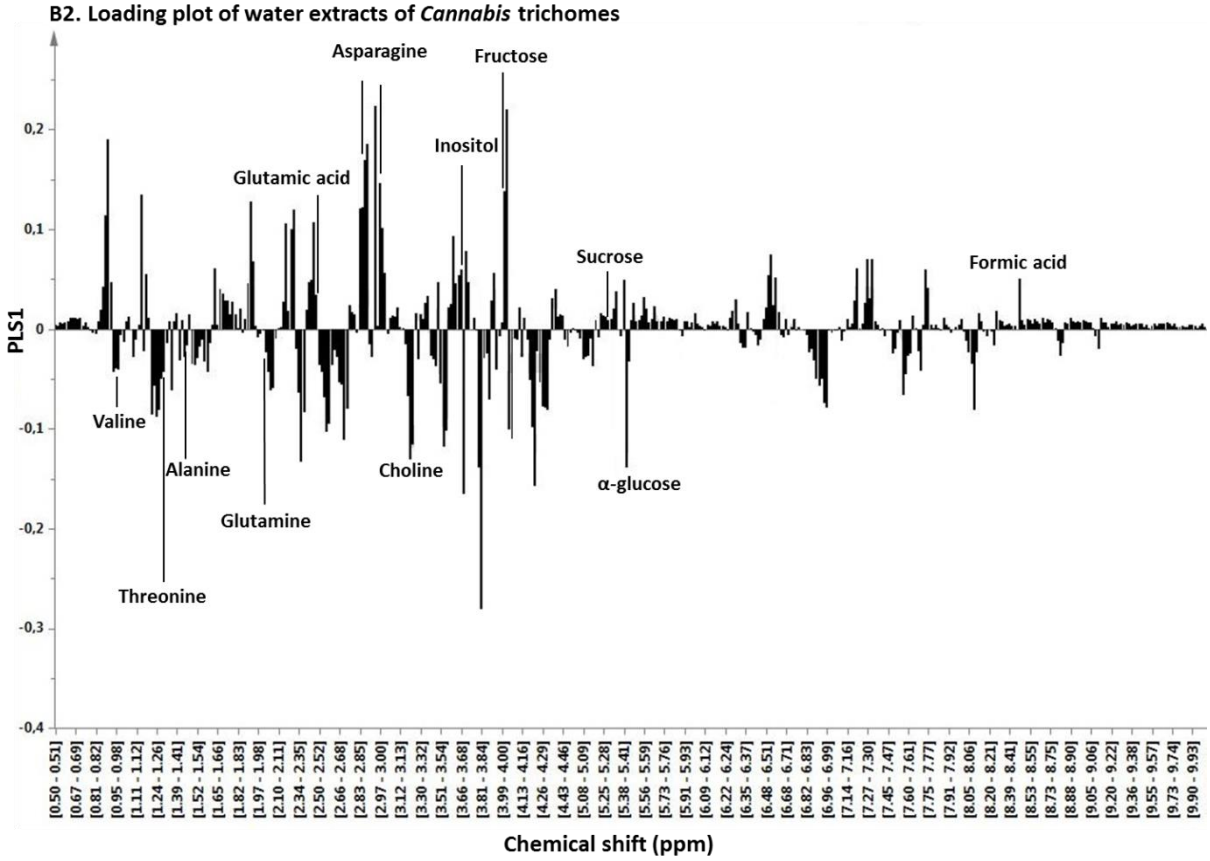
A2. Loading plot of chloroform extracts of *Cannabis* trichomes



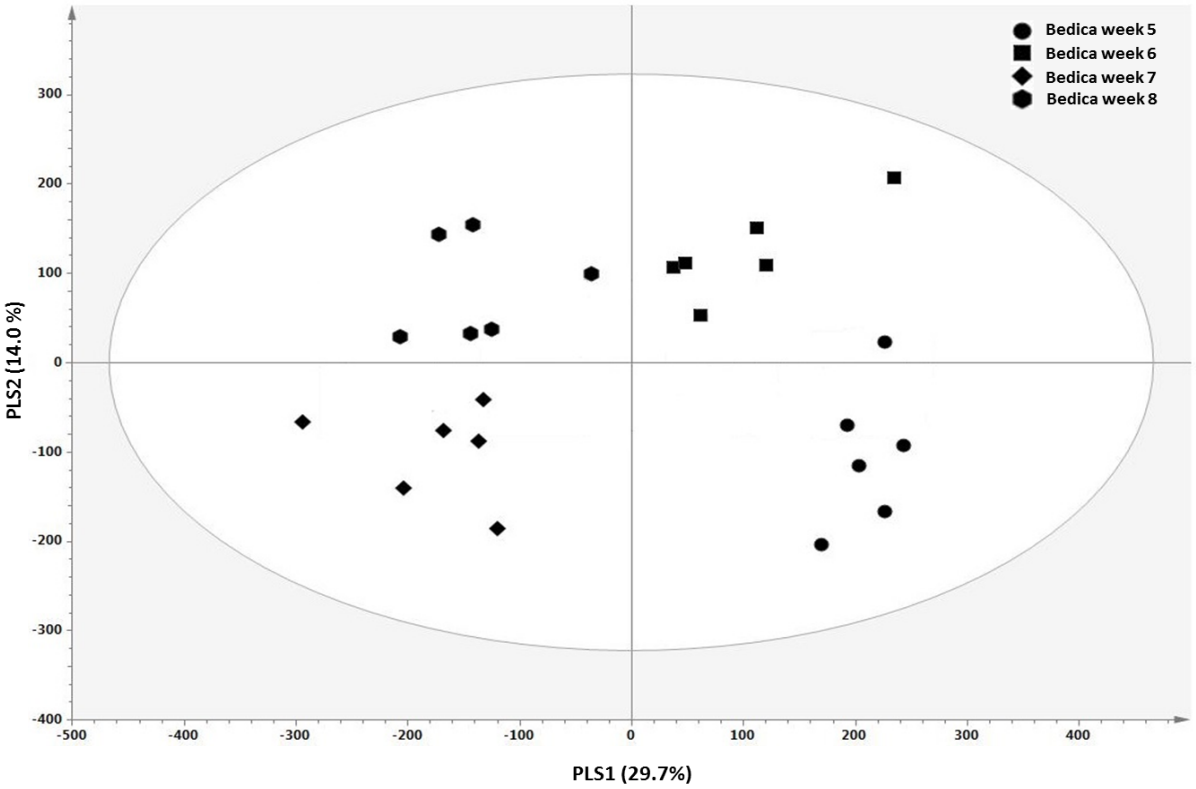
B1. Score plot of water extracts of *Cannabis* trichomes



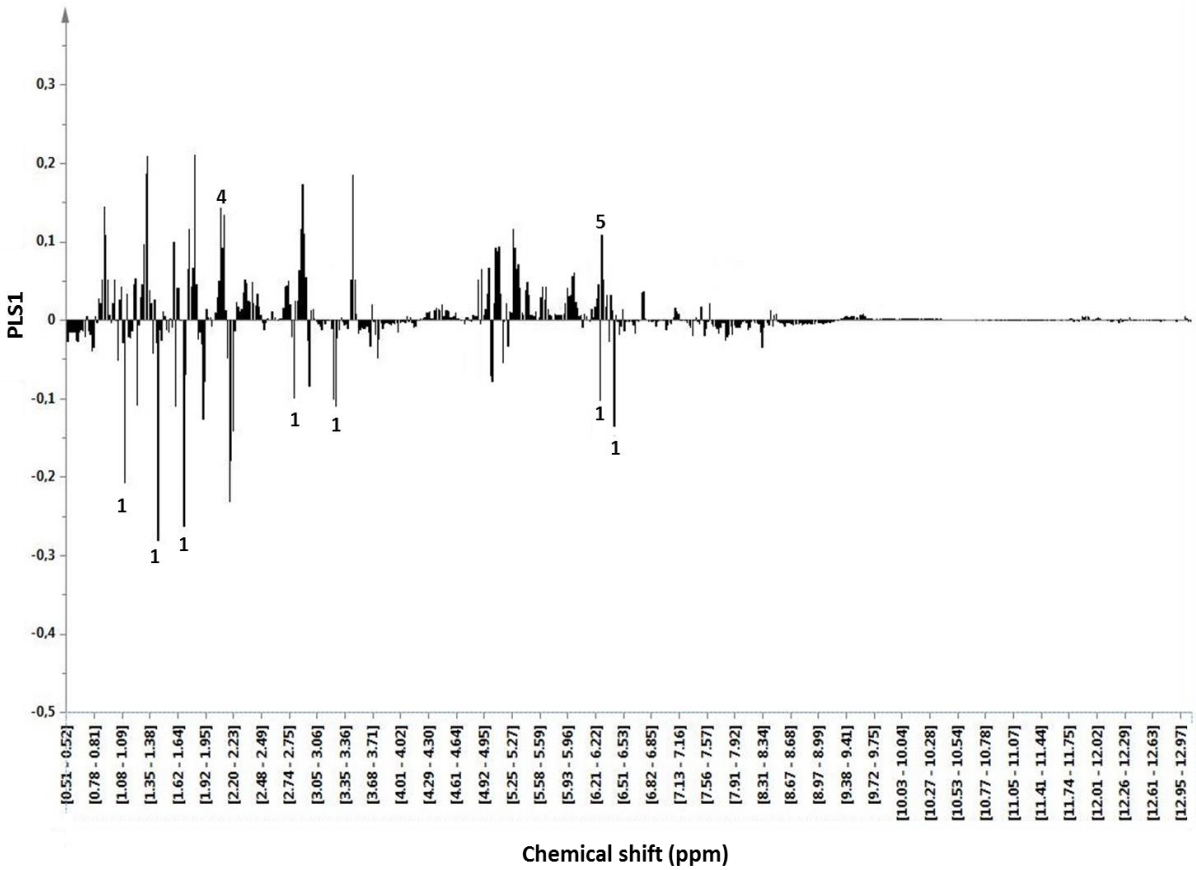
B2. Loading plot of water extracts of *Cannabis* trichomes



C1. Score plot of chloroform extracts of Bedica trichomes



C2. Loading plot of chloroform extracts of Bedica trichomes



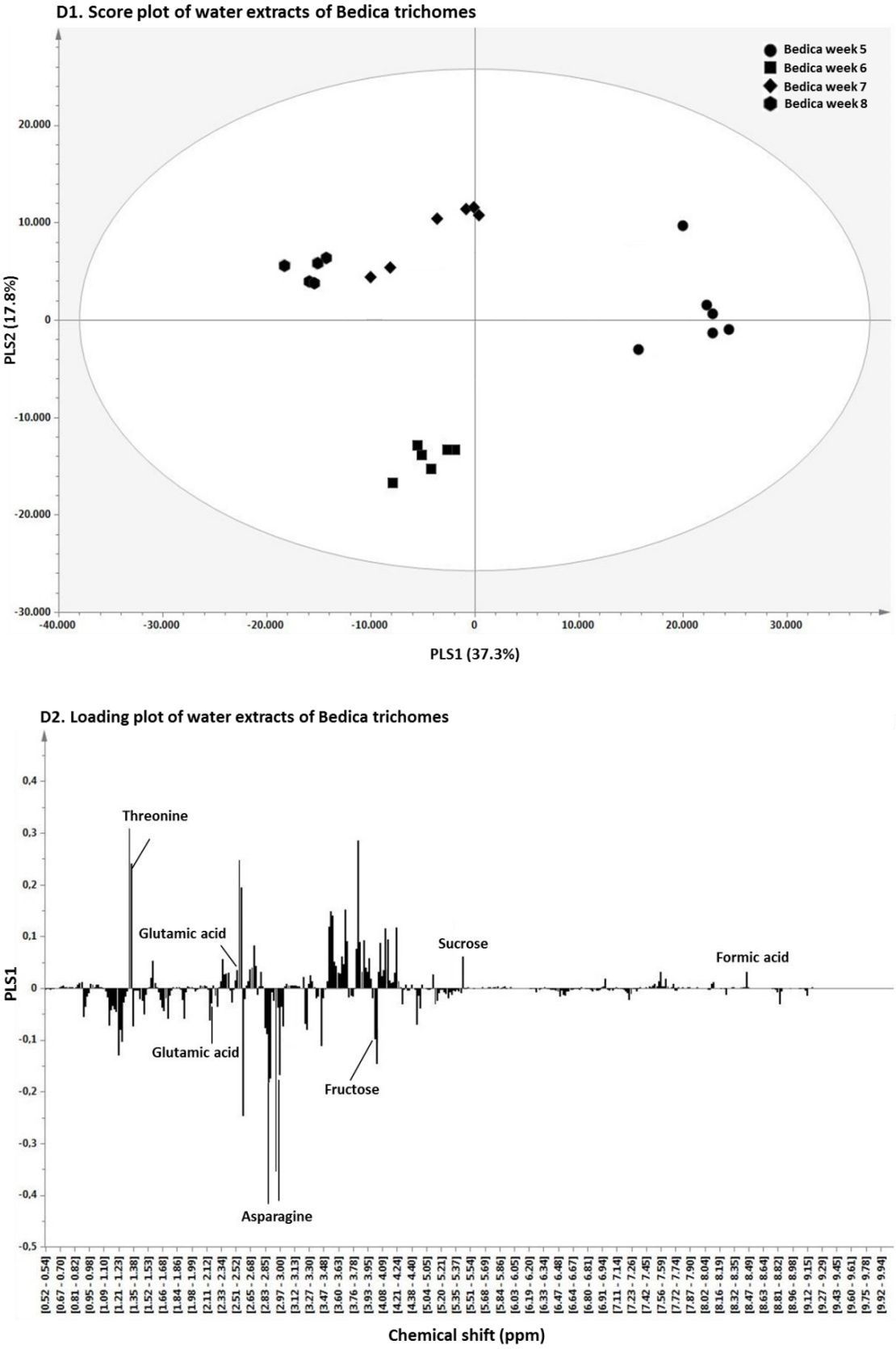


Figure 5A1-D2. Score and loading plots of the investigated trichome extracts. 1, THCA; 2, CBDA; 3,CBCA; 4, CBGA; 5, THC.

The subsequent PLSDA of metabolite biosynthesis in Bediol trichomes during the flowering period resulted in a 4 PLS component model (supplementary data 3) with R^2X = 69.9%, R^2Y = 88.4%, and Q^2 = 71.9% for the chloroform fractions. The score plot of PLS1 (29.8%) and PLS2 (14.4%) indicated successful separation of the investigated trichome samples based on the harvesting time-point, as shown in figure 5E1 (supplementary data 1). The position of the cluster corresponding to week 8 was significantly different from the others, indicating a unique biosynthetic profile of the trichomes in the final stage of monitoring. The PLS1 loading plot investigation revealed that THCA (1) and CBDA (2) were the compounds accounting for the observed differentiation (supplementary data 1, figure 5E2), with the signals of CBCA (3) at δ 6.23 (H-4) and δ 6.74 (H-5) being additional contributors. In case of the water extracts of the trichomes in question, the obtained PLSDA model comprised 6 PLS components (supplementary data 3) and explained 77.2% and 96.4% of total variations (R^2X and R^2Y, respectively), with the predictive ability (Q^2) of 82.9%. Well-separated clusters were once more obtained in the score plot of PLS1 (41.2%) and PLS2 (12.9%), as documented in figure 5F1 (supplementary data 1). The responsible compounds were identified through the investigation of the PLS1 loading plot (supplementary data 1, Figure 5F2); they were: threonine, asparagine, glutamic acid, sucrose, and fructose. Other amino acid signals at δ 1.05 (H-5) and δ 1.48 (H-3), corresponding to valine and alanine, respectively, were found to further contribute to the differentiation, as was the shift characteristic of choline at δ 3.21.

Next, the Bedrobinol trichome data set was subjected to PLSDA. The efficient biosynthetic profiling model of the chloroform extracts explained 85.3% and 86.1% of total variations (R^2X and R^2Y, respectively) using 5 PLS components (supplementary data 3) and had a predictive ability (Q^2) value of 71.8%. Combining PLS1 (39.1%) and PLS2 (21.6%) resulted in a score plot successfully illustrating metabolic discrimination, as seen in figure 5G1 (supplementary data 1). The compounds differentiating the Bedrobinol trichomes, as identified according to the loading plot of PLS1 (supplementary data 1, figure 5G2), comprised THCA (1), as the major discriminant, and its neutral form, THC (5), δ 6.13 (H-2). PLSDA modeling of Bedrobinol water-soluble metabolites involved 9 PLS (supplementary data 3) with R^2X = 62.0%, R^2Y = 99.6%, and Q^2 = 67.7%. The score plot combining PLS1 (13.1%) and PLS3 (5.1%) revealed four well-separated clusters (supplementary data 1, figure 5H1). The low variance values of both PLS components indicated that the Bedrobinol metabolite profiles recorded in the flowering period weeks 5 to 8 were similar. Yet, according to the loading plot analysis of PLS1 (supplementary data 1 figure 5H2), discriminative value could be assigned

to all quantified water-soluble compounds (alanine, valine, threonine, asparagine, glutamine, glutamic acid, α-glucose, sucrose, fructose, inositol, choline, and formic acid).

Finally, metabolic profiling of *C. sativa* Bedrocan trichoms over the final weeks of the flowering period was performed. The PLSDA model generated for the chloroform extract samples provided for their clear separation according to the harvesting time-point. The 6 PLS components (supplementary data 3) of this model explained 88.5% and 95.6% of total variations (R^2X and R^2Y, respectively) and gave a predictive ability (Q^2) value of 85.0%. The combination of PLS1 (44.6%) and PLS2 (18.4%) resulted in the definition of 4 well-separated clusters in the score plot, as shown in figure 5I1 (supplementary data 1). The characteristics signals of THCA (1) and CBGA (4) contributed to the differentiation observed in the PLS1 loading plot (supplementary data 1, figure 5I2). The aforementioned cannabinoids were also the metabolomic discriminants in Bedica trichomes, leading to the conclusion that the metabolic profiles of both varieties shared much similarity. The PLSDA model of the Bedrocan water extracts accounted for 77.7% and 97.9% of total variations (R^2X and R^2Y, respectively) and had a predictive ability (Q^2) of 82.5%. The score plot of the first 2 PLS components provided the best clustering preview of Bedrocan water-soluble trichome metabolites, as shown in figure 5J1 (supplementary data 1). As identified through PLS1 loading plot investigation (figure 5J2, supplementary data 1), threonine, glutamine, asparagine, choline, inositol, and fructose were the marker metabolites in the metabolomic differentiation of Bedrocan trichomes.

3.4 RT-PCR Analysis

Expression levels of cannabinoid genes in *Cannabis* trichomes were analyzed by RT-PCR in order to investigate their correlations with cannabinoid production, especially THCA and CBDA. Bediol trichomes have been chosen for this experiment since those have THCA and CBDA in almost same amount. Five genes involved in cannabinoid biosynthesis including the genes of THCA synthase (THCAS), CBDA synthase (CBDAS), olivetol synthase (OLS), olivetolic acid cyclase (OAC), and an enzyme that predicted as CBGA synthase from our grup (predicted-CBGAS) were analyzed by RT-PCR in Bediol trichomes over the last 4 weeks of flowering period. RT-PCR analysis showed that the expression levels of investigated genes did not differ significantly during the monitoring time as described in figure 6. This result seems different with quantitative analysis of cannabinoids in Bediol trichomes which shows increasing concentrations of THCA and CBDA over the monitoring time. Furthermore, these

findings probably might be a suggestion that there is no a positive direct correlation between the concentrations of cannabinoids with their gene expression levels in the trichomes during flowering period.

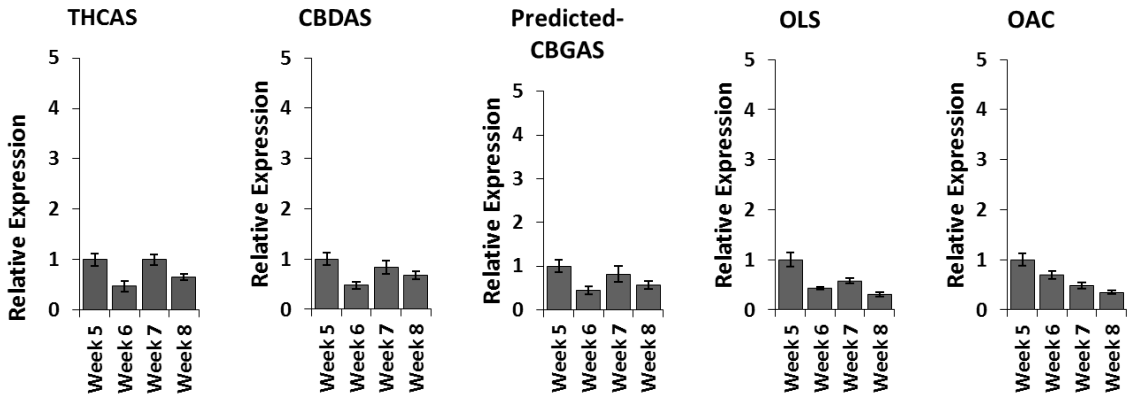


Figure 6. Expression level of cannabinoid genes in Bediol trichomes during the monitoring time.

4. Conclusions

Biosynthetic patterns characteristic of trichomes of 4 medicinal *C. sativa* varieties were investigated, using a quantitative ^{1}H NMR method and a metabolomic approach, over the final 4 weeks of the flowering period. Special emphasis was put on metabolic profiling of biologically active cannabinoids. The metabolomics approach, relying on PLSDA modeling, afforded differentiation of metabolite profiles within varieties. It further enabled efficient monitoring of metabolite production in trichomes of all individual varieties over the flowering period. THCA was identified as the vital discriminant cannabinoid in all investigated chloroform extracts. Furthermore, amino acids, such as threonine and asparagaine, and sugar compounds, represented by fructose and sucrose, were found to contribute significantly to the metabolomic discrimination of all the analyzed water extracts. This study indicated that, during the flowering period, *C. sativa* trichomes produced metabolites, particularly cannabinoids, in diverse amounts, depending on time and the plant variety. In this work, we ultimately showed that ^{1}H NMR-based metabolomics can be successfully applied for the time-dependent monitoring of biosynthetic processes. This method is simple, reproducible and can be coupled with ^{1}H NMR quantification.

Supplementary data of chapter 4

1. Supplementary data 1

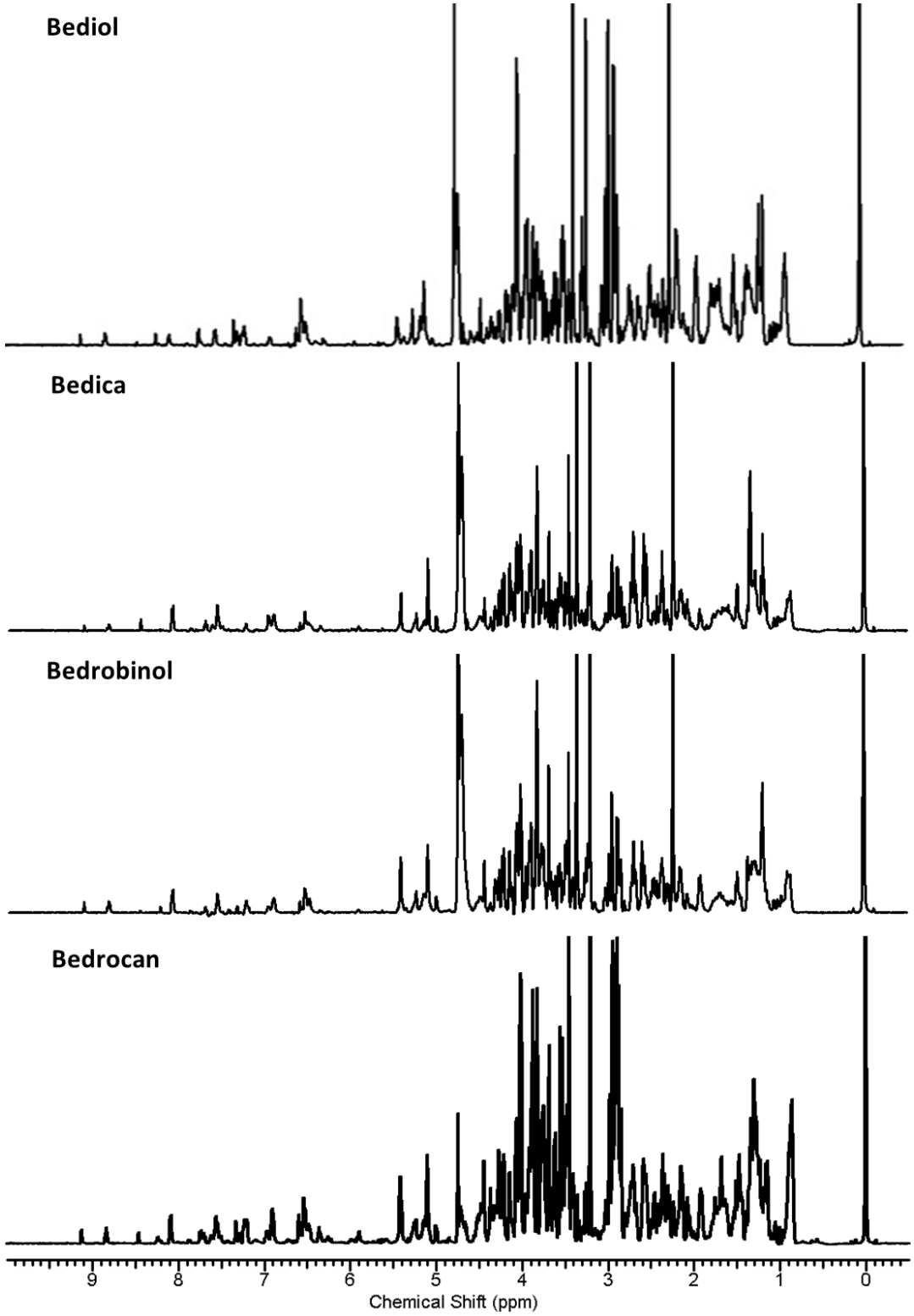
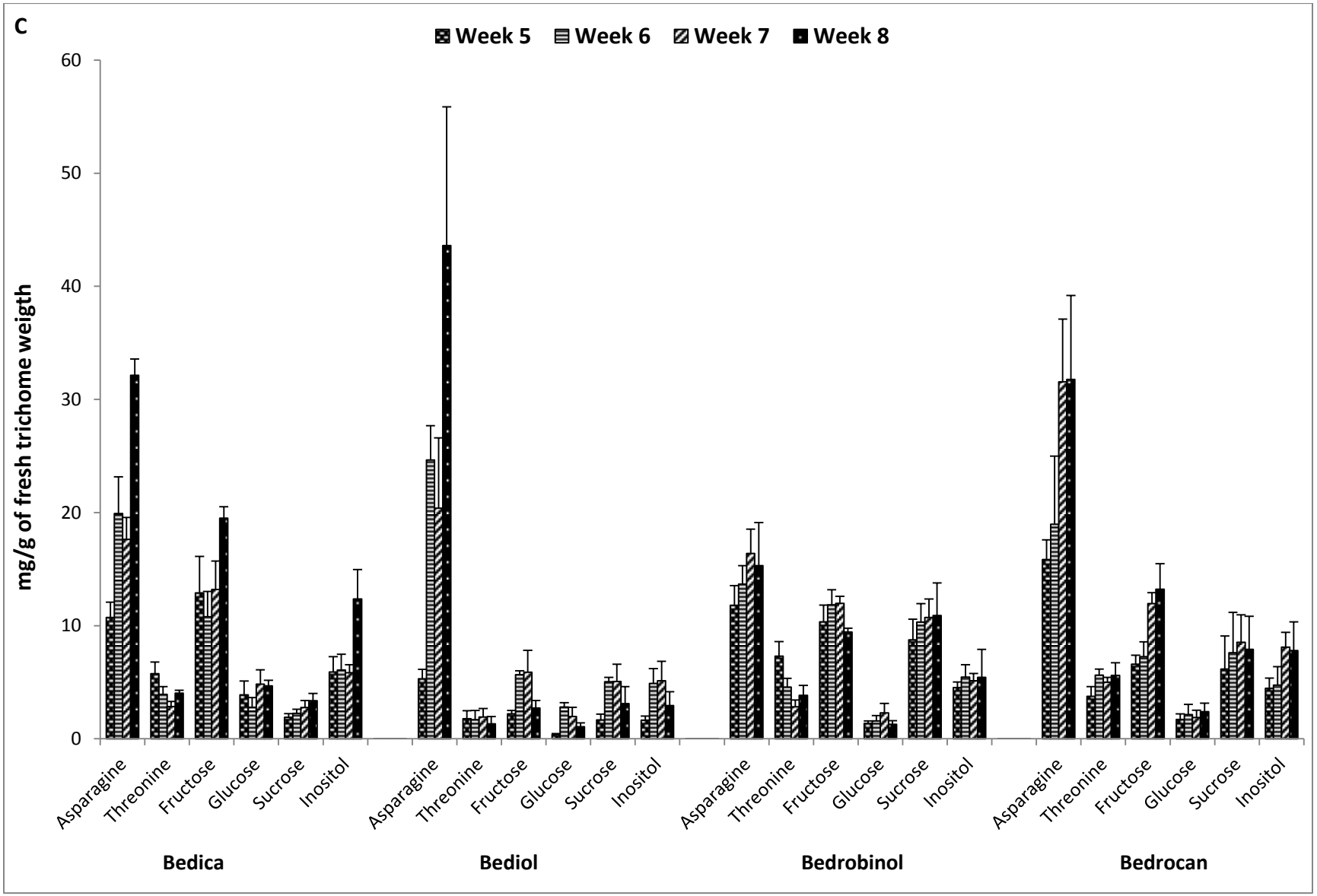


Figure 6. ^{1}H NMR spectra of the water extracts of *C. sativa* trichomes.


C
Week 5
Week 6
Week 7
Week 8
mg/g of fresh trichome weigth
60
50
40
30
20
10
0
Asparagine
Threonine
Fructose
Glucose
Sucrose
Inositol
Bedica
Bediol
Bedrobinol
Bedrocan

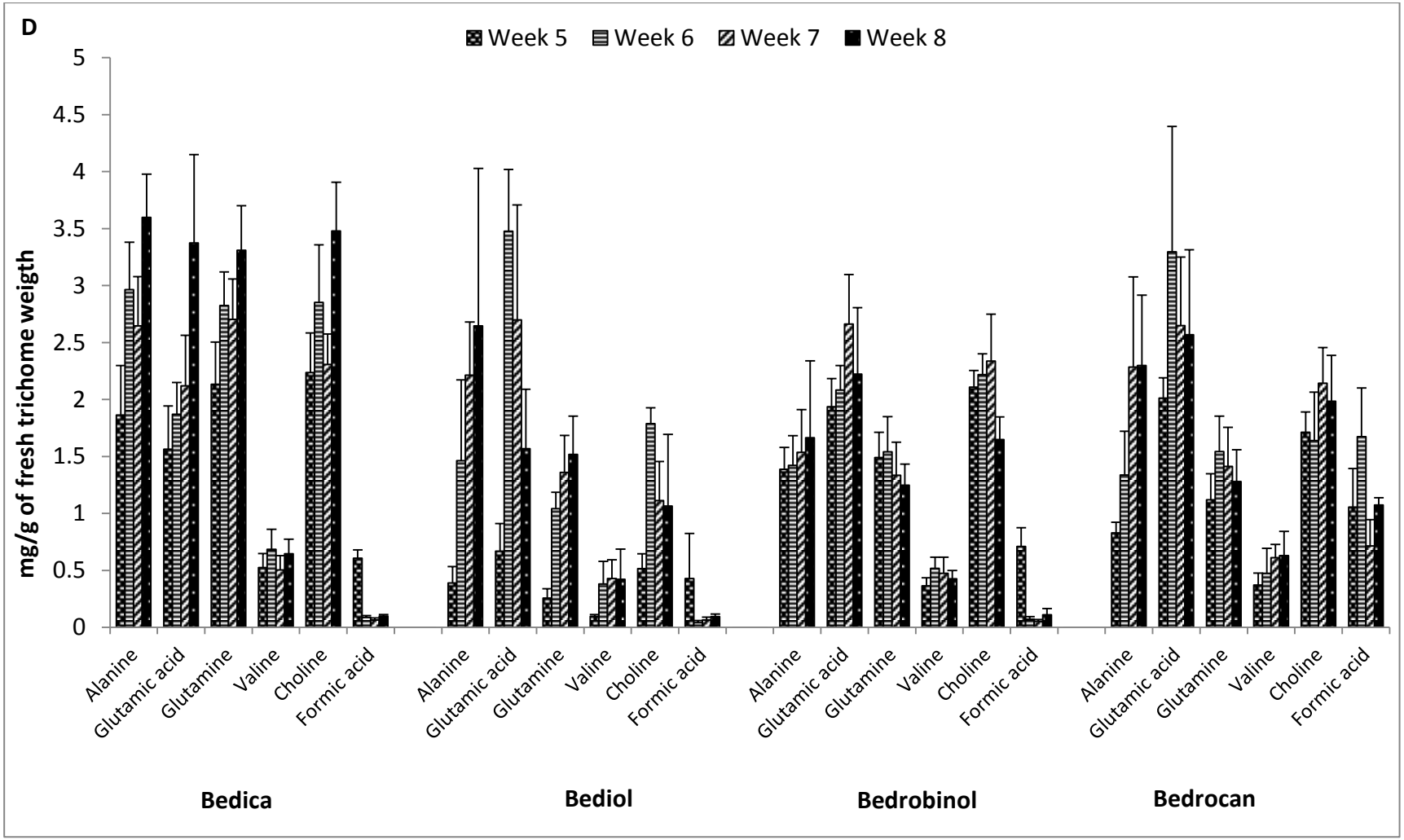
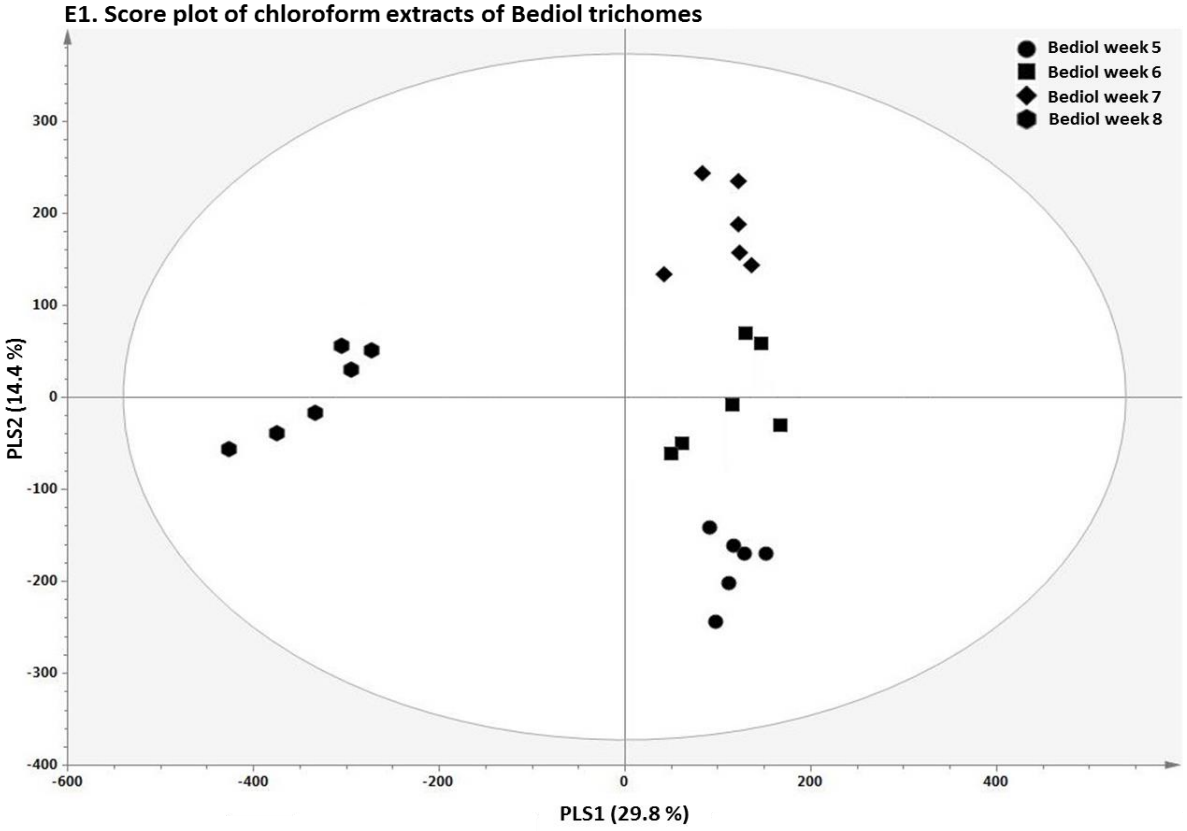
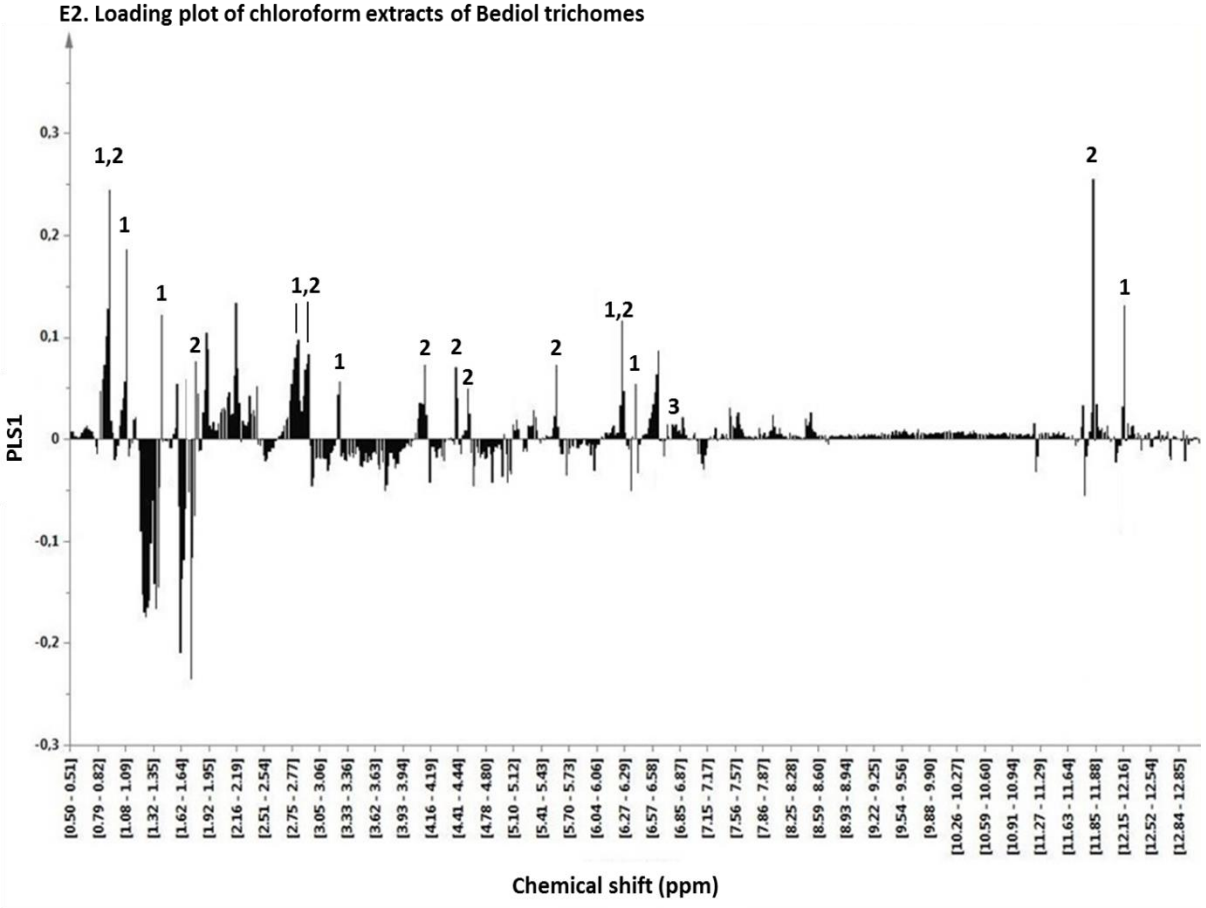


Figure 4C-D. Quantification of water-soluble compounds detected in the trichomes of 4 *C. sativa* varieties during the last 4 weeks of the flowering period.

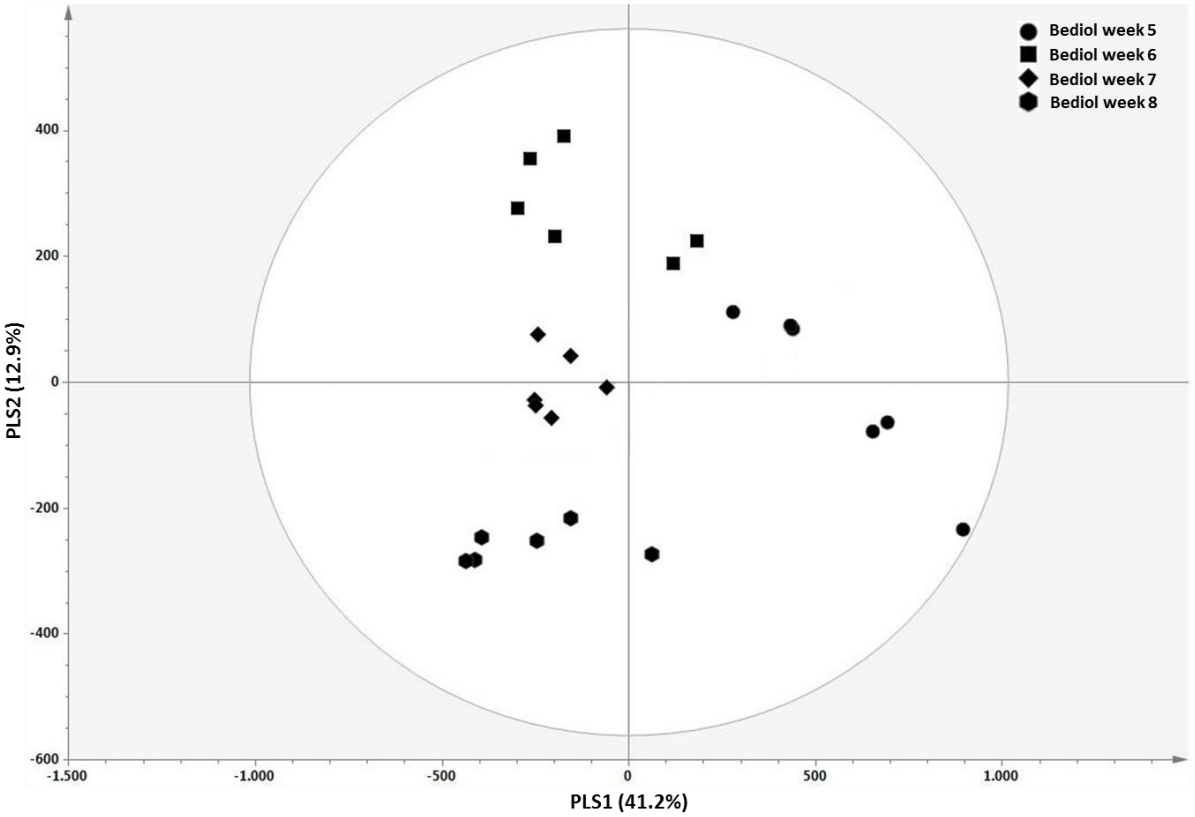
E1. Score plot of chloroform extracts of Bediol trichomes



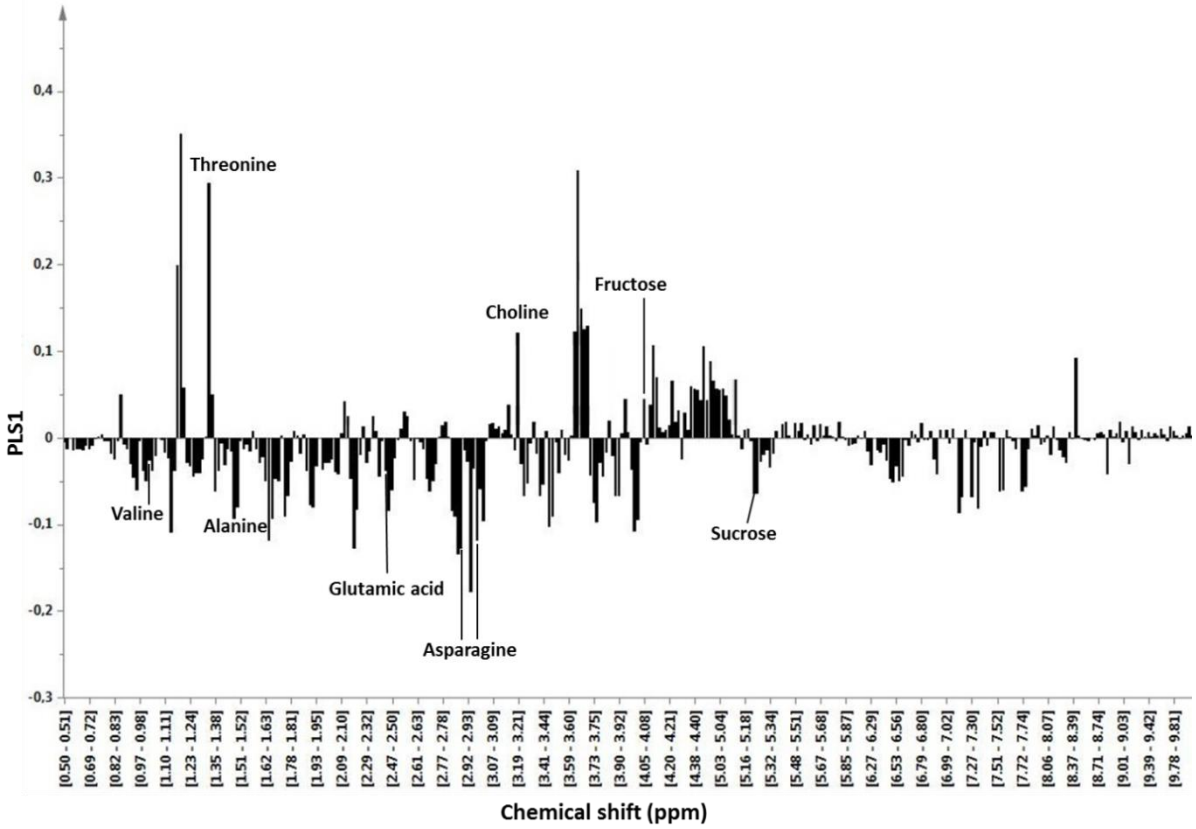
E2. Loading plot of chloroform extracts of Bediol trichomes



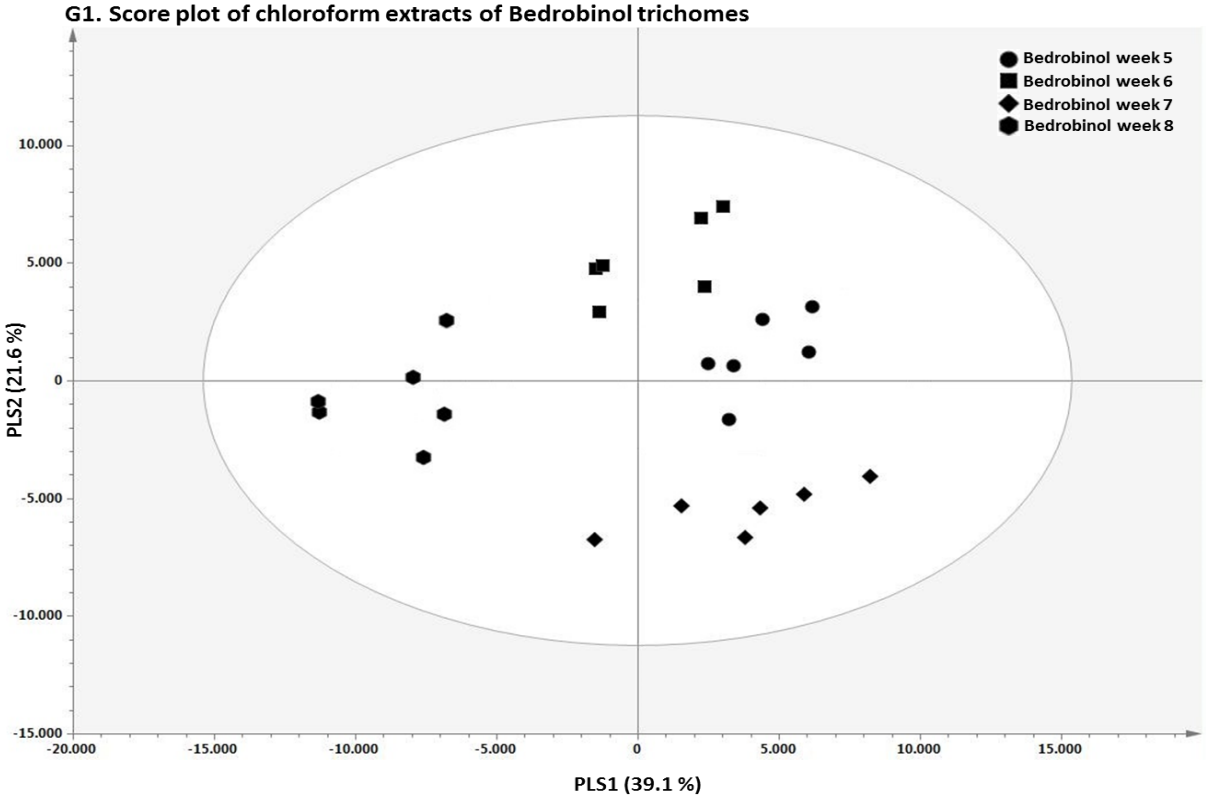
F1. Score plot of water extracts of Bediol trichomes



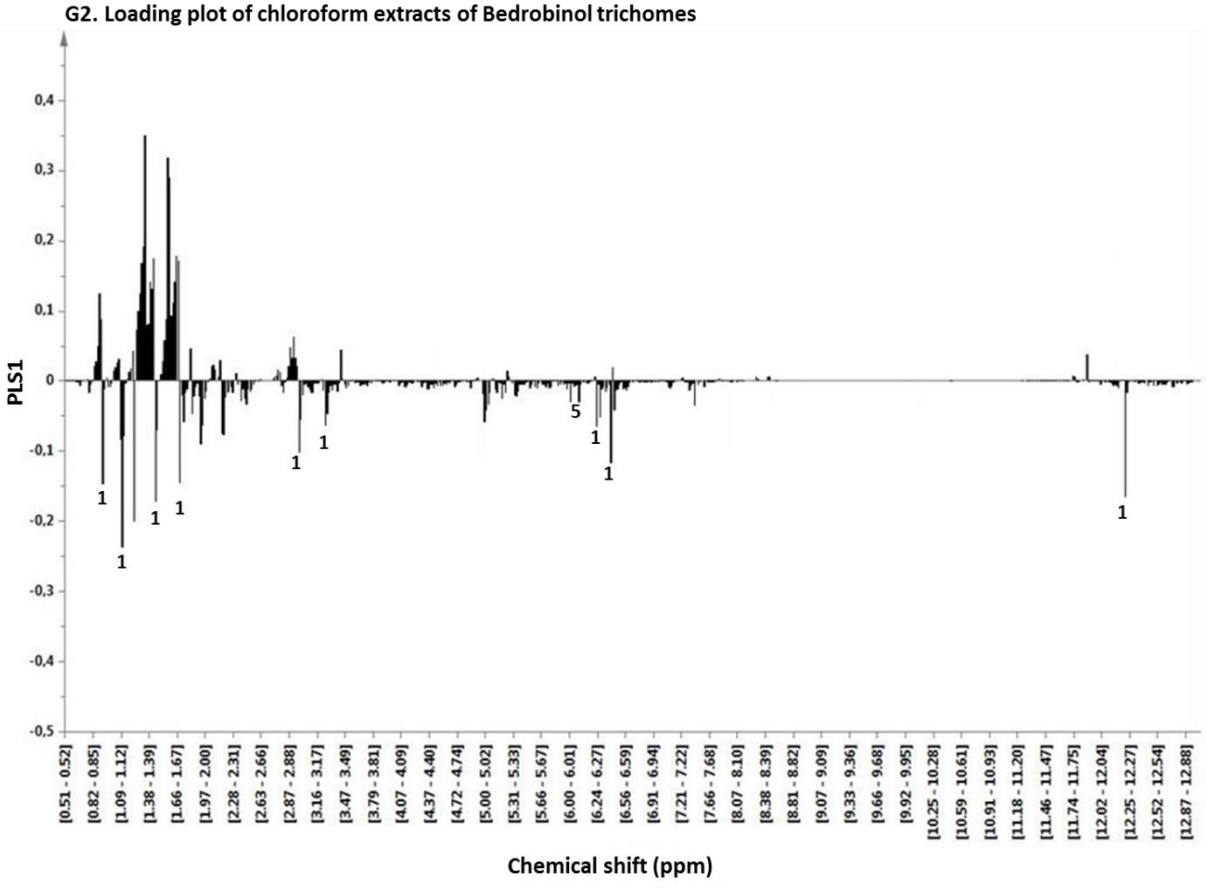
F2. Loading plot of water extracts of Bediol trichomes



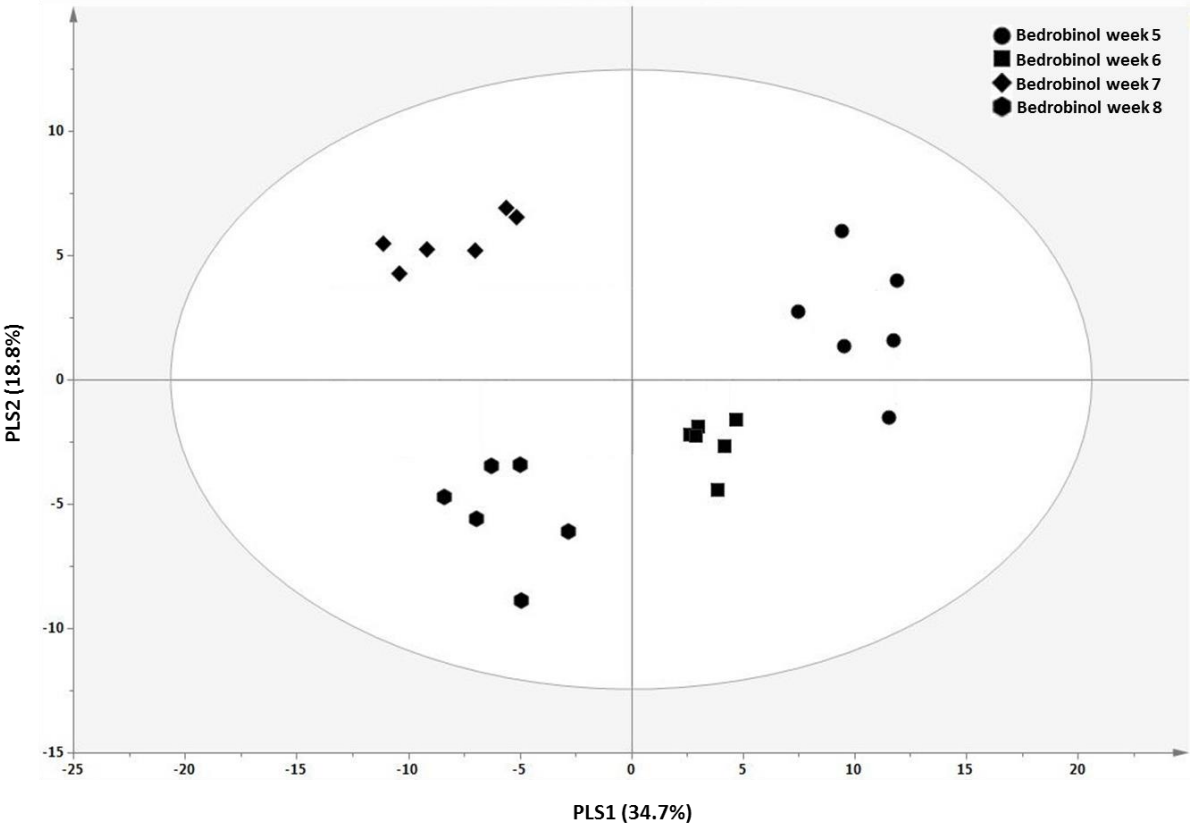
G1. Score plot of chloroform extracts of Bedrobinol trichomes



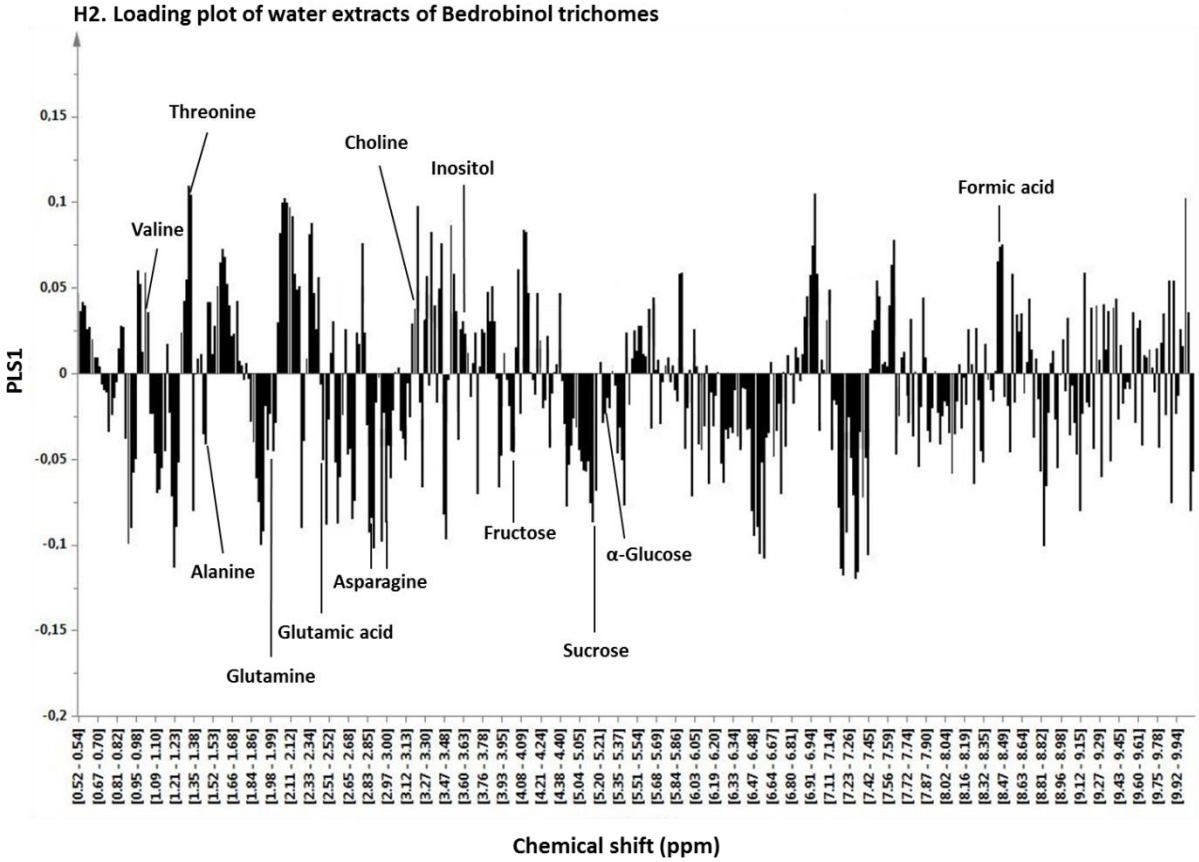
G2. Loading plot of chloroform extracts of Bedrobinol trichomes



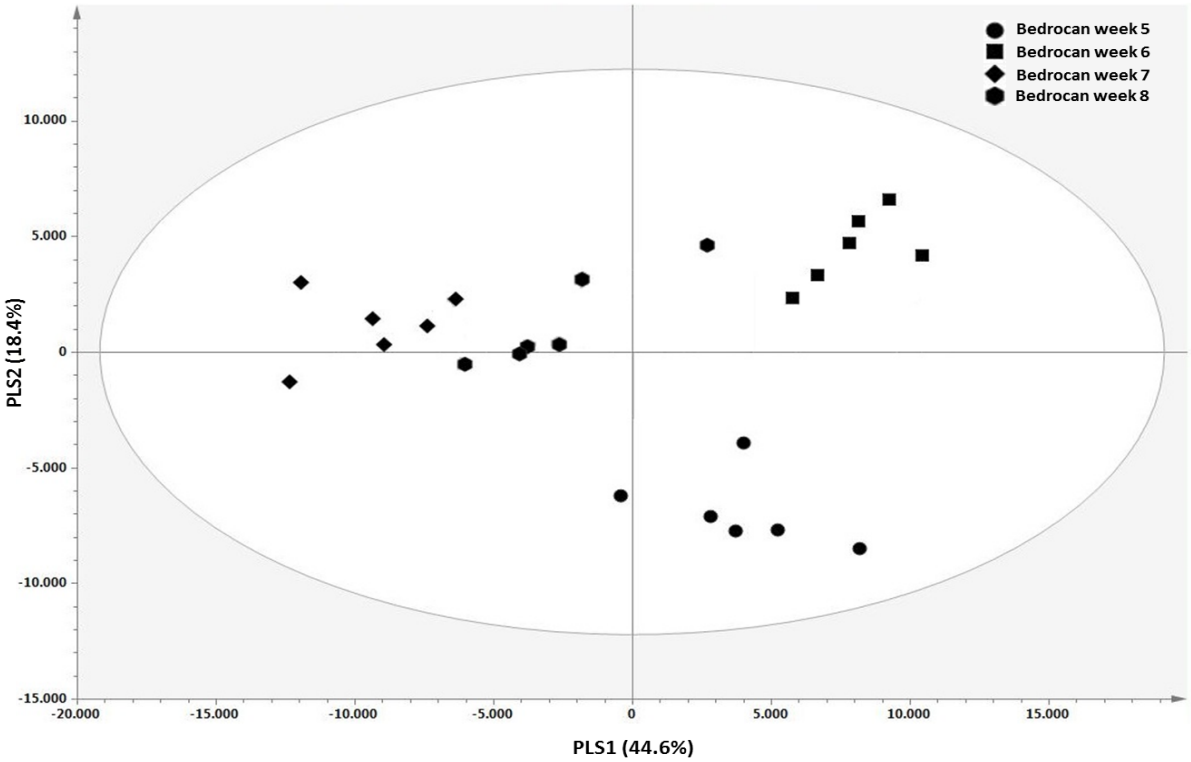
H1. Score plot water extracts of Bedrobinol trichomes



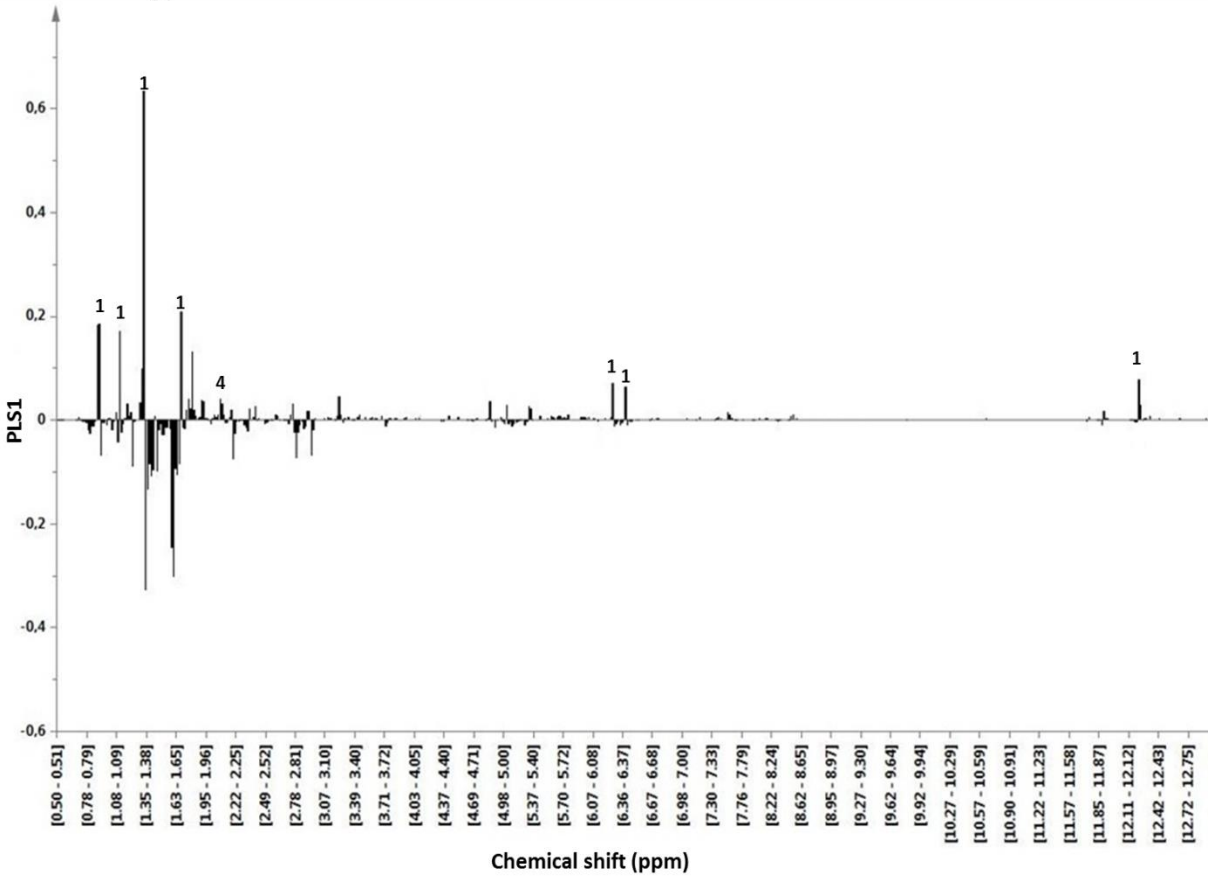
H2. Loading plot of water extracts of Bedrobinol trichomes

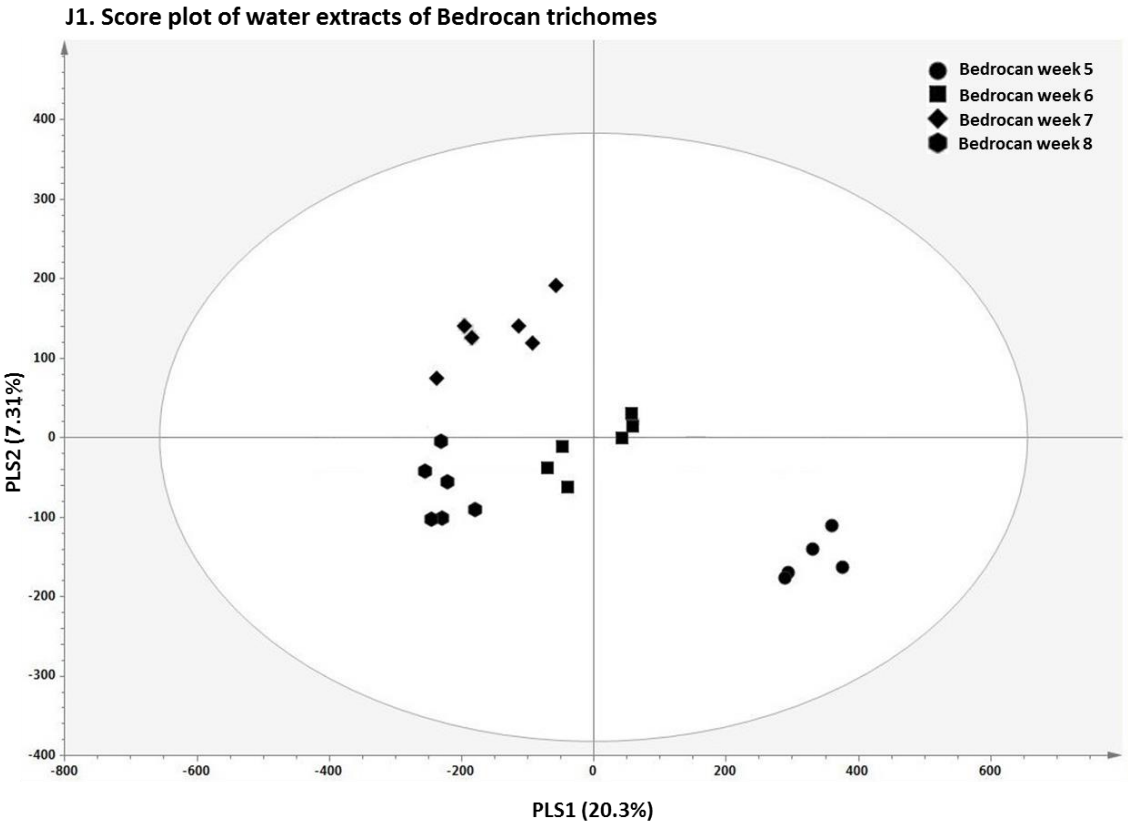


I1. Score plot of chloroform extracts of Bedrocan trichomes



I2. Loading plot of chloroform extracts of Bedrocan trichomes





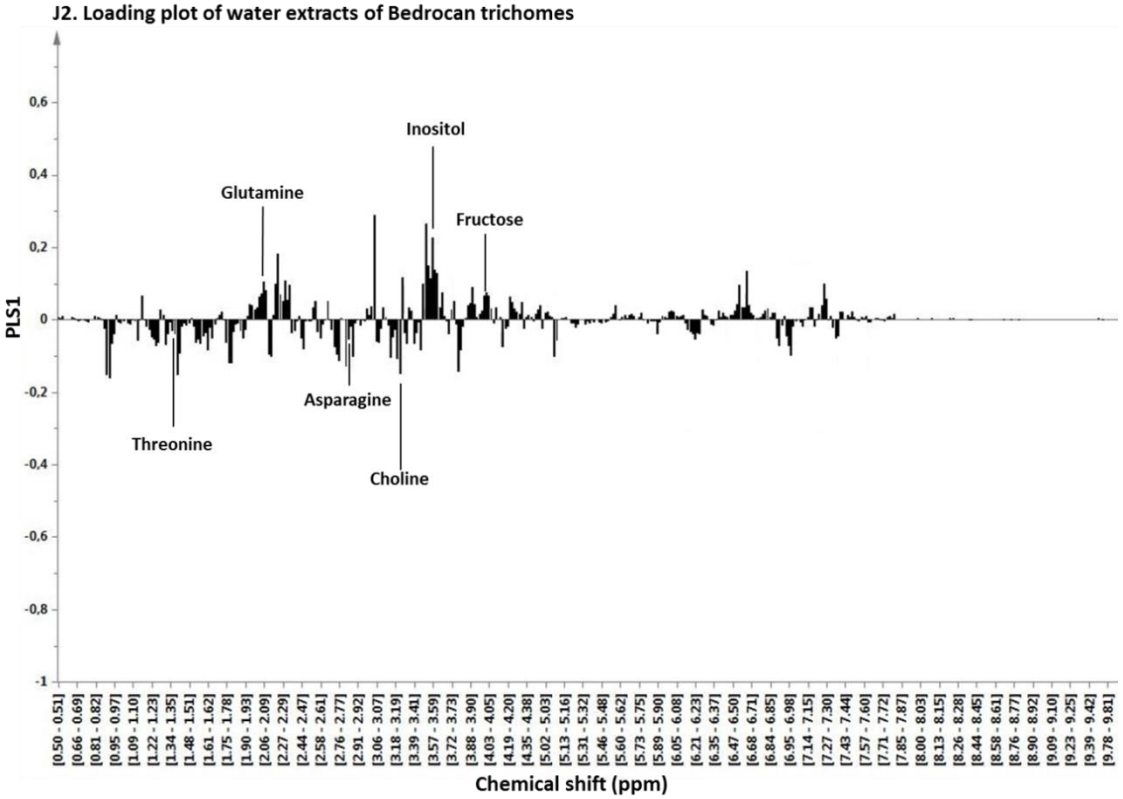
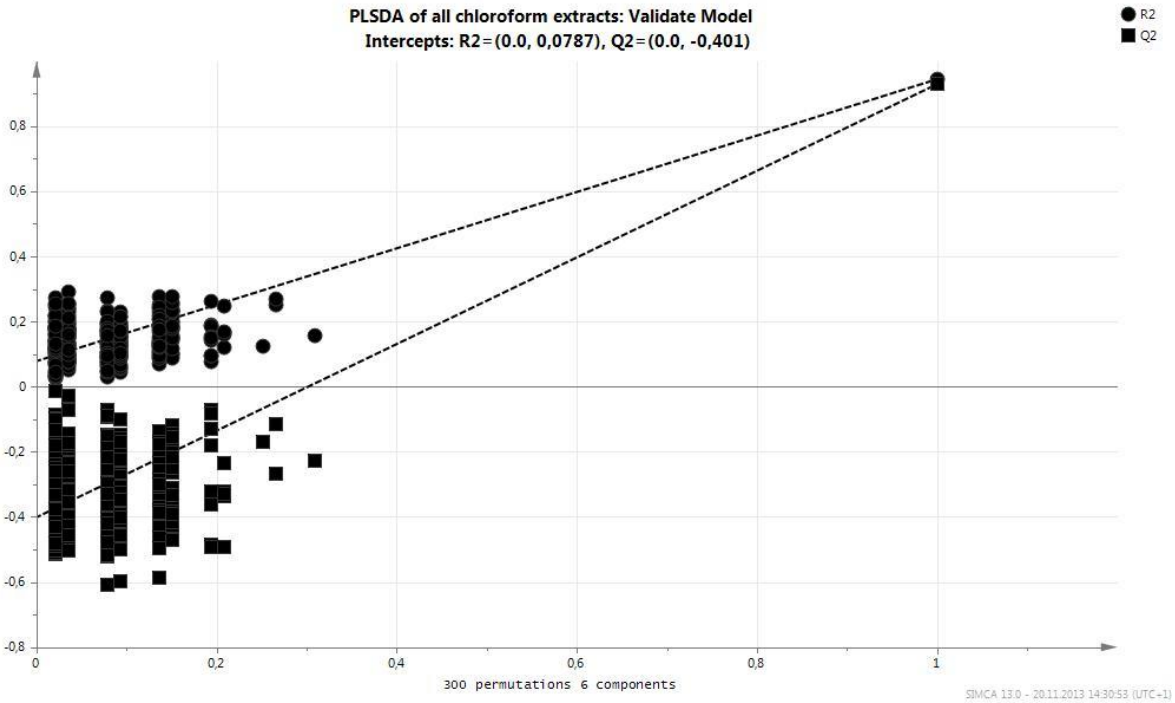


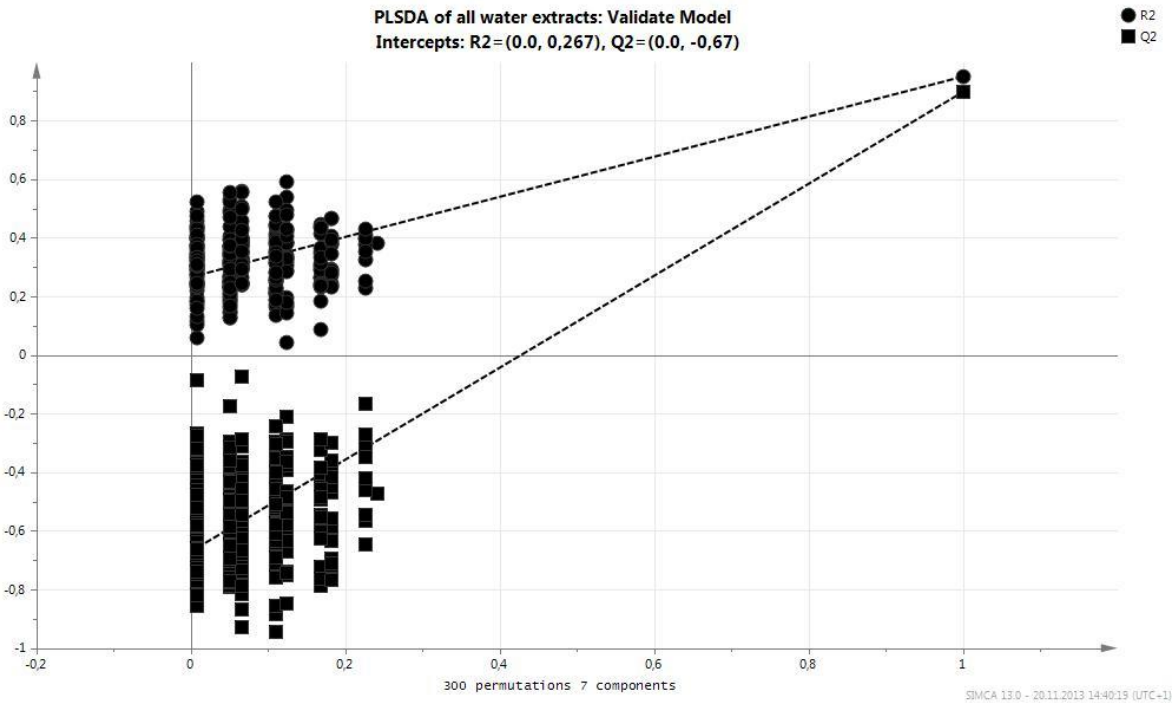
Figure 5E1-J2. Score and loading plots of *C. sativa var*. Bediol, Bedrobinol, and Bedrocan trichome extracts. 1, THCA; 2, CBDA; 3, CBCA; 4, CBGA; 5, THC.

2. Supplementary data 2

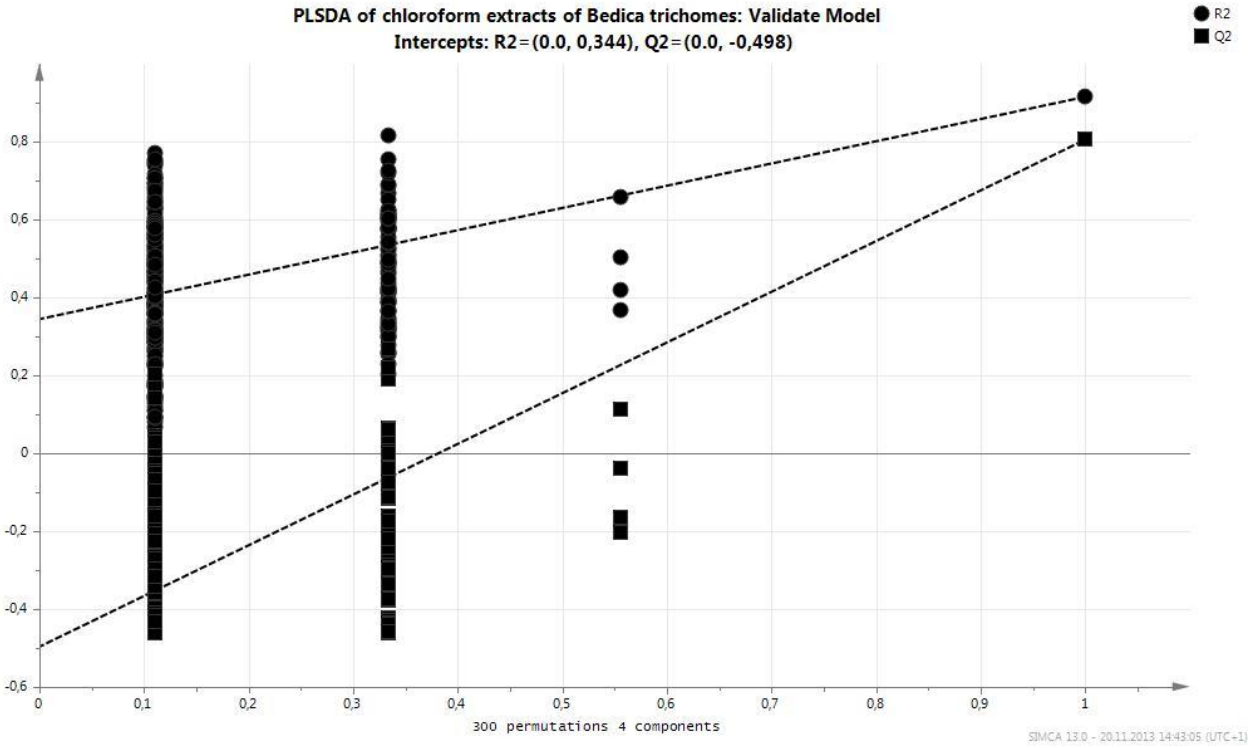
Permutation of all chloroform extracts



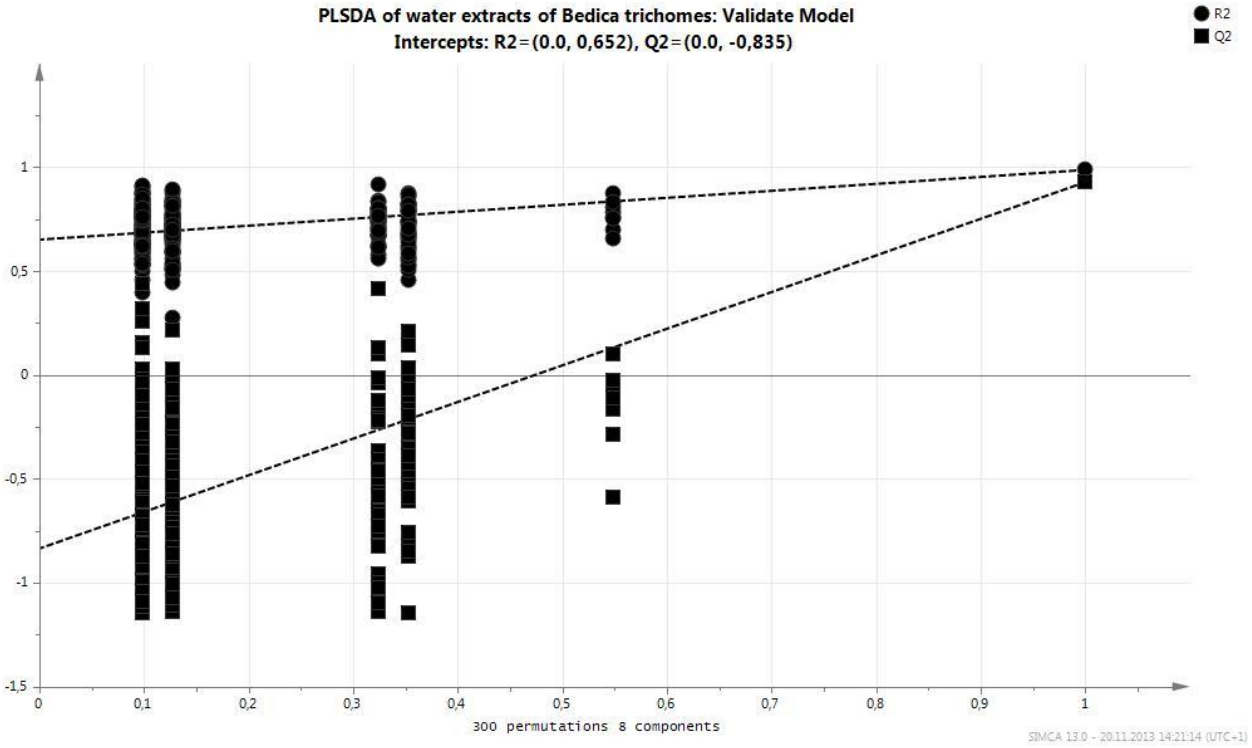
Permutation of all water extracts



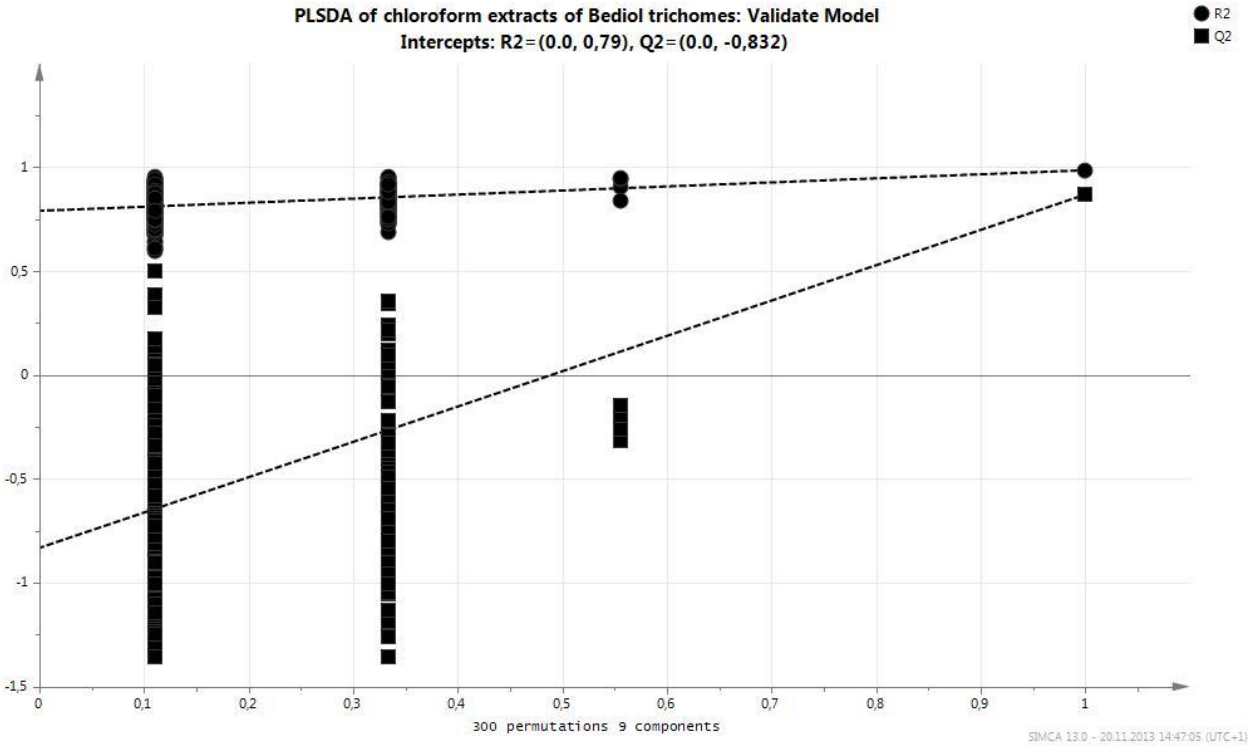
Permutation of chloroform extracts of Bedica trichomes



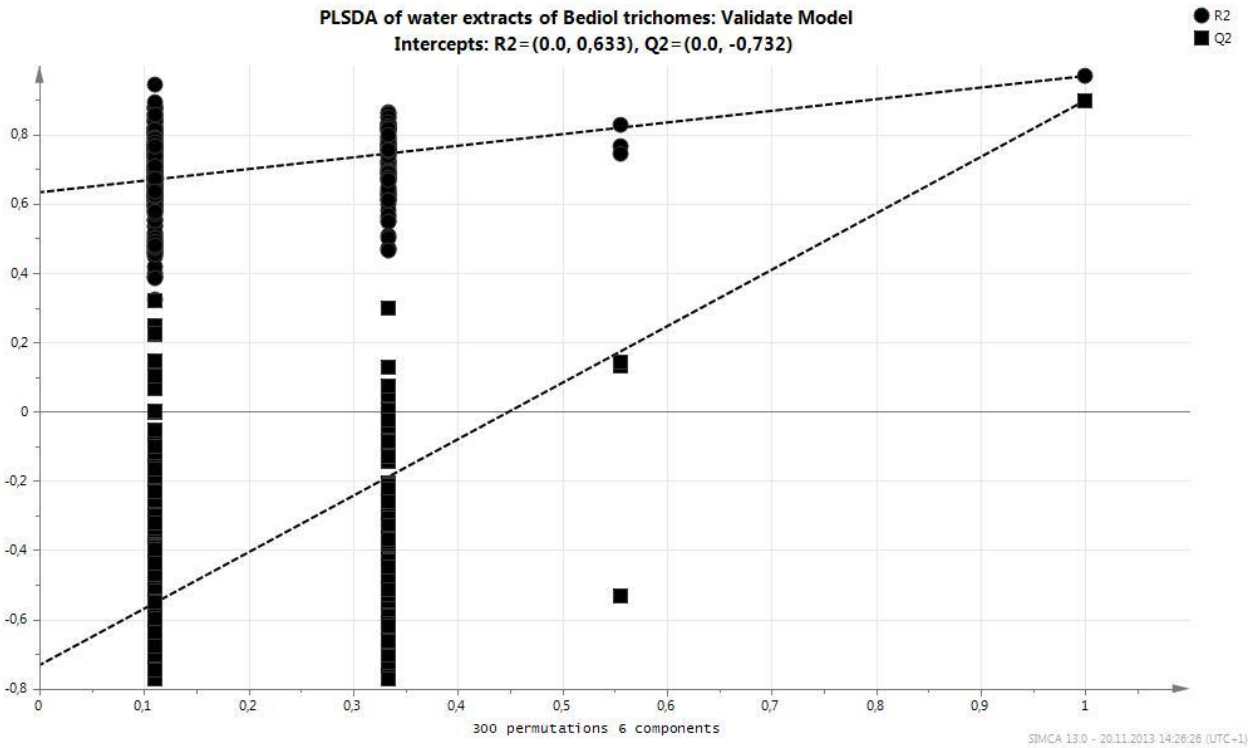
Permutation of water extracts of Bedica trichomes



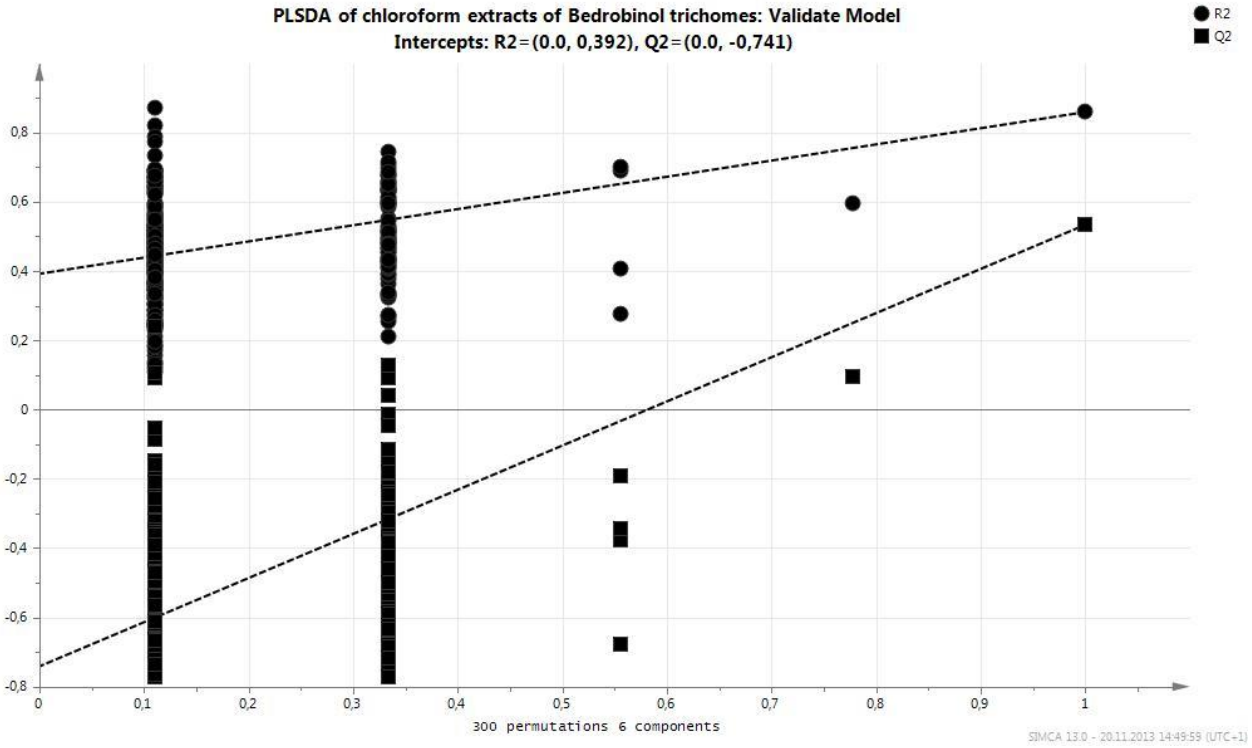
Permutation of chloroform extracts of Bediol trichomes



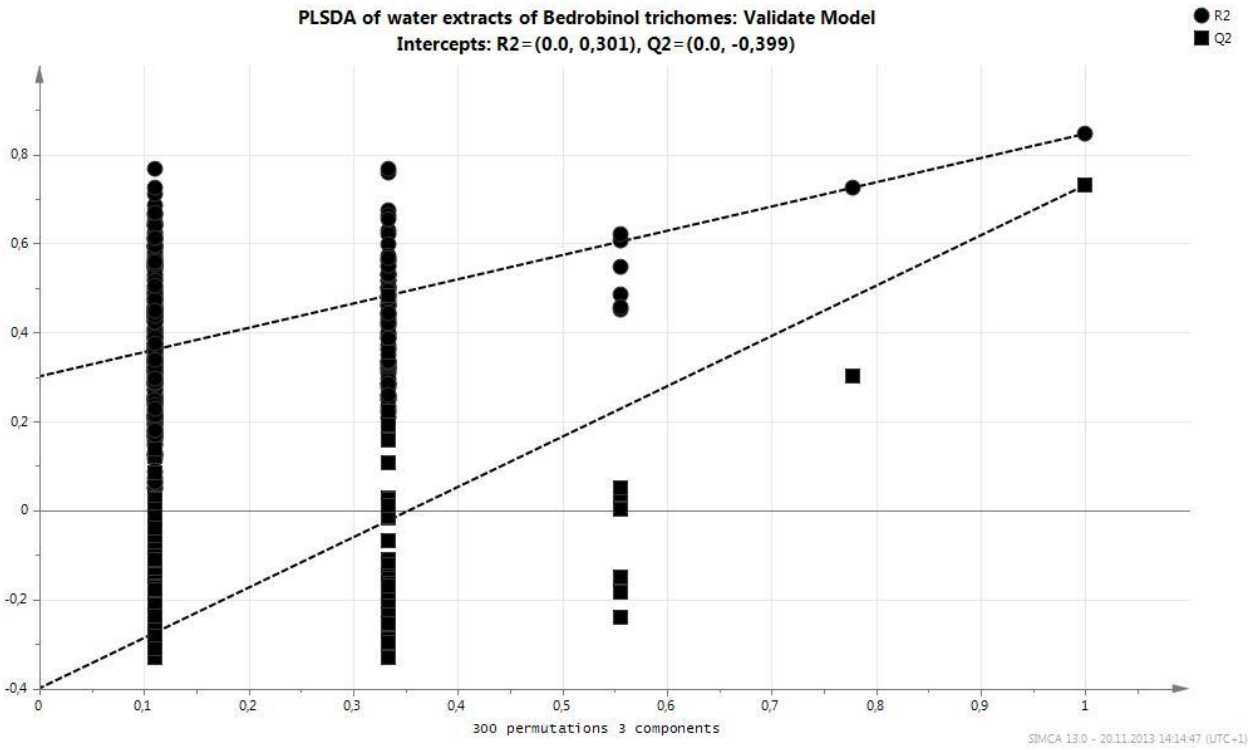
Permutation of water extracts of Bediol trichomes



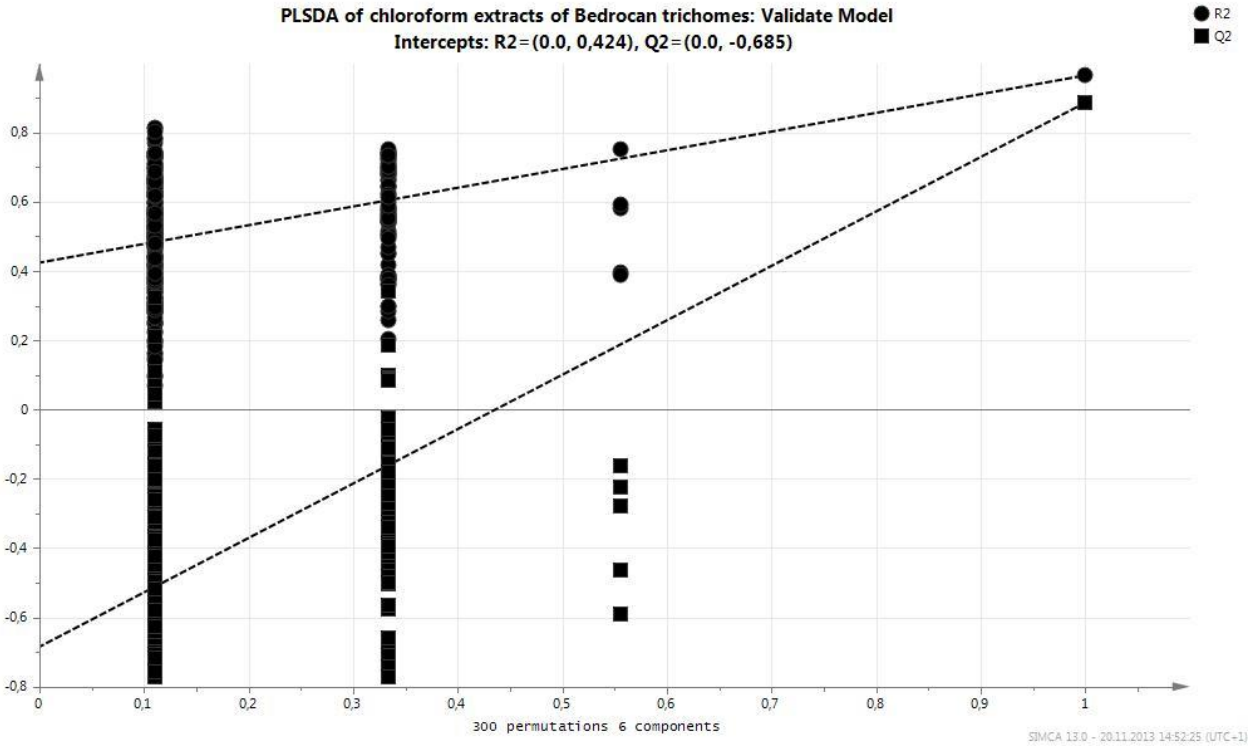
Permutation of chloroform extracts of Bedrobinol trichomes



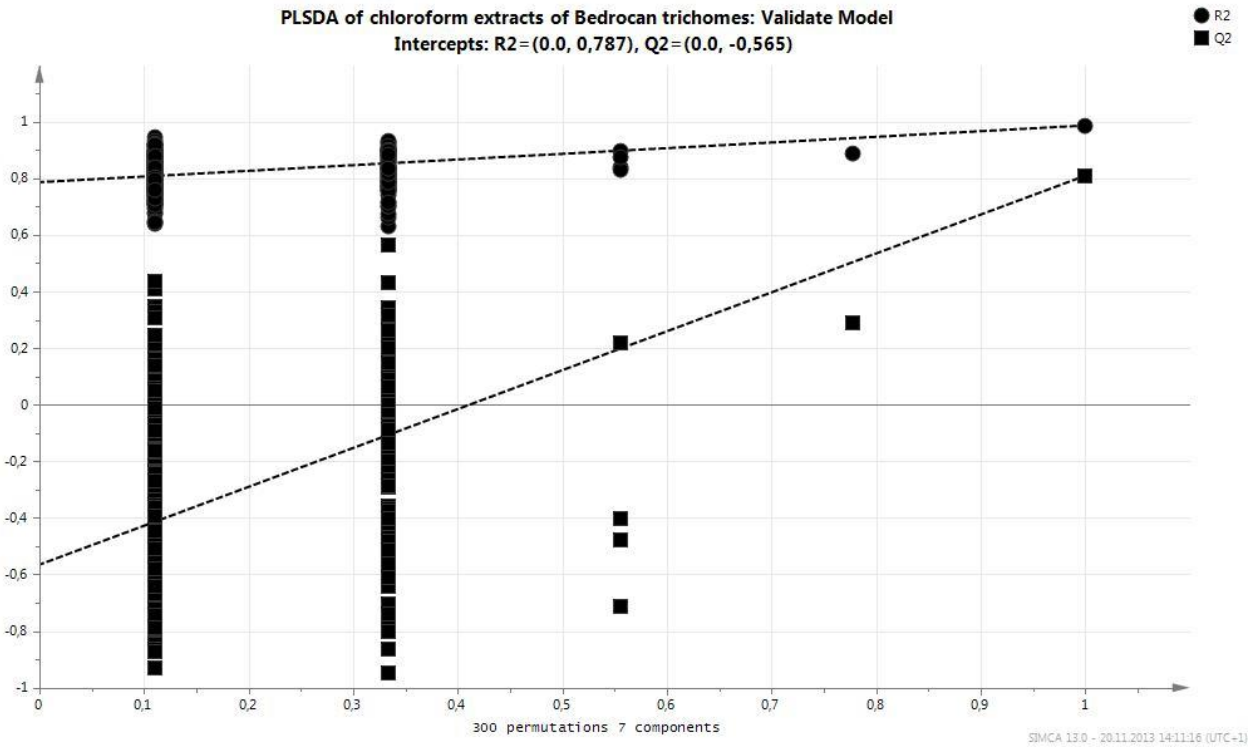
Permutation of water extracts of Bedrobinol trichomes



Permutation of chloroform extracts of Bedrocan trichomes



Permutation of water extracts of Bedrocan trichomes



3. Supplementary data 3

List of PLSDA components

PLSDA of all chloroform extract samples

Component	R2X	R2X(cum)	Eigenvalue	R2Y	R2Y(cum)	Q2	Limit	Q2(cum)
0	Cent.							
1	0.89	0.89	83.7	0.33	0.33	0.33	0.05	0.33
2	0.0242	0.914	2.27	0.265	0.595	0.384	0.05	0.587
3	0.0333	0.947	3.13	0.155	0.75	0.359	0.05	0.735
4	0.0168	0.964	1.57	0.109	0.859	0.416	0.05	0.845
5	0.00813	0.972	0.765	0.0782	0.937	0.537	0.05	0.928
6	0.00345	0.976	0.324	0.0242	0.961	0.269	0.05	0.948

PLSDA of all water extract samples

Component	R2X	R2X(cum)	Eigenvalue	R2Y	R2Y(cum)	Q2	Limit	Q2(cum)
0	Cent.							
1	0.222	0.222	19.9	0.299	0.299	0.29	0.05	0.29
2	0.165	0.386	14.8	0.271	0.569	0.379	0.05	0.559
3	0.103	0.49	9.3	0.27	0.839	0.607	0.05	0.827
4	0.0972	0.587	8.75	0.0368	0.876	0.205	0.05	0.862
5	0.052	0.639	4.68	0.0369	0.913	0.229	0.05	0.894
6	0.0378	0.677	3.4	0.025	0.938	0.238	0.05	0.919
7	0.0224	0.699	2.01	0.0165	0.954	0.0517	0.05	0.923

PLSDA of choloform extracts of Bedica trichomes

Component	R2X	R2X(cum)	Eigenvalue	R2Y	R2Y(cum)	Q2	Limit	Q2(cum)
0	Cent.							
1	0.297	0.297	7.12	0.301	0.301	0.26	0.05	0.26
2	0.14	0.436	3.35	0.258	0.559	0.218	0.05	0.421
3	0.116	0.552	2.79	0.265	0.824	0.375	0.05	0.638
4	0.2	0.753	4.81	0.0337	0.857	0.096	0.05	0.673

PLSDA of water extracts of Bedica trichomes

Component	R2X	R2X(cum)	Eigenvalue	R2Y	R2Y(cum)	Q2	Limit	Q2(cum)
0	Cent.							
1	0.373	0.373	8.57	0.319	0.319	0.284	0.05	0.284
2	0.178	0.55	4.08	0.309	0.628	0.435	0.05	0.595
3	0.116	0.666	2.67	0.216	0.843	0.481	0.05	0.79
4	0.127	0.793	2.91	0.0305	0.874	0.0638	0.05	0.803
5	0.0323	0.825	0.744	0.0662	0.94	0.141	0.05	0.831
6	0.0513	0.876	1.18	0.0188	0.959	0.182	0.05	0.862
7	0.0274	0.904	0.63	0.0117	0.971	0.072	0.05	0.872

8	0.016	0.92	0.368	0.0139	0.985	0.149	0.05	0.891

PLSDA of choloform extracts of Bediol trichomes

Component	R2X	R2X(cum)	Eigenvalue	R2Y	R2Y(cum)	Q2	Limit	Q2(cum)
0	Cent.							
1	0.298	0.298	7.14	0.324	0.324	0.306	0.05	0.306
2	0.144	0.441	3.44	0.304	0.628	0.368	0.05	0.562
3	0.156	0.597	3.74	0.158	0.786	0.118	0.05	0.613
4	0.102	0.699	2.45	0.098	0.884	0.273	0.05	0.719

PLSDA of water extracts of Bediol trichomes

Component	R2X	R2X(cum)	Eigenvalue	R2Y	R2Y(cum)	Q2	Limit	Q2(cum)
0	Cent.							
1	0.412	0.412	9.89	0.265	0.265	0.222	0.05	0.222
2	0.129	0.541	3.09	0.287	0.553	0.317	0.05	0.469
3	0.0938	0.635	2.25	0.226	0.779	0.355	0.05	0.657
4	0.0602	0.695	1.45	0.078	0.857	0.234	0.05	0.737
5	0.0351	0.73	0.843	0.0804	0.937	0.194	0.05	0.788
6	0.0425	0.772	1.02	0.0266	0.964	0.192	0.05	0.829

PLSDA of choloform extracts of Bedrobinol trichomes

Component	R2X	R2X(cum)	Eigenvalue	R2Y	R2Y(cum)	Q2	Limit	Q2(cum)
0	Cent.							
1	0.391	0.391	9.39	0.281	0.281	0.247	0.05	0.247
2	0.216	0.518	5.18	0.290	0.571	0.355	0.05	0.514
3	0.110	0.717	2.64	0.190	0.761	0.343	0.05	0.681
4	0.064	0.782	1.55	0.061	0.822	0.016	0.05	0.686
5	0.071	0.853	1.72	0.038	0.861	0.102	0.05	0.718

PLSDA of water extracts of Bedrobinol trichomes

Component	R2X	R2X(cum)	Eigenvalue	R2Y	R2Y(cum)	Q2	Limit	Q2(cum)
0	Cent.							
1	0.131	0.131	3.14	0.316	0.316	0.241	0.05	0.241
2	0.0821	0.213	1.97	0.289	0.606	0.181	0.05	0.379
3	0.0515	0.264	1.24	0.291	0.896	0.11	0.05	0.447
4	0.0612	0.326	1.47	0.0394	0.936	-0.0613	0.05	0.413
5	0.112	0.437	2.68	0.0162	0.952	-0.0781	0.05	0.367
6	0.0713	0.509	1.71	0.0204	0.972	0.157	0.05	0.467
7	0.04	0.548	0.959	0.0161	0.988	0.301	0.05	0.627
8	0.0412	0.59	0.988	0.00406	0.992	0.02	0.05	0.635
9	0.0306	0.62	0.735	0.00383	0.996	0.116	0.05	0.677

PLSDA of choloform extracts of Bedrocan trichomes

Component	R2X	R2X(cum)	Eigenvalue	R2Y	R2Y(cum)	Q2	Limit	Q2(cum)
0	Cent.							
1	0.446	0.446	10.7	0.297	0.297	0.274	0.05	0.274
2	0.184	0.630	4.42	0.293	0.59	0.377	0.05	0.548
3	0.143	0.773	3.44	0.228	0.818	0.481	0.05	0.765
4	0.0669	0.840	1.60	0.043	0.861	0.121	0.05	0.794
5	0.0166	0.856	0.399	0.0775	0.938	0.244	0.05	0.844
6	0.0282	0.885	0.677	0.0176	0.956	0.0400	0.05	0.850

PLSDA of water extracts of Bedrocan trichomes

Component	R2X	R2X(cum)	Eigenvalue	R2Y	R2Y(cum)	Q2	Limit	Q2(cum)
0	Cent.							
1	0.203	0.203	4.88	0.308	0.308	0.121	0.05	0.121
2	0.29	0.494	6.97	0.214	0.522	0.252	0.05	0.343
3	0.0828	0.576	1.99	0.224	0.747	0.26	0.05	0.513
4	0.0731	0.65	1.76	0.105	0.852	0.153	0.05	0.588
5	0.0536	0.703	1.29	0.0864	0.938	0.423	0.05	0.762
6	0.0439	0.747	1.05	0.0238	0.962	0.125	0.05	0.792
7	0.0299	0.777	0.717	0.0166	0.979	0.16	0.05	0.825

Chapter 5

Proteomics Profiling of Medicinal *Cannabis* Trichomes

Nizar Happyana, Stefan Loroch, Albert Sickmann, Oliver Kayser

Stefan Loroch and Albert Sickmann supported and coordinated proteomics experiments of *Cannabis* trichomes using nano-LC-MS/MS.

Abstract

Medicinal *Cannabis* trichomes are well known as the main site of biosynthesis of secondary metabolite, especially cannabinoids and terpenoids. However the information of proteom in this tissue is still limited. In this study, proteomic profiling of *Cannabis* trichomes variety Bediol, were performed by nano-LC-MS/MS analysis. In total, 481 proteins were identified and classified based on their biological function into 13 categories including proteins involved in energy (20.8), primary metabolism (18.1%), protein synthesis (17.3%), secondary metabolism (6.9%), and disease/defense (5.8%). Although no flavonoids have been reported from *Cannabis* trichomes, however most of enzymes related to the main biosynthetic pathway of these compounds were detected. Moreover, all published enzymes involved in the biosynthesis of cannabinoids were also identified. Besides that, the identification of proteins related to the defense system supported the function of trichomes in the plant protection. Furthermore, it is the first proteomic analysis of *Cannabis* trichomes using nano-LC-MS/MS method.

1. Introduction

Cannabis sativa L. (*Cannabaceae*), is an important flowering plants that has been cultivated and used by humankind as a source of fiber, food, oil, and medicine since prehistoric time (Chopra 1969; Fleming and Clarke 2008). Cannabinoids are well known as the most responsible compounds for the biological activities of this plant. Due to its psychoactivity, Δ^9-tetrahydrocannabinol (THC) is the most studied and interesting compound on this class. Based on the content of THC and cannabidiol (CBD), the plant can be classified in three chemotypes, namely medicinal type (THC > CBD), fiber type (THC < CBD), and intermediate type (THC ≈ CBD).

The main site of cannabinoid production on this plant is the trichomes, small protrusions of epidermal origin on the surfaces of leaves and other organs of the plant (Schilmiller et al. 2008). There are 3 trichomes types that can be found in this plant, namely capitate-stalked, capitate-sessile and bulbous trichomes. The former one contains abundant cannabinoids and mostly can be found in the flower part. Capitate-stalked trichomes consist of two parts, the gland (head) and the stem. We noted on our previous work (Happyana et al. 2012a) that the stem of capitate-stalked trichomes contributes in the cannabinoids production. Capitate-sessile trichomes which look like hairs are present on flowers, stems and leaves, and are found either in the vegetative or the flowering stages. Bulbous trichomes, shaped like a balloon, are the smallest trichomes on the plant. They are generally made up of 2 cells but never more than four cells (Dayanandan and Kaufman 1976).

Proteomics as a tool used to identify proteins expressed in an organism or cell has been applied for studying *Cannabis*. Protein profile of the trichomes was analyzed and compared with the profiles of leaves and flowers of *Cannabis* (Raharjo et al. 2004). However this report could not identify the proteins involved in cannabinoid biosynthesis. Other proteomic approaches have been performed to study copper effects on protein profile of *Cannabis* root (Bona et al. 2007) and to analyze protein profile of fiber-type seed of this plant (Park et al. 2012). However, the previous proteomics reports did not identify enzymes involved in secondary metabolite biosynthesis. Furthermore, it indicated that proteom data of *Cannabis* are still limited.

In this report, we performed proteomic analysis on trichomes of medicinal *Cannabis*, variety of Bediol using nano-LC-MS/MS analysis. The purpose of this research is to investigate protein profile of the trichomes and classify the proteins based on their biological function.

Our results revealed some proteins involved in secondary metabolites biosynthesis, such as proteins involved in flavonoids, terpenoids, and cannabinoid biosynthesis.

2. Material and methods

2.1 Plant material

Standardized medicinal *Cannabis sativa* L., variety Bediol® (THC level, approx. 6 %, CBD level, approx. 7.5 %) was supplied by Bedrocan BV (the Netherlands). *Cannabis* plants were grown indoors under standardized conditions. They were initially generated from cuttings of standardized mother plants and cultivated under controlled, long day-light conditions (18 h/day). After the vegetative growth phase, the flowering stage was induced under a shorter (12 h/day) light regime for 8 weeks. The trichomes from the 7 week plant were isolated and analyzed. This plant at week 7 contains high cannabinoids. Plant specimens were assigned voucher numbers (D7.09.07.2012) and deposited at Technical Biochemistry Department, TU Dortmund. All plant handling and experimental procedures were carried out under the license No. 4584989 issued by the Federal Institute for Drugs and Medical Devices (BfArM), Germany.

2.2 Trichomes isolation

Trichomes were isolated according to Yerger et al. (Yerger et al. 1992) with slightly modification. Fresh flowers of *Cannabis* were put in the liquid nitrogen. Floral leaves and stigma on the flowers were removed using a forceps with occasionally the flowers were put back into the liquid nitrogen. Afterward, a 5-10 gram of the flower was transferred into a 50 ml falcon tube and placed into liquid nitrogen. The tube containing the flower was removed from the liquid nitrogen tank and approximately 2 to 3 cm^3 of finely powdered dry ice (prepared by wrapping a piece in clean paper towels and crushing with a pestle) was added to the tube. Subsequently, the tube was loosely capped and vortexed at maximum speed for approximately one minute and flowers in the tube were removed. For obtaining trichomes, the content of each tube was sieved through a nylon net filter with 140 μm pore diameter (Merck Millipore, Germany) into a 500-mL beaker glass surrounded by dry ice. The trichomes in the beaker were transferred into a 2 ml frozen microcentrifuge tube with spatula and placed in the liquid nitrogen.

2.3 Protein Extraction

200 mg of frozen fresh trichomes is transferred into 2 ml Eppendorf tubes. 600 µl of chilled Tris pH 8.8 buffered phenol and 600 µl of chilled extraction buffer (0.1 M Tris-HCl pH 8.8, 10 mM EDTA, 0.4% β-mercaptoethanol, 0.9 M sucrose) are added in the tube. Afterward the tube is vortexed until all trichomes are wet, and followed with shaking for 30 min at 4 °C. The tube is centrifuged for 10 min at maximum speed, 4 °C. The phenol phase is transferred (the top phase) to a new 2 ml tube and placed on ice. 1.5 ml of cold 0.1 M ammonium acetate in 100% methanol (chilled at –20 °C) is added in the phenol phase for participating proteins and then it incubated at –20 °C for 4 hours or overnight. Afterward, it is centrifuged for 20 min at 4 °C and at maximum speed. The supernatant is removed and dispose in the non-chlorinated waste. The pellet protein is washed by adding 1.5 ml of COLD 0.1 M ammonium acetate in methanol (-20 °C) and placed at –20 °C for 5 min, and then centrifuged for 10 min, at maximum speed and 4 °C. The supernatant is removed. The washing step is repeated as mentioned above 1X more with cold 0.1 M ammonium acetate in methanol and 1X with cold acetone. The pellet protein is stored at 20 °C.

2.4 Proteolytic Digestion

Proteins were reduced with 10 mM DTT for 30 min at 56 °C and subsequently carbamidomethylated by incubation with 20 mM IAA for 30 min at room temperature in the dark. Samples were diluted 9-fold with ice-cold ethanol and stored at −40 °C for 1 h, followed by centrifugation at 4 °C at 20 000g for 30 min. The supernatant was carefully removed and protein pellets were resolubilized in 2 M GuHCl. For proteolytic digestion, trypsin was added in a 1:50 (w/w) ratio and samples were incubated for 15 h at 37 °C. Generated peptides were desalted using Hypersep C18 (Thermo Scientific) spin tips (10-200µl) equilibrated with three column volumes of 50% MeCN and three volumes of 0.1 % TFA. Samples were loaded 2x in 0.1% TFA, tips were washed with 3x 0.1% TFA, and elution was done in 3x 50% MeCN. Eluates were dried completely in a vacuum centrifuge and redissolved in 10 mM NH_4 acetate, pH6 to a concentration of approx. 1µg/µl.

2.5 Fractionation via high-pH-Reversed Phase

HPLC-based fractionation was performed using an U3000 HPLC (Dionex/Thermo Scientific) with a 150 × 0.5 mm Zorbax 300SB C18 column (Agilent Technologies, California, US, 3.5 µm of particle size) as stationary phase. The mobile phase was 10 mM NH_4 acetate, pH 6

(buffer A) and 10 mM NH_4 acetate, 84% MeCN pH 6 (buffer B). 15 µL of the resolubilized peptide mix was injected and separated at a flow rate of 5 µL/min via the following gradient: 3% B for 13 min, followed by 13% to 50% B in 65 min, 50% to 60% in 5 min, and 60% to 95% B in 4 min. Afterwards, B was kept at 95% for 5 min. Five concatenated fractions were collected in 30 s intervals starting at minute 15 till 75, so each fraction corresponded to 24 times 30 seconds. Collected fractions were dried in the vacuum centrifuge.

2.6 Nano-LC−MS/MS

For LC−MS analysis, fractions were dissolved in 100µl of 0.1% TFA. 15 % of each were analyzed by RPLC-ESI-MS/MS using an orbitrap mass spectrometer (Q-Exactive, Thermo Scientific) online coupled to a nanoHPLC system (U3000 RSLCnano, Dionex/Thermo Scientific). Samples were loaded in 15 µl onto a trap column (C18, 100 µm × 2 cm, 100Å pore size, Acclaim PepMap, Thermo Scientific) at a flow rate of 20 µL/min 0.1% TFA for 10 min and subsequently separated on the main column (C18, 75 µm × 50 cm, 100Å pore size, Acclaim PepMap RSLC, Thermo Scientific) using a linear gradient, at a flow rate of 230 nL/min at 60 °C with solvent A 0.1% FA and solvent B 0.1% FA, 84% ACN. The gradient was 5% to 45% B in 90 min, and from 45% to 95% in 5 min. Afterwards B was kept at 95% for 5 min and at 5% for 10 min. The Q-Exactive was operated in data-dependent acquisition mode with a full MS range from 300 to 1500 m/z at a resolution of 70,000. The 15 most intense ions were fragmented by HCD with normalized collision energy of 27. MS/MS spectra were recorded at a resolution of 17,500. Target values and maximum fill times were set to 3×10^6 and 120 ms for MS and 5×10^4 and 250 ms for MS/MS, respectively.

2.7 Data Interpretation

MS/MS spectra were searched using MASCOT 2.4 (Matrix Science), implemented in Proteome Discoverer 1.3 (Thermo Scientific), against the taxonomy *Viridiplantae of* SwissProt (UniProt and TrEMBl, 30th of July 2012), and Uniprot Cannabaceae (13th of January 2013) including 22 synthesis proteins manually added. The Cannabaceae/synthesis protein database was merged with a *Pyrococcus furiosis (Pfu) database* (3rd of August 2011) for false discovery rate (Vaudel et al. 2012) estimation. Carbamidomethylation of cysteines was set as fixed modification, oxidation of methionine and acetylation on protein N-termini as variable. Mass tolerances were set to 10 ppm for MS and 0.02 Da for MS/MS. Results were filtered using a false discovery rate of ≤1% and only rank 1 spectrum matches were

considered. For the searches against the viridiplantae database protein hits were only considered if more than 3 unique peptides were identified and for the search against Cannabaceae if 2 unique peptides were identified. Pfu proteins were removed from the list manually.

3. Results and Discussion

Totally 481 proteins were identified from protein extracts of *Cannabis* trichomes. The identified proteins were functionally classified into 13 categories using the UniProt database (http://www.uniprot.org) as described in figure 1 and supplementary table. The annotation was assigned mainly based on the dominant function even if the proteins may have various roles in the different compartments or different development stages. The largest functional category was proteins regulating energy (20.8%, 100 proteins). The next most abundant categories were other house-keeping proteins involved in primary metabolism (18.1%) and protein synthesis (17.3%). Unclear classification group (11.0%) including proteins with several potential functions or unknown function was the next abundant category, and followed by proteins involved in protein destination and storage (7.7%), secondary metabolism (6.9%), disease/defense (5.8%), transporters (2.9%), transcription (2.7%), intracellular traffic (2.5%), cell structures (2.1%), cell growth/division (1.9%), and signal transduction (0.4%). Furthermore, the list of 30 best identified proteins is shown in table 1. Not only house-keeping proteins but also enzymes participating in secondary metabolism were interestingly noted in this list, such as THCA synthase (THCAS) and acyl activating enzyme 1 that corresponding in cannabinoid biosynthesis. Besides that, (+)-alpha-pinene synthase and (-)-limonene synthase involved in monoterpene synthesis were recorded in the list as well.

3.1 Proteins involved in energy

Proteins participating in glycolysis were identified in *Cannabis* trichomes including hexokinase, phosphoglucose isomerase, phosphofructokinase, fructose-bisphosphate aldolase, triosephosphate isomerase, 3-glyceraldehyde phosphate dehydrogenase, enolase, and other glycolytic proteins. These proteins are considered as important suppliers of essential metabolic flexibility that facilitates plant development and acclimation to environmental stress (T. Wu et al. 2012). Some proteins involved in tricarboxylic acid cycle (TCA cycle) were also detected in the *Cannabis* trichomes such as citrate synthase, aconitase, isocitrate dehydrogenase, and succinyl-CoA synthetase. This TCA cycle releases high-energy phosphate

compounds that serve as the main source of cellular energy (Park et al. 2012). ATP synthase complex (alpha, beta, gamma subunits) and V-type ATPase (V-ATPase) that involved in transport electron chain were identified in the *Cannabis* trichomes. This ATPase family is important enzymes in energy production and in maintaining ion concentration balance in plant cells (T. Wu et al. 2012). Other proteins related to electron transport chain including NADH dehydrogenase, succinate dehydrogenase, ferredoxin-NADP reductase, and cytochrome c subunits were detected in the trichomes as well. The identification of proteins involved in glycolysis, TCA cycle, and transport electron chain, indicated that the trichomes produce a lot of energy. This energy is needed for powering cellular reactions in the *Cannabis* trichomes, such as fatty acid metabolisms, and secondary metabolite biosynthesis. Moreover this result is in accordance with previous report (G. J. Wagner 1991).

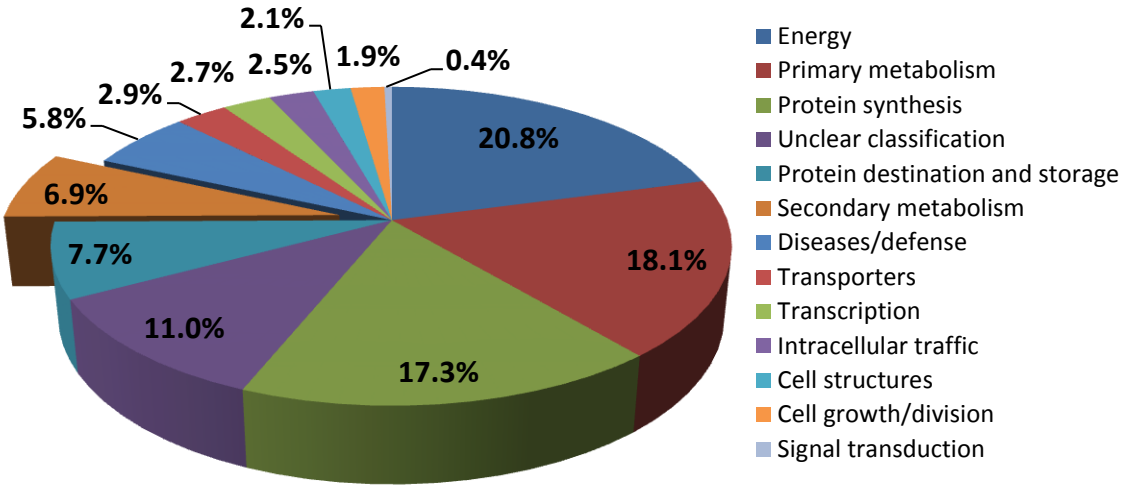


Figure 1. Functional group distribution of the identified proteins in the medicinal *Cannabis* trichomes variety Bediol. The pie chart describes the number of identified proteins classified by function. The secondary metabolism group is highlighted in bold print and shown protruding from the pie graph.

3.2 Proteins involved in primary metabolism

This category is dominated by proteins involved in the biosynthesis of amino acids, carbohydrates, and lipids. 27 proteins related to amino acids biosynthesis were detected in the trichomes including asparagine synthetase, threonine synthase, glutamate synthase, and glutamine synthetase. The products of these enzymes were also found in the *Cannabis* trichomes as shown in our metabolomics works. Moreover, the product of asparagine

synthetase was an important discriminant compound to distinguish metabolite profile of the trichomes from the leaves of *Cannabis*. In total, 19 proteins participating in carbohydrate metabolism were detected in *Cannabis* trichomes. One of them is sucrose-UDP glucosyltransferase that has an important role in providing UDP-glucose and fructose for various metabolic pathways. In *Cannabis*, some flavonoids are glycosylated and a UDP glycosyltransferase was predicted as the responsible enzyme for the glycosylation (Flores-Sanchez and Verpoorte 2008). Therefore, sucrose-UDP glucosyltransferase probably might be a candidate enzyme that supplies sucrose and glucose to biosynthesis pathway of *Cannabis* flavonoids. A total of 15 lipid metabolism-related proteins were recorded in the *Cannabis* trichomes, such as acetyl-CoA carboxylase. This enzyme catalyzes the carboxylation of acetyl-CoA, forming malonyl-CoA, which is used in the plastid for fatty acid synthesis and in the cytosol involved in various biosynthetic pathways including, cannabinoid, flavonoid biosynthesis and fatty acid elongation (Baud et al. 2003). Some studies have reported active lipid metabolism in trichomes of several plant species (Besser et al. 2009; Falara et al. 2008; Fridman et al. 2005; Aziz et al. 2005). They suggested that the lipid metabolism responsible for wax production in the trichomes.

3.3 Proteins involved in protein synthesis

This category is the third largest protein group detected in the *Cannabis* trichomes. This group consists of 83 identified proteins in total and is dominated by ribosomal proteins predominantly. These proteins cooperate with ribosomal RNA (rRNA) in the cellular process of translation and transcription. Besides that several proteins categorized as eukaryotic translation initiation factors and elongation factors were found belonging to this group as well. It has been reported that the changes in the level of these proteins can be correlated to active role of trichomes in the metabolite biosynthesis and protecting the plant against biotic and abiotic threats, which needs new proteins (T. Wu et al. 2012).

3.4 Proteins involved in secondary metabolism

In this study, many proteins participating in the biosynthesis of secondary metabolites were recorded, including flavonoid, terpenoid, and cannabinoid pathways. Interestingly, all transcripts of identified proteins related to these pathways were also detected as described in our transcriptomic work.

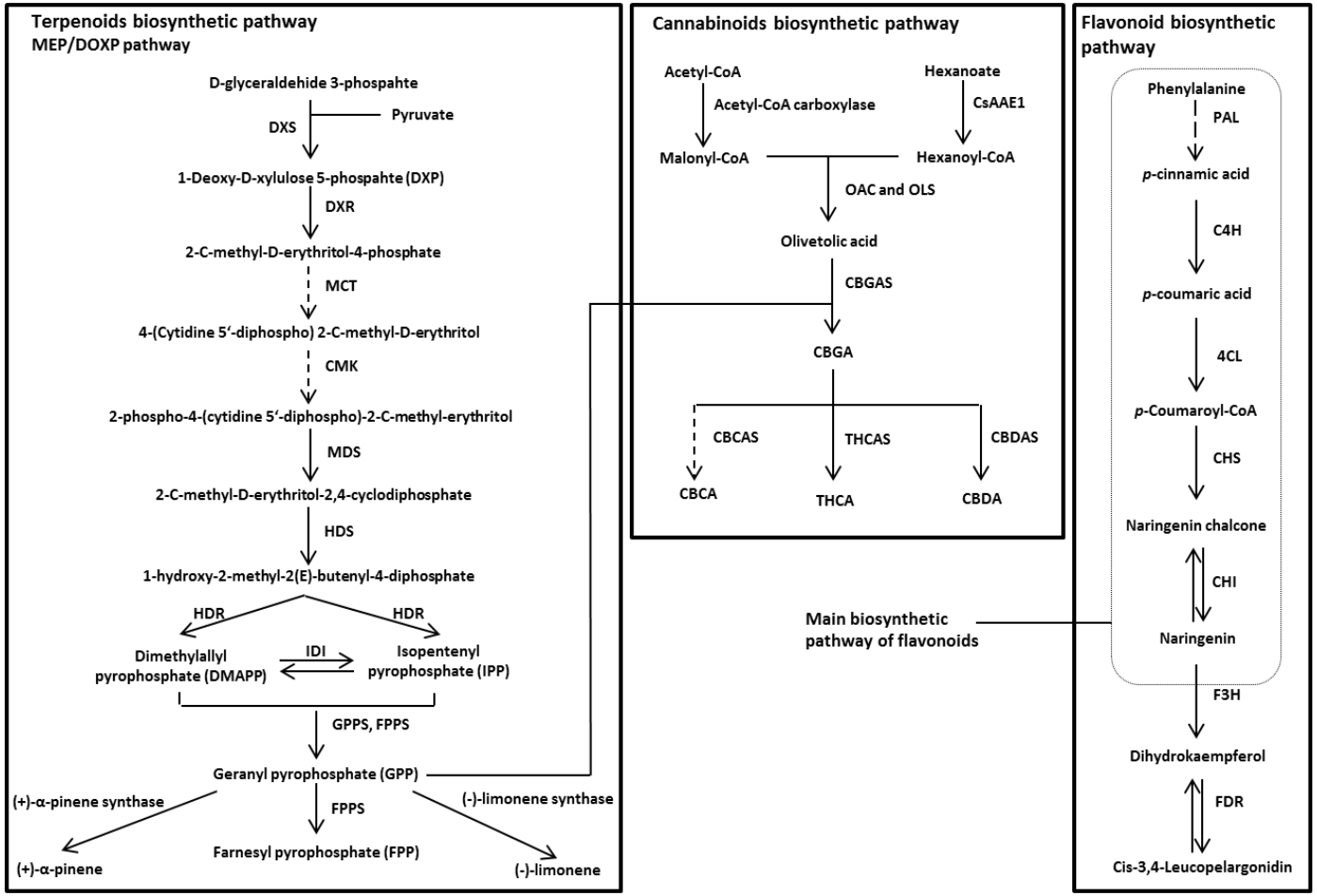
3.4.1 Proteins related to flavonoid biosynthesis

There are more than 20 flavonoids that have been reported in *Cannabis sativa* L., (ElSohly and Slade 2005) representing 7 chemical structures which can be glycosylated, prenylated or methylated (Flores-Sanchez and Verpoorte 2008). These flavonoids were only identified in flowers, leaves, twigs and pollen (Ross et al. 2005; Segelman et al. 1978; Vanhoenacker et al. 2002). Interestingly, most of enzymes involved in the main biosynthetic pathway of flavonoids were detected in the proteomics data of *Cannabis* trichomes as described in figure 2. The first enzyme involved in this pathway, phenylalanine ammonia lyase (PAL) which converts phenylalanine to *p*-cinnamic acid, unfortunately was not detected. Nevertheless, cinnamate 4-hydroxylase (C4H) that hydroxylates the product of the first enzyme to *p*-coumaric acid was identified. Besides that, 4-coumarate:CoA ligase (4CL) which adds a CoA thiol ester into *p*-coumaric acid yielding *p*-coumaroyl-CoA, was recorded as well. Moreover, other enzymes related to main biosynthetic pathway of flavonoids were also detected in the *Cannabis* trichomes including chalcone synthase (CHS) and chalcone isomerase (CHI). CHS condenses one molecule of *p*-coumaroyl-CoA and three molecules of malonyl-CoA that supplied by acetyl-CoA carboxylase yielding a naringenin chalcone. Meanwhile, CHI isomerizes naringenin chalcone to naringenin subsequently. Naringenin is a flavanone and a common precursor for the biosynthesis of other derivative flavonoids, such as flavones and flavonols. Overall, the identification of enzymes involved in this main biosynthetic pathway indicated that the *Cannabis* trichomes might synthesize flavonoids as well.

Other enzymes related to flavonoid biosynthesis were also identified in the trichomes, namely flavanone 3-hydroxylase (F3H), dihydroflavonol-4-reductase (FDR), and isoflavone reductase (IFR) (figure 2). F3H catalyzes the 3-beta-hydroxylation of 2S-flavanones to 2R,3R-dihydroflavonols which is intermediates in the biosynthesis of flavonols, anthocyanidins, catechins and proanthocyanidins (Britsch et al. 1993). In *Cannabis*, F3H converts naringenin into dihydrokaempferol. Meanwhile, FDR catalyzes the NADPH-dependent conversion of stereospecific dihydroflavonols such as dihydrokaempferol and dihydroquercetin into leucoanthocyanidins (or flavan-3,4-diols) (P. Zhang et al. 2008). Furthermore it is one of the key enzymes controlling metabolic flux into biosynthetic pathway branches leading to anthocyanins and proanthocyanidins (D. Y. Xie et al. 2004). The last enzyme catalyzes a stereospecific NADPH-dependent reduction of isoflavones to (3R)-isoflavanones.

Table 1. List of 30 best identified proteins in the *Cannabis* trichomes.

Identified protein	Uniprot accession number	Score	Coverage	Unique Peptides	MW [kDa]	calc. pI	organism	Biological Function
acyl-activating enzyme 2	AFD33346.1	7984.01	33.97	39	140.69	7.88	Cannabis sativa	Primary metabolism
ATP synthase subunit beta	F8TR83	9982.47	78.21	26	44.56	5.17	Cannabis sativa	Energy
Chaperone protein ClpC1	Q9FI56	3485.76	26.48	25	103.39	6.77	Arabidopsis thaliana	Intracellular traffic
Tetrahydrocannabinolic Acid Synthase	NH1319	7442.23	63.90	25	58.63	9.00	Cannabis sativa	Secondary metabolism
ATP synthase subunit beta	Q09X10	10029.02	65.46	23	53.76	5.59	Morus indica	Energy
ATP synthase subunit beta	P17614	11600.19	51.25	21	59.82	6.34	Nicotiana plumbaginifolia	Energy
Betv1-like protein	I6XT51	17844.62	95.65	20	17.60	5.39	Cannabis sativa	Unclear classification
ATP synthase subunit alpha	B0FA21	5940.99	50.61	19	44.44	6.93	Humulus lupulus	Energy
ATP synthase subunit alpha	P24459	6809.44	45.28	19	55.31	6.93	Phaseolus vulgaris	Energy
Polyketide synthase protein	F1LKH5	9461.61	49.61	18	42.54	6.54	Cannabis sativa	Secondary metabolism
Photosystem II CP47 chlorophyll apoprotein	A9XV91	2550.73	34.43	17	53.66	6.60	Cannabis sativa	Energy
Elongation factor 2	O23755	3141.92	18.03	15	93.74	6.30	Beta vulgaris	Protein synthesis
acyl-activating enzyme 1	AFD33345.1	1507.72	28.51	15	79.51	7.11	Cannabis sativa	Secondary metabolism
Serine hydroxymethyltransferase	Q9SZJ5	2330.02	34.24	14	57.36	8.13	Arabidopsis thaliana	Disease/defense
Photosystem II CP47 chlorophyll apoprotein	A4QKL8	1875.30	22.05	14	56.05	6.89	Capsella bursa-pastoris	Energy
26S protease regulatory subunit 6A homolog A	Q9SEI2	2251.85	40.33	14	47.45	5.03	Arabidopsis thaliana	Protein destination and storage
(+)-alpha-pinene synthase	A7IZZ2	1602.64	25.04	14	71.80	6.30	Cannabis sativa	Secondary metabolism
4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	F4K0E8	3579.60	17.95	14	82.20	6.43	Arabidopsis thaliana	Secondary metabolism
Isopentenyl-diphosphate isomerase	G9C073	2011.58	37.38	14	36.58	7.28	Humulus lupulus	Secondary metabolism
Heat shock 70 kDa protein	Q01899	5930.10	30.37	13	72.49	6.20	Phaseolus vulgaris	Disease/defence
Biotin carboxylase 2	B9N843	2508.86	30.61	13	57.46	6.93	Populus trichocarpa	Metabolism
Chalcone isomerase-like protein	I6WIE9	2265.31	61.68	13	23.70	8.16	Cannabis sativa	Secondary metabolism
Oxygen-evolving enhancer protein 1	Q40459	5021.26	31.33	12	35.21	5.96	Nicotiana tabacum	Energy
Phosphoglycerate kinase	Q42961	4009.28	30.98	12	50.15	8.38	Nicotiana tabacum	Energy
ATP-citrate synthase beta chain protein 1	Q93VT8	3155.48	22.37	12	66.03	7.68	Oryza sativa subsp. Japonica	Metabolism
26S protease regulatory subunit 7	O64982	1065.88	32.94	12	47.53	6.80	Prunus persica	Protein destination and storage
(-)-limonene synthase	A7IZZ1	792.21	21.54	12	72.34	6.92	Cannabis sativa	Secondary metabolism
Ribulose bisphosphate carboxylase large chain	P48698	7071.57	26.92	11	51.92	7.02	Datura stramonium	Energy
Ferredoxin-NADP reductase	Q41014	2268.55	27.59	11	42.29	8.66	Pisum sativum	Energy



DXS, 1-deoxyxylulose-5-phosphate synthase; DXR, 1-deoxy-D-xylulose-5- phosphate reductoisomerase; MCT, 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase; CMK, 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase; MDS, 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; HDS, 4-hydroxy-3-methylbut-2-enyl diphosphate synthase; HDR, 4-hydroxy-3-methylbut-2-enyl diphosphate reductase; IDI, IPP/DMAPP isomerase; GPPS, geranyl pyrophosphate synthase; FPPS, farnesyl pyrophosphate synthase; CsAAE1, acyl-activating enzyme 1; OAC, olivetolic acid cyclase; OLS, olivetol synthase; CBGAS, cannabigerolic acid synthase; THCAS, Δ^9-tetrahydrocannabinolic acid synthase; CBDAS, cannabidiolic acid synthase; CBCAS, cannabichromenic acid synthase; PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate:CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; FDR, dihydroflavonol-4-reductase.

Figure 2. Proteomic identification of pathways in secondary metabolite synthesis. Identified enzymes are represented by solid lines.

3.4.2 Proteins related to terpenoid biosynthesis

Isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) are building blocks of terpenoid compounds and called as isoprenoid. Both are synthesized in the cytosol via a well-characterized metabolic pathway, mevalonate (MVA) pathway. Meanwhile in the plastids, isoprenoids are synthesized with another pathway, 2-C-methyl-D-erythritol-4-phosphate/1-deoxy-D-xylulose-5-phosphate (MEP/DOXP). MVA pathway usually supplies isoprenoids for synthesis of sesquiterpenes and triterpens, while MEP/DOXP pathway mostly provides precursors for the formations of monoterpenes and diterpenes. However only enzymes involved in the MEP/DOXP pathway were detected in this study as seen in figure 2. Five of seven enzymes related to this pathway were detected successfully in the *Cannabis* trichomes, namely 1-deoxy-D-xylulose-5-phosphate synthase (DXS), 1-deoxy-D-xylulose-5-phosphate reductoisomerase (DXR), 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase (MDS), 4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase (HDS), and 4-hydroxy-3-methylbut-2-enyl diphosphate reductase (HDR). Meanwhile the undetected enzymes are 2-C-methyl-D-erythritol-4-phosphate cytidyltransferase (MCT) and 4-C-diphosphocytidyl-2-C-methyl-D-erythritol kinase (CMK).

In addition, isopenthenyl pyrophosphate isomerase (IDI), an enzyme catalyzes the isomerization of IPP to DMAPP, was also successfully detected in the trichomes (supplementary table). Moreover, farnesyl pyrophosphate synthase (FPPS), an enzyme catalyzes sequential condensation reactions of DMAPP with 2 units of IPP to form farnesyl pyrophosphate (FPP), was identified as well. The intermediate compound of this reaction is geranyl pyrophosphate (GPP), a condensation product of one molecule of IPP with one molecule of DMAPP. Furthermore, GPP is an important precursor in cannabinoid biosynthesis. Based on incorporation experiments using ^{13}C-labeled glucoses, GPP related to cannabinoid metabolism is provided entirely or predominantly (> 98%) from MEP/DOXP pathway (Fellermeier et al. 2001). Besides that, the content of cannabinoids is very dominant in the *Cannabis* trichomes as described in the metabolomics chapters. Therefore, these might be the reasons why terpenoid enzymes that are involved in the MEP/DOXP pathway could be detected easily in the proteomics data of *Cannabis* trichomes.

Other enzymes involved in the downstream steps of terpene metabolism were also identified, including (-)-limonene synthase and (+)-α-pinene synthase (see supplementary table). Both responsible for formation of monoterpenes, namely (-)-limonene and (+)-α-pinene

respectively. These enzymes previously were isolated and characterized from *Cannabis* trichomes variety Skunk (Günnewich et al. 2007). Two enzymes related to β-carotene biosynthesis were detected as well, namely phytoene dehydrogenase and ζ-carotene desaturase. The first enzyme converts phytoene into ζ-carotene via the intermediary of phytofluene by the symmetrical introduction of two double bonds at the C-11 and C-11' positions of phytoene (Bartley et al. 1999). Meanwhile the last enzyme catalyzes the conversion of ζ-carotene to lycopene via the intermediary of neurosporene (Bartley et al. 1999).

3.4.3 Proteins related to Cannabinoid biosynthesis

There are more than 100 cannabinoids that have been reported from *Cannabis sativa* L. Biosynthesis of these compounds is initiated by the formation of 3 important starting substrates, namely GPP, malonyl-CoA, and hexanoyl-CoA (figure 2). GPP is supplied from MEP/DOXP pathway as explained before, and malonyl-CoA is provided by acetyl-CoA carboxylase. Meanwhile, 2 enzymes were reported catalyzing the conversion of hexanoate to hexanoyl-CoA, namely acyl-activating enzyme 1 (CsAAE1) and acyl-activating enzyme 3 (CsAAE3) (Stout et al. 2012a). Both are enzymes related to fatty acid metabolism and were detected in the trichomes as well. However the last enzyme localized to peroxisome, while CsAAE1was found in the cytosol which is the same compartment where olivetol synthase (OLS), a cannabinoid enzyme, is localized. Therefore, CsAAE1 was suggested as the enzyme that responsible for supplying hexaonyl-CoA to the biosynthetic pathway of cannabinoids (Stout et al. 2012a).

In this study, OLS was also indetified in the *Cannabis* trichomes. It is a type III polyketide synthase (PKS) enzyme and catalyzes the formation of olivetol from malonyl-CoA and Hexaonyl-CoA (Taura et al. 2009). Olivetol is the decarboxylated form of olivetolic acid, an important precursor in cannabinoid biosynthesis. OLS requires the presence of olivetolic acid cyclase (OAC) for catalyzing the condensation of malonyl-CoA and hexaonyl-CoA to form olivetolic acid (Gagne et al. 2012). OAC is a polyketide cyclase enzyme and was detected in the trichomes as well. Moreover it is a dimeric α+β barrel (DABB) protein that is structurally similar to polyketide cyclases from *Streptomyces* species. Furthermore, the transcript analysis showed that this enzyme is present at high levels in the trichomes (Gagne et al. 2012).

In the next biosynthetic step, GPP and olivetolic acid are condensed to form cannabigerolic acid (CBGA). An enzyme isolated from young leaves of *Cannabis* has been reported catalyzing this condensation (Fellermeier and Zenk 1998). This enzyme is predicted to be a representative of the geranyltransferase group and noticed as geranylpyrophosphate: olivetolate geranyltransferase (GOT). However, the sequence of this enzyme has not yet been deposited into the public database, thus it could not be checked on our proteomic data. We also tried to check an enzyme from our college that predicted as CBGA synthase (unpublished), but this enzyme could not be detected as well.

CBGA is an important precursor for other cannabinoids. The compound is subsequently transformed to Δ^9-tetrahydrocannabinolic acid (THCA), cannabidiolic acid (CBDA) and cannabichromenic acid (CBCA) by THCAS (Taura et al. 1995), CBDA synthase (CBDAS) (Taura et al. 1996), and CBCA synthase (CBCAS) (Morimoto et al. 1998), respectively. THCAS and CBDAS were successfully identified on our proteomic data. THCAS is a monomeric and stereoselective protein, and catalyzes a unique monoterpene cyclase-like reaction coupled with two-electron oxidation. Moreover, the reaction of this enzyme does not require any metal ions, cofactors, coenzymes, and can complete the oxidocyclization by itself (Taura et al. 2007b). Besides that the recombinant THCAS has been overexpressed on a baculovirus-insect expression system, transgenic tobacco hairy roots (Sirikantaramas et al. 2004), and a methylotrophic yeast *Pichia pastoris* (Taura et al. 2007a). Properties of CBDAS reaction is similar to THCAS, it is stereoselective, does not need metal ions, cofactors, and coenzymes (Taura et al. 1996). This enzyme has been overexpressed in the insect cell cultures of *Spodoptera frugiperda* (Taura et al. 2007c). Although CBCAS was purified to homogeneity (Morimoto et al. 1998), its sequence has not yet been stored into the public databases, thus we could not check this enzyme. However the properties of this protein might be closed to the THCAS and CBDAS (Morimoto et al. 1998).

All cannabinoid enzymes that their sequences are available in the public databases have been successfully recorded in the proteomic data of *Cannabis* trichomes. This is one of the most important points in this study since cannabinoids are the most studied and interesting compounds in this plant, as a consequence of their biological activities. Moreover, it is confirmed that *Cannabis* trichomes are the main site of cannabinoid biosynthesis.

3.5 Proteins involved in disease/defense (stress responses)

Trichomes perform some important biological functions in the plant defense system. Some studies have reported that trichomes play important roles in protecting plants against destructive insects and herbivores (Romero et al. 2008; Van Dam and Hare 1998; Simmons and Gurr 2005). Moreover trichomes reduce leaf temperature, increase of light reflectance (G. J. Wagner 1991; G. J. Wagner et al. 2004), prevent water loss, and minimize leaf abrasion (Gonzalez et al. 2008). In accordance with these reports, some proteins involved in the detoxification, biotic and abiotic stress responses, and repairing enzymes, were identified in the *Cannabis* trichomes. These results supported the function of trichomes as an important self-defense system for plants.

Several heat shock proteins (HSPs) that associated with thermo-tolerance were detected in *Cannabis* trichomes (see supplementary table). HSPs prevent proteins from unfolding under heat stress condition. Heat shock 70 kDa protein 15 is an example of these proteins. The protein is likely to be a key factor in proper folding of cytosolic proteins, and may function as nucleotide exchange factor (Jungkunz et al. 2011). Another example is heat shock cognate protein 80, an enzyme that promotes the maturation, structural maintenance and proper regulation of specific target proteins involved for instance in cell cycle control and signal transduction (Koning et al. 1992).

Some enzymes related to detoxification processes were identified in the *Cannabis* trichomes as shown in supplementary table. Delta-1-pyrroline-5-carboxylate dehydrogenase 12A1 is a member of this category. This enzyme plays a role in the inhibition of programmed cell death by converting the toxic proline catabolism intermediate (s)-1-pyrroline-5-carboxylate (P5C) to glutamate (Deuschle et al. 2001). Other proteins involved in the detoxification are superoxide dismutase, phospholipid hydroperoxide glutathione peroxidase, lactoglutathione lyase, and aldo-keto reductase family 4 member C9. The first enzyme destroys radicals which are toxic to biological systems, and protects cellular components from being oxidized by reactive oxygen species (ROS) (Alscher et al. 2002). While, phospholipid hydroperoxide glutathione peroxidase protects cells and enzymes from oxidative damage, by catalyzing the reduction of hydrogen peroxide, lipid peroxides and organic hydroperoxide (Faltin et al. 2010). Lactoglutathione lyase participates in the detoxification of methylglyoxal and other reactive aldehydes that are produced by cells either during normal or stress conditions (Yadav et al.

2005). Meanwhile, the last protein reduces a range of toxic aldehydes and ketones produced during stress environments (Simpson et al. 2009).

Glutathione-ascorbate cycle plays a key role in detoxification hydrogen peroxide. The enzymes related to this cycle were detected in the trichomes, including ascorbate peroxidase, monodehydroascorbate reductase, and glutathione reductase. Ascorbate peroxidase catalyzes the reduction of hydrogen peroxide using ascorbate as the electron donor and yielding monodehydroascorbate and water (Raven 2000). Monodehydroascorbate reductase transforms back monodehydroascorbate to ascorbate involving NADH and H^+ (Wells and Xu 1994). However, monodehydroascorbate is a radical and if not rapidly reduced, it disproportionates into dehydroascorbate and ascorbate. Afterward, dehydroascorbate reductase catalyzes the reduction of dehydroascorbate with involving glutathione, and resulting ascorbate and oxidized glutathione. In the last step, glutathione is regenerated from its oxidized form by glutathione reductase (Meister 1988). Other proteins involved in responding hydrogen peroxide were detected as well, such as catalase-2, catalase isozyme 1, 2-Cys peroxiredoxin BAS1, and peroxiredoxin Q. The first 2 enzymes protect cells with catalyzes the decomposition of hydrogen peroxide to water and oxygen (Chelikani. P et al. 2004). Meanwhile the other proteins use thioredoxins as reducing substrates for removing hydrogen peroxide (Rhee et al. 2001).

Serine hydroxymethyltransferase 1 is one of proteins associated with repair enzymes and found in the *Cannabis* trichomes. This enzyme plays a critical role in controlling the cell damage provoked by abiotic stresses such as high light and salt and in restricting pathogen-induced cell death. Moreover, the enzyme participates also in supporting the photorespiration as the dissipative mechanisms of plants to minimize production of ROS at the chloroplast and to mitigate oxidative damage (Moreno et al. 2005). Another repair enzyme, peptide methionine sulfoxide reductase, was detected as well. It repairs proteins that have been inactivated by oxidation with catalyzing the reversible oxidation-reduction of methionine sulfoxide in proteins to methionine (Weissbach et al. 2002).

Iron is required for normal cell growth and proliferation. However, excess free iron is toxic to cells because it acts as a catalyst in the formation of free radicals from ROS via the Fenton Reaction (Linn 1998). Thus cells have developed a regulation for controlling iron levels. Ferritin is the protein that plays an important role in this regulation and was also identified in the trichomes. This protein binds ferrous iron and deposits it in the ferric state, a soluble, non-

toxic, and readily available form (Orino et al. 2001). In addition, a protein protecting the plant from bacterial infection, protein phosphatase 2C 59, was detected in *Cannabis* trichomes as well. The protein modulates defense response to pathogenic bacteria, conferring resistance and promoting salicylic acid accumulation (Lee et al. 2008).

4. Conclusions

In this work, proteomic profiling of medicinal *Cannabis* trichomes was performed with nano-LC-MS/MS. It shed more light on protein metabolism of *Cannabis* trichomes. This approach revealed not only housekeeping proteins but also a set proteins belonging to defense responses in the plants. It is consistent with the assigned functions of trichomes in protecting the plant against biotic and abiotic threats. Moreover this proteomic analysis is the first evidence of flavonoids from *Cannabis* trichomes in the protein level. Furthermore, the identification of enzymes involved in the biosynthetic pathway of cannabinoids, flavonoids, and terpenoids confirmed the function of *Cannabis* trichomes as the main site of secondary metabolite production.

Supplementary data of chapter 5

Supplementary table

Identified proteins of medicinal *Cannabis* trichomes variety Bediol

Identified Protein	Uniprot accession number	Score	Coverage	Unique Peptides	PSMs	AAs	MW [kDa]	calc. pI	organism
Energy									
2,3-bisphosphoglycerate-independent phosphoglycerate mutase 1	O04499	386.49	6.46	4	10	557	60.54	5.53	Arabidopsis thaliana
2-phosphoglycerate dehydratase	Q42971	4373.87	27.13	6	89	446	47.94	5.57	Oryza sativa subsp. japonica
2-phosphoglycerate dehydratase 3	Q9ZW34	129.31	7.37	3	3	475	51.57	5.52	Arabidopsis thaliana
6-phosphofructokinase 7	Q9C5J7	324.41	7.63	3	9	485	53.45	7.31	Arabidopsis thaliana
Aconitate hydratase 1	Q42560	1181.54	10.91	9	30	898	98.09	6.40	Arabidopsis thaliana
Alcohol dehydrogenase 1	P12886	512.00	9.74	5	15	380	41.13	6.52	Pisum sativum
Apocytochrome f	Q2QD76	937.28	20.63	6	25	320	35.13	8.56	Cucumis sativus
Aspartate aminotransferase	P46248	1230.14	18.54	9	32	453	49.80	8.15	Arabidopsis thaliana
ATP synthase gamma chain	P29790	389.53	8.22	3	7	377	41.42	8.02	Nicotiana tabacum
ATP synthase subunit alpha	A9L981	600.96	12.43	3	14	507	54.99	5.72	Lemna minor
ATP synthase subunit alpha	P24459	6809.44	45.28	19	145	508	55.31	6.93	Phaseolus vulgaris
ATP synthase subunit alpha (Fragment)	B0FA21	5940.99	50.61	19	137	413	44.44	6.93	Humulus lupulus
ATP synthase subunit beta	P17614	11600.19	51.25	21	255	560	59.82	6.34	Nicotiana plumbaginifolia
ATP synthase subunit beta	Q09X10	10029.02	65.46	23	205	498	53.76	5.59	Morus indica
ATP synthase subunit beta(Fragment)	F8TR83	9982.47	78.21	26	218	413	44.56	5.17	Cannabis sativa (Hemp)
Chlorophyll a-b binding protein 3	P09756	2785.12	40.68	9	67	263	27.84	5.64	Glycine max
Citrate synthase	P49299	414.62	12.60	6	9	516	56.72	8.79	Cucurbita maxima
Cytochrome c	P00051	251.63	29.73	3	7	111	12.11	9.44	Cucurbita maxima
Cytochrome c	P00053	318.75	51.35	5	12	111	12.04	9.86	Cannabis sativa (Hemp)
Cytochrome c oxidase subunit 2	P93285	334.36	18.46	4	11	260	29.66	5.24	Arabidopsis thaliana
Cytochrome c1-1	P25076	805.85	16.56	3	16	320	35.14	7.23	Solanum tuberosum

D-fructose-1,6-bisphosphate 1-phosphohydrolase	A2WXB2	614.78	16.52	5	12	339	37.01	5.77	Oryza sativa subsp. indica
Enolase	P42896	6682.45	38.43	10	121	445	47.88	5.78	Ricinus communis
Enolase 1	Q9C9C4	1246.51	14.68	5	22	477	51.44	6.13	Arabidopsis thaliana
Ferredoxin-NADP reductase	P41343	500.31	12.60	4	10	365	41.04	8.38	Mesembryanthemum crystallinum
Ferredoxin-NADP reductase	Q41014	2268.55	27.59	11	60	377	42.29	8.66	Pisum sativum
Formate dehydrogenase	Q07511	873.74	9.19	4	24	381	42.01	7.12	Solanum tuberosum
Fructose-1,6-bisphosphatase	P25851	506.82	9.35	4	12	417	45.13	5.40	Arabidopsis thaliana
Fructose-bisphosphate aldolase	P08440	1829.64	7.61	4	37	355	38.58	7.61	Zea mays
Fructose-bisphosphate aldolase	Q09X32	3653.80	20.51	7	70	507	55.41	5.21	Morus indica
Fructose-bisphosphate aldolase	Q9SZX3	612.56	7.09	6	19	494	53.81	6.67	Arabidopsis thaliana
Fumarate hydratase 1	P93033	637.67	12.60	4	11	492	52.97	7.88	Arabidopsis thaliana
Glucose-6-phosphate 1-dehydrogenase 2	Q9FY99	1104.34	17.11	10	20	596	67.12	8.29	Arabidopsis thaliana
Glucose-6-phosphate isomerase	P34795	191.38	11.25	4	5	560	61.68	6.65	Arabidopsis thaliana
Glyceraldehyde-3-phosphate dehydrogenase	P04796	7515.37	37.87	3	152	338	36.90	7.93	Sinapis alba
Glyceraldehyde-3-phosphate dehydrogenase	P26518	6924.35	44.57	6	160	341	36.96	7.55	Magnolia liliiflora
Glyceraldehyde-3-phosphate dehydrogenase 1	P08735	2264.47	30.27	4	61	337	36.50	6.96	Zea mays
Glycerate dehydrogenase (GDH)	P13443	1243.44	28.27	8	19	382	41.68	6.29	Cucumis sativus
Glycerol-3-phosphate dehydrogenase SDP6	Q9SS48	175.03	5.56	3	4	629	68.41	8.00	Arabidopsis thaliana
Hexokinase-1	O64390	176.78	8.63	3	4	498	54.10	6.99	Solanum tuberosum
Hexokinase-2	Q9SQ76	293.77	8.47	3	5	496	53.69	6.62	Solanum tuberosum
Isocitrate dehydrogenase [NADP] (IDH)	Q06197	1362.82	15.98	8	32	413	46.02	6.23	Glycine max
Isocitric dehydrogenase 5	Q945K7	460.15	17.38	3	12	374	40.60	6.79	Arabidopsis thaliana
Malate dehydrogenase	P19446	239.92	21.07	4	5	356	37.61	8.41	Citrullus lanatus
Malate dehydrogenase	Q08062	2139.85	28.31	6	52	332	35.57	6.09	Zea mays
Malate dehydrogenase	Q9SN86	1243.56	11.66	4	24	403	42.38	8.51	Arabidopsis thaliana
NADH dehydrogenase [ubiquinone] 1 alpha subcomplex subunit 13-B	O49313	727.51	23.78	3	14	143	16.11	9.36	Arabidopsis thaliana
NADH dehydrogenase [ubiquinone] 1 beta subcomplex subunit 9	Q945M1	746.39	24.79	4	18	117	13.61	8.68	Arabidopsis thaliana
NADH dehydrogenase [ubiquinone] flavoprotein 1	Q9FNN5	521.09	21.19	8	13	486	53.42	8.16	Arabidopsis thaliana
NADH dehydrogenase [ubiquinone] iron-sulfur protein 1	Q43644	420.40	7.45	5	7	738	79.92	6.24	Solanum tuberosum

NADH dehydrogenase [ubiquinone] iron-sulfur protein 2	P93306	471.54	17.26	5	8	394	44.93	7.12	Arabidopsis thaliana
NADH dehydrogenase [ubiquinone] iron-sulfur protein 3	P80261	313.06	25.79	4	7	190	22.87	6.55	Solanum tuberosum
NADH dehydrogenase [ubiquinone] iron-sulfur protein 8-A	Q42599	225.69	22.97	4	7	222	25.49	5.41	Arabidopsis thaliana
NADP-dependent glyceraldehyde-3-phosphate dehydrogenase	P93338	1057.89	23.59	9	20	496	53.11	7.37	Nicotiana plumbaginifolia
NADP-dependent glyceraldehydephosphate dehydrogenase subunit A	P09043	2686.06	29.34	9	51	392	41.84	7.02	Nicotiana tabacum
NADP-dependent glyceraldehydephosphate dehydrogenase subunit B	P12859	1340.24	19.73	3	33	451	48.07	7.66	Pisum sativum
Oxygen-evolving enhancer protein 2	P16059	540.62	7.72	3	19	259	28.03	8.19	Pisum sativum
Peroxisomal (S)-2-hydroxy-acid oxidase	P05414	342.39	20.05	3	12	369	40.26	9.13	Spinacia oleracea
Peroxisomal (S)-2-hydroxy-acid oxidase GLO1	Q9LRR9	1043.82	25.07	3	17	367	40.32	9.13	Arabidopsis thaliana
Phosphoenolpyruvate carboxylase	P29196	1415.92	14.61	5	34	965	110.24	5.77	Solanum tuberosum
Phosphoenolpyruvate carboxylase	Q02909	1256.28	12.93	3	30	967	110.62	5.96	Glycine max
Phosphoglycerate kinase	Q42961	4009.28	30.98	12	87	481	50.15	8.38	Nicotiana tabacum
Phosphoglycerate kinase	P12783	2973.55	25.44	6	71	401	42.10	5.86	Triticum aestivum
Phosphoribulokinase	P26302	786.97	14.60	3	19	404	45.11	6.05	Triticum aestivum
Photosystem I P700 chlorophyll a apoprotein A2	P05311	260.48	4.36	3	6	734	82.36	7.61	Pisum sativum
Photosystem I reaction center subunit II	P29302	2510.18	27.45	6	55	204	22.41	9.77	Nicotiana sylvestris
Photosystem II CP43 chlorophyll apoprotein (PSII 43 kDa protein)	Q09FW5	1454.26	25.81	10	39	461	50.32	6.71	Nandina domestica
Photosystem II CP47 chlorophyll apoprotein	A9XV91	2550.73	34.43	17	58	488	53.66	6.60	Cannabis sativa (Hemp)
Photosystem II CP47 chlorophyll apoprotein (PSII 47 kDa protein)	A4QKL8	1875.30	22.05	14	44	508	56.05	6.89	Capsella bursa-pastoris
Photosystem II D2 protein (PSII D2 protein)	A0A330	633.00	19.26	5	15	353	39.54	5.55	Coffea arabica
Photosystem II stability/assembly factor HCF136	Q5Z5A8	409.73	16.11	6	8	416	45.44	9.00	Oryza sativa subsp. japonica
Plastidial pyruvate kinase 2 (PKp2)	Q9FLW9	1284.57	18.65	7	32	579	63.48	7.03	Arabidopsis thaliana
Plastidial pyruvate kinase 3	Q93Z53	621.78	12.96	3	16	571	62.58	7.81	Arabidopsis thaliana
Protein translocase subunit SECA1	Q9SYI0	130.46	3.82	3	3	1022	115.11	6.39	Arabidopsis thaliana
Pyrophosphate-dependent 6-phosphofructose-1-kinase	P21342	416.13	13.15	8	12	616	67.11	7.31	Solanum tuberosum
Pyrophosphate-dependent 6-phosphofructose-1-kinase	Q41141	516.27	4.71	4	12	552	60.08	6.64	Ricinus communis
Pyruvate dehydrogenase E1 component subunit alpha	P52902	1398.53	16.88	7	41	397	43.50	7.90	Pisum sativum
Pyruvate dehydrogenase E1 component subunit beta	P52904	984.62	15.32	4	18	359	38.77	6.25	Pisum sativum

Pyruvate kinase	Q42806	764.62	13.89	5	14	511	55.27	7.56	Glycine max
Pyruvate kinase isozyme A	Q43117	143.64	8.92	4	4	583	64.05	5.57	Ricinus communis
Ribulose bisphosphate carboxylase large chain (RuBisCO large subunit)	P48698	7071.57	26.92	11	176	468	51.92	7.02	Datura stramonium
Ribulose bisphosphate carboxylase/oxygenase activase	O64981	3812.34	19.05	3	49	441	48.17	8.13	Phaseolus vulgaris
Ribulose-1,5-bisphosphate carboxylase large subunit	Q8HUZ0	4205.34	33.33	3	118	468	52.04	6.76	Cannabis sativa (Hemp)
Ribulose-1,5-bisphosphate carboxylase/oxygenase large subunit	Q597K6	6585.80	33.77	2	178	459	50.95	6.33	Celtis philippensis
RuBisCO large subunit-binding protein subunit alpha	P08824	2991.54	20.81	3	52	495	52.35	4.87	Ricinus communis
RuBisCO large subunit-binding protein subunit alpha	P21239	3800.73	20.51	4	61	546	57.66	4.93	Brassica napus
RuBisCO large subunit-binding protein subunit beta	P08927	6003.32	33.28	5	113	595	62.95	6.04	Pisum sativum
Sedoheptulose-1,7-bisphosphatase	O20252	562.23	12.92	4	13	387	42.05	6.20	Spinacia oleracea
Sedoheptulose-1,7-bisphosphatase	P46285	304.72	10.18	3	9	393	42.03	6.43	Triticum aestivum
Succinate dehydrogenase [ubiquinone] flavoprotein subunit	Q6ZDY8	272.26	6.51	4	8	630	68.81	7.08	Oryza sativa subsp. japonica
Succinyl-CoA synthetase beta chain	O82662	1159.11	14.49	6	25	421	45.32	6.71	Arabidopsis thaliana
Succinyl-CoA synthetase subunit alpha	Q6ZL94	2589.12	19.64	8	64	331	34.22	8.22	Oryza sativa subsp. japonica
Transketolase	Q7SIC9	291.66	5.93	3	14	675	72.95	5.72	Zea mays
Triosephosphate isomerase	Q9M4S8	1937.05	19.75	9	47	314	33.51	7.80	Fragaria ananassa
V-type proton ATPase catalytic subunit A (V-ATPase subunit A)	P13548	4843.87	40.77	3	83	623	68.64	5.44	Vigna radiata var. radiata
V-type proton ATPase subunit B 1 (V-ATPase subunit B 1)	Q43432	4277.61	40.16	4	88	488	54.17	5.10	Gossypium hirsutum
V-type proton ATPase subunit B2 (V-ATPase subunit B2)	Q9SZN1	4217.96	40.45	3	88	487	54.27	5.15	Arabidopsis thaliana
V-type proton ATPase subunit D (V-ATPase subunit D)	Q9XGM1	492.05	14.56	5	11	261	29.04	9.51	Arabidopsis thaliana
V-type proton ATPase subunit E (V-ATPase subunit E)	O23948	2846.97	29.11	3	56	237	27.15	7.01	Gossypium hirsutum
V-type proton ATPase subunit E (V-ATPase subunit E)	Q9MB46	2236.74	29.57	3	49	230	26.27	7.93	Citrus unshiu

Identified Protein	**Uniprot accession number**	**Score**	**Coverage**	**Unique Peptides**	**PSMs**	**AAs**	**MW [kDa]**	**calc. pI**	**organism**
Primary metabolism									
1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase 2	F6HDT7	537.12	14.07	3	9	199	23.42	5.05	Vitis vinifera
2-dehydro-3-deoxyphosphooctonate aldolase	O50044	313.60	12.76	3	6	290	31.70	7.40	Pisum sativum
2-isopropylmalate synthase 2	Q9C550	342.15	10.14	5	8	631	68.09	6.37	Arabidopsis thaliana
3-ketoacyl-CoA thiolase 2	Q56WD9	1140.53	11.26	5	13	462	48.55	8.34	Arabidopsis thaliana

3-oxoacyl-[acyl-carrier-protein] synthase I	P52410	690.28	6.77	3	16	473	50.38	8.06	Arabidopsis thaliana
5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase	O50008	2599.74	15.56	4	67	765	84.30	6.51	Arabidopsis thaliana
5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase	Q42662	3008.81	21.20	5	76	764	84.54	6.51	Plectranthus scutellarioides
Acetyl-CoA carboxylase 1	Q38970	467.20	5.15	5	10	2254	251.22	6.52	Arabidopsis thaliana
Acetyl-CoA carboxylase 1	Q8S6N5	298.76	3.35	3	7	2267	252.65	6.40	Oryza sativa subsp. japonica
Acetyl-coenzyme A carboxylase carboxyl transferase subunit alpha	Q9LD43	704.58	5.46	5	17	769	85.25	5.82	Arabidopsis thaliana
Acetyl-coenzyme A carboxylase carboxyl transferase subunit beta	Q1KXV0	290.01	7.29	3	6	480	54.25	5.07	Helianthus annuus
Acyl-[acyl-carrier-protein] desaturase	P22337	661.61	17.42	6	14	396	45.34	6.64	Ricinus communis
Acyl-activating enzyme 10	H9A1W2	910.78	29.96	11	23	564	61.50	6.49	Cannabis sativa (Hemp)
Acyl-activating enzyme 14	H9A8L3	489.36	15.13	7	10	727	80.30	7.18	Cannabis sativa (Hemp)
Acyl-activating enzyme 15	H9A8L4	171.35	5.80	4	4	776	86.67	6.32	Cannabis sativa (Hemp)
Acyl-activating enzyme 6	H9A1V8	238.50	13.01	4	6	569	62.27	7.85	Cannabis sativa (Hemp)
Acyl-activating enzyme 8	H9A1W0	744.33	18.82	10	17	526	56.16	6.54	Cannabis sativa (Hemp)
Acyl-activating enzyme 9	H9A1W1	170.74	14.97	5	5	561	61.46	7.91	Cannabis sativa (Hemp)
Adenosine kinase 1 (AK 1)	Q9SF85	565.74	6.10	4	10	344	37.81	5.45	Arabidopsis thaliana
Adenosylhomocysteinase	P68172	1630.75	18.76	4	40	485	53.07	5.77	Nicotiana sylvestris
Adenylosuccinate synthetase	B9IJ21	439.12	12.63	4	10	491	53.66	6.84	Populus trichocarpa
Alpha-1,4 glucan phosphorylase L isozyme	P27598	230.61	5.03	4	6	955	108.45	5.38	Ipomoea batatas
Alpha-1,4-glucan-protein synthase [UDP-forming]	P85413	446.25	58.33	4	12	60	6.84	8.16	Phoenix dactylifera
Anthranilate synthase component I-1	P32068	534.28	6.55	3	11	595	66.27	6.49	Arabidopsis thaliana
Arginase	P46637	286.30	13.45	4	6	342	37.32	6.55	Arabidopsis thaliana
Asparagine synthetase	O24338	184.85	4.95	4	4	525	59.57	6.52	Sandersonia aurantiaca
Aspartate carbamoyltransferase 3	Q43064	273.98	13.30	5	6	391	44.32	6.87	Pisum sativum
ATP sulfurylase 1	Q9LIK9	184.00	13.82	3	6	463	51.43	6.81	Arabidopsis thaliana
ATP sulfurylase 2	Q43870	227.19	11.97	3	7	476	53.60	6.61	Arabidopsis thaliana
Methylthioribulose-1-phosphate dehydratase/enolase-phosphatase E1	B9SQI7	225.52	4.36	3	6	527	57.70	6.00	Ricinus communis
Biotin carboxylase 2	B9N843	2508.86	30.61	13	57	526	57.46	6.93	Populus trichocarpa
Coproporphyrinogen-III oxidase	P35055	456.95	24.94	7	13	385	43.24	7.18	Glycine max

Cystathionine gamma-synthase	P55217	436.89	9.77	5	10	563	59.88	6.87	Arabidopsis thaliana
Cysteine synthase	Q43317	301.96	11.38	3	7	325	34.32	6.61	Citrullus lanatus
Cysteine synthase	Q9XEA8	233.86	16.62	3	7	325	34.29	5.41	Oryza sativa subsp. japonica
Delta-1-pyrroline-5-carboxylate synthase A	P54887	351.85	6.00	4	8	717	77.65	6.29	Arabidopsis thaliana
Delta-aminolevulinic acid dehydratase 1	Q9SFH9	220.01	13.49	4	6	430	46.66	7.31	Arabidopsis thaliana
Enoyl-[acyl-carrier-protein] reductase [NADH]	P80030	265.87	12.73	3	8	385	40.45	8.66	Brassica napus
Ferredoxin-dependent glutamate synthase	Q43155	543.21	5.54	5	12	1517	165.30	6.19	Spinacia oleracea
Ferrochelatase-1	P42043	142.37	7.73	3	4	466	52.00	5.82	Arabidopsis thaliana
Formyltetrahydrofolate synthetase	Q9SPK5	133.60	9.78	4	4	634	67.76	6.71	Arabidopsis thaliana
Gamma-aminobutyrate transaminase POP2	Q94CE5	248.19	6.35	3	8	504	55.15	7.94	Arabidopsis thaliana
GDP-mannose 3,5-epimerase 2	Q2R1V8	230.46	11.59	3	6	371	42.10	6.09	Oryza sativa subsp. japonica
GDP-mannose 4,6 dehydratase 2	P93031	142.87	9.12	3	5	373	41.94	6.11	Arabidopsis thaliana
Glucose phosphomutase	Q9ZSQ4	959.82	12.37	6	23	582	63.08	5.72	Populus tremula
Glucose-1-phosphate adenylyltransferase large subunit 1	P55229	118.73	5.56	3	4	522	57.64	7.91	Arabidopsis thaliana
Glucose-1-phosphate adenylyltransferase small subunit	Q9M462	347.94	18.85	8	10	520	57.01	6.24	Brassica napus
Glutamate dehydrogenase	P52596	759.97	18.73	3	19	411	44.52	6.76	Vitis vinifera
Glutamate dehydrogenase 1	Q43314	1031.11	24.82	5	24	411	44.50	6.86	Arabidopsis thaliana
Glutamate dehydrogenase 2	Q38946	1443.16	23.11	3	31	411	44.67	6.54	Arabidopsis thaliana
Glutamate synthase 1 [NADH]	Q0JKD0	162.14	2.86	4	4	2167	236.74	7.09	Oryza sativa subsp. japonica
Glutamate-1-semialdehyde aminotransferase	P45621	670.47	13.95	4	14	466	49.61	5.94	Glycine max
Glutamate-glyoxylate aminotransferase 1	Q9LR30	582.56	10.19	4	12	481	53.27	6.89	Arabidopsis thaliana
Glutamine synthetase	O22506	1562.49	11.57	7	46	432	47.73	5.91	Daucus carota
Glutamine synthetase cytosolic isozyme	P04078	307.72	13.48	4	9	356	39.08	5.73	Medicago sativa
Glyoxysomal fatty acid beta-oxidation multifunctional protein MFP-a	Q39659	212.06	6.48	3	5	725	79.12	9.07	Cucumis sativus
Histidinol dehydrogenase	A7KZQ1	71.52	21.74	2	2	92	10.19	6.79	Humulus lupulus
Ketol-acid reductoisomerase	Q65XK0	1573.70	14.01	4	22	578	62.34	6.43	Oryza sativa subsp. japonica
L-galactono-1,4-lactone dehydrogenase	Q9SU56	321.20	6.56	4	7	610	68.51	8.56	Arabidopsis thaliana
Magnesium-chelatase subunit ChlI	P93162	152.68	8.08	3	3	421	45.84	5.69	Glycine max
Magnesium-protoporphyrin IX monomethyl ester [oxidative] cyclase	Q945B7	237.81	10.86	4	5	405	47.28	8.65	Euphorbia esula

Mannose-1-phosphate guanylyltransferase 1	O22287	80.76	7.20	3	3	361	39.55	6.73	Arabidopsis thaliana
N-carbamoylputrescine amidase	Q93XI4	98.53	10.30	3	3	301	33.44	5.44	Oryza sativa subsp. japonica
Nucleoside diphosphate kinase 3	P81766	1334.82	33.33	5	34	153	17.11	8.31	Spinacia oleracea
Nucleoside diphosphate kinase B	P47920	802.97	17.57	3	17	148	16.19	6.95	Flaveria bidentis
Pectinesterase 1	O04886	328.82	7.02	4	7	584	63.47	8.88	Citrus sinensis
Pectinesterase/pectinesterase inhibitor PPE8B	Q43062	263.97	7.85	4	5	522	57.36	6.57	Prunus persica
Phosphoglucomutase	Q9SM59	301.31	9.58	6	8	626	68.53	6.25	Pisum sativum
Phosphomannomutase	O80840	187.32	13.01	4	6	246	27.74	5.54	Arabidopsis thaliana
Phosphoserine aminotransferase	Q96255	860.90	15.58	7	23	430	47.33	8.06	Arabidopsis thaliana
Probable aldo-keto reductase 1	C6TBN2	203.02	8.09	3	5	346	38.23	6.57	Glycine max
Protochlorophyllide reductase	Q01289	666.56	16.54	4	13	399	42.94	9.04	Pisum sativum
Protochlorophyllide reductase	Q41249	1048.45	13.82	3	17	398	43.05	8.91	Cucumis sativus
S-adenosylmethionine synthase 2	P50302	1637.09	31.03	3	43	390	42.58	6.48	Actinidia chinensis
S-adenosylmethionine synthase 5	A7Q0V4	2492.26	46.04	5	63	391	42.77	5.91	Vitis vinifera
Soluble inorganic pyrophosphatase 1	Q9LXC9	221.09	10.67	3	7	300	33.36	6.01	Arabidopsis thaliana
Sucrose synthase	A7KZQ6	181.70	6.80	3	5	309	35.64	6.54	Humulus lupulus
Sucrose-phosphate synthase	A7KZQ5	196.68	9.03	2	3	321	36.04	5.68	Humulus lupulus
Sucrose-UDP glucosyltransferase	O65026	228.98	7.70	3	7	805	92.27	6.24	Medicago sativa
Sucrose-UDP glucosyltransferase	P49034	391.51	7.97	5	10	803	91.57	6.81	Alnus glutinosa
Sucrose-UDP glucosyltransferase 2	Q00917	199.82	5.08	3	4	807	92.01	6.05	Arabidopsis thaliana
Threonine synthase	Q9MT28	104.29	4.82	3	3	519	57.38	7.01	Solanum tuberosum
Tryptophan synthase beta chain 1	P43283	172.89	10.28	3	4	389	42.49	6.57	Zea mays
UDP-glucose 6-dehydrogenase 1	Q96558	647.63	15.21	8	14	480	52.91	6.04	Glycine max
UDP-glucose pyrophosphorylase	P19595	452.58	8.18	3	11	477	51.84	6.00	Solanum tuberosum
Uridine 5'-monophosphate synthase	Q42942	167.64	8.03	3	4	461	49.73	7.72	Nicotiana tabacum
Uridine kinase-like protein 1	Q9FKS0	208.26	9.05	3	3	486	54.40	6.61	Arabidopsis thaliana
Identified Protein	**Uniprot accession number**	**Score**	**Coverage**	**Unique Peptides**	**PSMs**	**AAs**	**MW [kDa]**	**calc. pI**	**organism**
Disease/Defense (stress respons)									

2-Cys peroxiredoxin BAS1	Q6ER94	1328.03	8.81	3	27	261	28.08	6.00	Oryza sativa subsp. japonica
Aldo-keto reductase family 4 member C9	Q0PGJ6	121.07	9.52	3	4	315	35.11	8.12	Arabidopsis thaliana
Catalase isozyme 1	P48350	806.64	17.28	3	23	492	57.03	7.43	Cucurbita pepo
Catalase-2	P25819	994.01	15.04	3	28	492	56.90	7.12	Arabidopsis thaliana
Delta-1-pyrroline-5-carboxylate dehydrogenase 12A1	Q8VZC3	272.10	10.61	6	9	556	61.73	6.73	Arabidopsis thaliana
Ferritin-1	P19976	287.81	10.40	3	10	250	28.03	6.04	Glycine max
Glutathione reductase	P42770	278.50	9.03	3	6	565	60.81	7.87	Arabidopsis thaliana
Glutathione S-transferase parA	P25317	183.33	7.73	3	7	220	25.21	6.07	Nicotiana tabacum
Heat shock 70 kDa protein	P11143	5507.87	26.82	4	123	645	70.53	5.33	Zea mays
Heat shock 70 kDa protein	P37900	2438.89	16.44	4	43	675	72.26	6.02	Pisum sativum
Heat shock 70 kDa protein	Q01899	5930.10	30.37	13	106	675	72.49	6.20	Phaseolus vulgaris
Heat shock 70 kDa protein 15	F4HQD4	932.72	4.09	4	18	831	91.62	5.19	Arabidopsis thaliana
Heat shock cognate 70 kDa protein 2	P27322	6650.20	31.21	4	143	644	70.66	5.19	Solanum lycopersicum
Heat shock cognate protein 80	P36181	3066.96	22.17	3	63	699	80.09	5.03	Solanum lycopersicum
Heat shock protein 81-1 (HSP81-1)	A2YWQ1	3127.09	22.89	3	62	699	80.14	5.07	Oryza sativa subsp. indica
Heat shock protein 83	P51819	1720.54	27.31	3	37	703	80.77	5.05	Ipomoea nil
Lactoylglutathione lyase	O04885	178.74	16.76	3	4	185	20.77	5.85	Brassica juncea
L-ascorbate peroxidase 8	Q69SV0	786.58	11.92	7	18	478	51.16	5.53	Oryza sativa subsp. japonica
Monodehydroascorbate reductase (MDAR)	Q40977	546.41	13.16	3	10	433	47.28	6.11	Pisum sativum
Monodehydroascorbate reductase (MDAR)	Q43497	592.24	14.78	4	17	433	47.01	6.02	Solanum lycopersicum
Peptide methionine sulfoxide reductase	P54153	441.78	11.22	4	9	196	21.90	6.54	Solanum lycopersicum
Peroxiredoxin Q	Q9MB35	299.86	25.81	4	6	186	20.64	9.52	Sedum lineare
Phospholipid hydroperoxide glutathione peroxidase (PHGPx)	P30708	326.68	14.79	3	6	169	18.76	6.70	Nicotiana sylvestris
Probable protein phosphatase 2C 59 (AtPP2C59)	Q8RXV3	488.84	23.15	5	9	311	33.23	4.84	Arabidopsis thaliana
Serine hydroxymethyltransferase 1	Q9SZJ5	2330.02	34.24	14	53	517	57.36	8.13	Arabidopsis thaliana
Stromal 70 kDa heat shock-related protein	Q02028	4188.32	22.52	5	82	706	75.47	5.35	Pisum sativum
Succinate-semialdehyde dehydrogenase	B9F3B6	379.64	4.36	3	7	527	56.06	8.10	Oryza sativa subsp. japonica
Superoxide dismutase [Cu-Zn]	O65198	1320.45	13.86	3	29	202	20.81	6.51	Medicago sativa
Identified Protein	**Uniprot accession**	**Score**	**Coverage**	**Unique Peptides**	**PSMs**	**AAs**	**MW [kDa]**	**calc. pI**	**organism**

	number								
Secondary metabolism									
(-)-limonene synthase	A7IZZ1	792.21	21.54	12	21	622	72.34	6.92	Cannabis sativa (Hemp)
(+)-alpha-pinene synthase	A7IZZ2	1602.64	25.04	14	38	615	71.80	6.30	Cannabis sativa (Hemp)
1-deoxy-D-xylulose 5-phosphate reductoisomerase	Q8W250	1036.54	13.11	5	17	473	51.44	6.18	Oryza sativa subsp. japonica
1-deoxy-D-xylulose-5-phosphate synthase	O78328	1131.88	12.80	7	36	719	77.48	7.20	Capsicum annuum
2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	G9C075	338.37	27.35	5	9	245	25.93	7.14	Humulus lupulus
4-coumarate:CoA ligase	C0KKW5	134.52	7.48	3	3	548	59.87	5.92	Humulus lupulus
4-hydroxy-3-methylbut-2-en-1-yl diphosphate synthase	F4K0E8	3579.60	17.95	14	76	741	82.20	6.43	Arabidopsis thaliana
4-hydroxy-3-methylbut-2-enyl diphosphate reductase	Q6AVG6	1726.96	7.63	5	43	459	51.03	5.67	Oryza sativa subsp. japonica
acyl-activating enzyme 1	AFD33345.1	1507.72	28.51	15	37	719	79.51	7.11	Cannabis sativa (Hemp)
Acyl-activating enzyme 3	H9A1V5	252.40	9.21	3	4	543	59.46	8.59	Cannabis sativa (Hemp)
ATP-citrate synthase alpha chain protein 2	Q2QZ86	675.58	15.37	4	21	423	46.51	5.87	Oryza sativa subsp. japonica
ATP-citrate synthase beta chain protein 1	Q93VT8	3155.48	22.37	12	65	608	66.03	7.68	Oryza sativa subsp. japonica
Cannabidiolic acid synthase	A6P6W0	3632.98	16.51	2	87	545	62.35	8.37	Cannabis sativa (Hemp)
Chalcone isomerase-like protein	I6WIE9	2265.31	61.68	13	48	214	23.70	8.16	Cannabis sativa (Hemp)
Chalcone synthase 1A	P51081	154.74	7.97	3	6	389	42.89	6.62	Pisum sativum
Chalcone synthase-like protein 1	C6KI62	228.14	14.80	4	7	392	43.15	6.99	Cannabis sativa (Hemp)
Cinnamic acid 4-hydroxylase	C0KKW6	272.96	13.07	6	8	505	57.84	9.19	Humulus lupulus
Cinnamic acid 4-hydroxylase) (C4H)	P48522	160.93	6.73	3	4	505	58.24	9.09	Catharanthus roseus
Dihydroflavonol 4-reductase	A2A245	521.31	16.29	6	13	350	39.33	6.33	Humulus lupulus
Dihydroflavonol-4-reductase (DFR)	P51110	132.90	5.04	3	4	337	37.73	6.61	Vitis vinifera
Farnesyl pyrophosphate synthase	Q94G65	332.60	10.82	3	6	342	39.31	5.62	Humulus lupulus
Flavanone-3-hydroxylase (F3H)	Q05963	477.95	13.48	3	9	356	40.18	5.87	Callistephus chinensis
Geranylgeranyl diphosphate reductase	Q9ZS34	281.46	11.42	4	6	464	51.27	8.97	Nicotiana tabacum
Isoflavone reductase homolog	P52578	647.11	16.88	6	17	308	33.83	6.62	Solanum tuberosum
Isopentenyl pyrophosphate isomerase II	Q42553	402.99	20.42	6	13	284	32.59	6.55	Arabidopsis thaliana
Isopentenyl-diphosphate isomerase	G9C073	2011.58	37.38	14	57	321	36.58	7.28	Humulus lupulus
Naringenin-chalcone synthase	Q8RVK9	201.61	18.77	4	8	389	42.69	6.47	Cannabis sativa (Hemp)

Olivetol synthase	B1Q2B6	9461.61	49.61	1	267	385	42.56	6.54	Cannabis sativa (Hemp)
Olivetolic acid cyclase	I6WU39	1985.11	57.43	9	51	101	11.99	6.16	Cannabis sativa (Hemp)
Phytoene dehydrogenase	Q07356	141.52	6.71	3	3	566	62.92	6.48	Arabidopsis thaliana
Polyketide synthase 3	F1LKH5	9461.61	49.61	18	267	385	42.54	6.54	Cannabis sativa (Hemp)
Tetrahydrocannabinolic Acid Synthase	NH1319	7442.23	63.90	25	236	518	58.63	9.00	Cannabis sativa (Hemp)
Zeta-carotene desaturase	Q9SE20	129.88	8.33	4	4	588	64.69	8.28	Solanum lycopersicum

Identified Protein	**Uniprot accession number**	**Score**	**Coverage**	**Unique Peptides**	**PSMs**	**AAs**	**MW [kDa]**	**calc. pI**	**organism**
Protein synthesis									
30S ribosomal protein S1	P29344	244.79	11.19	4	6	411	44.76	5.55	Spinacia oleracea
30S ribosomal protein S11	Q4VZK2	172.20	26.81	3	4	138	14.96	12.22	Cucumis sativus
30S ribosomal protein S15	Q09WW1	273.58	24.44	4	7	90	10.85	11.03	Morus indica
30S ribosomal protein S3	A4GG84	149.35	10.19	3	4	216	24.89	9.92	Phaseolus vulgaris
30S ribosomal protein S5	P93014	342.69	14.85	3	7	303	32.62	8.97	Arabidopsis thaliana
30S ribosomal protein S7	A4QJG0	396.82	36.77	6	11	155	17.35	11.28	Aethionema cordifolium
40S ribosomal protein S13-1	P59223	1530.29	38.41	9	43	151	17.08	10.39	Arabidopsis thaliana
40S ribosomal protein S14-3	P42036	2132.64	35.33	9	53	150	16.23	10.59	Arabidopsis thaliana
40S ribosomal protein S15a-1	P42798	982.68	50.77	6	19	130	14.79	9.89	Arabidopsis thaliana
40S ribosomal protein S16	P16149	598.03	35.86	6	14	145	16.18	10.32	Lupinus polyphyllus
40S ribosomal protein S17-2	Q9SJ36	2130.73	27.14	8	55	140	15.94	10.04	Arabidopsis thaliana
40S ribosomal protein S18	P34788	1802.50	31.58	6	34	152	17.53	10.54	Arabidopsis thaliana
40S ribosomal protein S19-3	Q9FNP8	380.81	17.48	3	11	143	15.69	10.21	Arabidopsis thaliana
40S ribosomal protein S23 (S12)	P46297	918.89	28.87	4	19	142	15.77	10.45	Fragaria ananassa
40S ribosomal protein S24-1	Q9SS17	809.95	19.55	4	16	133	15.36	10.70	Arabidopsis thaliana
40S ribosomal protein S25-3	Q8GYL5	1928.29	30.56	5	39	108	12.01	10.70	Arabidopsis thaliana
40S ribosomal protein S27-1	O64650	984.50	36.90	3	20	84	9.40	8.94	Arabidopsis thaliana
40S ribosomal protein S30	P49689	484.78	19.35	3	13	62	6.88	12.23	Arabidopsis thaliana
40S ribosomal protein S3-3	Q9FJA6	733.60	20.97	7	23	248	27.44	9.54	Arabidopsis thaliana

40S ribosomal protein S3a	Q285L8	1704.79	26.05	8	42	261	29.71	9.72	Nicotiana tabacum
40S ribosomal protein S4	O81363	1197.15	33.33	4	31	261	29.55	10.27	Prunus armeniaca
40S ribosomal protein S4	P46299	1021.70	29.39	3	26	262	29.62	10.21	Gossypium hirsutum
40S ribosomal protein S5	O65731	781.20	22.84	8	20	197	22.00	9.89	Cicer arietinum
40S ribosomal protein S6-1	O48549	649.59	14.40	4	12	250	28.35	10.61	Arabidopsis thaliana
40S ribosomal protein S7	Q9XET4	623.97	7.29	4	11	192	22.17	9.74	Secale cereale
40S ribosomal protein S8-2	Q9FIF3	1389.91	20.95	5	23	210	23.76	10.49	Arabidopsis thaliana
40S ribosomal protein S9-1	Q9LXG1	244.78	21.72	5	9	198	23.02	10.17	Arabidopsis thaliana
40S ribosomal protein SA (p40)	O80377	844.06	15.49	6	20	297	32.43	5.02	Daucus carota
50S ribosomal protein L1	P49208	580.57	23.56	5	12	208	23.48	10.23	Pisum sativum
50S ribosomal protein L15	P31165	208.73	13.18	6	7	258	27.58	10.40	Pisum sativum
50S ribosomal protein L16	Q4VZN1	605.06	31.11	3	10	135	15.28	11.69	Cucumis sativus
50S ribosomal protein L16	C6L8D8	619.02	39.83	4	11	118	13.17	11.11	Humulus lupulus
50S ribosomal protein L2	A4QJF6	741.46	29.56	5	15	274	29.87	10.86	Aethionema cordifolium
50S ribosomal protein L3	O80360	474.99	14.67	4	14	259	28.35	10.64	Nicotiana tabacum
60S acidic ribosomal protein P0-1	O04204	1124.50	24.61	11	29	317	33.65	5.29	Arabidopsis thaliana
60S ribosomal protein L11 (L5)	P46287	600.51	36.46	3	19	181	20.67	9.96	Medicago sativa
60S ribosomal protein L12	O50003	860.95	22.29	3	18	166	17.87	8.92	Prunus armeniaca
60S ribosomal protein L13-3	Q9FF90	1546.67	22.33	7	41	206	23.47	11.00	Arabidopsis thaliana
60S ribosomal protein L13a-2	Q9LRX8	521.06	16.99	4	12	206	23.44	10.35	Arabidopsis thaliana
60S ribosomal protein L17	O48557	142.66	11.70	3	5	171	19.49	10.26	Zea mays
60S ribosomal protein L18-2	P42791	627.75	12.83	3	14	187	20.91	10.98	Arabidopsis thaliana
60S ribosomal protein L18a-2	P51418	1028.27	25.84	4	21	178	21.29	10.48	Arabidopsis thaliana
60S ribosomal protein L23	P49690	1686.22	37.14	8	37	140	15.02	10.48	Arabidopsis thaliana
60S ribosomal protein L23a-1	Q8LD46	3384.61	39.61	11	81	154	17.43	10.20	Arabidopsis thaliana
60S ribosomal protein L24	Q9FUL4	559.69	17.74	4	13	186	21.31	10.62	Prunus avium
60S ribosomal protein L26-2	Q9FJX2	663.96	13.70	6	26	146	16.78	11.12	Arabidopsis thaliana
60S ribosomal protein L30	O49884	410.32	16.07	3	8	112	12.32	9.54	Lupinus luteus
60S ribosomal protein L32-1	P49211	101.66	12.03	3	4	133	15.49	10.89	Arabidopsis thaliana

60S ribosomal protein L34-2	Q9FE65	341.87	28.57	4	13	119	13.64	11.59	Arabidopsis thaliana
60S ribosomal protein L35a-3	Q9C912	295.98	14.29	3	11	112	12.91	10.68	Arabidopsis thaliana
60S ribosomal protein L36	P52866	715.10	30.19	4	17	106	12.24	9.89	Daucus carota
60S ribosomal protein L4 (L1)	Q9XF97	757.87	10.54	4	18	408	44.84	10.23	Prunus armeniaca
60S ribosomal protein L4-1 (L1)	Q9SF40	367.45	7.64	3	10	406	44.67	10.35	Arabidopsis thaliana
60S ribosomal protein L5-2	P49227	1012.15	12.29	4	25	301	34.42	9.17	Arabidopsis thaliana
60S ribosomal protein L6-1	Q9FZ76	966.65	24.03	6	24	233	26.14	10.10	Arabidopsis thaliana
60S ribosomal protein L7-4	Q9LHP1	299.82	15.57	5	10	244	28.42	9.95	Arabidopsis thaliana
60S ribosomal protein L7a-2	Q9LZH9	658.79	14.45	5	16	256	29.02	10.15	Arabidopsis thaliana
60S ribosomal protein L8-1	P46286	195.96	9.69	3	7	258	27.84	10.90	Arabidopsis thaliana
60S ribosomal protein L9	P30707	983.79	13.47	3	20	193	21.74	9.28	Pisum sativum
Alanyl-tRNA synthetase	P36428	203.64	3.59	3	5	1003	110.42	6.46	Arabidopsis thaliana
Diadenosine tetraphosphate synthetase	O23627	300.33	5.90	4	7	729	81.89	7.01	Arabidopsis thaliana
Elongation factor 1-alpha (EF-1-alpha)	P25698	3023.42	34.23	3	97	447	49.35	9.06	Glycine max
Elongation factor 1-alpha (EF-1-alpha)	P29521	1892.23	35.41	5	54	449	49.27	9.17	Daucus carota
Elongation factor 1-delta (EF-1-delta)	P93447	248.39	7.96	3	7	226	24.48	4.56	Pimpinella brachycarpa
Elongation factor 1-gamma 3 (EF-1-gamma 3)	Q5Z627	427.73	10.10	5	14	416	47.36	6.47	Oryza sativa subsp. japonica
Elongation factor 2 (EF-2)	O23755	3141.92	18.03	15	63	843	93.74	6.30	Beta vulgaris
Elongation factor G (EF-G)	P34811	483.72	12.31	8	12	788	86.90	5.77	Glycine max
Elongation factor Tu	Q9ZT91	1423.23	16.96	7	30	454	49.38	6.70	Arabidopsis thaliana
Elongation factor Tu (EF-Tu)	Q43467	1763.65	20.46	6	26	479	52.06	6.68	Glycine max
Eukaryotic initiation factor 4A-15	Q40468	3070.64	33.41	10	65	413	46.69	5.58	Nicotiana tabacum
Eukaryotic peptide chain release factor subunit 1-1	Q39097	183.34	11.01	4	4	436	48.69	5.30	Arabidopsis thaliana
Eukaryotic translation initiation factor 1A (eIF-1A)	P56331	335.33	19.31	3	6	145	16.37	5.35	Onobrychis viciifolia
Eukaryotic translation initiation factor 3 subunit D	P56820	158.50	7.78	3	6	591	66.68	5.77	Arabidopsis thaliana
Eukaryotic translation initiation factor 3 subunit I	Q38884	261.72	11.59	3	6	328	36.37	6.99	Arabidopsis thaliana
Eukaryotic translation initiation factor 5A-1	P26564	906.48	40.37	3	26	161	17.65	5.78	Medicago sativa
Polyadenylate-binding protein 2	P42731	348.77	6.68	3	8	629	68.63	8.21	Arabidopsis thaliana
Protein decapping 5	Q9C658	329.15	5.07	3	8	611	64.33	7.71	Arabidopsis thaliana

Protein translation factor SUI1 homolog	O48650	78.80	9.73	3	3	113	12.60	8.43	Salix bakko
Ribosomal protein L2	G0YCW8	732.60	32.48	6	16	274	29.82	10.80	Celtis occidentalis
Ribosomal protein S4	A9XUW3	188.87	24.24	4	5	198	23.08	10.46	Cannabis sativa (Hemp)
Ribosomal protein S7	G0YCX2	354.35	36.77	6	11	155	17.39	11.28	Celtis occidentalis
Seryl-tRNA synthetase	O81983	116.33	7.53	3	3	438	50.07	6.93	Helianthus annuus
Signal recognition particle 54 kDa protein 1	P49968	390.95	10.26	5	10	497	54.48	9.29	Hordeum vulgare
Identified Protein	**Uniprot accession number**	**Score**	**Coverage**	**Unique Peptides**	**PSMs**	**AAs**	**MW [kDa]**	**calc. pI**	**organism**
Cell growth/division									
Cell division control protein 48 homolog D (AtCDC48d)	Q9SCN8	2344.75	26.26	5	54	815	90.28	5.16	Arabidopsis thaliana
Cell division cycle protein 48 homolog	P54774	2524.66	26.02	6	57	807	89.71	5.31	Glycine max
Cell division protein FtsZ homolog 1	Q42545	309.42	18.24	5	6	433	45.54	7.40	Arabidopsis thaliana
Cell division protein FtsZ homolog 2-1	O82533	584.79	16.53	6	11	478	50.69	5.71	Arabidopsis thaliana
DnaJ protein homolog (DNAJ-1)	Q04960	421.12	13.32	4	10	413	46.03	6.48	Cucumis sativus
Guanine nucleotide-binding protein subunit beta-like protein	O24076	144.22	9.85	3	5	325	35.64	7.44	Medicago sativa
Linoleate 9S-lipoxygenase 1	P27480	679.33	3.71	3	18	862	97.09	6.96	Phaseolus vulgaris
Phospholipase D alpha 1 (PLD 1)	O82549	1028.75	8.64	6	20	810	91.78	5.86	Brassica oleracea
Proliferating cell nuclear antigen (PCNA)	O82797	237.05	20.83	3	5	264	29.25	4.79	Nicotiana tabacum
Identified Protein	**Uniprot accession number**	**Score**	**Coverage**	**Unique Peptides**	**PSMs**	**AAs**	**MW [kDa]**	**calc. pI**	**organism**
Cell structures									
Actin-7	P53492	6916.95	50.40	5	183	377	41.71	5.49	Arabidopsis thaliana
Actin-related protein 7	Q8L4Y5	115.83	6.89	3	3	363	39.88	4.89	Arabidopsis thaliana
Dynamin-2B	Q9LQ55	500.71	9.02	6	9	920	100.17	9.04	Arabidopsis thaliana
Dynamin-related protein 12A (SDL12A)	Q39821	220.58	9.84	4	4	610	68.31	7.94	Glycine max
Dynamin-related protein 1E	Q9FNX5	174.11	5.29	3	3	624	69.76	7.52	Arabidopsis thaliana
Microtubule-associated protein 70-2 (AtMAP70-2)	Q8L7S4	166.56	4.89	3	3	634	70.16	7.09	Arabidopsis thaliana
Tubulin alpha chain	Q9FT36	923.25	23.95	7	17	451	49.61	5.01	Daucus carota
Tubulin beta-1 chain	P29500	442.52	24.22	7	10	450	50.53	4.86	Pisum sativum
Vacuolar protein sorting-associated protein 2 homolog 1	Q9SKI2	111.53	15.11	4	4	225	25.28	5.63	Arabidopsis thaliana

(AtVPS2-1)									
Vacuolar protein sorting-associated protein 32 homolog 1 (AtVPS32-1)	O82197	702.15	25.82	6	18	213	24.00	4.83	Arabidopsis thaliana
Identified Protein	**Uniprot accession number**	**Score**	**Coverage**	**Unique Peptides**	**PSMs**	**AAs**	**MW [kDa]**	**calc. pI**	**organism**
Intracellular traffic									
ADP-ribosylation factor 1	O48649	526.44	40.88	7	11	181	20.57	6.95	Salix bakko
Chaperone protein ClpC1	Q9FI56	3485.76	26.48	25	71	929	103.39	6.77	Arabidopsis thaliana
Heat shock 70 kDa protein 6	Q9STW6	4560.43	23.96	6	102	718	76.46	5.20	Arabidopsis thaliana
Importin subunit alpha-1a	Q71VM4	392.03	14.07	5	9	526	57.53	5.22	Oryza sativa subsp. japonica
Karyopherin subunit alpha-2	O04294	136.93	3.95	3	4	531	58.58	4.89	Arabidopsis thaliana
Mitochondrial outer membrane protein porin of 34 kDa	P42055	129.19	7.25	3	5	276	29.56	8.60	Solanum tuberosum
Outer plastidial membrane protein porin	P42054	1381.08	9.06	4	38	276	29.58	9.11	Pisum sativum
Protein TIC110	Q8LPR9	439.44	5.31	4	9	1016	112.05	5.97	Arabidopsis thaliana
Protein TOC75-3	Q9STE8	204.43	4.52	3	5	818	89.13	8.79	Arabidopsis thaliana
Ras-related protein RABA2a (AtRABA2a)	O04486	254.45	24.42	5	7	217	24.09	6.60	Arabidopsis thaliana
Ras-related protein RABB1c (AtRABB1c)	P92963	368.91	25.59	5	8	211	23.15	7.42	Arabidopsis thaliana
Trigger factor-like protein TIG	Q8S9L5	201.51	8.41	4	4	547	61.69	5.33	Arabidopsis thaliana
Identified Protein	**Uniprot accession number**	**Score**	**Coverage**	**Unique Peptides**	**PSMs**	**AAs**	**MW [kDa]**	**calc. pI**	**organism**
Protein destination and storage									
26S protease regulatory subunit 4 homolog	P46466	1429.47	33.93	5	27	448	49.56	6.21	Oryza sativa subsp. japonica
26S protease regulatory subunit 6A homolog A	Q9SEI2	2251.85	40.33	14	55	424	47.45	5.03	Arabidopsis thaliana
26S protease regulatory subunit 7	O64982	1065.88	32.94	12	21	425	47.53	6.80	Prunus persica
26S protease regulatory subunit S10B homolog B	Q9MAK9	1533.74	32.33	9	34	399	44.73	8.19	Arabidopsis thaliana
26S protease subunit 6B homolog	Q9SEI4	1621.49	41.91	11	38	408	45.72	5.60	Arabidopsis thaliana
26S proteasome AAA-ATPase subunit RPT2b	Q9SL67	1190.38	33.41	4	26	443	49.32	6.10	Arabidopsis thaliana
26S proteasome AAA-ATPase subunit RPT6b	Q94BQ2	1035.28	32.22	10	25	419	47.13	8.59	Arabidopsis thaliana
26S proteasome non-ATPase regulatory subunit 3 homolog A	Q9LNU4	213.57	12.70	6	7	488	55.55	8.19	Arabidopsis thaliana
26S proteasome regulatory subunit RPN7	Q93Y35	304.24	12.14	5	6	387	44.25	5.99	Arabidopsis thaliana

Identified Protein	Uniprot accession	Score	Coverage	Unique Peptides	PSMs	AAs	MW [kDa]	calc. pI	organism
Aminomethyltransferase	O23936	927.74	22.60	8	21	407	44.26	8.68	Flaveria trinervia
ATP-dependent zinc metalloprotease FTSH	Q39444	1607.57	20.69	5	35	662	71.02	6.96	Capsicum annuum
ATP-dependent zinc metalloprotease FTSH 2	Q655S1	2372.64	26.48	11	42	676	72.49	5.72	Oryza sativa subsp. japonica
ATP-dependent zinc metalloprotease FTSH 3	Q84WU8	209.50	4.57	4	6	809	89.30	7.27	Arabidopsis thaliana
ATP-dependent zinc metalloprotease FTSH 5	Q8LQJ8	539.55	8.53	3	11	715	77.39	8.05	Oryza sativa subsp. japonica
Chaperonin 60 subunit beta 2	Q9LJE4	3645.85	24.16	4	73	596	63.30	5.73	Arabidopsis thaliana
Chaperonin 60 subunit beta 3	C0Z361	5674.97	29.31	3	105	597	63.29	5.87	Arabidopsis thaliana
Chaperonin CPN60	P29197	2480.80	25.82	4	50	577	61.24	5.78	Arabidopsis thaliana
Chaperonin CPN60-1	Q05045	2324.97	27.65	3	49	575	61.02	5.77	Cucurbita maxima
Glycine dehydrogenase	P26969	554.27	7.85	9	15	1057	114.61	7.50	Pisum sativum
Lon protease homolog	P93648	307.46	6.33	5	6	964	105.59	5.66	Zea mays
Peptidyl-prolyl cis-trans isomerase	O49939	897.40	19.15	7	18	449	49.84	5.41	Spinacia oleracea
Peptidyl-prolyl cis-trans isomerase FKBP62	Q38931	427.12	9.26	6	13	551	61.41	5.31	Arabidopsis thaliana
Presequence protease 1	Q9LJL3	501.57	10.28	9	13	1080	120.94	5.68	Arabidopsis thaliana
Proteasome subunit alpha type-2	A2YVR7	1268.50	26.81	5	20	235	25.83	5.48	Oryza sativa subsp. indica
Proteasome subunit alpha type-3	O24362	1133.57	20.08	4	18	249	27.27	6.54	Spinacia oleracea
Proteasome subunit alpha type-4	O82530	450.39	17.27	4	10	249	27.21	5.74	Petunia hybrida
Proteasome subunit alpha type-5	Q9M4T8	1207.38	37.97	11	24	237	25.96	4.75	Glycine max
Proteasome subunit alpha type-6	O48551	517.70	26.83	4	12	246	27.37	6.23	Glycine max
Proteasome subunit alpha type-7	Q9SXU1	825.42	23.29	4	15	249	27.08	7.37	Cicer arietinum
Proteasome subunit beta type-5	O24361	821.95	20.59	4	16	272	29.60	6.80	Spinacia oleracea
SKP1-like protein 1B	Q9FHW7	308.43	10.53	4	8	171	19.09	4.61	Arabidopsis thaliana
T-complex protein 1 subunit alpha	P28769	869.31	20.92	10	18	545	59.19	6.30	Arabidopsis thaliana
T-complex protein 1 subunit epsilon	O04450	251.01	10.84	4	6	535	59.35	5.66	Arabidopsis thaliana
Ubiquitin carboxyl-terminal hydrolase 7	Q84WC6	167.53	5.87	3	3	477	53.48	6.27	Arabidopsis thaliana
Ubiquitin receptor RAD23d	Q84L30	797.64	12.43	4	23	378	40.04	4.64	Arabidopsis thaliana
Ubiquitin-conjugating enzyme E2 35	Q94A97	323.26	32.68	3	6	153	17.18	7.34	Arabidopsis thaliana
Ubiquitin-NEDD8-like protein RUB1	P0C030	2838.86	54.25	11	59	153	17.12	6.06	Oryza sativa subsp. japonica

	number								
Signal transduction									
Calcium-dependent protein kinase 2	P49101	224.18	8.77	3	5	513	58.04	6.55	Zea mays
Serine/threonine-protein kinase SAPK6	Q6ZI44	270.87	11.78	3	5	365	41.78	6.04	Oryza sativa subsp. japonica
Identified Protein	**Uniprot accession number**	**Score**	**Coverage**	**Unique Peptides**	**PSMs**	**AAs**	**MW [kDa]**	**calc. pI**	**organism**
Transcription									
ATP-dependent helicase UPF1	Q9FJR0	182.38	2.55	3	3	1254	136.78	6.48	Arabidopsis thaliana
Chloroplast stem-loop binding protein of 41 kDa b	Q9SA52	467.22	22.22	7	14	378	42.59	8.16	Arabidopsis thaliana
DEAD-box ATP-dependent RNA helicase 12	Q109G2	146.53	8.83	3	4	521	57.91	8.16	Oryza sativa subsp. japonica
DEAD-box ATP-dependent RNA helicase 15	Q56XG6	396.47	15.69	7	11	427	48.31	5.64	Arabidopsis thaliana
DEAD-box ATP-dependent RNA helicase 21	Q53RK8	146.77	5.43	3	3	736	84.47	8.84	Oryza sativa subsp. japonica
DEAD-box ATP-dependent RNA helicase 34	Q10I26	541.24	22.77	7	14	404	45.60	6.37	Oryza sativa subsp. japonica
DEAD-box ATP-dependent RNA helicase 52C	Q2R1M8	407.16	10.43	6	9	623	65.95	8.09	Oryza sativa subsp. japonica
Histone H2A variant 1	O23628	1215.80	29.41	4	33	136	14.53	10.32	Arabidopsis thaliana
Histone H2B.3	O65819	3197.47	38.69	6	74	137	15.04	10.11	Solanum lycopersicum
Histone H3-like 5	Q9FKQ3	305.10	17.27	3	10	139	15.58	11.25	Arabidopsis thaliana
Histone H4 variant TH011	P62785	2687.76	44.66	9	75	103	11.40	11.47	Triticum aestivum
Myb transcription factor	G7ZLA3	104.86	6.85	2	2	336	36.26	6.65	Humulus lupulus
SWI/SNF complex component SNF12 homolog	Q9FMT4	155.23	8.24	3	3	534	59.23	9.48	Arabidopsis thaliana
Identified Protein	**Uniprot accession number**	**Score**	**Coverage**	**Unique Peptides**	**PSMs**	**AAs**	**MW [kDa]**	**calc. pI**	**organism**
Transporter									
ABC transporter B family member 10	Q9SGY1	524.73	4.89	4	7	1227	134.39	7.21	Arabidopsis thaliana
ADP/ATP translocase	P25083	764.06	17.36	4	21	386	42.03	9.79	Solanum tuberosum
ADP/ATP translocase (Fragment)	P27081	939.40	14.77	3	25	386	41.80	9.73	Solanum tuberosum
ATPase 10	Q43128	786.57	8.24	3	16	947	104.75	6.43	Arabidopsis thaliana
ATPase 4	Q9SU58	1133.28	10.21	4	21	960	105.65	6.52	Arabidopsis thaliana
Calcium-transporting ATPase 1	P92939	171.83	4.52	3	3	1061	116.29	5.11	Arabidopsis thaliana
GTP-binding nuclear protein Ran1	P38546	922.47	23.08	6	19	221	25.17	6.71	Solanum lycopersicum

GTP-binding protein SAR1A	O04266	444.70	41.45	6	12	193	21.95	7.53	Brassica campestris
GTP-binding protein YPTM1	P16976	246.84	19.71	3	4	208	23.30	5.76	Zea mays
Plasma membrane ATPase 1	P22180	1058.77	12.45	3	21	956	105.04	6.86	Solanum lycopersicum
Plasma membrane ATPase 4	Q03194	992.20	11.66	3	22	952	105.12	6.96	Nicotiana plumbaginifolia
Pyrophosphate-energized vacuolar membrane proton pump	P21616	583.10	9.02	6	10	765	79.93	5.49	Vigna radiata var. radiata
Ras-related protein Rab7	Q43463	241.27	25.24	5	7	206	23.10	5.68	Glycine max
Vesicle-fusing ATPase	Q9M0Y8	157.83	3.23	3	3	742	81.44	6.10	Arabidopsis thaliana
Identified Protein	**Uniprot accession number**	**Score**	**Coverage**	**Unique Peptides**	**PSMs**	**AAs**	**MW [kDa]**	**calc. pI**	**organism**
Unclear classification									
14-3-3 protein 1	P93206	802.36	22.09	3	19	249	28.20	4.82	Solanum lycopersicum
14-3-3 protein 6	P93211	3516.71	33.72	3	82	258	28.95	4.75	Solanum lycopersicum
14-3-3-like protein B	P42654	1051.19	17.24	3	23	261	29.51	4.88	Vicia faba
14-3-3-like protein C	P93343	3811.03	34.23	3	83	260	29.35	4.84	Nicotiana tabacum
14-3-3-like protein E	O49997	3299.65	33.09	4	75	272	30.55	5.05	Nicotiana tabacum
acyl-activating enzyme 2	AFD33346.1	7984.01	33.97	39	284	1248	140.69	7.88	Cannabis sativa (Hemp)
Acyl-CoA-binding domain-containing protein 5	Q8RWD9	91.70	2.78	3	3	648	70.96	6.32	Arabidopsis thaliana
Adenylate kinase B	Q08480	652.62	25.51	5	14	243	26.66	7.80	Oryza sativa subsp. japonica
Aldehyde dehydrogenase family 7 member A1	Q41247	112.21	4.45	3	3	494	52.65	5.67	Brassica napus
Allene oxide cyclase C4	Q68IP6	173.99	5.91	2	4	254	27.65	8.37	Humulus lupulus
Beta-adaptin-like protein C	O81742	184.69	5.26	3	5	893	99.03	5.02	Arabidopsis thaliana
Beta-xylosidase/alpha-L-arabinofuranosidase 1	A5JTQ2	256.96	5.04	5	7	774	83.67	6.68	Medicago sativa subsp. varia
Betv1-like protein	I6XT51	17844.62	95.65	20	416	161	17.60	5.39	Cannabis sativa (Hemp)
Calmodulin-2	P0DH97	2221.83	74.50	10	42	149	16.81	4.27	Arabidopsis thaliana
Casein kinase I isoform delta-like	P42158	142.08	8.00	3	4	450	50.91	9.45	Arabidopsis thaliana
Chaperone protein ClpB2	Q75GT3	497.98	7.36	3	13	978	108.92	6.62	Oryza sativa subsp. japonica
Chaperone protein ClpB3	Q9LF37	937.66	14.05	6	20	968	108.88	6.23	Arabidopsis thaliana
Clathrin heavy chain 1	Q0WNJ6	807.20	6.51	5	21	1705	193.12	5.40	Arabidopsis thaliana
Clathrin heavy chain 2	Q2QYW2	583.97	4.57	3	15	1708	193.22	5.39	Oryza sativa subsp. japonica

Endoplasmin homolog	P35016	331.27	10.65	6	7	817	93.43	4.97	Catharanthus roseus
Glutamate decarboxylase 4	Q9ZPS3	1462.65	13.79	5	23	493	55.97	6.34	Arabidopsis thaliana
Glutamyl endopeptidase	Q10MJ1	407.61	7.04	5	9	938	103.83	5.95	Oryza sativa subsp. japonica
Glutamyl-tRNA(Gln) amidotransferase subunit B	A2ZF53	184.96	8.46	3	3	544	60.06	6.40	Oryza sativa subsp. indica
Leucine aminopeptidase 3	Q944P7	486.12	7.72	4	11	583	61.27	7.08	Arabidopsis thaliana
Luminal-binding protein	P49118	3008.52	23.72	5	74	666	73.19	5.21	Solanum lycopersicum
Luminal-binding protein 4	Q03684	3591.97	23.84	3	76	667	73.48	5.19	Nicotiana tabacum
Maf-like protein	G9C079	544.57	31.03	6	11	203	22.18	5.49	Humulus lupulus
Mitogen activated protein kinase	G9C080	150.07	14.47	4	4	380	43.60	6.16	Humulus lupulus
NAD-dependent malic enzyme 59 kDa isoform	P37225	228.79	7.32	3	4	601	66.23	6.34	Solanum tuberosum
NAD-dependent malic enzyme 62 kDa isoform	P37221	708.65	9.11	5	12	626	69.91	5.94	Solanum tuberosum
NADP-dependent malic enzyme	P34105	403.65	6.09	3	6	591	65.18	6.96	Populus trichocarpa
O-methyltransferase 5	B0ZB59	402.45	33.60	5	6	247	27.71	5.36	Humulus lupulus
Outer envelope protein 80	Q9C5J8	260.78	6.56	4	4	732	79.89	8.16	Arabidopsis thaliana
Oxygen-evolving enhancer protein 1	Q40459	5021.26	31.33	12	103	332	35.21	5.96	Nicotiana tabacum
Plastid allene oxide synthase	Q5XNP6	116.30	14.38	4	4	299	33.90	6.96	Humulus lupulus
Polyubiquitin	B3GK02	1826.89	77.63	8	38	76	8.58	9.19	Humulus lupulus
Probable aldo-keto reductase 5	Q9ASZ9	798.81	12.17	4	17	345	37.83	6.30	Arabidopsis thaliana
Probable calcium-binding protein CML13	Q94AZ4	298.96	27.70	4	7	148	16.51	4.94	Arabidopsis thaliana
Probable methyltransferase PMT1	Q8H118	82.49	7.53	3	3	611	69.31	6.73	Arabidopsis thaliana
Protein phosphatase 2A 65 kDa regulatory subunit	A7KZQ4	435.55	8.38	2	5	334	37.16	5.25	Humulus lupulus
Putative MO25-like protein At5g47540	Q9FGK3	165.01	5.54	3	4	343	39.43	7.03	Arabidopsis thaliana
Quinone oxidoreductase-like protein At1g23740	Q9ZUC1	225.65	11.14	3	5	386	40.96	8.35	Arabidopsis thaliana
Receptor-like protein kinase At2g47060	O80719	183.21	15.89	4	5	365	39.91	7.40	Arabidopsis thaliana
Serine/threonine-protein phosphatase 2A	Q38950	238.75	9.88	4	6	587	65.56	5.14	Arabidopsis thaliana
SKP1 component-like 1	A7KZP6	335.33	11.46	4	11	157	17.65	4.74	Humulus lupulus
Transketolas	Q42676	1525.34	8.09	4	30	519	56.15	6.15	Craterostigma plantagineum
Translationally-controlled tumor protein homolog (TCTP)	Q9ZRX0	849.47	23.95	3	14	167	18.83	4.82	Pseudotsuga menziesii
U6 snRNA-associated Sm-like protein LSm4	Q9ZRU9	164.40	19.59	3	4	148	16.34	10.13	Fagus sylvatica

Ubiquitin-conjugating enzyme E2 variant 1D	Q9SVD7	1212.91	45.89	5	21	146	16.52	6.68	Arabidopsis thaliana
Ubiquitin-like protein SMT3	Q5TIQ0	368.40	35.53	3	8	76	8.73	4.97	Cannabis sativa (Hemp)
Uncharacterized protein At2g34160	O22969	315.97	16.15	3	7	130	14.61	5.36	Arabidopsis thaliana
Villin-2	O81644	355.83	3.79	3	6	976	107.78	5.29	Arabidopsis thaliana
Villin-3	O81645	277.09	4.25	3	5	965	106.28	5.85	Arabidopsis thaliana

Chapter 6

Transcriptome Analysis of Medicinal *Cannabis* Trichomes: Assembly, Annotation, and Elucidation of Biosynthetic Pathways of Secondary Metabolites

Nizar Happyana, Kathleen Pamplaniyil, Oliver Rupp, Jörn Kalinowski, Oliver Kayser

Kathleen Pamplaniyil supported the construction of cDNA library of *Cannabis* trichomes.
Oliver Rupp and Jörn Kalinowski provided SAMS for analyzing the cDNA library.
Oliver Kayser coordinated and supervised the project.

Abstact

Despite the recent completion of *Cannabis sativa* L. genome sequencing, the biosynthetic pathways in this plant, in particular secondary metabolites, are not fully understood. In this study, a cDNA library of *Cannabis* trichomes was assembled, annotated, and applied for elucidating biosynthetic pathways of secondary metabolites. The cDNA library was constructed by 454 GS FLX pyrosequencing system. The sequencing produced 880,032 reads with an average read length of 333 nucleotides, and then reduced to 735,261 cleaned reads. These reads were further assembled into 14695 contigs, 12892 isotigs, and 11300 isogroups. Assembled sequences were annotated functionally using the Sequence Analysis and Management System (SAMS) for Gene Ontology (GO) analysis and pathway assignments with the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Putative genes which participate in the biosynthetic pathway of the secondary metabolites were successfully identified in the cDNA library. All genes related to mevalonate and non-mevalonate pathways, and other genes involved in the downstream steps of terpenoid biosynthesis were successfully elucidated. Furthermore, all published genes corresponding to cannabinoid biosynthesis were recorded in the cDNA library as well. Although there is no flavonoids reported from *Cannabis* trichomes, however many transcripts including all genes participating in the main biosynthetic pathway of flavonoids putatively noted in the cDNA library. Thus it is most likely a signature of the flavonoids existence in the *Cannabis* trichomes. In summary, our findings suggest that there is a unique relation between the putative transcripts of the cDNA library and secondary metabolite pathways. Hence, it provides valuable information to reveal the governing biosynthetic mechanism of these medicinal compounds.

Keywords: *Cannabis*, transcriptome, trichomes, cDNA library, SAMS, KEGG, GO, cannabinoids, terpenoids, flavonoids

1. Introduction

Cannabis sativa L. is an interesting annual plant as a consequence of its cannabinoids content. These natural products are a unique group of terpenophenolics possessing alkylresorcinol and monoterpene moieties in their molecular structures. There are more than 100 cannabinoids that have been fully identified and structurally elucidated from this plant. Cannabinoids are responsible compounds for most biological activities of this plant. Δ^9-Tetrahydrocannabinol (THC) is well known as the psychoactive component of *Cannabis* (Gaoni and Mechoulam 1964) and mainly activates cannabinoids receptors CB1 and CB2 (Matsuda et al. 1990; Munro et al. 1993). Another example of active cannabinoid, namely, cannabidiol (CBD) has been reported with its anti-oxidative and anti-inflammatory properties (Hampson et al. 1998; Lastres-Becker et al. 2005).

Beside cannabinoids, terpenoids and flavonoids are other interesting compounds found in *Cannabis*. Currently, there are 120 terpenoids that have been reported from this plant, including 61 monoterpenes, 52 sesquiterpenes, 2 triterpenes, one diterpene, and 4 terpenoid derivatives (ElSohly and Slade 2005). Different composition of these compounds leads to the variation of flavor and aroma of *Cannabis* varieties (Flores-Sanchez and Verpoorte 2008). Some *Cannabis* terpenoids have been recorded possess biological activities and can synergize with the effects of the cannabinoids (Burstein and Varanelli 1975; E. B. McPartland and Mediavilla 2002). Meanwhile, there are more than 20 flavonoids that have been identified from this plant (ElSohly and Slade 2005). Pharmological effects of *Cannabis* flavonoids have been reported, such as cannflavin A and B with their inhibition properties of prostaglandin E2 production (Barrett et al. 1986).

Cannabis trichomes are well known as the main site of cannabinoids and terpenoids production (Lanyon et al. 1981; Malingre et al. 1975; Petri et al. 1988; Turner et al. 1978). Trichomes are usually found in the surfaces of *Cannabis*. There are 3 types of trichomes can be found in this plant, namely capitate-stalked, capitate-sessile and bulbous trichomes. The former one contains abundant cannabinoids and generally consists of two parts, the gland (head) and the stem. We noted on our previous work (Happyana et al. 2012a) that the stem of capitate-stalked trichomes contributes in the cannabinoids production.

Recently cDNA libraries of *Cannabis* trichomes have been constructed in order to obtain better understanding of the secondary metabolite biosynthesis in this plant. Mark et al. analyzed the cDNA library for identifying the candidate genes that is affecting Δ^9-

tetrahydrocannabinolic acid (THCA) biosynthesis (Marks et al. 2009a). In addition, a cDNA library has been constructed to identify an acyl activating enzyme that is responsible for formation hexanoyl-CoA in cannabinoid biosynthesis (Stout et al. 2012b). Moreover, the ESTs from this cDNA library have been used for discovering olivetolic acid cyclase which cooperates with olivetol synthase to form olivetolic acid, an important percursor of cannabinoids (Gagne et al. 2012). However the previous reports only focused on analysis of cannabinoid biosynthesis, even though other secondary metabolites, such as flavonoids and terpenoids have been discovered on this plant.

In this work, we report transcriptome analysis of medicinal *Cannabis* trichomes, Bediol variety, in order to further advance the understanding of molecular metabolism, especially secondary metabolite biosynthesis. The cDNA library of *Cannabis* trichomes was constructed using 454 GS FLX pyrosequencing system. The assembled sequences were analyzed with SAMS for GO annotation and KEGG pathway assignments, respectively. Furthermore, real time-PCR (RT-PCR) was applied to analyze expression level of some genes involved in biosynthesis of secondary metabolites in different organs of this plant.

2. Material and methods

2.1 Plant material

Standardized medicinal *Cannabis sativa* L., variety Bediol® (THC level, approx. 6 %, CBD level, approx. 7.5 %) was supplied by Bedrocan BV (the Netherlands). *Cannabis* plants were grown indoors under standardized conditions. They were initially generated from cuttings of standardized mother plants and cultivated under controlled, long day-light conditions (18 h/day). After the vegetative growth phase, the flowering stage was induced under a shorter (12 h/day) light regime for 8 weeks. The trichomes from the 7 week plant were isolated and analyzed. Plant specimens were assigned voucher numbers (D7.09.07.2012) and deposited at Technical Biochemistry Department, TU Dortmund. All plant handling and experimental procedures were carried out under the license No. 4584989 issued by the Federal Institute for Drugs and Medical Devices (BfArM), Germany.

2.2 Trichomes isolation

Trichomes were isolated according to a modified procedure based on the previous findings by Yerger et al. (Yerger et al. 1992). Fresh flowers of *Cannabis* were put in the liquid nitrogen.

Floral leaves and stigma on the flowers were removed using a forceps with occasionally the flowers were put back into the liquid nitrogen. Afterward, a 5-10 gram of the flower was transferred into a 50 ml falcon tube and placed into liquid nitrogen. The tube containing the flower was removed from the liquid nitrogen tank and approximately 2 to 3 cm^3 of finely powdered dry ice (prepared by wrapping a piece in clean paper towels and crushing with a pestle) was added to the tube. Subsequently, the tube was loosely capped and vortexed at maximum speed for approximately one minute and flowers in the tube were removed. In order to collect the trichomes, the content of each tube was sieved through a nylon net filter with 140 µm pore diameter (Merck Millipore, Germany) into a 500-mL beaker glass surrounded by dry ice. The trichomes in the beaker were transferred into a 2 ml frozen microcentrifuge tube with spatula and placed in the liquid nitrogen.

2.3 RNA isolation

Fresh plant materials were ground to a fine powder using a pestle and a mortar under a cold condition (with adding liquid nitrogen). Afterward, around 90 mg of fresh fine powder of plant materials was prepared for RNA isolation using RNeasy™ Plant Mini Kit (Qiagen N.V., The Netherlands). The isolation procedure followed manufacturer's instructions. Concentration of isolated RNA was determined with a Qubit® RNA Assay Kit (Life Technologies, California, USA). mRNA was purified from 5µg total RNA by exonuclease digestion followed by LiCl precipitation (mRNA-Only Eucaryotic mRNA Isolation Kit, Epicentre, Madison, WI, USA).

2.4 cDNA synthesis and 454 pyrosequencing

Synthesis and amplification of first-strand cDNA from1µg mRNA was conducted with the Mint-Universal cDNA Synthesis Kit according to user manual (Evrogen, Moscow, Russia). 800 ng amplified cDNA were used as starting material in the normalization reaction using the Trimmer Kit (Evrogen, Moscow, Russia). Normalized material was re-amplified for 18 cycles. 2 µg of normalized cDNA was digested with 10 Units SfiI for 2hours at 48°C. Fragments larger than 800bp were isolated from a LMP Agarose Gel and purified using the MinElute Gel Extraction Kit (Qiagen, Hilden, Germany). 200ng purified cDNA fragments were ligated to 100ng Sfi cut and dephosphorylated pDNR-lib Vector (Clontech) in 10 µl volume using the Fast Ligation Kit (NEB ,Ipswich, MA,USA). Ligations were desalted by Ethanol precipitation, and re-dissolved in 10µl water. 3 times of 1.5 µl desalted Ligation was

used to transform NEB10b competent cells (NEB, Ipswich, MA, USA). 96 clones were randomly choosen for sequencing to verify successful normalization. Roughly a million clones were plated on LB-Cm plates, scrapped off the plates and stored as glycerol stocks at -70°C. One half of the cells were used to inoculate a 300ml Terrific Broth/Cm culture, which was grown for 5 hours at 30°C. Plasmid DNA was prepared using standard methods (Qiagen, Hilden, Germany). 200 µg of purified Plasmid DNA was digested with 100 Units SfiI for 2 hours at 48°C. cDNA Inserts were gel-purified (LMP-Agarose/MinElute Gel Extraction Kit) and ligated to high-molecular-weight DNA using a proprietary Sfi-linker. Library generation for the 454 FLX sequencing was carried out according to the manufacturer's standard protocols (Roche/454 life sciences, Branford, CT 06405, USA). In short, the concatenated inserts were sheared randomly by nebulization to fragments ranging in size from 400 bp to 900 bp. These fragments were end polished and the 454 A and B adaptors that are required for the emulsion PCR and sequencing were added to the the ends of the fragments by ligation. The resulting fragment library was sequenced on half picotiterplate (PTP) on the GS FLX using the Roche/454 Titanium chemistry. A total of 880,032 sequence reads with an average read length of 333 nucleotides were obtained.

2.5 Sequence assembly and functional annotation

Prior to assembly the sequence reads were screened for the Sfi-linker that was used for concatenation. The linker sequences were clipped out of the reads and the clipped reads assembled to individual transcripts using the Roche/454 Newbler software at default settings [454 Life Sciences Corporation, Software Release: 2.6 (20110517_1502)]. The transcript data were further analyzed using the Sequence Analysis and Management System (SAMS), which is based on GenDB (Bekel et al. 2009). Afterward the sequences were functional annotated using an automated function prediction (Metanor). This method uses a combination of standard bioinformatics tolls such as BLAST and Interpro. Meanwhile, other predicted genes involved in biosynthetic pathways of cannabinoids were annotated manually. SAMS also has been applied for Gene Ontology (GO) annotation and pathway assignments with the Kyoto Encyclopedia of Genes and Genomes (KEGG) database.

2.6 RT-PCR analysis

Equivalent quantities of RNA isolated from leaves, flowers, and *Cannabis* trichomes were reverse-transcribed using iScript^{TM} cDNA synthesis kit (BIO-RAD Laboratories, California,

USA) according to the manufacturer's instructions. Quantitative real-time PCR experiments were performed with Applied Biosystems StepOnePlus™ Real-Time PCR Systems (Applied Biosystem, Foster City, Canada) and SYBR® Green PCR Master Mix (Applied Biosystem). These experiments were carried out using comparative CT method (ΔΔCT) (Livak and Schmittgen 2001). S18 housekeeping gene was used as an endogenous control gene. Putative transcripts of 2-C-methyl-D-erythritol-2,4-cyclodiphosphate synthase (MDS), mevalonate-5-pyrophospate decarboxylase (MVAPPD), (+)-α-pinene synthase, (-)-limonene synthase, THCA synthase (THCAS), CBDA synthase (CBDAS), olivetol synthase (OLS), olivetolic acid cyclase (OAC), predicted cannabigerolic acid synthase (predicted-CBGAS) and chalcone synthase (CHS) were used as the templates for designing corresponding primers using Primers Express Software v.3.0.1 (Applied Biosystem). The list of primers can be seen in table 1. Real-time PCR was run with conditions: pre-incubation at 95 ºC for 10 min, followed by 40 cycles of 95 ºC for 15 s and 60 ºC for 1 min, and then melt curve stage at 95 ºC for 15 s, 60 ºC for 1 min, and 95 ºC for 15 min.

Table 1. Primers of investigated genes for RT-PCR analysis.

Gene	Forward Primer	Reverse Primer	Isotig number
18S	GAGAAACGGCTACCACATCCA	CCGTGTCAGGATTGGGTAATTT	-
MDS	GCGACCACCGCTTTGAAC	CCTTTGAGGGAGTAGCAGATACCT	CSAI_isotig06922
MVAPPD	GCCTCCGGAAAGGGAGATT	GTCCCACCTTTGACCACGAT	CSAI_isotig03787
(+)-α-pinene synthase	GAGGGTCATCACGCTGCAT	CCTAGATCATCTGCAAGTCGTAACA	CSAI_isotig12293
(-)-limonene synthase	TCGATAATATTACGACTTGCAGATGA	GATTTAGGAACATCGCCTCTTTTC	CSAI_isotig1124
THCAS	AAGTTGGCTTGCAGATTCGAA	TGTAGGACATACCCTCAGCATCA	CSAI_isotig00569
Predicted-CBGAS	TTGGGAAGGCATGTTGGAA	AATCCACAAGCGCATGAAGTAA	CSAI_isotig04415
OLS	GGGCTGCTGCGGTGATT	TATCGGCCTTTCCCCAACT	CSAI_isotig00943
OAC	TGTTGGATTTGGAGATGTCTATCG	TTCGTGGTGTGTAGTCAAAAATGA	CSAI_isotig08341

3. Results and Discussions

3.1 Sequencing and assembly of cDNA library

A normalized cDNA library of *Cannabis* trichomes was sequenced with a 454 GS FLX Titanium system in order to obtain a global view of the trichomes transcriptome. A total of 880,032 sequence reads with an average read length of 333 nucleotides were generated. After removing adaptors, primers, and low quality sequences, the raw reads were reduced to 735,261 cleaned reads with 49,836 singletons. SAMS was used for further assembling the reads, yielding 14,695 contigs with an average length of 810.99 bp. and a maximum length of

4,643 bp. The assembly also produced 12,892 isotigs with an average length of 1,033.5 bp. The number of isotigs with one contigs was 10,046, while the average number of contigs per isotigs was 1.38. Meanwhile, a total of 11,300 isogroups with an average number of isotigs of 1.14 were obtained as well. Besides that, 10,004 isogroups were assembled by one contig, whereas the mean number of contigs per isogroup was 1.30. The results of sequencing and assembly of the cDNA library could be seen in the table 2.

Table 2. Summary of sequencing and assembly of cDNA library of *Cannabis* trichomes.

	Number
Sequencing	
Total number of reads	880,032
Average Read Length	333 bp
Total Number of Bases	293 Mb
Full Assembled Reads	735,261
Singletons	49,836
Contigs	
Number of contigs	14,695
Maximum length of contigs	4,643 bp
Minimum length of contigs	2 bp
Average length of contigs	810.99 bp
Isotigs	
Number of isotigs	12,892
Maximum length of isotigs	6,809 bp
Minimum length of isotigs	70 bp
Average length of isotigs	1,033.05 bp
Maximum number of contigs	7
Minimum number of contigs	1
Average number of contigs	1.38
Number of isotigs with one contig	10,046
Isogroups	
Number of isogroups	11,300
Maximum number of isogroups	10
Minimum number of isogroups	1
Average number of isogroups	1.14
Number of isogroups with one isotig	10,060
Maximum number of contigs	15
Minimum number of contigs	1
Average number of contigs	1.30
Number of isogroups with one contig	10,046

3.2 Functional annotation

SAMS was applied for functional annotation of unique sequences of cDNA library of *Cannabis* trichomes. It was originally designed and implemented for quality control of sequence data that was obtained during the high throughput phase of genome sequencing projects (Szczepanowski et al. 2008). In SAMS individual sequences are analyzed and functionally annotated. For instance, an automatic functional annotation was computed using a combination of standard bioinformatics tools such as BLAST and KEGG. This approach leads to consistent gene annotations, assigning gene names, gene products, Enzyme Commission (EC) numbers, functional protein categories (COGs) and GO numbers (Tauch et al. 2006). As the result, a total of 12,892 unique sequences have been functionally annotated with SAMS.

3.2.1 Gene ontology annotation

The unique sequences were annotated using GO analysis provided by SAMS. The sequences were functionally classified into 3 GO categories, namely biological process, molecular function, and cellular component. However, it must be noted that many sequences had more than one assignment within a GO category. A total of 4,500 unique sequences were classified in the biological process category. A large number of sequences related to metabolic process (32.69%), transport (10.09%), translation (9.10%), protein amino acid phosphorylation (8.04%), and transcription DNA dependent (7.01%) were the well-represented in this category (see figure 1A). Besides that, some unique transcripts involved in defense and stress responds (2.75%) were detected in this category as well. Therefore, this identification supported the function of thrichomes in plant defense system. In the molecular function group, 6,348 unique sequences were successfully annotated. In this category, the majority of the sequences were involved in metal ion binding (10.98%), protein binding (10.59%), transferase activity (9.20%), and kinase activity (6.16%) as shown in figure 1B. Meanwhile, a total of 2,219 unique sequences were identified in the category of cellular component, including the sequences related to membarane (36.34%), ribosome (12.58%), intracellular (10.93%), nucleus (10.92%), and cytoplasm (8.12%) as seen in the figure 1C. Overall, these GO annotation provide valuable information on the function of unique transcripts of *Cannabis* trichomes.

3.2.2 KEGG assignment

SAMS is also facilitated with KEGG assignment tool that provides an alternative functional annotation of identified putative genes for the establishment of the pathway association. In this KEGG annotation, unique transcripts of *Cannabis* trichomes were placed to the specific biosynthetic pathway based on the corresponding EC distribution in the KEGG database. A total of 2829 unique transcripts were successfully mapped to KEGG pathways. The KEGG profiles of unique transcripts of *Cannabis* trichomes are shown in figure 2. A large number of putative genes corresponded to carbohydrate metabolism (352) dominated the KEGG profiles. Interestingly, 236 unique transcripts involved in phenylpropanoid biosynthesis were mapped in the KEGG annotation. Phenylpropanoid is a diverse family of secondary metabolites synthesized by plants from the amino acid phenylalanine. Flavonoids, stilbenoids, chalcone, and stilbenoids are members of phenylpropanoid family. Besides that, some putative genes related to other secondary metabolism were also identified, including terpenoid and alkaloid pathways. However, identification of putative genes involved in cannabinoid biosynthesis was performed manually since its pathway is not available in KEGG database. Other pathways such as amino acid metabolism (179 transcripts), T cell receptor signaling pathway (170 transcripts), fatty acid metabolism (152 transcripts), mTOR signaling pathway (113 transcripts), oxidative phosphorylation (96 transcripts), and purine pyrimidine metabolism (95 transcripts) were also detected in the KEGG annotation as shown in figure 2.

3.3 Putative Genes participating in biosynthesis of secondary metabolites

A wide range of secondary metabolites were identified in *Cannabis*, including cannabinoids, terpenoids, flavonoids, and alkaloids. Cannabinoids and terpenoids were found almost in the whole parts of the plant and accumulated especially in the trichomes. Around 23 flavonoids have been isolated and identified from flowers, twigs and pollens of *Cannabis* (ElSohly and Slade 2005). Meanwhile, 10 alkaloids have been isolated and detected from roots, leaves, stems, pollens and seeds of *Cannabis*. There are no references for the presence of alkaloids in the trichomes of this plant. However, the putative transcripts involved in biosynthetic pathway of alkaloids were detected in this study. The presence of alkaloids in the trichomes can be indicated due to the detection of the putative transcripts in this biosynthetic pathway. Therefore, it is needed further investigation. Overall, the putative genes related to secondary metabolites provide an important tool for the elucidation of the pathways.

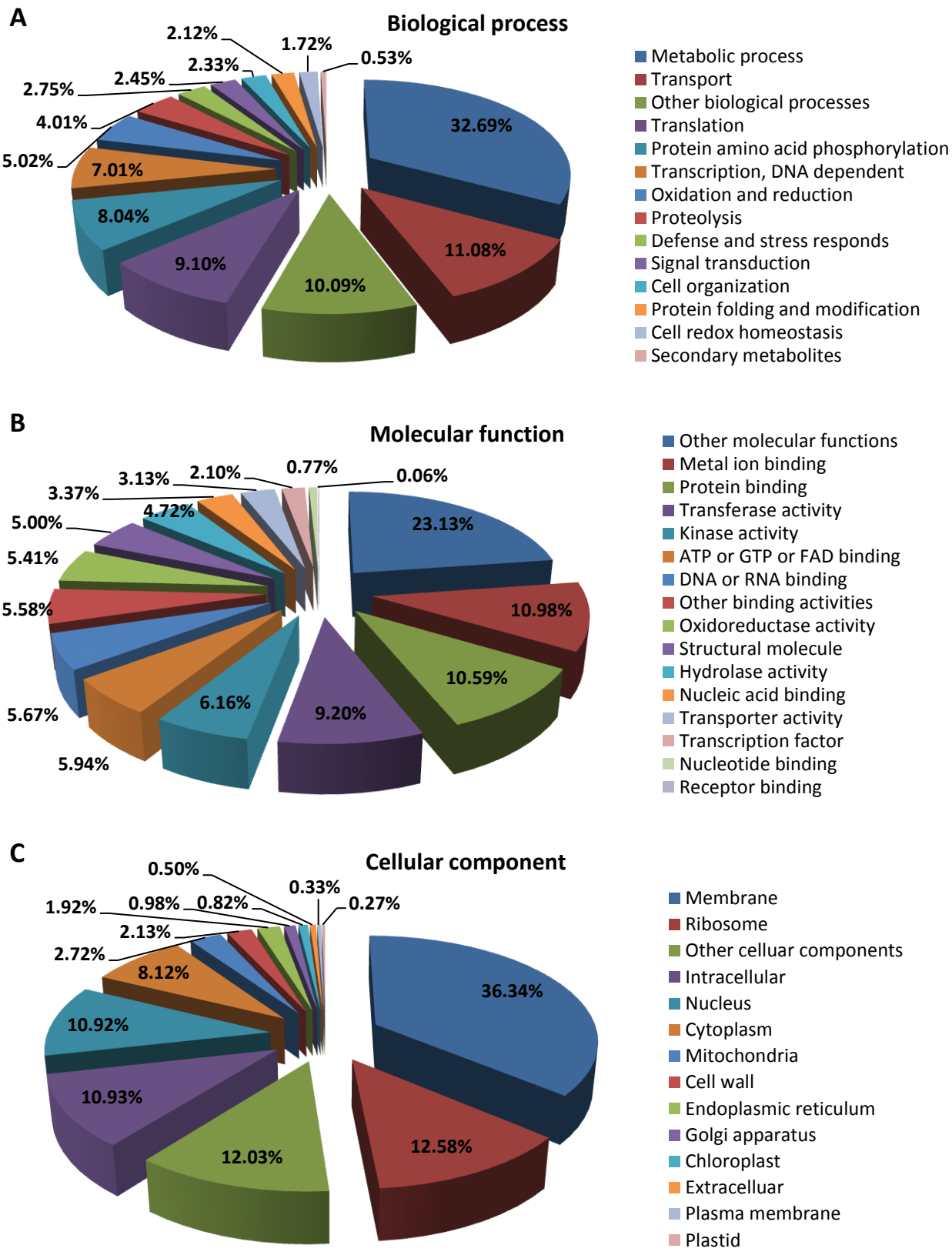


Figure 1. GO analysis of unique sequences of medicinal *Cannabis* trichomes based on biological process (A), molecular function (B), and cellular component (C).

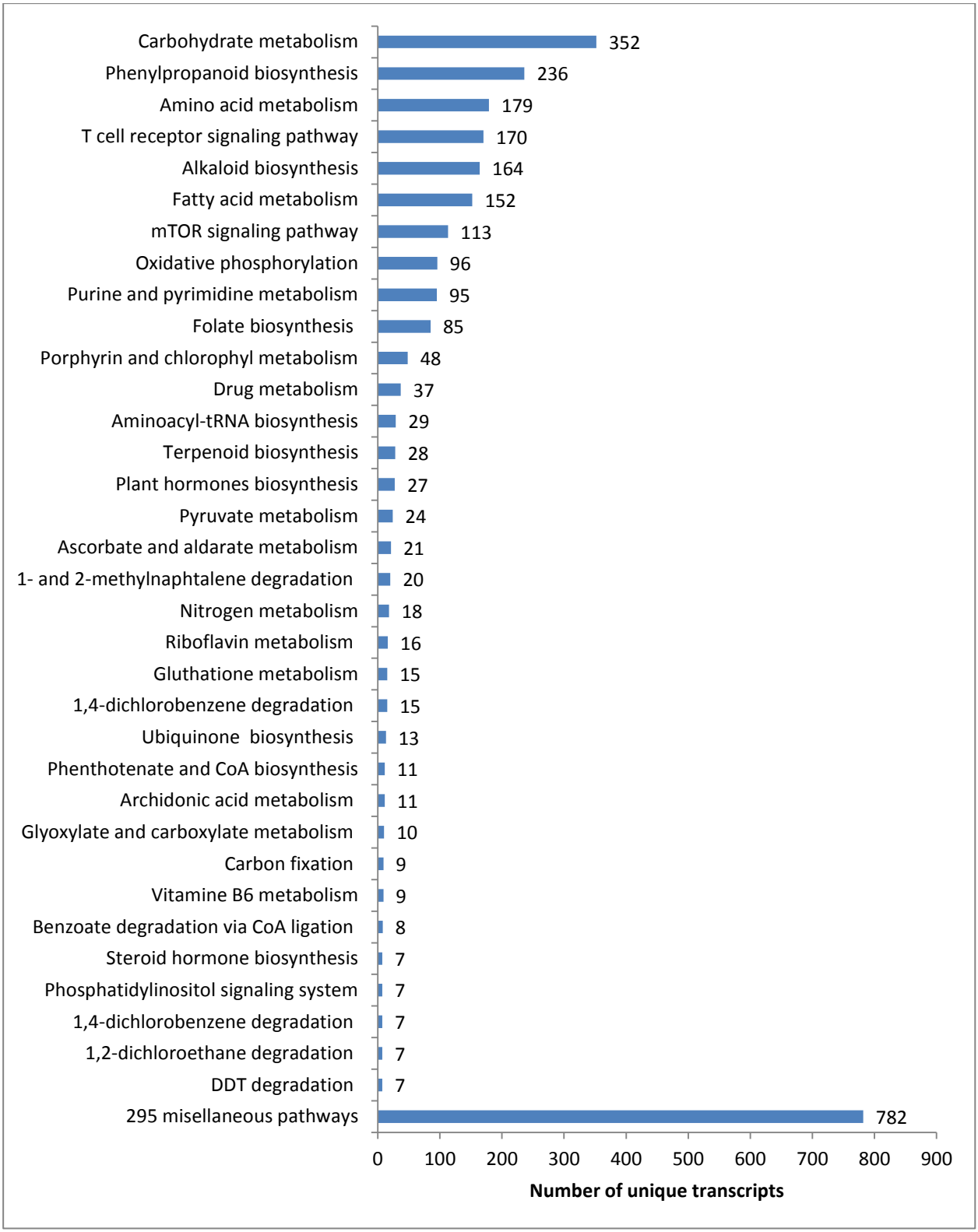


Figure 2. Distribution of annotated transcripts of *Cannabis* trichomes in various metabolic pathway based on KEGG assigment.

3.3.1 Putative genes related to terpenoid biosynthesis

Terpenoids are organic compounds which are derived biosynthetically from units of isoprenoid (C_5H_8), isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). There are 2 known pathways for the biosynthesis of IPP or DMAPP, namely mevalonate (MVA) and non- mevalonate [2-C-methyl-D-erythritol-4-phosphate/1-deoxy-D-xylulose-5-phosphate (MEP/DOXP)] pathway. The first one produces isoprenoids in the cytosol for the synthesis of sesquiterpenoids and triterpenoids normally. Meanwhile MEP/DOXP pathway synthesizes isopreneoids in the plastid for the formation of monoterpenoids and diterpenoids commonly. In this study, all putative genes corresponding in both the plastidic and cytosolic metabolic pathways were identified in the *Cannabis* trichomes (see figure 3). Thus, these results supported the trichomes as the main site of terpenoid biosynthesize.

Putative genes involved in the next steps of terpenoid biosynthesis were also detected, including the transcripts of isopenthenylpyrophosphate isomerase (IPPI), geranyl pyrophosphate synthase (GPPS), farnesyl pyrophosphate synthetase (FPPS), and geranylgeranyl pyrophosphate synthase (GGPPS). The first enzyme catalyzes the isomerization of IPP to DMAPP, while GPPS condenses IPP and DMAPP to form geranyl pyrophosphate (GPP), a precursor of monoterpenoids. FPPS catalyzes sequential condensation reactions of DMAPP with 2 units of IPP to form farnesyl pyrophosphate (FPP), a precursor compound in sesquiterpene biosynthesis. Meanwhile, a precursor of diterpenoids, geranylgeranyl pyrophosphate (GGPP), is synthesized by GGPPS from IPP and FPP.

Several putative genes participating in the downstream steps of terpenoid biosynthesis were detected in the *Cannabis* trichomes as well (see figure 4). The transcripts of monoterpene synthases were identified including (-)-limonene synthase, (+)-α-pinene synthase, and myrcene synthase. These enzymes are responsible for synthesizing (+)-limonene, (+)-α-pinene, and myrcene, respectively. In addition, (-)-limonene synthase and (+)-α-pinene synthase previously were isolated and characterized from *Cannabis* trichomes variety Skunk (Günnewich et al. 2007). Some putative genes involved in the sesquiterpenoid biosynthesis have been identified in the *Cannabis* trichomes, such as transcripts of (+)-δ-cadinene synthase and γ-humulene synthase. Furthermore several putative transcripts related to diterpenoid and carotenoid pathways were recorded as well, such as casbene synthase and phytoene synthase, respectively.

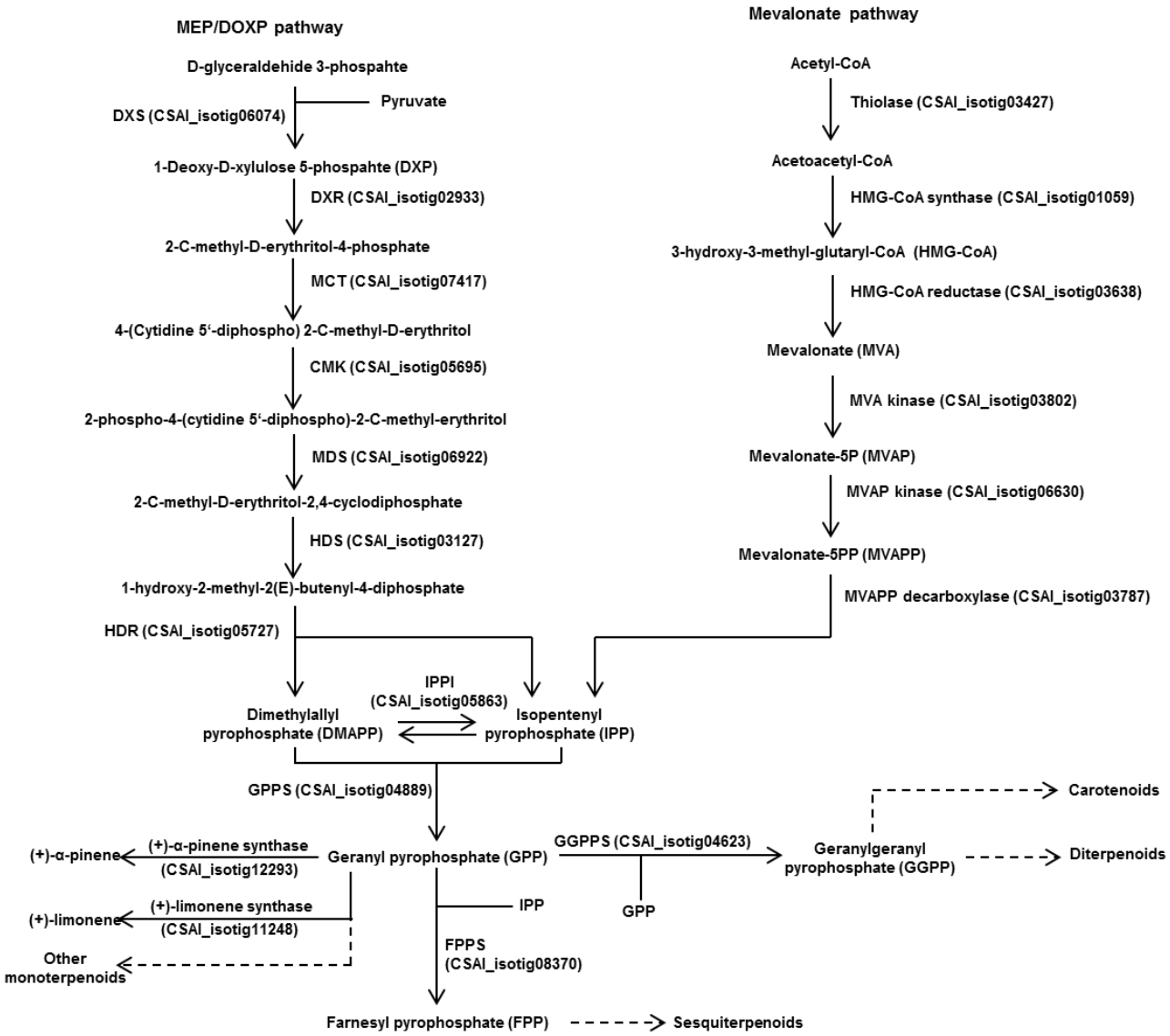


Figure 3. Biosynthetic pathway of terpenoid in *Cannabis* showing putative transcripts encoding enzymes in the pathway. DXS, 1-deoxyxylulose-5-phosphate synthase; DXR, 1-deoxy-D-xylulose-5- phosphate reductoisomerase; MCT, 2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase; CMK, 4-diphosphocytidyl-2-C-methyl-D-erythritol kinase; MDS, 2C-methyl-D-erythritol 2,4-cyclodiphosphate synthase; HDS, 4-hydroxy-3-methylbut-2-enyl diphosphate synthase; HDR, 4-hydroxy-3-methylbut-2-enyl diphosphate reductase; IPPI, IPP/DMAPP isomerase; GPPS, geranyl pyrophosphate synthase; FPPS, farnesyl pyrophosphate synthase; GGPPS, geranylgeranyl pyrophosphate synthase;.

In order to determine expression level of terpenoid genes in different organs of *Cannabis* (leaves, trichomes, and flowers) a RT-PCR analysis was applied. Expression levels of 2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase (MDS), mevalonate-5-pyrophospate decarboxylase (MVAPPD), (+)-α-pinene synthase, and (-)-limonene synthase were analyzed by RT-PCR. As the result, MDS, a gene participating in MEP/DOXP pathway, was expressed almost in the same level in different organs of *Cannabis* (see figure 4A). The same case also has been found on MVAPPD, a gene involved in MVA pathway. The expression levels of this gene did not differ significantly as described in figure 4B. Meanwhile, the quantity of (+)-α-pinene synthase in the trichomes was 20 times higher than in the leaves (figure 4C). In contrast, the highest expression level of (-)-limonene synthase was obtained in the leaves as seen in figure 4D.

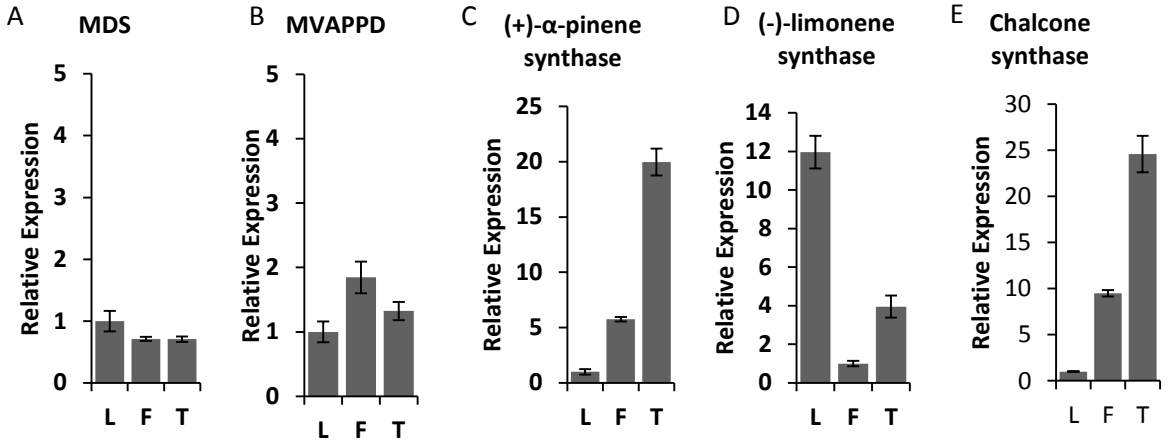


Figure 4. A-D: expression levels of terpenoids genes in the different organs of *Cannabis*; E: expression level of chalcone synthase in the different organs of *Cannabis*. L: Leaves; F: Flowers; T: Trichomes.

3.3.2 Putative genes related to cannabinoid biosynthesis

Putative transcripts of enzymes supplying 3 starting substrates in cannabinoid biosynthesis were identified in the cDNA libaray of *Cannabis* trichomes, namely acetyl-CoA carboxylase, GPPS that explained before, and acyl-activating enzyme 1 (CsAAE1) as shown in figure 5. The first enzyme provides malonyl-CoA from acetyl-CoA, while CsAAE1 supplies hexanoyl-CoA from hexanoate (Stout et al. 2012a). GPPS is responsible for supplying GPP to the biosynthetic pathway of cannabinoids. Furthermore, based on incorporation experiments using ^{13}C-labeled glucoses, this GPP is provided entirely or predominantly (> 98%) from MEP/DOXP pathway (Fellermeier et al. 2001). Besides that, putative genes involved in

formation of olivetolic acid, an important precursor of cannabinoids, were also detected, namely olivetol synthase (OLS) and olivetolic acid cyclase (OAC). Both cooperate in catalyzing condensation of malony-CoA and hexanoyl-CoA to form olivetolic acid (Gagne et al. 2012)

In the next step, cannabigerolic acid (CBGA) is produced from a condensation reaction of GPP and olivetolic acid. It has been reported that a geranyltransferase isolated from young leaves of *Cannabis* has successfully catalyzed this reaction (Fellermeier and Zenk 1998). Unfortunately, gene sequence of this enzyme has not been deposited into the public database yet, thus we could not check it on our cDNA library. However, transcript of an enzyme predicted as CBGA synthase (predicted-CBGAS) from our group (unpublished) was detected in the *Cannabis* trichomes as described in figure 5. Furthermore 2 putative genes participating in transformation of CBGA were also detected, namely THCA synthase (THCAS) and cannabidiolic acid synthase (CBDAS) (see figure 5). The first gene is responsible for converting CBGA into THCA (Taura et al. 1995), while the second catalyzes transformation of CBGA to cannabidiolic acid (CBDA) (Taura et al. 1996). CBGA is also transformed into cannabichromenic acid (CBCA) by CBCA synthase (CBCAS). Nevertheless its sequence has not yet been stored into the public databases even CBCAS was purified to homogeneity (Morimoto et al. 1998). Therefore it could not be checked in the cDNA library of *Cannabis* trichomes.

In this study, the gene expression levels of THCAS, CBDAS, OLS, OAC and the predicted-CBGAS in the different organs of *Cannabis* were investigated with RT-PCR analysis. Surprisingly the highest expression level of THCAS gene was found in the leaves as documented in figure 6A. It was almost 8 times and 3 times higher in the leaves than in the flowers and the trichomes, respectively. The same case was obtained on the expression levels of CBDAS (figure 6B). The quantity of CBDAS gene in the leaves was around 5 fold and 2 fold higher than in the flowers and the trichomes, respectively. In contrast, OLS and OAC in the trichomes were more expressed around 12 times and 50 times than in the leaves of *Cannabis*, respectively (figure 6C-D). The last analyzed gene, predicted-CBGAS, was expressed almost in the same level in all *Cannabis* organs as described in figure 6E.

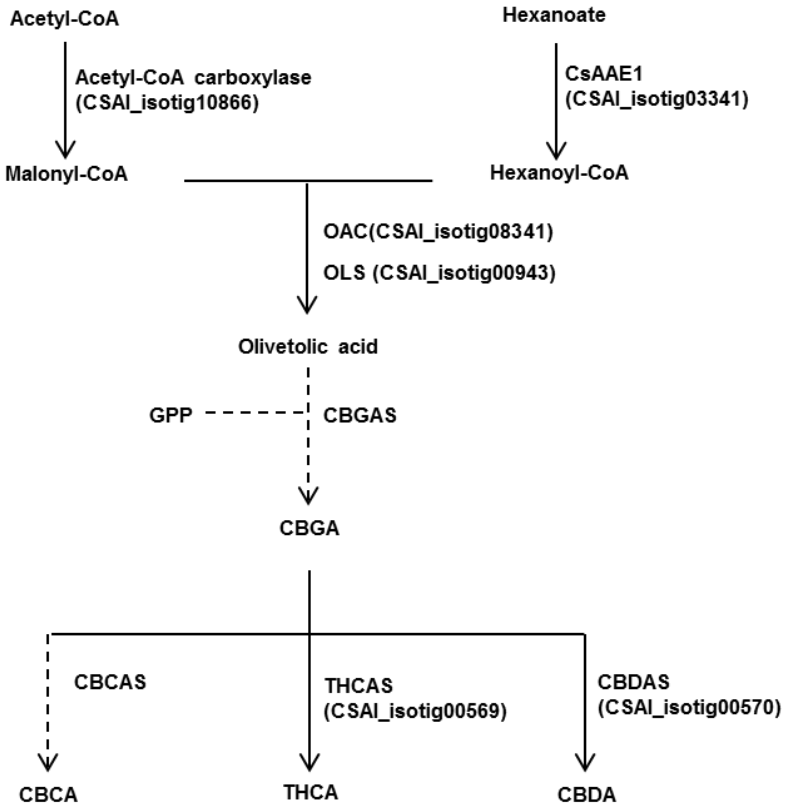


Figure 5. Biosynthetic pathway of cannabinoids showing putative transcripts encoding enzymes in the pathway. Identified enzymes are represented with solid lines. CsAAE1, acyl-activating enzyme 1; OAC, olivetolic acid cyclase; OLS, olivetol synthase; CBGAS, cannabigerolic acid synthase; THCAS, Δ^9-tetrahydrocannabinolic acid synthase; CBDAS, cannabidiolic acid synthase; CBCAS, cannabichromenic acid synthase.

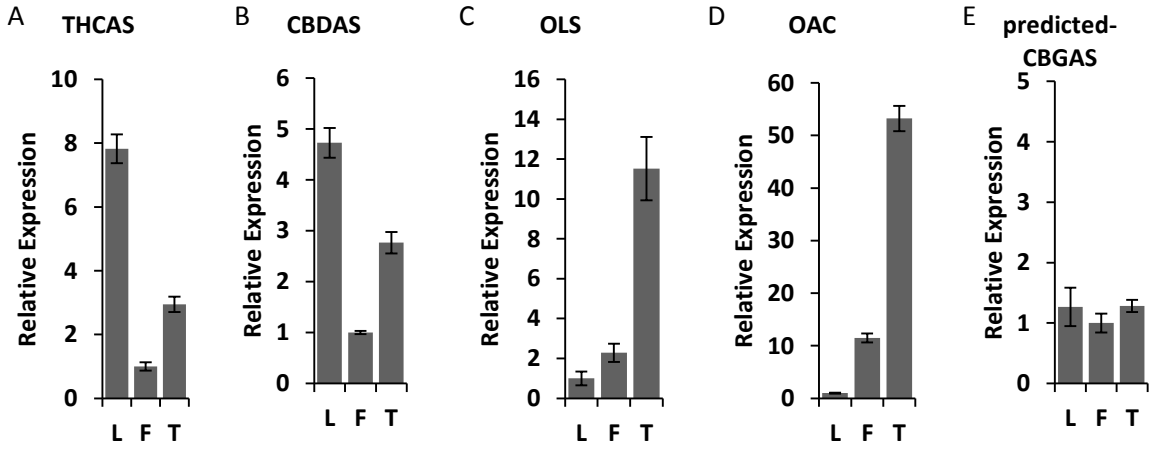


Figure 6. Expression levels of cannabinoids genes in the different organs of *Cannabis*. L: Leaves; F: Flowers; T: Trichomes.

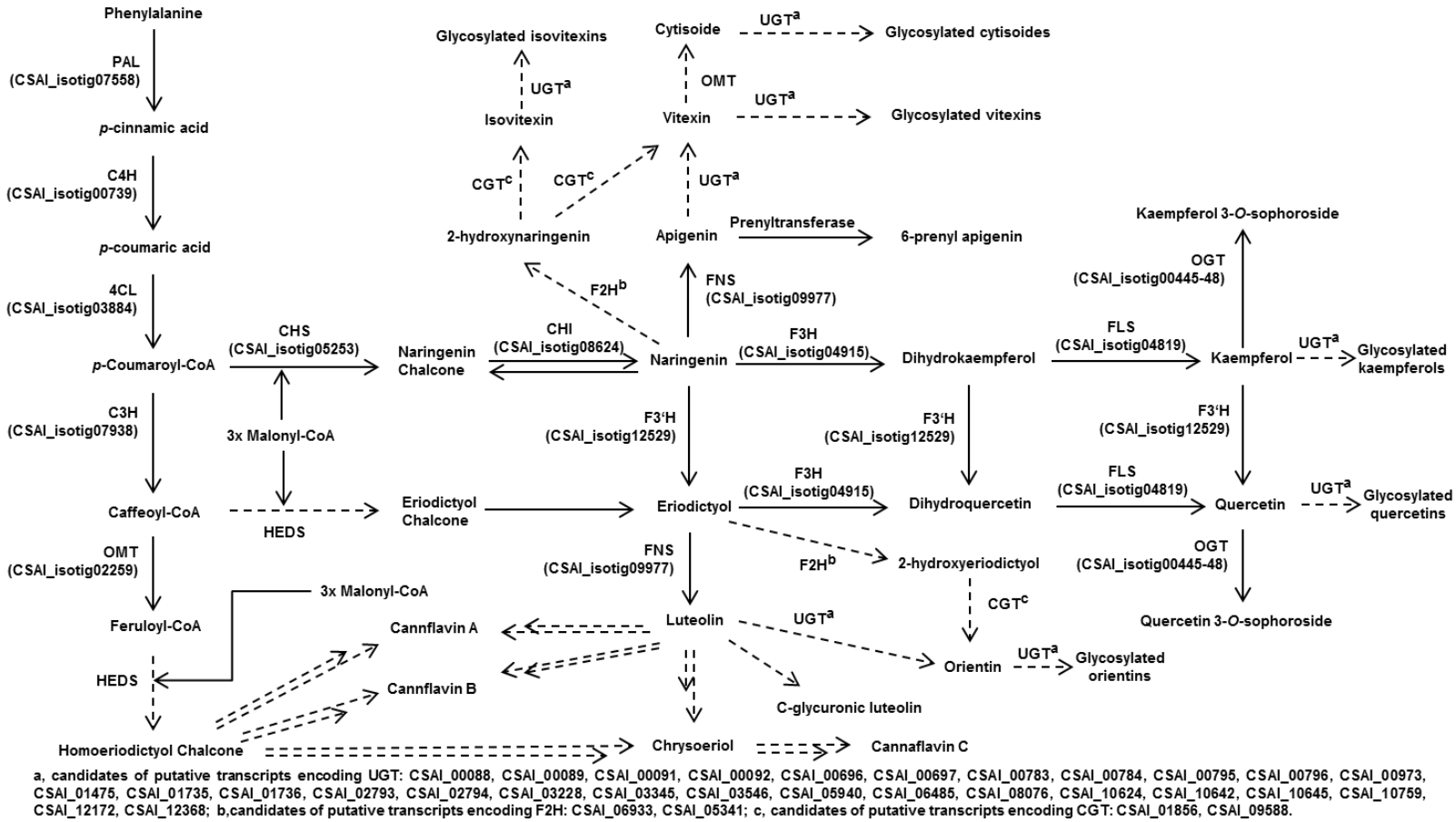


Figure 7. Proposed general flavonoid biosynthetic pathway in *Cannabis* and putative transcripts encoding enzymes in the pathway. Identified enzymes are represented with solid lines. PAL, phenylalanine ammonia lyase; C4H, cinnamate 4-hydroxylase; 4CL, 4-coumarate:CoA ligase; CHS, chalcone synthase; CHI, chalcone isomerase; F3H, flavanone 3-hydroxylase; F'3H, flavanone 3'-hydroxylase; FNS, flavone synthase; FLS, flavonolsynthase; OGT, *O*-glucosyltrasnferase; UGT, UDP-glycosyltransferases; F2H, flavanone 2-hydroxylase; CGT, *C*-glucosyltransferase; C3H, p-coumaryl-CoA 3-hydroxylase; HEDS, homoeriodictyol/eriodictyol synthase; OMT, *O*-methyltransferase.

3.3.3 Putative genes related to flavonoid biosynthesis

In this study, all putative genes that are participating in the main backbone of flavonoids were detected in the cDNA library of *Cannabis* trichomes as seen in figure 7, including phenylalanine ammonia lyase (PAL), cinnamate 4-hydroxylase (C4H), 4-coumarate:CoA ligase (4CL), chalcone synthase (CHS), and chalcone isomerase (CHI). PAL catalyzes conversion phenylalanine into *p*-cinnamic acid, while C4H hydroxylates the product of the first enzyme into *p*-coumaric acid. 4CL adds a CoA thiol ester into *p*-coumaric acid yielding *p*-coumaroyl-CoA, while CHS condenses one molecule of *p*-coumaroyl-CoA and three molecules of malonyl-CoA that is supplied by acetyl-CoA carboxylase yielding a naringenin chalcone. Furthermore, expression level of CHS in the trichomes was around 25 times higher than in the leaves of *Cannabis* as described in figure 4E. CHI isomerizes subsequently naringenin chalcone to naringenin (flavanone), a common precursor of other derivative flavonoids, such as flavonols and flavones.

Some putative genes involved in the downstream steps of flavonoid biosynthesis were also identified in the *Cannabis* trichomes such as flavanone 3-hydroxylase (F3H) and flavanone 3'-hydroxylase (F3'H). Both catalyze the formation of dihydroflavonols from flavanones. In particular, F3H adds a hydroxyl molecule on carbon number 3 of flavanone, such as converting naringenin into dihydrokaempferol, while F3'H puts a hydroxyl group on the carbon number 3', such as modifying naringenin to eriodictyol. Moreover, the transcript of flavonol synthase (FLS), an enzyme that transforms dihydroflavonols into flavonols, was detected as well. This enzyme converts dihydrokaempferol into kaempferol. Besides that, transcript of flavone synthase (FNS) was also identified in the cDNA library of *Cannabis* trichomes. This enzyme catalyzes the conversion of flavanones into flavone, such as converting naringenin into apigenine and transforming eriodictyol to luteolin.

Cannaflavin A and cannaflavin B are unique flavonoids belonging to flavone compounds and found only in *Cannabis*. It was suggested that both compounds biosynthetically originated from luteolin that has been prenylated by a prenyltransferase and methylated by an *O*-methyltransferase (OMT) (Flores-Sanchez and Verpoorte 2008). An alternative biosynthetic route for cannaflavin A and B was also proposed via feruloyl-CoA (Flores-Sanchez and Verpoorte 2008). This route is initiated by the conversion of p-coumaryl-CoA into caffeoyl-CoA by p-coumaryl-CoA 3-hydroxylase (C3H), followed with methylation by an OMT to

form feruloyl-CoA. It is continued with the conversion of feruloyl-CoA and 3x malonyl-CoA to form homoeriodictyol chalcone by homoeriodictyol/eriodictyol synthase (HEDS). In the next biosynthetic steps, homoeriodictyol chalcone is modified with some reactions by unknown enzymes to form cannaflavin A and B. Interestingly, putative genes of C3H and OMT that are participating in this route were detected in the cDNA library of *Cannabis* trcihomes as described in figure 7.

Some glycosylated flavonoids have been isolated from *Cannabis* including kaempferol 3-*O*-glucoside and quercetin 3-*O*-sophoroside (Ross et al. 2005). Both compounds are products from glycosylation of kaempferol and quercetin, respectively, by an UDP-glycosyltransferase (OGT). Fortunately, some unique transcripts that are annotated participating in these glycosylations, were recorded in the cDNA library of *Cannabis* trichomes as described in figure 7. In addition, several *Cannabis* flavonoids are glycosylated through carbon-carbon bonds such as vitexin, isovitexin, and orientin. Biosynthetically, vitexin and isovitexin are suggested as products from glucosylation of apigenin by an UDP-glucosyltransferase (UGT), while orientin is a glucosylation form of luteolin (Flores-Sanchez and Verpoorte 2008). This route was supported by biotransformation of apigenin into vitexin using cell cultures of *Cannabis* (Braemer et al. 1987). Interestingly, some transcripts annotated as UGT were recorded in the cDNA library and probably might be candidates of UGTs involved in glycosylation of *Cannabis* flavonoids. An alternative route for biosynthesis of *C*-glycosylated flavonoids that has been described in cereals (Brazier-Hicks et al. 2009), might be occurred in *Cannabis*. In this pathway, a flavanone is hydroxylated by flavanone 2-hydroxylase (F2H) to form 2-hydroxyflavanone. In the next step, glycosylation occurs on the open-chain form of 2-hydroxyflavanone by *C*-glucosyltransferase (CGT) to produce 2-hydroxyflavanone *C*-glucoside, and then followed subsequently by dehydration to yield flavone-6*C*-glucoside. The gene sequences of F2H and CGT were checked in cDNA library of *Cannabis* trichomes. As the result, some unique transcripts hit the sequences (see figure 7) but with low scores.

4. Conclusion

In this study, cDNA library of *Cannabis* trichomes was constructed using 454 GS FLX pyrosequencing system. SAMS has been successfully applied on the assembled sequences for GO annotation and KEGG pathway assignments. A wide range of putative genes involved in the biosynthetic pathways of secondary metabolites, including terpenoids, cannabinoids, flavonoids, and alkaloids, were identified in the cDNA library. Furthermore, the identification

of transcripts annotated participating in the biosynthesis of flavonoids and alkaloids probably might be a signature for their presences in the *Cannabis* trichomes. Moreover, cDNA library of *Cannabis* trichomes provided valuable information for elucidating the biosynthetic pathways of secondary metabolites in this plant.

Chapter 7

Analysis of cannabinoids in laser-microdissected trichomes of medicinal *Cannabis sativa* using LCMS and cryogenic NMR

Nizar Happyana, Sara Agnolet, Remco Muntendam, Annie Van Dam, Bernd Schneider, Oliver Kayser

Sara Agnolet and Bernd Schneider facilitated cryogenic NMR measurements and corrected the manuscript.
Annie Van Dam supported LCMS analysis.
Remco Muntendam supported sample preparations and corrected the manuscript.
Oliver Kayser coordinated and supervised the project and corrected the manuscript.

Published in Phytochemistry, 2012, 87, 51-59

Abstract

Trichomes, especially the capitate-stalked glandular hairs, are well known as the main sites of cannabinoid and essential oil production of *Cannabis sativa*. In this study the distribution and density of various types of *Cannabis sativa* L. trichomes, have been investigated by scanning electron microscopy (SEM). Furthermore, glandular trichomes were isolated over the flowering period (8 weeks) by laser microdissection (LMD) and the cannabinoid profile analyzed by LCMS. Cannabinoids were detected in extracts of 25 - 143 collected cells of capitate-sessile and capitate stalked trichomes and separately in the gland (head) and the stem of the latter. Δ^9-Tetrahydrocannabinolic acid [THCA (1)], cannabidiolic acid [CBDA (2)], and cannabigerolic acid [CBGA (3)] were identified as most-abundant compounds in all analyzed samples while their decarboxylated derivatives, Δ^9-tetrahydrocannabinol [THC (4)], cannabidiol [CBD (5)], and cannabigerol [CBG (6)], co-detected in all samples, were present at significantly lower levels. Cannabichromene [CBC (8)] along with cannabinol (CBN (9)) were identified as minor compounds only in the samples of intact capitate-stalked trichomes and their heads harvested from 8-week old plants. Cryogenic nuclear magnetic resonance spectroscopy (NMR) was used to confirm the occurrence of major cannabinoids, THCA (1) and CBDA (2), in capitate-stalked and capitate-sessile trichomes. Cryogenic NMR enabled the additional identification of cannabichromenic acid [CBCA (7)] in the dissected trichomes, which was not possible by LCMS as standard was not available. The hereby documented detection of metabolites in the stems of capitate-stalked trichomes indicates a complex biosynthesis and localization over the trichome cells forming the glandular secretion unit.

Keywords: *Cannabis sativa*, laser microdissection (LMD), liquid chromatography – mass spectrometry (LCMS), cryogenic nuclear magnetic resonance (NMR), cannabinoid biosynthesis, capitate-stalked trichome, capitate-sessile trichome, gland, scanning electron microscopy (SEM)

1. Introduction

Cannabis sativa L. is an annual, dioecious herb (Flemming et al. 2007; Pacifico et al. 2006), belonging to the family of Cannabaceae and originating from Eastern and Central Asia (Candolle 1886; de Barge 1860). It has long been used in traditional Asian medicine, mainly in India, before the Christian era (Zuardi 2006). Phytocannabinoids (cannabinoids), a unique group of terpenophenolics possessing alkylresorcinol and monoterpene moieties in their molecular structure (Fig. 1) are considered the most responsible compounds for the biological activities of *Cannabis sativa* L. More than 100 cannabinoids have been identified and structurally elucidated, including recently isolated new entities (Ahmed et al. 2008a; Ahmed et al. 2008b; ElSohly and Slade 2005; Radwan et al. 2009; Radwan et al. 2008a; Radwan et al. 2008b). Because of their psychoactivity, Δ^9-tetrahydrocannabinol (THC (4)) and cannabidiol (CBD (5)) are the most studied and interesting compounds of the class.

Despite the recent completion of *C. sativa* genome sequencing (van Bakel et al. 2011), the metabolic pathway of cannabinoids is not fully understood so far. The biosynthesis starts with the formation of two cannabinoid precursors, namely geranyldiphosphate (GPP), originating predominantly from the non-mevalonate pathway (MEP) (Fellermeier et al. 2001), and olivetolic acid, derived from the not yet fully elucidated polyketide pathway. GPP and olivetolic acid are condensed to form cannabigerolic acid (CBGA (3)) by an enzyme predicted to be a representative of the geranyltransferase group (Fellermeier and Zenk 1998). CBGA (3) is subsequently transformed to tetrahydrocannabinolic acid (THCA (1)), cannabidiolic acid (CBDA (2)) and cannabichromenic acid (CBCA (7)) by THCA synthase (Taura et al. 1995), CBDA synthase (Taura et al. 1996), and CBCA synthase (Morimoto et al. 1998), respectively. The protein structures and enzymatic activities of THCA and CBDA synthases have been elucidated comprehensively (Sirikantaramas et al. 2004; Taura et al. 2007c). Although CBCA synthase was purified to homogeneity (Morimoto et al. 1998), its sequence has not yet been deposited into the public databases. Finally, THC (4), CBD (5), and cannabichromene (CBC (8)) are formed by decarboxylation of their acidic forms during storage or through interaction with heat and light (Lewis and Turner 1978).

Recently, new insights into the cellular localization of the cannabinoids biosynthesis were revealed, focusing on the capitate-stalked trichomes (E. S. Kim and Mahlberg 1997; Marks et al. 2009a; Sirikantaramas et al. 2005) as the main site of their storage (Lanyon et al. 1981; Petri et al. 1988; Turner et al. 1978). Capitate-stalked trichomes consist of two parts, the gland

(head) and the stem (Fig. 1B). The head contains disc cells which are surrounded by the storage cavity. Disc cells are presumed to be the site of cannabinoid production (E. S. Kim and Mahlberg 1997, 2003, 1991). The stem is formed by stipe cells and basal cells (E. S. Kim and Mahlberg 1997) and is not yet functionally characterized. Kim and Mahlberg (E. S. Kim and Mahlberg 1997) reported that cannabinoids, represented by THC, in capitate-stalked trichomes are secreted particularly from disc cells and accumulated in the cell wall, the fibrillar matrix – a surface feature of vesicles in the storage cavity, the subcuticular wall, and the cuticle. Furthermore, Marks *et al.* (Marks et al. 2009a) confirmed the head of the capitate-stalked trichome to be the major site of cannabinoids production by proving the presence of cDNAs encoding for three possible polyketide (only one provided olivetol), MEP pathway, and THCA synthases.

Each tissue and cell type of plants has a specific task that is driven by its own unique transcriptome, proteome, and metabolome. Identifying the functions of specific tissues and cell types from plants requires accurate and efficient methods for collecting the populations of the material of interest. Recently, the combination of laser-assisted microdissection technique with diverse range of molecular technologies has allowed this purpose to be achieved. (Day et al. 2005) Laser microdissection (LMD) and molecular biology techniques have been used successfully to localize enzymes of artemesinin biosynthesis in the apical cells of trichomes of *Artemisia annua* L. (Olsson et al. 2009) Furthermore, this alliance method has been applied for constructing a cDNA library from isolated phloem cells of rice leaf tissue (Asano et al. 2002) and comprehensive proteome analysis of specific plant tissues (Schad et al. 2005a). LMD has not only been combined with molecular biology techniques but also with metabolites analysis techniques. LMD together with GC-MS has been used to analyze a set of metabolites from vascular bundles of *Arabidopsis thaliana* (Schad et al. 2005b) and its combination with cryogenic NMR and HPLC or MS has enabled the comprehensive analysis of metabolites from the secretory cavities of *Dilantris pillansii* leaves (Schneider and Hölscher 2007) or the identification of two phenolic compounds in the stone cells of Norway spruce respectively (S.-H. Li et al. 2007).

In this study, the distribution and density of various types of *Cannabis sativa* L. trichomes, have been determined by scanning electron microscopy (SEM) analysis. Moreover to further advance the understanding of cannabinoids production in trichomes, we hereby report on metabolite profiles as analyzed in specific cells of the glandular hairs as a secreting plant

organ, particularly in the intact capitate-sessile and capitate-stalked trichomes as well as in the heads and stems of the latter. The specific cells of trichomes were separated and collected using LMD. Cannabinoids in the dissected samples were analyzed by means of liquid chromatography-mass spectrometry (LCMS) and cryogenic nuclear magnetic resonance (NMR). Our results suggest that cannabinoid biosynthesis is not only limited to the expected head cells, but also stems of *Cannabis* capitate-stalked trichomes might play a role in cannabinoids production. Main difference is the quantitative ratio bringing up the question of biosynthesis regulation in Cannabis trichomes.

2. Material and methods

2.1 Chemicals and reagents

Reference compounds of cannabinoids for Δ^9-tetrahydrocannabinol [Δ^9-THC (4)], Δ^8-tetrahydrocannabinol (Δ^8-THC), cannabigerol [CBG (6)], cannabidiol [CBD (5)], cannabichromene [CBC (8)], cannabinol [CBN (9)], Δ^9-tetrahydrocannabinolic acid [THCA (1)], cannabigerolic acid [CBGA (3)], cannabidiolic acid [CBDA (2)], olivetol, and olivetolic acid were purchased from THC Pharm GmbH (Frankfurt, Germany). (±)-11-nor-Δ^9-THC carboxylic acid-D3 (THCA-D3) was acquired from Lipomed (Arleisheim, Switzerland). Analytical reagent-grade and HPLC-grade solvents were purchased from Merck Biosolve Ltd. (Valkenswaard, the Netherlands).

2.2 Plant material

Medicinal *Cannabis sativa* L., variety Bediol® (THC level, approx. 6 %, CBD level, approx. 7.5 %) was supplied by Bedrocan BV (the Netherlands). *Cannabis* plants were grown indoors under standardised conditions. They were initially generated from cuttings of standardized mother plants and cultivated under controlled, long day-light conditions (18 h/day). After the vegetative growth phase, the flowering stage was induced under a shorter (12 h/day) light regime for 8 weeks. The trichomes were isolated and analyzed at 4 to 8 weeks of the flowering period. Plant specimens were assigned voucher numbers (A1-05.41.050710) and deposited at Bedrocan BV. All plant handling and experimental procedures were carried out under the license No. 4584989 issued by the Federal Institute for Drugs and Medical Devices (BfArM), Germany.

2.3 Scanning electron microscopy (SEM)

Slices of bracts, leaf-flowers, leaves, and stems were fixed in 2% buffered glutaraldehyde solution and incubated overnight at 4 ˚C. The samples were washed with 0.1 M cacodylate buffer, incubated with 1% OsO_4 in 0.1 M cacodylate buffer for 1 h at room temperature, washed with water, and subsequently dehydrated with 30%, 50%, and 70% ethanol, consecutively for 15 min, and 3 × 96% ethanol for 30 min. After complete dehydration by critical point drying, the samples were mounted on specimen holders and examined and photographed by means of JEOL 6301F SEM.

2.4 Laser microdissection (LMD) microscopy

The samples were fixed according to Olsson *et al.* (Olsson et al. 2009) with slight modifications. Bracts and floral flowers were cut in small slices with a razor blade (length 0.2 - 0.5 cm), fixed in 4% phosphate buffered formaldehyde, exposed to vacuum for 10 min, and infiltrated overnight at 4 ˚C. The specimens were then put on steel frame (PET-membrane, 25 mm x 76 mm) slides and subjected to microdissection using a Leica LMD 6000 microscope. Dissections were performed at the following setting: draw + cut or move + cut mode, 4 or 3 revolver, 10x or 20x magnification, 32 aperture, and 13 speed. We dissected intact capitate-stalked trichomes, their heads and stems, and intact capitate-sessile trichomes. Thus prepared trichomes samples were collected by gravity in Eppendorf tubes, vortexed for 1 min (New Heidolph Reax Top Shaker, Heidolph Brinkmann, USA), and centrifuged for 5 min (Microcentrifuge 1-14 (10014), Sartorius Stedim Biotech, Germany). The dissected samples were stored at -20 ˚C.

2.5 Extractions

In this research, extractions of microdissected samples were carried out with two different methods depend on the analysis methods. First, for LCMS measurements, all microdissected samples from week 4, 5, 6, 7, and 8 in the tubes were added with 60 µL of methanol, followed by sonication for 10 min and incubation for overnight at room temperature. The samples were diluted in a H_2O/MeOH (2/1, v/v) + 0.1% formic acid solution. For cryogenic NMR measurements, dissected capitate-stalked and capitate-sessile trichomes harvested at week 8 were added with the minimum amount of $CDCl_3$, sonicated for 1 min, and transferred to 2.5 mm-diameter NMR tubes for analysis.

2.6 Liquid chromatography – mass spectrometry (LCMS)

HPLC was performed using a Shimadzu LC system consisting of a SIL-20AC autosampler, two LC-20AD gradient pumps, and an LC-20AD pump for the post-column addition of ammonia. Chromatographic separation was achieved at room temperature on an Alltima C18 column (2.1 × 150 mm, 5 μm) from Grace Davison Discovery Sciences. The mobile phase was composed of solvent A: H_2O + 0.1% formic acid and solvent B: MeOH + 0.1% formic acid. The flow rate was 250 μL/min. The elution was performed starting at 30% B, which was increased linearly to 70% B in 1 min after which a linear gradient was applied to 90% B in 30 min. The column was subsequently washed with 95% B for 10 min, before returning to the starting conditions. The injection volume was 60 μL for the measurements in positive mode and 10 μL for the measurements in negative mode. The HPLC system was coupled with an API 3000 triple-quadrupole mass spectrometer (Applied Biosystems/MDS Sciex) equipped with a TurboIon Spray source. Depending on the properties of the components, either positive or negative electrospray ionization was applied. The ionization spray voltage was 5.2 kV in the positive mode and -4.5 kV in the negative mode. The TurboIon Spray temperature was 450 ˚C. Nitrogen was used as a turbo heating gas, nebulizing gas, and curtain gas. For negative ionization, post-column addition of a 1% ammonia solution in H_2O/MeOH (1/1, v/v) was utilized at a flow rate of 50 μL/min. Tuning parameters were optimized for each individual component in order to maximize the signals in multiple reaction monitoring (MRM) (Tab. 2). Data were collected and analyzed with Analyst 1.5.1 software (Applied Biosystems/MDS Sciex). Δ^8-Tetrahydrocannabinol was added as an internal standard to all samples in the positive ionization measurements and (±)-11-nor-Δ^9-THC carboxylic acid-D3 (THCA-D3) – in the negative ionization measurements. For validation of semi-quantitative analysis, calibration curves of internal standards were prepared.

2.7 Cryogenic NMR

^{1}H NMR spectra of the laser-microdissected trichome samples were measured with a Bruker Avance 500 NMR spectrometer operating at 500.13 MHz and equipped with a 5 mm TCI cryoprobe. The spectra were recorded as 32 k data points with the spectral width of 20 ppm, using a sequence with water suppression and a total of 1 k scans. TMS was used as an internal standard. The spectra were manually calibrated, and the phase and baseline were corrected by means of TopSpin 2.1 software.

^{1}H NMR spectra of the standard compounds were recorded on a Bruker Avance DRX 500 spectrometer operating at 500.13 MHz and equipped with a 5 mm ^{1}H(^{13}C) probe and a BACS 60 sample changer. The spectra were acquired automatically at ambient temperature and a total of 128 transient signals were recorded as 64 k data points with a spectral width of 15 ppm using a sequence with water suppression. Cannabigerolic acid (3) reference compound was dissolved in DMSO-d_6, and the solvent residual peak was used as an internal reference. Δ^9-Tetrahydrocannabinol (1) was measured in methanol-d_4, and the remaining standard compounds in $CDCl_3$, with the TMS signal serving as an internal standard.

3. Results and discussion

3.1 Distribution and density of trichomes

Trichomes are basically divided into two general categories: non-glandular and glandular (Schilmiller et al. 2008). There are three different trichome classes on *Cannabis sativa* L., namely capitate-stalked, capitate-sessile and bulbous trichomes and those are categorized as glandular trichomes (Fig. 2). To study the distribution and density of trichomes on medicinal *Cannabis sativa* L., we subjected slices of bracts, floral leaves, stems, and leaves to scanning electron microscopy (SEM) analysis and calculated the numbers of all trichome types according to the monography of the European Pharmacopoea.

The appearance of capitate-stalked trichomes was champignon-like, characterized by a gland (head) and a stem (Fig. 2B). The SEM studies revealed their presence only on the flowers during the flowering period, with particularly high density observed on the bracts and the floral leaves. Moreover, capitate-stalked trichomes on the bracts were more crowded than those on the floral leaves (Tab. 1). The presence of cannabinoid-rich capitate-stalked trichomes on the bracts and the floral leaves accounts for the abundance of cannabinoids in the flower.

Subsequent analyses revealed presence of hair-like capitate-sessile trichomes (Fig. 2C) on flowers, stems, and leaves in both vegetative and flowering periods. Two types of capitate-sessile trichomes could be distinguished according to their size: big, found only on the flowers (Fig. 2G) and small (Fig. 2H), present on flowers, leaves, and stems, confirming the results of previous studies (Dayanandan and Kaufman 1976; Fairbair.Jw 1972). As shown in Tab. 1, the highest population density of big capitate-sessile trichomes was present on the bracts, while the small variety could predominantly be found on the abaxial surfaces of the leaves. Small

capitate-sessile trichomes on the flower could not be counted, as they were overshadowed by other trichome types.

Bulbous trichomes, shaped like balloons (Fig. 2D), were the smallest trichomes found on the plant surface. They are generally made up of two, but never more than four cells (Dayanandan and Kaufman 1976). According to our observations, bulbous trichomes were present on flowers, leaves, and stems, with the highest population density detected on the stems and the lowest on the bract (Tab. 1).

3.2 Laser microdissection

For the analysis of the cannabinoid profiles, our study focused on capitate-stalked and big capitate-sessile trichomes. Both are reported to accumulate high levels of cannabinoids (especially the capitate-stalked glandular hairs (Turner et al. 1978)) and represent the majority of trichomes visually traceable by a microscope. According to the densities reported above, the floral parts of the plants were used to isolate the selected trichomes. Specific dissections of trichome cells were conducted using laser microdissection microscopy (LMD). Based on our continuing work (not published), cannabinoids in bred plants are produced mostly at the last five weeks of the total flowering cyclus of eight weeks. Therefore in this experiment, we focused to analyse cannabinoids production on trichomes from week 4 to 8 of the flowering period.

Using LMD, we dissected intact capitate-stalked trichomes, heads and stems thereof, as well as intact capitate-sessile hairs. To avoid decarboxylation and degradation of cannabinoids, the dissected samples were stored at -20˚C directly after LMD. The process of trichome dissection is illustrated in Fig. 3. Thus obtained samples were extracted as described at experimental section and subjected to LCMS or cryogenic NMR for cannabinoid analysis.

3.3 Cannabinoid analysis with LCMS

The LCMS analysis were carried out in negative ion mode for identifying acidic compounds such as cannabinoid acids and positive ion mode for detecting neutral cannabinoids. Δ^8-Tetrahydrocannabinol was used as internal standard in the positive ionization measurements, meanwhile (±)-11-nor-Δ^9-THC carboxylic acid-D3 (THCA-D3) was used as internal standard in the negative ion mode.

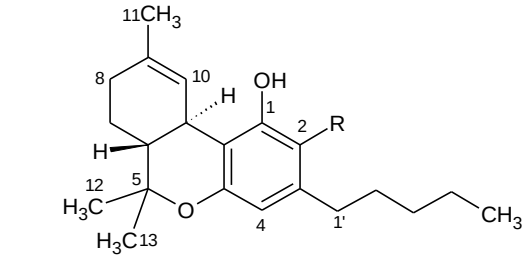
Table 1. Density and distribution of trichomes on flowers, leaves, and stems of *Cannabis sativa* L. during flowering period. T: trichomes, NT: number of trichomes, D: density of trichomes, B: bract, Fd: adaxial surface of a floral leaf, Fb: abaxial surface of a floral leaf, Ld: adaxial surface of a leaf, Lb: abaxial surface of a leaf, S: stem, A: average, NF: not found, NC: not counted.

Sample	Area (mm^2)	Capitate-stalked		Big Capitate-sessile		Small Capitate-sessile		Bulbous	
		NT	D (T/mm^2)	NT	D (T/mm^2)	NT	D (T/mm^2)	NT	D (T/mm^2)
B1	5.4	269	49.7	15	2.7	NC	-	15	2.7
B2	10.4	467	44.9	23	2.2	NC	-	23	2.2
B3	8.2	345	41.8	18	2.2	NC	-	18	2.2
Fd1	6.6	88	13.3	6	0.9	NC	-	6	0.9
Fd2	4.9	86	17.7	5	1.0	NC	-	5	1.0
Fd3	4.2	66	15.6	4	0.9	NC	-	4	0.9
Fb1	6.8	151	22.2	23	3.4	NC	-	23	3.4
Fb2	8.3	246	29.6	8	0.9	NC		24	2.9
Fb3	10.8	219	20.3	23	2.1	NC	-	44	4.0
Ld1	3.6	NF	-	NF	-	113	31.0	3	0.8
Ld2	4.2	NF	-	NF	-	152	35.9	4	0.9
Ld3	4.5	NF	-	NF	-	126	27.7	5	1.1
Lb1	1.5	NF	-	NF	-	430	286.7	13	8.7
Lb2	1.4	NF	-	NF	-	554	395.7	12	8.6
Lb3	1.9	NF	-	NF	-	668	351.5	16	8.4
S1	5.4	NF	-	NF	-	326	60.8	82	15.3
S2	5.8	NF		NF		310	53.1	78	13.3
S3	4.8	NF	-	NF	-	294	61.0	84	17.4

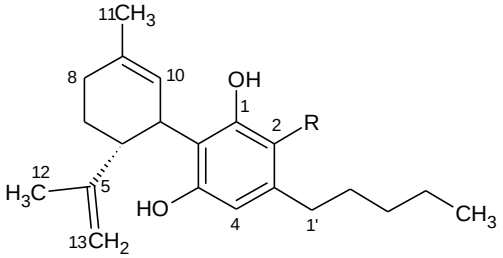
Microdissected trichomes were analyzed for the presence of olivetolic acid, acidic and neutral cannabinoids. Previous studies reported about cannabinoids in the trichomes, but they did not discriminate between the trichome type and detected compounds were neutral cannabinoids, such as THC (4), CBD (5), CBC (8), and CBN (9) (E. S. Kim and Mahlberg 2003, 1991; Lanyon et al. 1981; Mahlberg and Kim 2004; Turner et al. 1981; J. C. Turner et al. 1980; Turner et al. 1978). In this LCMS study the spectrum was extended to cannabinoid acids such as THCA (1), CBDA (2), and CBGA (3), together with neutral cannabinoids, THC (4), CBD (5), and CBG (6) as shown in the Fig. 4. To our knowledge, this is the first report regarding cannabinoid acids in *Cannabis* trichomes. Identified compounds and structures are given in Fig. 1. The unprecedented detection of cannabinoids in the stems of capitate-stalked trichomes

was of high interest. Although olivetolic acid or its decarboxylation product was expected to be present in studied samples, their concentrations were below the LCMS detection limits.

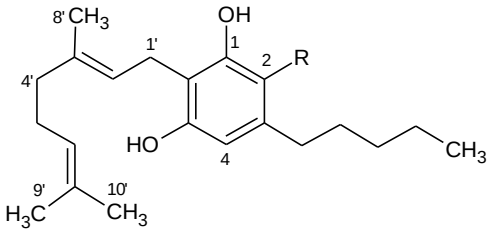
Comparative study of cannabinoids profiles of microdissected trichomes collected at the last 5 weeks of the total 8-week growth cycle was carried out qualitatively based on LCMS analysis. As shown in Fig. 4, the cannabinoid profiles in all of the investigated samples were qualitatively similar: THCA (1), CBDA (2), CBGA (3), THC (4), CBD (5), and CBG (6) were detected in all samples during the flowering stages, while CBC (8) and CBN (9) were only identified in the intact capitate-stalked trichomes, and their heads at week 8.



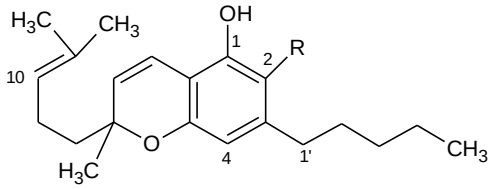
1 R=COOH tetrahydrocannabinolic acid (THCA)
4 R=H tetrahydrocannabinol (THC)



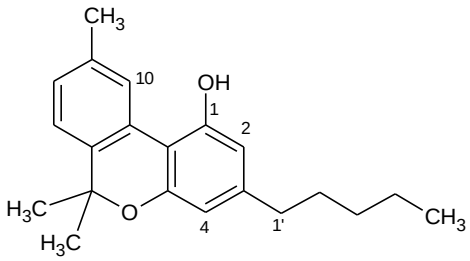
2 R=COOH cannabidiolic acid (CBDA)
5 R=H cannabidiol (CBD)



3 R=COOH cannabigerolic acid (CBGA)
6 R=H cannabigerol (CBG)



7 R=COOH cannabichromenic acid (CBCA)
8 R=H cannabichromen (CBC)



9 cannabinol (CBN)

Figure 1. Structures of identified cannabinoids

Table 2. Multiple reaction monitoring (MRM) transitions of the components measured by electrospray ionization; positive voltages in positive ionization, negative voltages in negative ionization.

Compounds	Retention time (min)	MRM ionpair	Declustering Potential (V)	Focusing Potential (V)	Collision Energy (V)	Collision Cell Exit Potential (V)
CBG (6)	19.3	317.2/193.2	36	130	23	12
CBD (5)	18.7	315.2/193.2	46	160	29	10
CBN (9)	24.6	311.2/223.2	56	190	29	10
THC (4)	26.6	315.2/193.2	46	160	29	10
$^{8}\Delta$-THC (IS)	27.8	315.2/193.2	46	160	29	10
CBC (9)	30.1	315.2/193.2	46	160	29	10
THCA-D3 (IS)	18.4	346.2/302.1	-56	-170	-30	-13
OA	8.0	223.1/179.0	-36	-110	-26	-9
THCA (1)	34.6	357.2/313.2	-61	-190	-34	-13
CBDA (2)	20.9	357.2/339.2	-56	-170	-30	-17

In the next step, the content of cannabinoids in the dissected trichomes was determined semi-quantitatively. As shown in Tab. 3, the concentration of cannabinoids in the capitate-stalked trichomes was higher as compared to the capitate-sessile hairs, e.g., the concentration of THCA (1)/number dissected entities at week 7 in the capitate-stalked trichomes was 388 ng/mL, while that in their capitate-sessile counterparts was 93 ng/mL. This result stands along with the previous publication by Turner *et al.* (Turner et al. 1978). In the last week of harvesting (week 8), cannabinoids were produced in significantly higher amounts, e.g., in the capitate-sessile trichomes, the concentration of THCA (1) at week 7 was 93 ng/mL and at week 8, 313 ng/mL. Furthermore, THCA (1), CBDA (2), and CBGA (3) were present in high concentrations in all tested samples relative to other cannabinoids. Meanwhile, THC (4), CBD (5), and CBG (6) were detected at low concentrations.

Cannabinoid acids, such as THCA (1), CBDA (2), and CBGA (3), are decarboxylated into neutral cannabinoids in the living plant. The decarboxylation rate increases after harvesting, during storage or through heating and exposure to light (Lewis and Turner 1978). Moreover, heat, light (Lewis and Turner 1978), and auto-oxidation (Razdan et al. 1972) are the contributing factors of further cannabinoid deterioration. CBN (9) is an artifact of THC (4) and can be an indicator of the degradation process (Flemming et al. 2007). As LMD might generate heat in the plant tissue at the site of laser cutting, we were concerned that its application would enforce decarboxylation of acids and degrade other cannabinoids.

Nevertheless, according to the results of LCMS analysis, the concentrations of THC, CBD (5), and CBG (6) were lower than those of their acidic counterparts (Tab. 2). Moreover, CBN (9) as the indicator of degradation could only be identified at low concentration at week 8. Therefore, we conclude that the use of LMD for trichomes dissection does not influence the composition of cannabinoids in laser-microdissected samples.

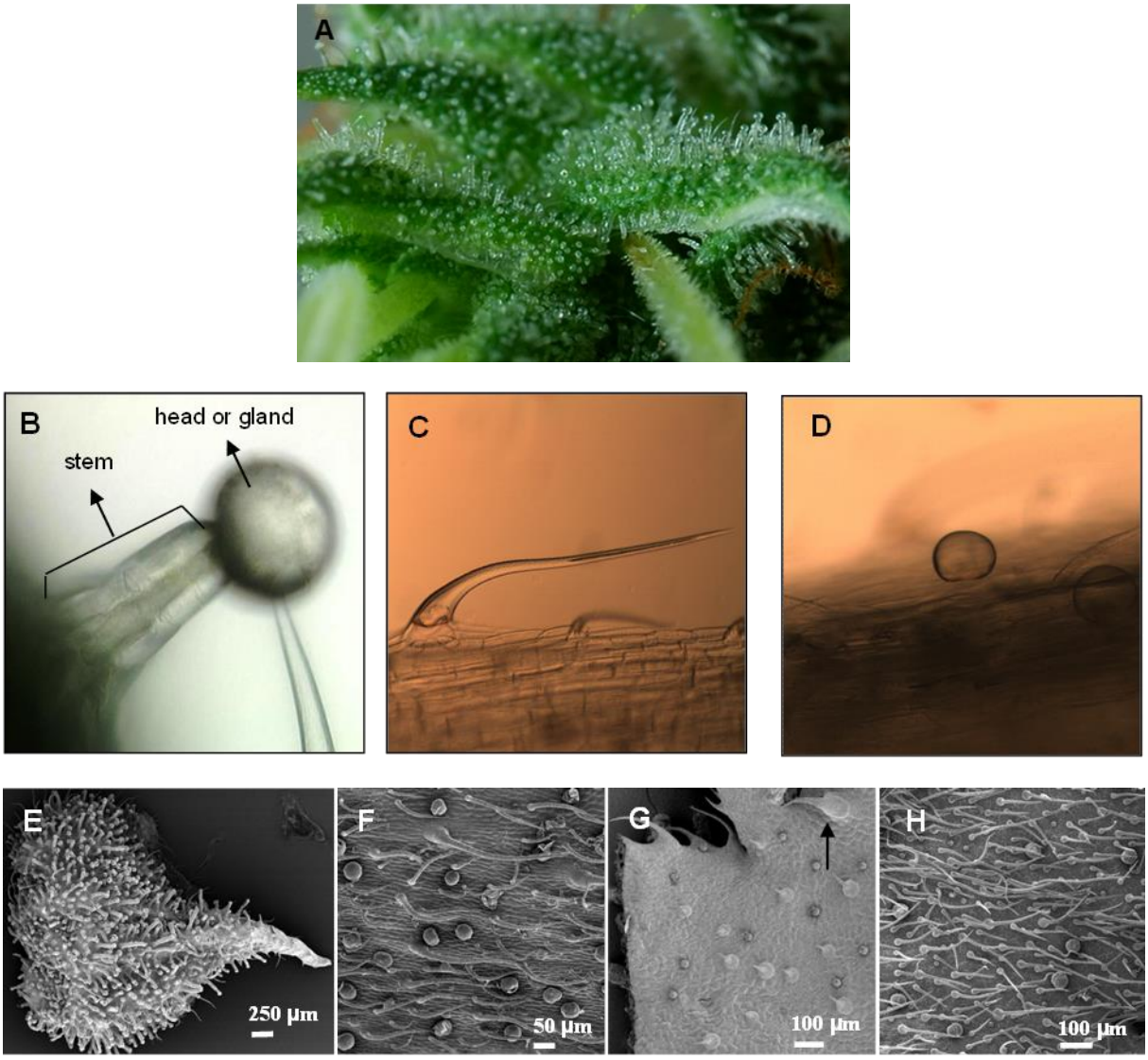


Figure 2. Trichomes of *Cannabis sativa* L.. A: trichomes on the flower, B: capitate-stalked trichome, C: capitate-sessile trichome, D: bulbous trichome, E: trichomes on the bract, F: trichomes on the stem, G: trichomes on the adaxial surface of a floral leaf; a big capitate-sessile trichome is indicated with an arrow, H: trichomes on the abaxial surface of a leaf; present abundant small capitate-sessile and bulbous trichomes.

Previous research confirmed the localization of cannabinoid production in the heads of capitate-stalked trichomes by identifying candidate biosynthetic genes (Marks et al. 2009a; Sirikantaramas et al. 2005). However, our analysis of the capitate-stalked trichome stems indicated the presence of cannabinoids therein as well. Thus, cannabinoid production might not be exclusive to the head of the glandular trichomes. Nonetheless, the source of cannabinoids in the stems of capitate-stalked trichomes is still unclear – they could be derived from the stem itself or translocated from surrounding tissues, like leaves and flowers. Although the function of the capitate-stalked trichome stems in cannabinoid biosynthesis is not yet defined, the detection of cannabinoids suggests that they could play a role in the production of these valuable natural products.

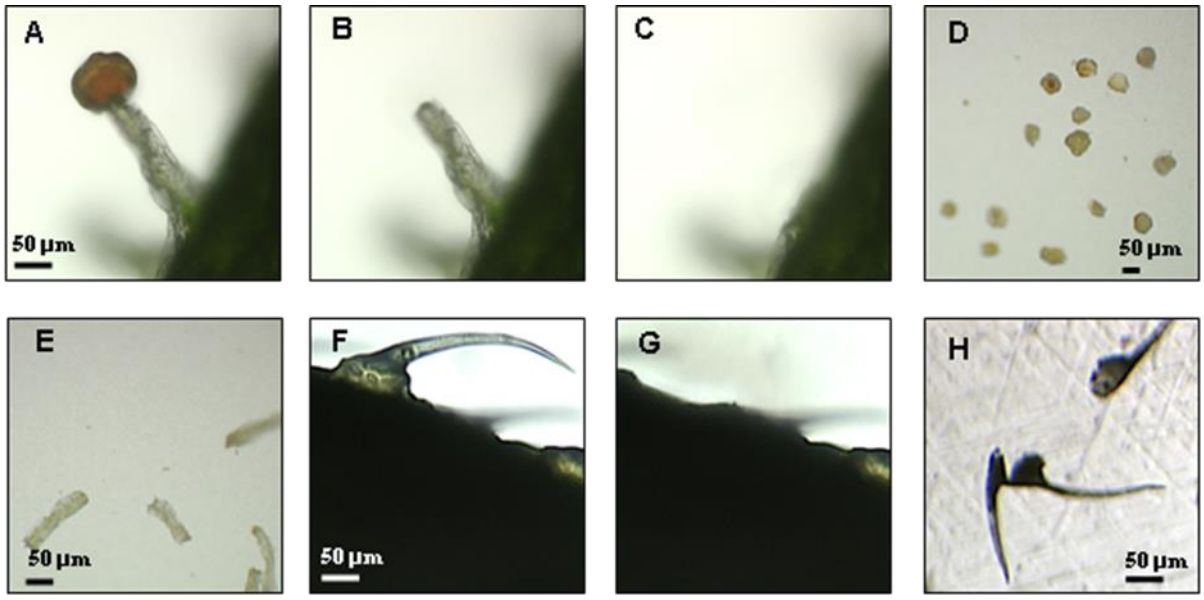


Figure 3. Microdissection of trichomes of medicinal *Cannabis sativa* L. A: intact capitate-stalked trichome before dissection, B: capitate-stalked trichome after dissection of the head cells, C: capitate-stalked trichome – complete dissection, D: dissected head cells, E: stem cells after dissection, F: capitate-sessile trichome before dissection, G: view after dissection of the capitate-sessile trichome, H: dissected capitate-sessile trichomes.

The function of the storage cavity in the capitate-stalked trichome head as a metabolite repository (cannabinoids as well as other compounds, e.g. essential oil components) is well documented (E. S. Kim and Mahlberg 1997). In spite of its cytotoxic effect on plant cells, THCA (1) is accumulated in the storage cavity (Sirikantaramas et al. 2005), presumably due

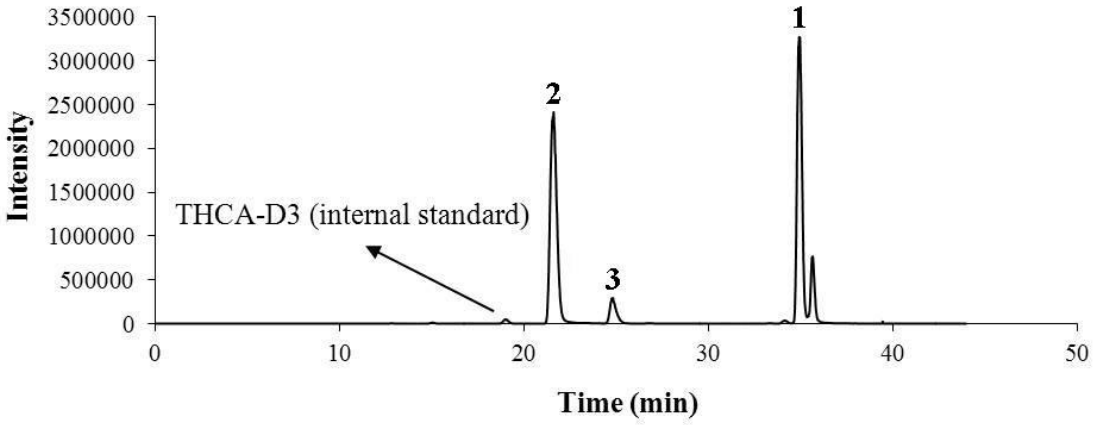
to its ecological role in protection against insects (Taura et al. 2007b). In addition, it was reported that THCA (1) and CBCA (7) effectively induce cell death in the leaf cells or suspension-cultured cells of *Cannabis sativa* (Morimoto et al. 2007). Therefore, to prevent cellular damage, cannabinoid secretion into the storage cavity is necessary (Sirikantaramas et al. 2005; Taura et al. 2007b). It seems reasonable to assume that the toxic cannabinoids potentially synthesized in the stem of a capitate-stalked trichome would also have to be transported to the storage cavity.

3.4 Cannabinoids analysis with cryogenic NMR

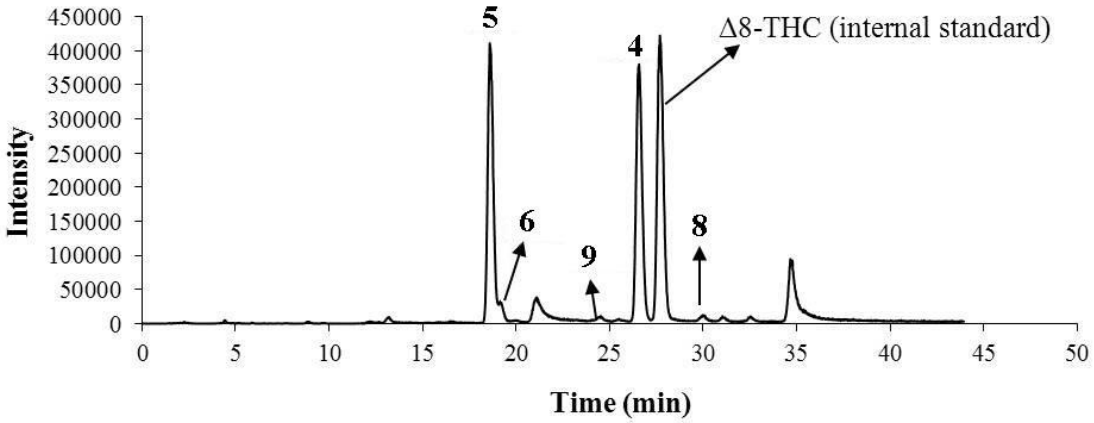
Cryogenic NMR generally is useful to identify compounds in mass-limited samples thanks to sensitivity improving and noise reducing probing capability coupled with a cryogenic cooling system for the receiver coil and preamplifiers, thus providing spectra of improved quality. Therefore, beside LCMS, laser-microdissected samples of capitate-stalked and capitate-sessile trichomes from week 8 were also analysed with cryogenic ^{1}H NMR. Fig. 5A and B depicts the spectra obtained at week 8. Signals labeled **1** in the spectrum shown in Fig. 5a nicely match the spectrum of THCA (1) (Fig. 5C), confirming the presence of THCA (1) in the capitate-stalked trichomes. The signals of THCA (1) are also visible in the ^{1}H NMR spectrum of capitate-sessile trichomes harvested at week 8 (Fig. 5B). Moreover, signals labeled **2** are matching the spectrum of CBDA (2) (Fig. 5D) and established the presence of CBDA (2) in capitate-stalked trichomes. These ^{1}H NMR data are in accord with previously obtained LCMS data, identifying THCA (1) along with CBDA (2) as the predominant compounds of the capitate-stalked and capitate-sessile trichomes. In contrast, ^{1}H NMR signals of the neutral cannabinoids could not be detected in the trichome samples due to their aforementioned low concentrations.

Application of the alternative analytical method of cryogenic ^{1}H NMR (vs. LCMS) enabled identification of signals tentatively assigned to CBCA (7) protons, in addition to ^{1}H NMR resonances of THCA (1) and CBDA (2), in the capitate-stalked trichome samples: H-5, H-4, and H-6 were detected at δ 6.74 (1H, d, J = 10.1 Hz), δ 6.23 (1H, s), and δ 5.49 (1H, d, J = 10.1 Hz), respectively, and are in accordance with reported data (R. Y. Lee and X. Wang 2005).

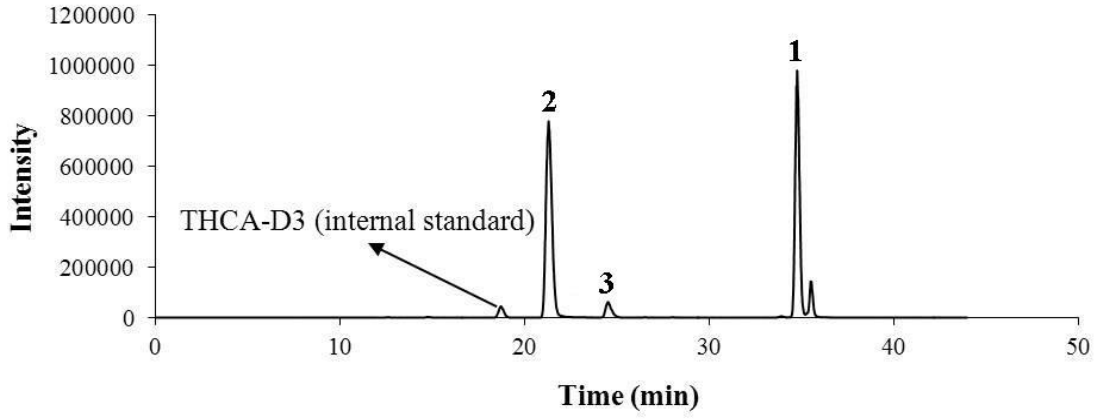
MRM ion pair chromatogram of ST8 in negative mode measurement



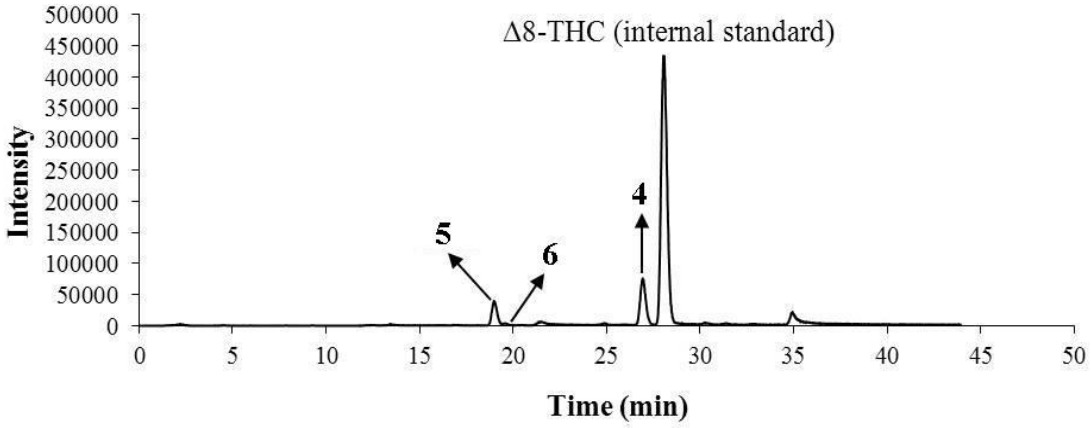
MRM ion pair chromatogram of ST8 in positive mode measurement



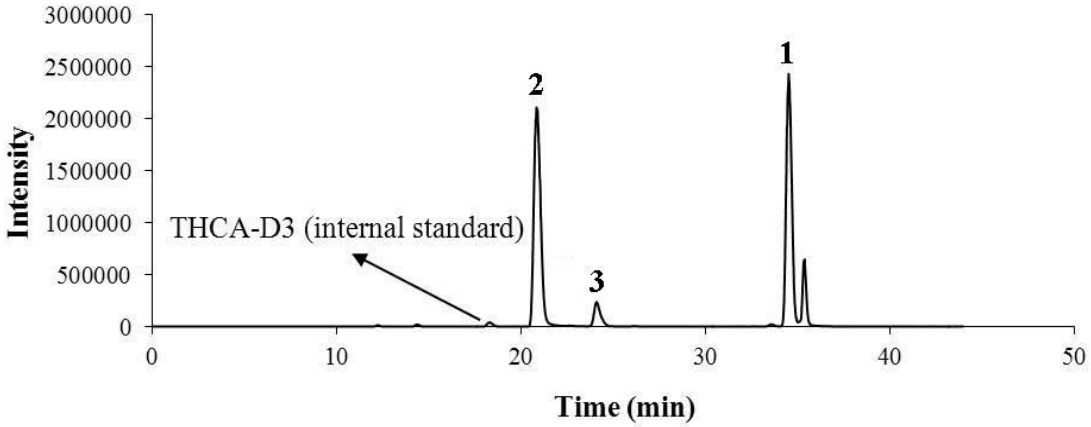
MRM ion pair chromatogram of SE8 in negative mode measurement



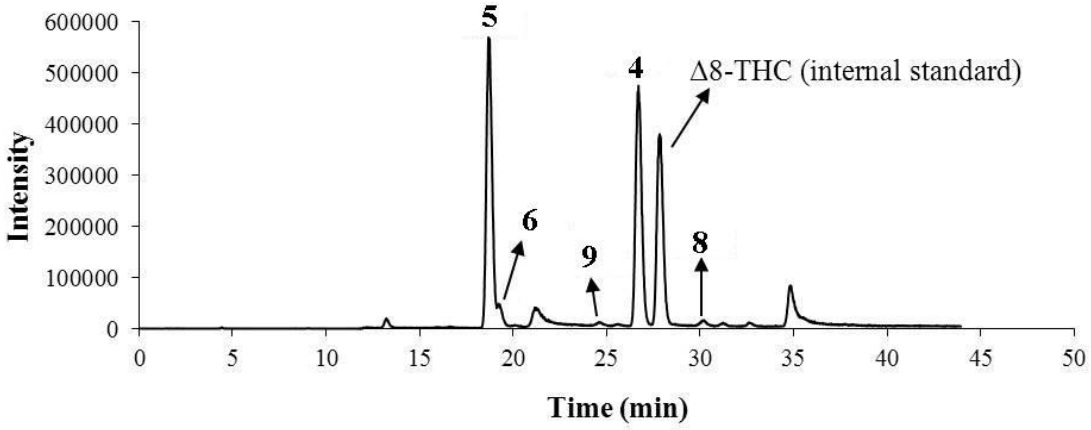
MRM ion pair chromatogram of SE8 in positive mode measurement



MRM ion pair chromatogram of D8 in negative mode measurement



MRM ion pair chromatogram of D8 in positive mode measurement



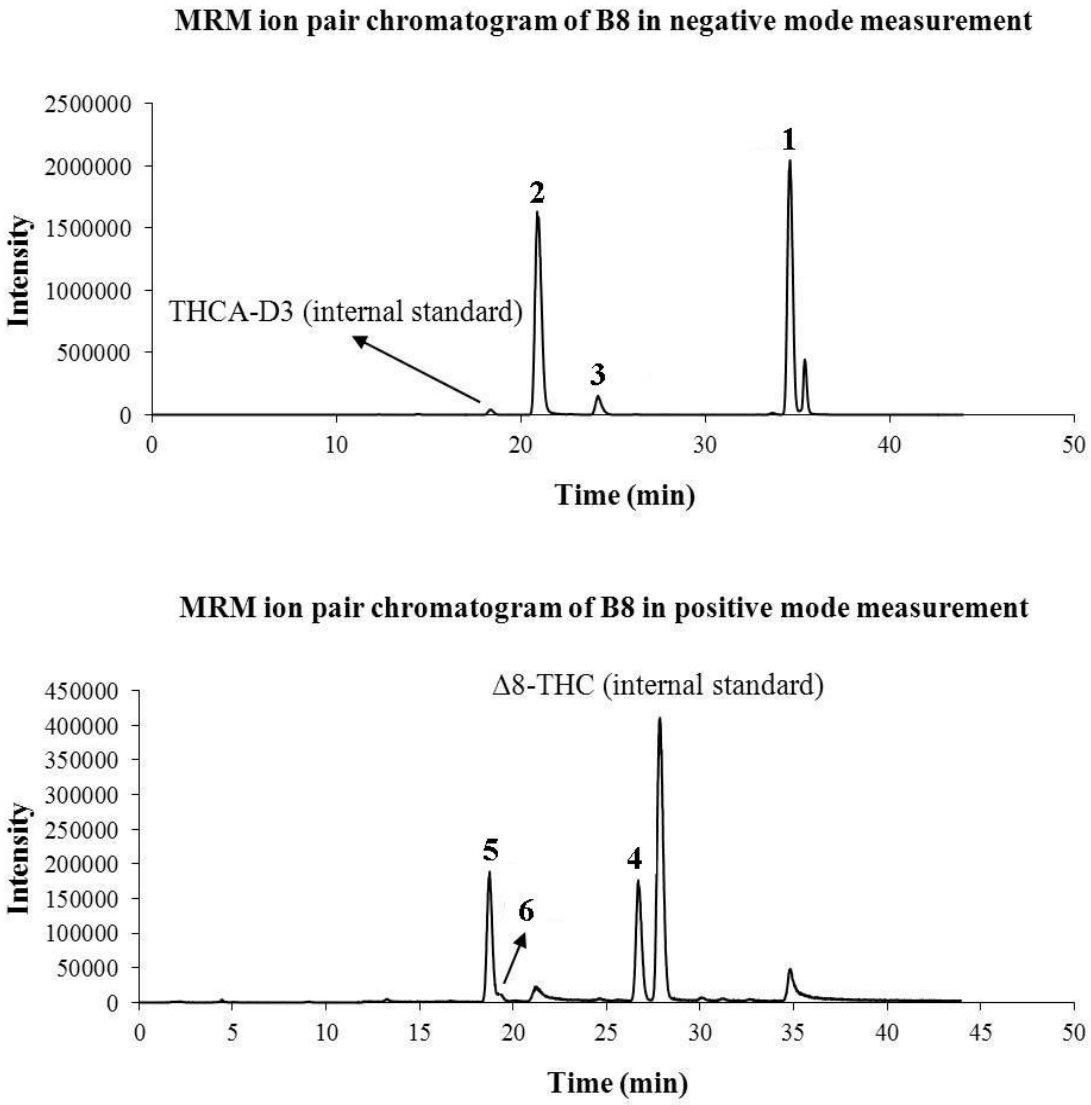


Figure 4. Total ion chromatograms of ion pair multiple reaction monitoring (MRM) of identified cannabinoids and internal standards, LCMS analysis at week 8. ST: intact capitate-stalked trichomes, D: heads of capitate-stalked trichomes, B: stems of capitate-stalked trichomes, SE: intact capitate-sessile trichomes; 1: THCA, 2: CBDA, 3: CBGA , 4: THC, 5: CBD, 6: CBG, 8: CBC, 9: CBN.

Table 3. Concentrations of cannabinoids in dissected trichomes samples based on LCMS analysis. N: number of dissected trichomes, ST: intact capitate-stalked trichomes, D: heads of capitate-stalked trichomes, B: stems of capitate-stalked trichomes, SE: intact capitate-sessile trichomes; number in the sample name represents the collection week; ND for CBD, CBG, CBN, OA: < 3 ng/mL in measured sample, ND for CBC: < 30 ng/mL in measured sample.

Sample	N	Concentrations of cannabinoids/ number dissected trichomes (ng/mL)								
		THCA (1)	CBDA (2)	CBGA (3)	THC (4)	CBD (5)	CBG (6)	CBC (8)	CBN (9)	OA
ST4	76	250	100	31	3.5	1.5	0.2	ND	ND	ND
ST5	83	38	46	6	1.8	1.4	0.1	ND	ND	ND
ST6	92	733	416	88	7	7	0.6	ND	ND	ND
ST7	93	388	156	31	2.0	2.9	0.2	ND	ND	ND
ST8	90	866	571	58	11	14	0.6	0.3	0.1	ND
D4	121	126	58	18	0.8	0.4	0.1	ND	ND	ND
D5	101	220	145	38	2.0	1.2	0.1	ND	ND	ND
D6	126	272	164	43	3.0	2.0	0.3	ND	ND	ND
D7	143	141	55	13	1.5	0.7	0.1	ND	ND	ND
D8	95	841	665	71	15	21	0.9	0.9	0.1	ND
B4	43	196	116	24	2.2	1.3	0.1	ND	ND	ND
B5	88	292	164	41	3.0	1.8	0.2	ND	ND	ND
B6	74	207	87	10	2.4	1.1	0.1	ND	ND	ND
B7	94	349	114	27	3.5	1.4	0.1	ND	ND	ND
B8	75	783	530	49	6	7	0.3	ND	ND	ND
SE4	25	53	11	6	0.4	ND	ND	ND	ND	ND
SE5	45	40	13	6	0.3	ND	ND	ND	ND	ND
SE6	52	74	23	4	1.6	0.5	ND	ND	ND	ND
SE7	53	93	45	11	0.5	0.4	ND	ND	ND	ND
SE8	60	313	214	33	3	1.9	ND	ND	ND	ND

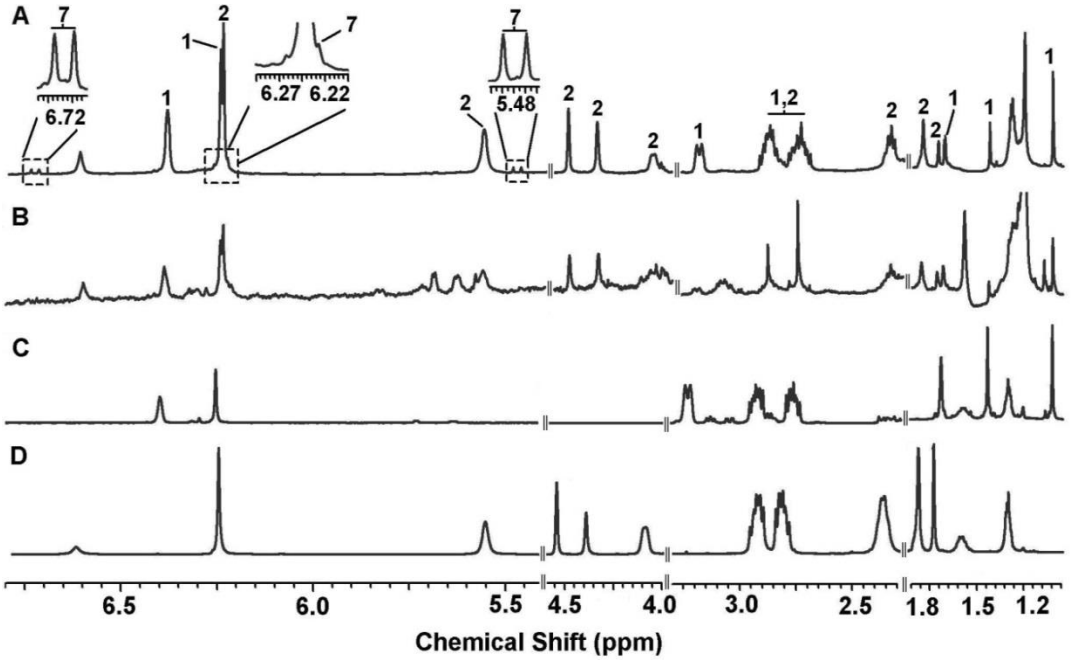


Figure 5. ^{1}H NMR spectra of dissected trichomes and reference compounds. A: capitate-stalked trichomes at week 8, B: capitate-sessile trichomes at week 8, C: reference THCA, D: reference CBDA. Signals in spectrum A are labeled with compound numbers **1**, **2**, and **7**.

4. Conclusions

Here we report cannabiniod analysis in trichomes of medicinal *Cannabis sativa* L. combining LMD, LCMS, and cryogenic NMR. LMD allowed isolating specific trichomes and their individual parts without contaminations from neighbor cells or tissues. Subsequent application of LCMS provided for effective detection and measurement of low concentrated cannabinoids in a low number of various cells of dissected trichomes. Cryogenic NMR was applied to confirm the major findings of LCMS and the technique enabled the successful detection of CBCA (7) showing its unique advantage in metabolites identification even in absence of appropriate reference compounds. From the data obtained, we report that for capitate-stalked glandular trichomes as a complex cell network for cannabinoid production assumed disk cells do not have a unique role in the biosynthesis and that stem cells glandular trichomes play an important role as well. Interestingly, in capitate sessile trichomes the biosynthesis towards THCA (1) is dominating, where in capitates stalked trichomes minor cannabinoids were detected as well. Although the microdissection process generates heat, threatening the stability of cannabinoid acids, we clearly demonstrated the applicability of LMD for trichome isolation in the following analyses of labile cannabinoid acids and other unstable cannabinoids, without

detecting artifacts or unwanted by-products. Therefore, we contend that this method of sample preparation facilitates the investigation of reactive and labile compounds in specific cells or tissues of plants.

Chapter 8

General Discussion

Nizar Happyana

Although often misused as a drug, *Cannabis sativa* L. is still an interesting medicinal plant to be investigated. This plant has a huge potency to treat a wide range of diseases. Many patients medically consume *Cannabis* for relieving suffering from various diseases, including multiple sclerosis, cancer, and HIV/AIDS. Valuable compounds that mostly responsible for the therapeutic properties of *Cannabis* are cannabinoids with THC as the leading compound. These compounds are synthesized and stored mainly in the trichomes. However, the biosynthetic pathways of cannabinoids in particular and secondary metabolites in general, in this plant are still not fully understood, even though its complete genome sequencing has been published.

Regarding the special function of *Cannabis* trichomes in the synthesis of secondary metabolites, especially cannabinoids, in this thesis we focused our works on this organ. Metabolomics, proteomics, and transcriptomics have been applied in this research as analytical approaches in order to study metabolites (especially secondary metabolites), proteins, and genes that are synthesized in the trichomes. Besides that, a review of applications of metabolomics in medicinal plants is described in **chapter 2**. This section highlights metabolomics as a powerful quality control of cultivars and products of medicinal plants by identifying the marker compounds for discriminating. Moreover, it provides an overview of application metabolomics for monitoring elicitation process in medicinal plant cell cultures and for identifying effects of different genetic modification approaches in metabolite production of transgenic plants. In addition, combination of metabolomics with bioactivity or (and) clinical assays for identifying compounds responsible for the bioactivities, analyzing synergistic effects, predicting bioactivity, studying efficacy, and investigating toxicity of medicinal plant products are explained in this chapter.

Some metabolomics works on *Cannabis sativa* L. have been reported previously. Gas chromatography (GC) together with principal component analysis (PCA), an unsupervised multivariate data analysis of metabolomics has been successfully applied for chemotaxonomic discrimination of *Cannabis* varieties and quality control of plant materials (Fischedick et al. 2010). Meanwhile, ^{1}H NMR coupled with PCA has been used for differentiation of *Cannabis* cultivars as well (Choi et al. 2004). In addition, ^{1}H NMR-based metabolomics along with real time PCR (RT-PCR) analysis has been performed for investigating cannabinoid production of *Cannabis* cultivars during flowering period (Muntendam et al. 2012). However, the previous reports focused their metabolomics studies on flowers and leaves of *Cannabis*. Therefore,

application of metabolomics for investigating metabolites, especially cannabinoids in the *Cannabis* trichomes is still interesting to be performed.

In this research, ^{1}H NMR-based metabolomics differentiation together with RT-PCR analysis was applied in order to study metabolite profiles and cannabinoid biosynthesis in the different *Cannabis* organs (**chapter 3**). Trichomes, flowers, and leaves of 2 medicinal *Cannabis* varieties, Bedrocan and Bedica were collected and analyzed. Metabolomics analysis showed that the differences of metabolite profiles within organs of Bedrocan and Bedica were bigger than the differences of same organs between both varieties. THCA, CBCA, CBGA, THC, and 15 identified compounds in the water extracts were detected contributing in the classification of metabolite profiles of *Cannabis* samples. Furthermore, THCA and asparagine were found as the most important discriminant compounds in the differentiations. In addition, quantitative analysis showed that THCA was found in a higher concentration in the trichomes than in the leaves.

Different results were discovered from analyzing gene expression level of THCA synthase (THCAS), the gene that is responsible for formation of THCA, by RT-PCR analysis. This gene was less expressed in the trichomes than in the leaves. Besides that, the expression level of THCAS in Bediol trichomes during the last 4 weeks of flowering period did not differ significantly as described in the **chapter 4**. These results probably might be a suggestion that there is no a positive direct-correlation between the abundance of THCA with expression level of THCAS gene in the trichomes. Previous report revealed that olivetolic acid synthase (OLS) and olivetolic acid cyclase (OAC) cooperate in the condensation of malonyl-CoA and Hexaonyl-CoA to form olivetolic acid, a starting compound in THCA biosynthesis (Gagne et al. 2012). RT-PCR analysis showed that both were expressed higher in the trichomes than in the leaves. Thus, an alternative opinion likely might be proposed, that is the high gene expression levels of OLS and OAC in the trichomes trigger the increasing of production of olivetolic acid and then lead to the abundance of THCA in the trichomes.

Discovering the highest level of THCAS gene in the leaves of Bedrocan and Bedica was interesting since a previous report mentioned that this gene was more highly expressed in the trichomes than in the leaves (Marks et al. 2009). However, the leaves used in this report were young inflorescence-associated leaves. In order to confirm this report with our finding, expression level of THCAS gene in young inflorescence-associated leaves of Bedrocan was

analyzed. We found that the amount of THCAS transcript in this organ was 11 times lower than in the trichomes, thus the finding was accordance with the previous report.

In the next metabolomics work, metabolite profiles, especially cannabinoid production in the trichomes of 4 medicinal *Cannabis* varieties (Bediol, Bedica, Bedrobinol, and Bedrocan) were monitored during the last weeks of flowering period (**chapter 4**). PLSDA, a supervised approach of multivariate data analysis, was applied in this metabolomics experiment. PLSDA models have successfully revealed the differentiation of metabolite profiles within the varieties. Furthermore it enabled efficient monitoring of metabolite production, particularly cannabinoids in trichomes of all individual varieties over the flowering period. THCA was detected as the most important discriminant constituent in all investigated non-polar extracts. Meanwhile, some amino acids including threonine and asparagine, and sugar compounds, represented by fructose and sucrose, were identified providing significant contribution to the metabolomic discrimination of all the analyzed polar extracts. According to this study, *Cannabis* trichomes during the flowering period synthesize metabolites, particularly cannabinoids, in various amounts, depending on time and the plant variety. Additionally, we have shown in this study, that ^{1}H NMR-based metabolomics has been successfully performed for the time-dependent monitoring of metabolite production in the *Cannabis* trichomes.

Proteomics as a method that can be used for investigating enzymes involved in the metabolic pathways has been performed in *Cannabis* study. Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS) together with 2-dimensional gel electrophoresis has been applied as a proteomics method for investigating proteins of *Cannabis*. Raharjo et al. used this method for revealing protein profiles in the trichomes, the leaves, and the flowers of this plant (Raharjo et al. 2004). Nevertheless, they could not discover enzymes involved in the cannabinoid biosynthesis. This method was also applied to study copper effects on protein profile of *Cannabis* root (Bona et al. 2007) and to analyze protein profile of fiber-type seed of this plant (Park et al. 2012). However, the previous reports did not identify enzymes involved in the biosynthesis of secondary metabolites.

In this research, proteomics analysis of *Cannabis* trichomes variety Bediol (**chapter 5**) has been performed in order to analyze protein profiles and to reveal enzymes that are participating in the metabolic pathways of cannabinoids, terpenoids, and flavonoids in this plant. The proteomics analysis was carried out by nano-LC-MS/MS method and then the identified proteins were classified based on their biological function. As the results, many

enzymes corresponding to the biosynthesis of secondary metabolites, including cannabinoids, flavonoids, and terpenoids were recorded. This finding confirmed the function of *Cannabis* trichomes as the main site of secondary metabolite production. In addition, house-keeping proteins participating in primary metabolism, energy production, protein synthesis, and other metabolism have been also identified in the trichomes. Furthermore, the identification of a set proteins involved in the disease/defense responses supported the functions of trichomes in protecting the plant against biotic and abiotic threats.

Recently, some cDNA libraries of *Cannabis* trichomes have been reported. Mark et al. constructed the cDNA library for identifying the candidate genes that is affecting THCA biosynthesis (Marks et al. 2009). Meanwhile, a cDNA library of *Cannabis* trichomes has been also assembled to identify an acyl activating enzyme that is responsible for the formation of hexanoyl-CoA in cannabinoid biosynthesis (Stout et al. 2012a). Moreover, the ESTs from this cDNA library were used for discovering olivetolic acid cyclase (Gagne et al. 2012). Nevertheless, the previous reports only focused on transcript analysis of cannabinoid biosynthesis, even though other secondary metabolites, such as flavonoids and terpenoids have been discovered on this plant.

In this study, a cDNA library of *Cannabis* trichomes variety Bediol was constructed by 454 GS FLX pyrosequencing system (**chapter 6**). The assembled sequenced were annotated functionally by the Sequence Analysis and Management System (SAMS) for Gene Ontology (GO) analysis and pathway assignments with the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. As the results, putative transcripts of enzymes that are involved in metabolic pathways of secondary metabolites, particularly cannabinoids, terpenoids, and flavonoids in the *Cannabis* trichomes were successfully identified. Thus, the cDNA library provided valuable information for elucidating the biosynthetic pathways of secondary metabolites in this plant. Furthermore, putative transcripts participating in the biosynthetic pathways of alkaloids were identified in the cDNA library as well. This finding probably might be a signature for the presence of alkaloids in the *Cannabis* trichomes.

Cannabinoid biosynthesis is initiated by the formation of 3 starting substrates, namely malonyl-CoA, hexanoyl-CoA, and geranyl pyrophosphate (GPP). The first substrate is provided by acetyl-CoA carboxylase from the reaction of carboxylation of acetyl-CoA. The presence of this enzyme in the *Cannabis* trichomes was recorded in the proteomics data and supported by the identification of its transcripts in the cDNA library. Meanwhile, it has been

reported that acyl-activating enzyme 1 (CsAAE1) and acyl-activating enzyme 3 (CsAAE3) can supply hexanoyl-CoA via the reaction of hexanoate conversion (Stout et al. 2012b). Both were identified in the proteomic data of *Cannabis* trichomes and their transcripts were detected in the cDNA library as well. Nevertheless, CsAAE3 is localized to peroxisome, while CsAAE1 is found in the cytosol which is the same compartment where OLS (cannabinoid enzyme) is localized. Therefore, CsAAE1 was suggested as the enzyme that responsible for supplying hexaonyl-CoA to the biosynthetic pathway of cannabinoids (Stout et al. 2012b). The last starting substrate, GPP, is prepared by GPP synthase (GPPS) via condensation reaction of isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP). This enzyme was recorded in the proteomics data and its transcript was noted in the cDNA library as well.

In the next step of cannabinoid biosynthesis, malonyl-CoA and hexaonyl-CoA are condensed by OLS and OAC to form olivetolic acid (Gagne et al. 2012). Both enzymes were recorded in the proteomics data and confirmed by the identification of their transcripts in the cDNA library of *Cannabis* trichomes. Subsequently, the product of these enzymes is combined with GPP by CBGA synthase (CBGAS) to form CBGA. This enzyme is predicted as a member of geranyltransferase family (Fellermeier and Zenk 1998). Unfortunately, the sequence of CBGAS has not yet been deposited into public databases, thus it could not be checked either on our proteomics data or our cDNA library. Recently our group identified an enzyme that is predicted as CBGAS (predicted-CBGAS). Interestingly, although predicted-CBGAS was not noted in the proteomics data of *Cannabis* trichomes, nevertheless the putative transcript of this enzyme was recorded in the cDNA library. Thus, it needs further investigations for confirming the existence of predicted-CBGAS in the *Cannabis* trichomes. CBGA is transformed subsequently into other cannabinoids by THCAS, CBDAS, and CBCAS. The first enzyme converts CBGA into THCA, acidic form of THC (Taura et al. 1995), and the second enzyme transforms CBGA into CBDA (Taura et al. 1996). Both enzymes were recorded in the proteomics data and their transcripts were noted in the cDNA library as well. Meanwhile, the last enzymes, CBCAS, converts CBGA into CBCA (Morimoto et al. 1998). Nevertheless, this enzyme could not be checked on the proteomics data and the cDNA library of *Cannabis* trichomes since its sequence has not yet been stored into public databases even CBCAS has been purified to homogeneity (Morimoto et al. 1998).

There are 2 biosynthetic pathways for producing building blocks of terpenoids, IPP and DMAPP, namely mevalonate (MVA) and non-mevalonate [2-C-methyl-D-erythritol-4-phosphate/1-deoxy-D-xylulose-5-phosphate (MEP/DOXP)] pathway. The first pathway produces the building blocks in the cytosol for the synthesis of sesquiterpenoids and triterpenoids. Meanwhile, MEP/DOXP pathway provides IPP and DMAPP in the plastid for the formation of monoterpenoids and diterpenoids. Interestingly, although all putative genes corresponding in both metabolic pathways were identified in the cDNA library of *Cannabis* trichomes, nevertheless only enzymes involved in the MEP/DOXP pathway were recorded in the proteomics data. This finding likely correlates with the biosynthesis of GPP that is synthesized for the biosynthetic pathway of cannabinoids. Based on incorporation experiments using ^{13}C-labeled glucoses, GPP related to cannabinoid metabolism is provided entirely or predominantly (> 98%) from MEP/DOXP pathway (Fellermeier et al. 2001). Besides that, the content of cannabinoids is very dominant in the *Cannabis* trichomes as described in the metabolomics chapters. These findings trigger the abundace of terpenoid enzymes that are involved in the MEP/DOXP pathway, in the *Cannabis* trichomes. Thus, these enzymes could be recorded easily by proteomics analysis.

Some enzymes involved in the next steps of terpenoid biosynthesis were also recorded in the proteomics data, including isopenthenyl pyrophospahate isomerase (IPPI) and farnesyl pyrophospahate synthase (FPPS). The first enzyme catalyzes the isomerization of IPP to DMAPP, while FPPS catalyzes sequential condensation reactions of DMAPP with 2 units of IPP to form farnesyl pyrophosphate (FPP), a precursor compound in the sesquiterpene biosynthesis. Furthermore, the putative transcripts of both enzymes were detected in the cDNA library. Thus, it confirmed the existences of IPPI and FPPS in the *Cannabis* trichomes. Other enzymes involved in the downstream steps of terpene metabolism were also identified by proteomic analysis, including (-)-limonene synthase and (+)-α-pinene synthase. Both responsible for formation of monoterpenes, namely (-)-limonene and (+)-α-pinene respectively. Interestingly, the presence of both enzymes in the *Cannabis* trichomes was supported by the identification of their transcripts in the cDNA library. In contrast, some terpenoid enzymes that their putative transcripts were recorded in the cDNA library were not detected in the proteomics data, such as myrcene synthase and (+)-δ-cadinene synthase. These enzymes are responsible for synthesizing myrcene and (+)-δ-cadinene, respectively. The absence of myrcene synthase and (+)-δ-cadinene synthase in the proteomics data might be caused by the low expression levels of both enzymes in the *Cannabis* trichomes. Overall, the

identification of terpenoid enzymes in the proteomics analysis and the detection of their putative transcripts in the cDNA library confirmed that the *Cannabis* trichomes are the mine place not only for producing cannabinoid but also for synthesizing terpenoids.

Although there are no flavonoids have been reported from *Cannabis* trichomes, nevertheless almost all enzymes involved in the main biosynthetic pathway of flavonoids were recorded in the proteomics data. The first enzyme involved in this pathway, phenylalanine ammonia lyase (PAL) which converts phenylalanine to *p*-cinnamic acid, unfortunately was not detected. Nevertheless the rest enzymes including cinnamate 4-hydroxylase (C4H), 4-coumarate:CoA ligase (4CL), chalcone synthase (CHS), and chalcone isomerase (CHI) were successfully identified in the proteomics data. C4H hydroxylates the product of the first enzyme into *p*-coumaric acid. 4CL adds a CoA thiol ester into *p*-coumaric acid yielding *p*-coumaroyl-CoA, while CHS condenses one molecule of *p*-coumaroyl-CoA and three molecules of malonyl-CoA that is supplied by acetyl-CoA carboxylase yielding a naringenin chalcone. Subsequently, CHI isomerizes naringenin chalcone to naringenin (flavanone), a common precursor of other derivative flavonoids, such as flavonols and flavones. Other enzymes related to flavonoid biosynthesis were also identified in the proteomics data, such as flavanone 3-hydroxylase (F3H) and dihydroflavonol-4-reductase (FDR). F3H catalyzes the formation of dihydroflavonols from flavanones, in particular, this enzyme adds a hydroxyl molecule on carbon number 3 of flavanone, such as converting naringenin into dihydrokaempferol. Meanwhile, FDR catalyzes the NADPH-dependent conversion of stereospecific dihydroflavonols such as dihydrokaempferol and dihydroquercetin into leucoanthocyanidins (or flavan-3,4-diols) (Zhang et al. 2008). Interestingly, the putative genes of flavonoids enzymes that mentioned before were recorded in the cDNA library as well. Thus, these findings not only confirm the presence of these enzymes but also support the existence of flavonoids in the *Cannabis* trichomes.

In this study, some putative transcripts of important enzymes that are involved in the downstream steps of flavonoid biosynthesis were identified in the cDNA library of *Cannabis* trichomes, even though their existences could not be detected with proteomics analysis. Flavanone 3'-hydroxylase (F3'H), flavonol synthase (FLS), and flavone synthase (FNS) are the enzymes belonging to this category. F3'H catalyzes the formation of dihydroflavonols from flavanones, in particular, F3'H puts a hydroxyl group on the carbon number 3', such as modifying naringenin into eriodictyol. FLS transforms dihydroflavonols into flavonols, such

as converting dihydrokaempferol into kaempferol. Meanwhile FNS catalyzes the conversion of flavanones into flavone, such as converting naringenin into apigenine and transforming eriodictyol into luteolin.

Cannaflavin A and cannaflavin B are unique flavonoids that are found only in *Cannabis* sativa L. There are 2 biosynthetic pathways that have been proposed for producing these compounds (Flores-Sanchez and Verpoorte 2008). The first pathway suggested that both are produced through the formation of luteolin which then modified with prenylation and methylatation by prenyltransferase and *O*-methyltransferase (OMT), respectively. Meanwhile the second pathway proposed that cannaflavin A and cannaflavin B are synthesized via the formation of homoeriodictyol chalcone which then modified with some reactions by unknown enzymes. Interestingly, most of enzymes involved in the biosynthesis of luteolin and homoeriodictyol chalcone were recorded in the proteomics data and their putative transcripts were detected in the cDNA library of *Cannabis* trichomes. Nevertheless, for the enzymes involved in the modification of luteolin and homoeriodictyol chalcone, only putative transcripts of OMT were successfully detected in the cDNA library. Therefore, it still needs more investigations for elucidating the biosynthetic pathway of cannaflavin A and cannaflavin B.

Based on the bond type of glycosylation, glycosylated flavonoids of *Cannabis* are divided into 2 groups, namely *O*-glycosylated flavonoids (glycosylation via carbon-oxygen bond) and *C*-glycosylated flavonoids (glycosylation via carbon-carbon bond). Kaempferol 3-*O*-glucoside and quercetin 3-*O*-sophoroside that have been identified in *Cannabis*, are example of the first group. Both compounds are products from glycosylation of kaempferol and quercetin, respectively, by UDP-glycosyltransferases (OGTs). Fortunately, some putative transcripts of OGTs that involved in these glycosylation were successfully recorded in the cDNA library of *Cannabis* trichomes. Vitexin, isovitexin, and orientin are *Cannabis* flavonoids that are glycosylated through carbon-carbon bond. Biosynthetically, vitexin and isovitexin are suggested as products from glucosylation of apigenin by an UDP-glucosyltransferase (UGT), while orientin is proposed as a result from glucosylation of luteolin by an UGT (Flores-Sanchez and Verpoorte 2008). Interestingly, some putative transcripts that are annotated as UGTs were noted in the cDNA library. Thus, these transcripts might be gene candidates of UGTs that are involved in the glycosylation of *Cannabis* flavonoids. An alternative route for biosynthesis of *C*-glycosylated flavonoids that has been described in cereals (Brazier-Hicks et al. 2009), might be occurred in *Cannabis*. In this pathway, a flavanone is hydroxylated by

flavanone 2-hydroxylase (F2H) to form 2-hydroxyflavanone. In the next step, glycosylation occurs on the open-chain form of 2-hydroxyflavanone by *C*-glucosyltransferase (CGT) to produce 2-hydroxyflavanone *C*-glucoside, and then followed subsequently by dehydration to yield flavone-6*C*-glucoside. The gene sequences of F2H and CGT were checked in the cDNA library of *Cannabis* trichomes. As the result, some unique transcripts hit the sequences but with low scores.

There are 3 types of trichomes can be found in *Cannabis* sativa L., namely capitate-stalked, capitate-sessile and bulbous trichomes. Capitate-stalked trichomes contain most cannabinoids and generally consist of two parts, the gland (head) and the stem. The head contains disc cells which are surrounded by the storage cavity. Disc cells are presumed to be the site of cannabinoid production (Kim and Mahlberg 1997, 2003, 1991). The stem is formed by stipe cells and basal cells (Kim and Mahlberg 1997) and is not yet functionally characterized. Identifying the functions of specific tissues and cell types from plants requires accurate and efficient methods for collecting the populations of the material of interest. Recently, the combination of laser-assisted microdissection (LMD) technique with diverse range of analytical technologies has allowed this purpose to be achieved.

In this study, analysis of cannabiniods in the microdissected-trichomes of *Cannabis* variety Bediol was performed (**chapter 7**). Specific trichomes and their individual parts without contaminations were collected with LMD. Low concentrated cannabinoids in a low number of various cells of dissected trichomes were effectively detected with subsequent application of LCMS. Cryogenic NMR was applied to confirm the major findings of LCMS. Moreover this method allowed the detection of CBCA showing its unique advantage in metabolites identification even in absence of appropriate reference compounds. The quantitative analysis revealed that the concentration of cannabinoids in the capitate-stalked trichomes is higher than that in the capitate-sessile trichomes. Moreover, cannabinoids were detected not only in the head of capitate-stalked trichomes but also in its stem part.

Previous research confirmed the localization of cannabinoid production in the heads of capitate-stalked trichomes by identifying candidate biosynthetic genes (Marks et al. 2009; Sirikantaramas et al. 2005). However, our analysis of the capitate-stalked trichome stems indicated the presence of cannabinoids therein as well. Thus, cannabinoid production might not be exclusive to the head of the glandular trichomes. Nonetheless, the source of cannabinoids in the stems of capitate-stalked trichomes is still unclear – they could be derived

from the stem itself or translocated from surrounding tissues, like leaves and flowers. Although the function of the capitate-stalked trichome stems in cannabinoid biosynthesis is not yet defined, the detection of cannabinoids suggests that they could play a role in the production of these valuable natural products.

The function of the storage cavity in the capitate-stalked trichome head as a metabolite repository is well documented (Kim and Mahlberg 1997). In spite of its cytotoxic effect on plant cells, THCA is accumulated in the storage cavity (Sirikantaramas et al. 2005), presumably due to its ecological role in protection against insects (Taura et al. 2007). In addition, it was reported that THCA and CBCA effectively induce cell death in the leaf cells or suspension-cultured cells of *Cannabis sativa* (Morimoto et al. 2007). Therefore, to prevent cellular damage, cannabinoid secretion into the storage cavity is necessary (Sirikantaramas et al. 2005; Taura et al. 2007). It seems reasonable to assume that the toxic cannabinoids potentially synthesized in the stem of a capitate-stalked trichome would also have to be transported to the storage cavity.

References

Abhyankar, G., Reddy, V. D., Giri, C. C., Rao, K. V., Lakshmi, V. V. S., Prabhakar, S., et al. (2005). Amplified fragment length polymorphism and metabolomic profiles of hairy roots of Psoralea corylifolia L. *Phytochemistry, 66*(20), 2441-2457, doi:DOI 10.1016/j.phytochem.2005.08.003.

Abreu, I. N., Choi, Y. H., Sawaya, A. C. H. F., Eberlin, M. N., Mazzafera, P., & Verpoorte, R. (2011). Metabolic Alterations in Different Developmental Stages of Pilocarpus microphyllus. *Planta medica, 77*(3), 293-300, doi:DOI 10.1055/s-0030-1250314.

Agurell, S., & Nilsson, J. L. G. (1972). Chemistry and Biological-Activity of Cannabis - Review of Symposium, Held in Stockholm in 1971. *Bulletin on Narcotics, 24*(2), 35-37.

Ahmed, S. A., Ross, S. A., Slade, D., Radwan, M. M., Khan, I. A., & Elsohly, M. A. (2008a). Structure determination and absolute configuration of cannabichromanone derivatives from high potency Cannabis sativa. *Tetrahedron Letters, 49*(42), 6050-6053, doi:DOI 10.1016/j.tetlet.2008.07.178.

Ahmed, S. A., Ross, S. A., Slade, D., Radwan, M. M., Zulfiqar, F., & ElSohly, M. A. (2008b). Cannabinoid ester constituents from high-potency Cannabis sativa. *Journal of Natural Products, 71*(4), 536-542, doi:Doi 10.1021/Np070454a.

Ali, K., Maltese, F., Fortes, A. M., Pais, M. S., Choi, Y. H., & Verpoorte, R. (2011). Monitoring biochemical changes during grape berry development in Portuguese cultivars by NMR spectroscopy. *Food Chemistry, 124*(4), 1760-1769, doi:DOI 10.1016/j.foodchem.2010.08.015.

Allwood, J., William., Ellis, D., I. , & Goodacre, R. (2008). Metabolomic technologies and their application to the study of plants and plant–host interactions. *Physiologia Plantarum, 132*(117–135).

Allwood, J. W., & Goodacre, R. (2010). An Introduction to Liquid Chromatography-Mass Spectrometry Instrumentation Applied in Plant Metabolomic Analyses. *Phytochemical Analysis, 21*(1), 33-47, doi:10.1002/pca.1187.

Alscher, R. G., Erturk, N., & Heath, L. S. (2002). Role of superoxide dismutases (SODs) in controlling oxidative stress in plants. *Journal of Experimental Botany, 53*(372), 1331-1341, doi:10.1093/jexbot/53.372.1331.

Anderson, L. C. (1974). A study of systematic wood anatomy in Cannabis. *Harvard University Botanical Museum Leaflets, 24*, 29–36.

Anderson, L. C. (1980). Leaf variation among Cannabis species from a controlled garden. *Harvard University Botanical Museum Leaflets, 28*, 61–69.

Asano, T., Masumura, T., Kusano, H., Kikuchi, S., Kurita, A., Shimada, H., et al. (2002). Construction of a specialized cDNA library from plant cells isolated by laser capture microdissection: toward comprehensive analysis of the genes expressed in the rice phloem. *Plant Journal, 32*(3), 401-408, doi:1423 [pii].

Aziz, N., Paiva, N. L., May, G. D., & Dixon, R. A. (2005). Transcriptome analysis of alfalfa glandular trichomes. *Planta, 221*(1), 28-38.

Bacher, A., Rieder, C., Eichinger, D., Arigoni, D., Fuchs, G., & Eisenreich, W. (1998). Elucidation of novel biosynthetic pathways and metabolite flux patterns by retrobiosynthetic NMR analysis. *Fems Microbiology Reviews, 22*(5), 567-598.

Bailey, N. J. C., Sampson, J., Hylands, P. J., Nicholson, J. K., & Holmes, E. (2002). Multi-component metabolic classification of commercial feverfew preparations via high-field H-1-NMR spectroscopy and chemometrics. *Planta medica, 68*(8), 734-738.

Bailey, N. J. C., Wang, Y. L., Sampson, J., Davis, W., Whitcombe, I., Hylands, P. J., et al. (2004). Prediction of anti-plasmodial activity of Artemisia annua extracts: application of H-1 NMR spectroscopy and chernometrics. *Journal of Pharmaceutical and Biomedical Analysis, 35*(1), 117-126, doi:DOI 10.1016/j.jpba.2003.12.024.

Barrett, M. L., Scutt, A. M., & Evans, F. J. (1986). Cannflavin-a and Cannflavin-B, Prenylated Flavones from Cannabis-Sativa L. *Experientia, 42*(4), 452-453, doi:Doi 10.1007/Bf02118655.

Bartley, G. E., P.A., S., & P., B. (1999). Two Arabidopsis thaliana carotene desaturases, phytoene desaturase and zeta-carotene desaturase, expressed in Escherichia coli, catalyze a poly-cis pathway to yield pro-lycopene. *European Journal of Biochemistry, 259*, 396-403.

Baud, S., Guyon, V., Kronenberger, J., Wuilleme, S., Miquel, M., Caboche, M., et al. (2003). Multifunctional acetyl-CoA carboxylase 1 is essential for very long chain fatty acid elongation and embryo development in Arabidopsis. *The Plant Journal, 33*(1), 75–86.

Beal, J. E., Olson, R., Laubenstein, L., Morales, J. O., Bellman, P., Yangco, B., et al. (1995). Dronabinol as a Treatment for Anorexia Associated with Weight-Loss in Patients with Aids. *Journal of Pain and Symptom Management, 10*(2), 89-97, doi:Doi 10.1016/0885-3924(94)00117-4.

Beal, J. E., Olson, R., Lefkowitz, L., Larenstein, L., Bellman, P., Yangco, B., et al. (1997). Long-term efficacy and safety of dronabinol for acquired immunodeficiency syndrome-associated anorexia. *Journal of Pain and Symptom Management, 14*(1), 7-14, doi:Doi 10.1016/S0885-3924(97)00038-9.

Bekel, T., Henckel, K., Küster, H., Meyer, F., Mittard, R. V., Neuweger, H., et al. (2009). The Sequence Analysis and Management System -- SAMS-2.0: data management and sequence analysis adapted to changing requirements from traditional sanger sequencing to ultrafast sequencing technologies. *Journal of Biotechnology, 140*, 3-12.

Besser, K., Harper, A., Welsby, N., Schauvinhold, I., Slocombe, S., Li, Y., et al. (2009). Divergent Regulation of Terpenoid Metabolism in the Trichomes of Wild and Cultivated Tomato Species. *Plant Physiology, 149*(1), 499-514.

Bona, E., Marsano, F., Cavaletto, M., & Berta, G. (2007). Proteomic characterization of copper stress response in Cannabis sativa roots. [Research Support, Non-U.S. Gov't]. *Proteomics, 7*(7), 1121-1130, doi:10.1002/pmic.200600712.

Braemer, R., Tsoutsias, Y., Hurabielle, M., & Paris, M. (1987). Biotransformations of Quercetin and Apigenin by a Cell Suspension Culture of Cannabis sativa. *Planta medica, 53*(3), 225-226.

Brazier-Hicks, M., Evans, K. M., Gershater, M. C., Puschmann, H., Steel, P. G., & Edwards, R. (2009). The C-glycosylation of flavonoids in cereals. *The Journal of Biological Chemistry, 284*(27), 17926-17934, doi:10.1074/jbc.M109.009258.

Brenneisen, R., Egli, A., Elsohly, M. A., Henn, V., & Spiess, Y. (1996). The effect of orally and rectally administered Delta(9)-tetrahydrocannabinol on spasticity: A pilot study with 2 patients. *International Journal of Clinical Pharmacology and Therapeutics, 34*(10), 446-452.

Britsch, L., Dedio, J., Saedler, H., & Forkmann, G. (1993). Molecular characterization of flavanone 3 beta-hydroxylases. Consensus sequence, comparison with related enzymes and the role of conserved histidine residues. *European Journal of Biochemistry, 217*, 745-754.

Broeckling, C. D., Huhman, D. V., Farag, M. A., Smith, J. T., May, G. D., Mendes, P., et al. (2005). Metabolic profiling of Medicago truncatula cell cultures reveals the effects of biotic and abiotic elicitors on metabolism. *Journal of Experimental Botany, 56*(410), 323-336, doi:Doi 10.1093/Jxb/Eri058.

Broyart, C., Fontaine, J. X., Molinie, R., Cailleu, D., Terce-Laforgue, T., Dubois, F., et al. (2010). Metabolic profiling of maize mutants deficient for two glutamine synthetase isoenzymes using 1H-NMR-based metabolomics. *Phytochemical analysis : PCA, 21*(1), 102-109, doi:10.1002/pca.1177.

Burstein, S., & Varanelli, C. (1975). Prostaglandins and cannabis-III: inhibition of biosynthesis by essential oil components of marijuana. *Biochemical Pharmacology, 24*, 1053-1054.

Bylesjo, M., Rantalainen, M., Cloarec, O., Nicholson, J. K., Holmes, E., & Trygg, J. (2006). OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *Journal of Chemometrics, 20*(8-10), 341-351, doi:Doi 10.1002/Cem.1006.

Candolle, A. d. (Ed.). (1886). *Origin of cultivated plants.* (Vol. 49, International scientific series). London: Paul Trench.

Chan, W., & Cai, Z. W. (2008). Aristolochic acid induced changes in the metabolic profile of rat urine. *Journal of Pharmaceutical and Biomedical Analysis, 46*(4), 757-762, doi:DOI 10.1016/j.jpba.2007.11.042.

Chelikani. P, Fita, I., & Loewen, P. C. (2004). Diversity of structures and properties among catalases. *Cellular and Molecular Life Sciences, 61*(2), 192-208.

Chen, M. J., Su, M. M., Zhao, L. P., Jiang, J., Liu, P., Cheng, J. Y., et al. (2006). Metabonomic study of aristolochic acid-induced nephrotoxicity in rats. *Journal of Proteome Research, 5*(4), 995-1002, doi:Doi 10.1021/Pr050404w.

Choi, H. K., Choi, Y. H., Verberne, M., Lefeber, A. W. M., Erkelens, C., & Verpoorte, R. (2004). Metabolic fingerprinting of wild type and transgenic tobacco plants by H-1 NMR and multivariate analysis technique. *Phytochemistry, 65*(7), 857-864, doi:DOI 10.1016/j.phytochem.2004.01.019.

Choi, H. K., & Yoon, J. H. (2007). Metabolomic profiling of Vitis vinifera cell suspension culture elicited with silver nitrate by H-1 NMR spectrometry and principal components analysis. *Process Biochemistry, 42*(2), 271-274, doi:DOI 10.1016/j.procbio.2006.07.007.

Choi, Y. H., Kim, H. K., Hazekamp, A., Erkelens, C., Lefeber, A. W. M., & Verpoorte, R. (2004a). Metabolomic differentiation of Cannabis sativa cultivars using H-1 NMR spectroscopy and principal component analysis. *Journal of Natural Products, 67*(6), 953-957, doi:Doi 10.1021/Np049919c.

Choi, Y. H., Kim, H. K., Linthorst, H. J. M., Hollander, J. G., Lefeber, A. W. M., Erkelens, C., et al. (2006). NMR metabolomics to revisit the tobacco mosaic virus infection in Nicotiana tabacum leaves. *Journal of Natural Products, 69*(5), 742-748, doi:Doi 10.1021/Np050535b.

Choi, Y. H., Kim, H. K., & Verpoorte, R. (2008). Metabolomics. In O. Kayser, & W. Quax (Eds.), *Medicinal Plant Biotechnology: From Basic Research to Industrial Applications* (pp. 9-28): Wiley.

Choi, Y. H., Tapias, E. C., Kim, H. K., Lefeber, A. W. M., Erkelens, C., Verhoeven, J. T. J., et al. (2004b). Metabolic discrimination of Catharanthus roseus leaves infected by phytoplasma using H-1-NMR spectroscopy and multivariate data analysis. *Plant Physiology, 135*(4), 2398-2410, doi:DOI 10.1104/pp.104.041012.

Chopra, G. S. (1969). Man and Marijuana. *International Journal of the Addictions, 4*(2), 215-247.

Dai, Y. T., Li, Z. Y., Xue, L. M., Dou, C. Y., Zhou, Y. Z., Zhang, L. Z., et al. (2010). Metabolomics study on the anti-depression effect of xiaoyaosan on rat model of chronic unpredictable mild stress. *Journal of Ethnopharmacology, 128*(2), 482-489, doi:DOI 10.1016/j.jep.2010.01.016.

Day, R. C., Grossniklaus, U., & Macknight, R. C. (2005). Be more specific! Laser-assisted microdissection of plant cells. *Trends in Plant Science, 10*(8), 397-406, doi:DOI 10.1016/j.tplants.2005.06.006.

Dayanandan, P., & Kaufman, P. B. (1976). Trichomes of Cannabis-Sativa L (Cannabaceae). *American Journal of Botany, 63*(5), 578-591.

de Barge, A. (Ed.). (1860). *Lettre de M. Alex. de Bunge à M. Decaisne. Botanique de France* (Vol. 7, Annals of Applied Biology).

de Lamarck, J. B. (1785). *Encyclopedique Methodique de Botanique* (Vol. 1). Paris. France.

DESA (2009). World Population Prospects, The 2008 Revision, Highlights (D. o. E. a. S. A. (DESA), Trans.). New York: United Nations, Department of Economic and Social Affairs (DESA).

Dettmer, K., Aronov, P. A., & Hammock, B. D. (2007). Mass spectrometry-based metabolomics. *Mass Spectrometry Reviews, 26*(1), 51-78, doi:Doi 10.1002/Mas.20108.

Deuschle, K., Funck, D., Hellmann, H., Daeschner, K., Binder, S., & Frommer, W. B. (2001). A nuclear gene encoding mitochondrial Δ1-pyrroline-5-carboxylate dehydrogenase and its potential role in protection from proline toxicity. *The Plant Journal, 27*(4), 345–356.

Duke, S. O., Canel, C., Rimando, A. M., Tellez, M. R., Duke, M. V., & Paul, R. N. (2000). Current and potential exploitation of plant glandular trichome productivity. *Advances in Botanical Research Incorporating Advances in Plant Pathology, Vol 31 2000, 31*, 121-151, doi:Doi 10.1016/S0065-2296(00)31008-4.

Dunn, W. B., Bailey, N. J. C., & Johnson, H. E. (2005). Measuring the metabolome: current analytical technologies. *Analyst, 130*(5), 606-625, doi:Doi 10.1039/B418288j.

Edwards, R. (2004). No remedy in sight for herbal ransack. *New Scientist, 181*(2429), 10-11.

ElSohly, M. A., & Slade, D. (2005). Chemical constituents of marijuana: The complex mixture of natural cannabinoids. *Life Sciences, 78*(5), 539-548, doi:DOI 10.1016/j.lfs.2005.09.011.

Fairbair.Jw (1972). Trichomes and Glands of Cannabis Sativa L. *Bulletin on Narcotics, 24*(4), 29-33.

Falara, V., Fotopoulos, V., Margaritis, T., Anastasaki, T., Pateraki, I., Bosabalidis, M. A., et al. (2008). Transcriptome analysis approaches for the isolation 1of trichome-specific genes from the medicinal plant Cistus creticus subsp. creticus. *Plant Molecular Biology, 68*, 633-651, doi:10.1007/s11103-008-9399-0.

Faltin, Z., Holland, D., Velcheva, M., Tsapovetsky, M., Roeckel-Drevet, P., Handa, A. K., et al. (2010). Glutathione Peroxidase Regulation of Reactive Oxygen Species Level is Crucial for In Vitro Plant Differentiation *Plant and Cell Physiology, 51*(7), 1151–1162, doi:10.1093/pcp/pcq082

Fan, T. W. M., & Lane, A. N. (2008). Structure-based profiling of metabolites and isotopomers by NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy, 52*(2-3), 69-117, doi:DOI 10.1016/j.pnmrs.2007.03.002.

Fellermeier, M., Eisenreich, W., Bacher, A., & Zenk, M. H. (2001). Biosynthesis of cannabinoids - Incorporation experiments with C-13-labeled glucoses. *European Journal of Biochemistry, 268*(6), 1596-1604.

Fellermeier, M., & Zenk, M. H. (1998). Prenylation of olivetolate by a hemp transferase yields cannabigerolic acid, the precursor of tetrahydrocannabinol. *Febs Letters, 427*(2), 283-285.

Fischedick, J. T., Hazekamp, A., Erkelens, T., Choi, Y. H., & Verpoorte, R. (2010). Metabolic fingerprinting of Cannabis sativa L, cannabinoids and terpenoids for chemotaxonomic and drug standardization purposes. *Phytochemistry, 71*(17-18), 2058-2073, doi:DOI 10.1016/j.phytochem.2010.10.001.

Fleming, M. P., & Clarke, R. C. (2008). Physical evidence for the antiquity of Cannabis sativa L. *Journal of the International Hemp Association, 5*(2), 80-92.

Flemming, T., Muntendam, R., Steup, C., & Kayser, O. (2007). Chemistry and Biological Activity of Tetrahydrocannabinol and its Derivatives. *Top Heterocycl Chem, 10*, 1-42, doi:10.1007/7081_2007_084.

Flores-Sanchez, I. J., Pec, J., Fei, J. N., Choi, Y. H., Dusek, J., & Verpoorte, R. (2009). Elicitation studies in cell suspension cultures of Cannabis sativa L. *Journal of Biotechnology, 143*(2), 157-168, doi:DOI 10.1016/j.jbiotec.2009.05.006.

Flores-Sanchez, I. J., & Verpoorte, R. (2008). Secondary metabolism in cannabis. *Phytochemistry Reviews, 7*, 615-639.

Forshed, J., Torgrip, R. J. O., Aberg, K. M., Karlberg, B., Lindberg, J., & Jacobsson, S. P. (2005). A comparison of methods for alignment of NMR peaks in the context of cluster analysis. *Journal of Pharmaceutical and Biomedical Analysis, 38*(5), 824-832, doi:DOI 10.1016/j.jpba.2005.01.042.

Frederich, M., Choi, Y. H., Angenot, L., Harnischfeger, G., Lefeber, A. W. M., & Verpoorte, R. (2004). Metabolomic analysis of Strychnos nux-vomica, Strychnos icaja and Strychnos ignatii extracts by H-1 nuclear magnetic resonance spectrometry and multivariate analysis tectiniques. *Phytochemistry, 65*(13), 1993-2001, doi:DOI 10.1016/j.phytochem.2004.06.015.

Fridman, E., Wang, J., Iijima, Y., Froehlich, J. E., Gang, D. R., Ohlrogge, J., et al. (2005). Metabolic, Genomic, and Biochemical Analyses of Glandular Trichomes from the Wild Tomato Species Lycopersicon hirsutum Identify a Key Enzyme in the Biosynthesis of Methylketones. *The Plant Cell, 17*(4), 1252-1267.

Gagne, S. J., Stout, J. M., Liu, E. W., Boubakir, Z., Clark, S. M., & Page, J. E. (2012). Identification of olivetolic acid cyclase from Cannabis sativa reveals a unique catalytic route to plant polyketides. *Proceedings of the National Academy of Sciences of the United States of America, 109*(31), 12811-12816, doi:DOI 10.1073/pnas.1200330109.

Gaoni, Y., & Mechoulam, R. (1964). Isolation Structure + Partial Synthesis of Active Constituent of Hashish. *Journal of the American Chemical Society, 86*(8), 1646-&.

Garthwaite, A. J., Bothmer, R. V., & Colmer, T. D. (2005). Salt tolerance in wild Hordeum species is associated with restricted entry of Na^+ and Cl^- into the shoots. *Journal of Experimental Botany, 56*, 2365-2378, doi:10.1093/jxb/eri229.

Gonzalez, W. L., Negritto, M. A., Suarez, L. H., & Gianoli, E. (2008). Induction of glandular and non-glandular trichomes by damage in leaves of Madia sativa under contrasting water regimes. *Acta Oecologica-International Journal of Ecology, 33*(1), 128-132, doi:DOI 10.1016/j.actao.2007.10.004.

Grlic, L. (1962). A Comparative-Study on Some Chemical and Biological Characteristics of Various Samples of Cannabis Resin. *Bulletin on Narcotics, 14*(3), 37-46.

Grlic, L. (1968). Combined Spectrophotometric Differentiation of Samples of Cannabis. *Bulletin on Narcotics, 20*(3), 25-29.

Grlic, L., & Andrec, A. (1961). Content of Acid Fraction in Cannabis Resin of Various Age and Provenance. *Experientia, 17*(7), 325-&, doi:Doi 10.1007/Bf02158184.

Günnewich, N., Page, J. E., Köllner, T. G., Degenhardt, J., & Kutchan, T. M. (2007). Functional expression and characterization of trichome-specific (-)-limonene synthase and (+)-α-pinene synthase from Cannabis sativa. *Natural Product Communications, 2*(3), 223-232.

Hampson, A. J., Grimaldi, M., Axelrod, J., & Wink, D. (1998). Cannabidiol and (-)Delta(9)-tetrahydrocannabinol are neuroprotective antioxidants. *Proceedings of the National Academy of Sciences of the United States of America, 95*(14), 8268-8273.

Happyana, N., Agnolet, S., Muntendam, R., Van Dam, A., Schneider, B., & Kayser, O. (2012a). Cannabinoid analysis of laser-microdissected trichomes of Cannabis sativa L. BY LC-MS and cryogenic NMR. *Planta medica, 78*(11), 1134-1134.

Happyana, N., Agnolet, S., Muntendam, R., Van Dam, A., Schneider, B., & Kayser, O. (2013). Analysis of cannabinoids in laser-microdissected trichomes of medicinal Cannabis sativa using LCMS and cryogenic NMR. *Phytochemistry, 87*, 51-59, doi:DOI 10.1016/j.phytochem.2012.11.001.

Happyana, N., Muntendam, R., & Kayser, O. (2012b). Metabolomics as a Bionalytical Tool for Characterization of Medicinal Plants and Their Phytomedical Preparations. In O. Kayser, & H. Warzecha (Eds.), *Pharmaceutical Biotechnology* (pp. 527-552). Weinheim: WILEY-BLACKWELL.

Happyana, N., Muntendam, R., & Kayser, O. (2012c). Metabolomics as Bioanalytical Tool for Characterization of Medicinal Plants and their Phytomedical Preparations. In O. Kayser, & H. Warzecha (Eds.), *Pharmaceutical Biology*. Weinheim: Wiley-Blackwell.

Hazekamp, A., Choi, Y. H., & Verpoorte, R. (2004). Quantitative analysis of cannabinoids from Cannabis sativa using H-1-NMR. *Chemical & Pharmaceutical Bulletin, 52*(6), 718-721, doi:Doi 10.1248/Cpb.52.718.

Heinrich, M. (2008). Ethnopharmacy and natural product research-Multidisciplinary opportunities for research in the metabolomic age. *Phytochemistry Letters, 1*(1), 1-5, doi:DOI 10.1016/j.phytol.2007.11.003.

Hendrawati, O., Yao, Q. Q., Kim, H. K., Linthorst, H. J. M., Erkelens, C., Lefeber, A. W. M., et al. (2006). Metabolic differentiation of Arabidopsis treated with methyl jasmonate using nuclear magnetic resonance spectroscopy. *Plant Science, 170*(6), 1118-1124, doi:DOI 10.1016/j.plantsci.2006.01.017.

Hillig, K. W. (2005). Genetic evidence for speciation in Cannabis (Cannabaceae). *Genetic Resources and Crop Evolution, 52*(2), 161-180, doi:DOI 10.1007/s10722-003-4452-y.

Hulskamp, M. (2000). Cell morphogenesis: How plants split hairs. *Current Biology, 10*(8), R308-R310, doi:Doi 10.1016/S0960-9822(00)00437-1.

Hulskamp, M., & Kirik, V. (2000). Trichome differentiation and morphogenesis in Arabidopsis. *Advances in Botanical Research Incorporating Advances in Plant Pathology, Vol 31 2000, 31*, 237-260, doi:Doi 10.1016/S0065-2296(00)31013-8.

Ioset, K. N., Nyberg, N. T., Van Diermen, D., Malnoe, P., Hostettmann, K., Shikov, A. N., et al. (2011). Metabolic Profiling of Rhodiola rosea Rhizomes by H-1 NMR Spectroscopy. *Phytochemical Analysis, 22*(2), 158-165, doi:Doi 10.1002/Pca.1262.

J. L. Izquierdo-García, Villa, P., Kyriazis, A., del Puerto-Nevado, L., Sandra Pérez-Rial c, Rodriguez, I., et al. (2011). Descriptive review of current NMR-based metabolomic data analysis packages. *Progress in Nuclear Magnetic Resonance Spectroscopy*, doi:doi:10.1016/j.pnmrs.2011.02.001

Jackson, J. E. (1991). *A User's Guide to Principal Components*. New York: Wiley-Interscience.

Jansen, J. J., Smit, S., Hoefsloot, H. C. J., & Smilde, A. K. (2010). The Photographer and the Greenhouse: How to Analyse Plant Metabolomics Data. *Phytochemical Analysis, 21*(1), 48-60, doi:Doi 10.1002/Pca.1181.

Jungkunz, I., Link, K., Vogel, F., Voll, L. M., Sonnewald, S., & Sonnewald, U. (2011). AtHsp70-15-deficient Arabidopsis plants are characterized by reduced growth, a constitutive cytosolic protein response and enhanced resistance to TuMV. *The Plant Journal, 66*, 983-995, doi:10.1111/j.1365-313X.2011.04558.x.

Kim, E. S., & Mahlberg, P. G. (1991). Secretory Cavity Development in Glandular Trichomes of Cannabis-Sativa L (Cannabaceae). *American Journal of Botany, 78*(2), 220-229.

Kim, E. S., & Mahlberg, P. G. (1997). Immunochemical localization of tetrahydrocannabinol (THC) in cryofixed glandular trichomes of Cannabis (Cannabaceae). *American Journal of Botany, 84*(3), 336-342.

Kim, E. S., & Mahlberg, P. G. (2003). Secretory vesicle formation in the secretory cavity of glandular trichomes of Cannabis sativa L. (Cannabaceae). *Molecules and Cells, 15*(3), 387-395.

Kim, H. K., Choi, Y. H., Erkelens, C., Lefeber, A. W. M., & Verpoorte, R. (2005). Metabolic fingerprinting of Ephedra species using H-1-NMR spectroscopy and principal component analysis. *Chemical & Pharmaceutical Bulletin, 53*(1), 105-109.

Kim, H. K., & Verpoorte, R. (2010). Sample Preparation for Plant Metabolomics. *Phytochemical Analysis, 21*(1), 4-13, doi:Doi 10.1002/Pca.1188.

Kim, H. K., Wilson, E. G., Choi, Y. H., & Verpoorte, R. (2010). Metabolomics: A Tool for Anticancer Lead-Finding from Natural Products. *Planta medica, 76*(11), 1094-1102, doi:DOI 10.1055/s-0030-1249898.

Kirk, H., Cheng, D., Choi, Y. H., Vrieling, K., & Klinkhamer, P. G. (2012). Transgressive segregation of primary and secondary metabolites in F(2) hybrids between Jacobaea aquatica and J. vulgaris. *Metabolomics : Official journal of the Metabolomic Society, 8*(2), 211-219, doi:10.1007/s11306-011-0301-8.

Koning, A. J., Rose, R., & Comai, L. (1992). Developmental expression of tomato heat-shock cognate protein 80. *Plant Physiology, 100*, 801-811.

Krishnan, P., Kruger, N. J., & Ratcliffe, R. G. (2005). Metabolite fingerprinting and profiling in plants using NMR. *Journal of Experimental Botany, 56*(410), 255-265, doi:Doi 10.1093/Jxb/Eri010.

Kruger, N. J., Ratcliffe, R. G., & Roscher, A. (2003). Quantitative approaches for analysing fluxes through plant metabolic networks using NMR and stable isotope labelling. *Phytochemistry Reviews, 2*, 17-30.

Kusaka, M., Ohta, M., & Fujimura, T. (2005). Contribution of inorganic components to osmotic adjustment and leaf folding for drought tolerance in pearl millet. *Physiologia Plantarum, 125*(4), 474-489, doi:10.1111/j.1399-3054.2005.00578.x.

Lanyon, V. S., Turner, J. C., & Mahlberg, P. G. (1981). Quantitative-Analysis of Cannabinoids in the Secretory Product from Capitate-Stalked Glands of Cannabis-Sativa L (Cannabaceae). *Botanical Gazette, 142*(3), 316-319.

Larkin, J. C., Brown, M. L., & Schiefelbein, J. (2003). How do cells know what they want to be when they grow up? Lessons from epidermal patterning in Arabidopsis. *Annual Review of Plant Biology, 54*, 403-430, doi:DOI 10.1146/annurev.arplant.54.031902.134823.

Lastres-Becker, I., Molina-Holgado, F., Ramos, J. A., Mechoulam, R., & Fernandez-Ruiz, J. (2005). Cannabinoids provide neuroprotection against 6-hydroxydopamine toxicity in vivo and in vitro: Relevance to Parkinson's disease. *Neurobiology of Disease, 19*(1-2), 96-107, doi:DOI 10.1016/j.nbd.2004.11.009.

Lea, P. J., Sodek, L., Parry, M. A. J., Shewry, P. R., & Halford, N. G. (2007). Asparagines in plants. *Annals of Applied Biology, 150*(1), 1-26, doi:doi:10.1111/j.1744-7348.2006.00104.x.

Lee, M. W., Jelenska, J., & Greenberg, J. T. (2008). Arabidopsis proteins important for modulating defense responses to Pseudomonas syringae that secrete HopW1-1. *The Plant Journal, 54*, 452-465, doi:10.1111/j.1365-313X.2008.03439.x.

Lee, R. Y., & Wang, X. (2005). Concise Synthesis of Biologically Interesting (±)-Cannabichromene, (±)-Cannabichromenic Acid, and (±)-Daurichromenic Acid. *Bulletin Korean Chemical Society, 26*(12), 1933-1936.

Lee, Y. R., & Wang, X. (2005). Concise synthesis of biologically interesting (+/-)-cannabichromene, (+/-)-cannabichromenic acid, and (+/-)-daurichromenic acid. *Bulletin of the Korean Chemical Society, 26*(12), 1933-1936.

Lewis, G. S., & Turner, C. E. (1978). Constituents of Cannabis-Sativa L .13. Stability of Dosage Form Prepared by Impregnating Synthetic (-)-Delta-9-Trans-Tetrahydrocannabinol on Placebo Cannabis Plant Material. *Journal of Pharmaceutical Sciences, 67*(6), 876-878.

Li, F. M., Lu, X. M., Liu, H. P., Liu, M., & Xiong, Z. L. (2007). A pharmaco-metabonomic study on the therapeutic basis and metabolic effects of Epimedium brevicornum Maxim. on hydrocortisone-induced rat using UPLC-MS. *Biomedical Chromatography, 21*(4), 397-405, doi:Doi 10.1002/Bmc.770.

Li, H.-L. (1974). An archeological and historical account of cannabis in China. *Economic Botany, 28*(4), 437-448.

Li, L., Sun, B., Zhang, Q., Fang, J. J., Ma, K. P., Li, Y., et al. (2008). Metabonomic study on the toxicity of Hei-Shun-Pian, the processed lateral root of Aconitum carmichaelii Debx. (Ranunculaceae). *Journal of Ethnopharmacology, 116*(3), 561-568, doi:DOI 10.1016/j.jep.2008.01.014.

Li, P., Wei, Y. H., Qi, L. W., Luo, H. W., Yi, L., & Sheng, L. H. (2007). Improved quality control method for Fufang Danshen preparations through simultaneous determination of phenolic acids, saponins and diterpenoid quinones by HPLC coupled with diode array and evaporative light scattering detectors. *Journal of Pharmaceutical and Biomedical Analysis, 45*(5), 775-784, doi:10.1016/j.jpba.2007.07.013.

Li, S.-H., Schneider, B., & Gershenzon, J. (2007). Microchemical analysis of laser-microdissected stone cells of Norway spruce by cryogenic nuclear magnetic resonance spectroscopy. *Planta, 225*, 771-779, doi:10.1007/s00425-006-0376-z.

Li, Y. P., Hu, Z., & He, L. C. (2007). An approach to develop binary chromatographic fingerprints of the total alkaloids from Caulophyllum robustum by high performance liquid chromatography/diode array detector and gas chromatography/mass spectrometry. *Journal of Pharmaceutical and Biomedical Analysis, 43*(5), 1667-1672, doi:DOI 10.1016/j.jpba.2006.12.028.

Liang, Y. Z., Zeng, Z. D., Zhang, T., Chau, F. T., & Wang, Y. L. (2006). Orthogonal projection (OP) technique applied to pattern recognition of fingerprints of the herbal medicine houttuynia cordata Thunb. and its final injection products. *Analytical and Bioanalytical Chemistry, 385*(2), 392-400, doi:10.1007/s00216-006-0405-6.

Linn, S. (1998). DNA damage by iron and hydrogen peroxide in vitro and in vivo. *Drug metabolism reviews, 30*, 313-326.

Linnaeus, C. (1753). *Species plantarum T. I-II*.

Liu, N. Q., Cao, M., Frederich, M., Choi, Y. H., Verpoorte, R., & van der Kooy, F. (2010). Metabolomic investigation of the ethnopharmacological use of Artemisia afra with NMR spectroscopy and multivariate data analysis. *Journal of Ethnopharmacology, 128*(1), 230-235, doi:DOI 10.1016/j.jep.2010.01.020.

Liu, Q., Li, H., & Hang, T. (2007). Identification of chemical components illegally mixed in traditional Chinese medicine preparations. . *Zhongguo Yaoshi, 10*, 735-737.

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods, 25*(4), 402-408, doi:10.1006/meth.2001.1262.

Lu, Y. H., Liu, X. R., Liang, X., Xiang, L., & Zhang, W. D. (2011). Metabolomic strategy to study therapeutic and synergistic effects of tanshinone IIA, salvianolic acid B and ginsenoside Rb1 in myocardial ischemia rats. *Journal of Ethnopharmacology, 134*(1), 45-49, doi:DOI 10.1016/j.jep.2010.11.048.

Ludwig, C., & Viant, M. R. (2010). Two-dimensional J-resolved NMR Spectroscopy: Review of a Key Methodology in the Metabolomics Toolbox. *Phytochemical Analysis, 21*(1), 22-32, doi:Doi 10.1002/Pca.1186.

Ma, C. F., Wang, H. H., Lu, X., Wang, H., Xu, G. W., & Liu, B. Y. (2009). Terpenoid metabolic profiling analysis of transgenic Artemisia annua L. by comprehensive two-dimensional gas chromatography time-of-flight mass spectrometry. *Metabolomics, 5*(4), 497-506, doi:DOI 10.1007/s11306-009-0170-6.

Ma, C. F., Wang, H. H., Lu, X., Xu, G. W., & Liu, B. Y. (2008). Metabolic fingerprinting investigation of Artemisia annua L. in different stages of development by gas chromatography and gas chromatography-mass spectrometry. *Journal of Chromatography A, 1186*(1-2), 412-419, doi:DOI 10.1016/j.chroma.2007.09.023.

Mahlberg, P. G., & Kim, E. S. (2004). Accumulation of Cannabinoids in Glandular Trichomes of *Cannabis* (Cannabaceae). *Journal of Industrial Hemp, 9*(1), 15-36, doi:http://dx.doi.org/10.1300/J237v09n01_04.

Malingre, T., Hendriks, H., Batterman, S., Bos, R., & Visser, J. (1975). The essential oil of Cannabis sativa. *Planta medica, 28*(1), 56-61, doi:10.1055/s-0028-1097829.

Marks, M. D., Tian, L., Wenger, J. P., Omburo, S. N., Soto-Fuentes, W., He, J., et al. (2009a). Identification of candidate genes affecting delta(9)-tetrahydrocannabinol biosynthesis in Cannabis sativa. *Journal of Experimental Botany, 60*(13), 3715-3726, doi:Doi 10.1093/Jxb/Erp210.

Marks, M. D., Wenger, J. P., Gilding, E., Jilk, R., & Dixon, R. A. (2009b). Transcriptome Analysis of Arabidopsis Wild-Type and gl3-sst sim Trichomes Identifies Four Additional Genes Required for Trichome Development. *Molecular Plant, 2*(4), 803-822, doi:Doi 10.1093/Mp/Ssp037.

Matsuda, L. A., Lolait, S. J., Brownstein, M. J., Young, A. C., & Bonner, T. I. (1990). Structure of a Cannabinoid Receptor and Functional Expression of the Cloned Cdna. *Nature, 346*(6284), 561-564.

Maurer, M., Henn, V., Dittrich, A., & Hofmann, A. (1990). Delta-9-Tetrahydrocannabinol Shows Antispastic and Analgesic Effects in a Single Case Double-Blind Trial. *European Archives of Psychiatry and Clinical Neuroscience, 240*(1), 1-4.

McPartland, E. B., & Mediavilla, V. (2002). Noncannabinoid Components. In E. Russo, & F. Grotenhermen (Eds.), *Cannabis and cannabinoids: pharmacology, toxicology and therapeutic potential.* (pp. 401-409). New York: The Haworth Integrative Healing Press.

McPartland, J. M., Clarke, R. C., Watson, D. P., & C.A.B. International. (2000). *Hemp diseases and pests : management and biological control : an advanced treatise*. Wallingford: CABI.

Mechoulam, R., Parker, L. A., & Gallily, R. (2002). Cannabidiol: An overview of some pharmacological aspects. *Journal of Clinical Pharmacology, 42*(11), 11s-19s, doi:Doi 10.1177/0091270002238789.

Meister, A. (1988). Glutathione Metabolism and Its Selective Modification. *The Journal of Biological Chemistry, 263*(33), 17205-17208.

Merritt, J. C., Crawford, W. J., Alexander, P. C., Anduze, A. L., & Gelbart, S. S. (1980). Effect of Marihuana on Intraocular and Blood-Pressure in Glaucoma. *Ophthalmology, 87*(3), 222-228.

Militante, J. D., & Lombardini, J. B. (2001). Increased cardiac levels of taurine in cardiomyopathy: the paradoxical benefits of oral taurine treatment. *Nutrition Research, 21*(1-2), 93-102.

Moreno, J. I., Martin, R., & Castresana, C. (2005). Arabidopsis SHMT1, a serine hydroxymethyltransferase that functions in the photorespiratory pathway influences resistance to biotic and abiotic stress. *The Plant Journal, 41*, 451-463.

Morimoto, S., Komatsu, K., Taura, F., & Shoyama, Y. (1998). Purification and characterization of cannabichromenic acid synthase from Cannabis sativa. *Phytochemistry, 49*(6), 1525-1529.

Morimoto, S., Tanaka, Y., Sasaki, K., Tanaka, H., Fukamizu, T., Shoyama, Y., et al. (2007). Identification and characterization of cannabinoids that induce cell death through mitochondrial permeability transition in Cannabis leaf cells. *Journal of Biological Chemistry, 282*(28), 20739-20751, doi:DOI 10.1074/jbc.M700133200.

Mortensen, J., Eriksen, J., & Nielsen, J. D. (1993). Sulphur deficiency and amino acid composition in seeds and grass. *Phyton – Annales Rei Botanicae, 32*(85-90).

Moyano, F., Cocucci, A., & Sersic, A. (2003). Accessory pollen adhesive from glandular trichomes on the anthers of Leonurus sibiricus L. (Lamiaceae). *Plant Biology, 5*(4), 411-418, doi:Doi 10.1055/S-2003-42707.

Muller-Vahl, K. R., Schneider, U., Koblenz, A., Jobges, M., Kolbe, H., Daldrup, T., et al. (2002). Treatment of Tourette's syndrome with Delta(9)-tetrahydrocannabinol (THC): A randomized crossover trial. *Pharmacopsychiatry, 35*(2), 57-61.

Muller-Vahl, K. R., Schneider, U., Kolbe, H., & Emrich, H. M. (1999). Treatment of Tourette's syndrome with delta-9-tetrahydrocannabinol. *American Journal of Psychiatry, 156*(3), 495-495.

Munro, S., Thomas, K. L., & Abushaar, M. (1993). Molecular Characterization of a Peripheral Receptor for Cannabinoids. *Nature, 365*(6441), 61-65.

Muntendam, R., Happyana, N., Erkelens, C., Bruining, F., & Kayser, O. (2012). Time Dependent Metabolomics and Transcriptional Analysis of Cannabinoid Biosynthesis in Cannabis sativa var. Bedrobinol and Bediol Grown under Standardized Condition and with Genetic Homogeneity. *Online International Journal of Medicinal Plants Research, 1*(2), 31-40.

Neff, G. W., O'Brien, C. B., Reddy, K. R., Bergasa, N. V., Regev, A., Molina, E., et al. (2002). Preliminary observation with dronabinol in patients with intractable pruritus secondary to cholestatic liver disease. *American Journal of Gastroenterology, 97*(8), 2117-2119, doi:Pii S0002-9270(02)04208-9.

Ni, Y., Su, M. M., Qiu, Y. P., Chen, M. J., Liu, Y. M., Zhao, A. H., et al. (2007). Metabolic profiling using combined GC-MS and LC-MS provides a systems understanding of aristolochic acid-induced nephrotoxicity in rat. *Febs Letters, 581*(4), 707-711, doi:DOI 10.1016/j.febslet.2007.01.036.

Nuringtyas, T. R., Choi, Y. H., Verpoorte, R., Klinkhamer, P. G., & Leiss, K. A. (2012). Differential tissue distribution of metabolites in Jacobaea vulgaris, Jacobaea aquatica and their crosses. [Research Support, Non-U.S. Gov't]. *Phytochemistry, 78*, 89-97, doi:10.1016/j.phytochem.2012.03.011.

Olsson, M. E., Olofsson, L. M., Lindahl, A. L., Lundgren, A., Brodelius, M., & Brodelius, P. E. (2009). Localization of enzymes of artemisinin biosynthesis to the apical cells of glandular secretory trichomes of Artemisia annua L. *Phytochemistry, 70*(9), 1123-1128, doi:DOI 10.1016/j.phytochem.2009.07.009.

Orino, K., Lehman, L., Tsuji, Y., Ayaki, H., Torti, S. V., & Torti, F. M. (2001). Ferritin and the response to oxidative stress. *Biochemical Journal, 357*, 241-247.

Pacifico, D., Miselli, F., Micheler, M., Carboni, A., Ranalli, P., & Mandolino, G. (2006). Genetics and marker-assisted selection of the chemotype in Cannabis sativa L. *Molecular Breeding, 17*(3), 257-268, doi:DOI 10.1007/s11032-005-5681-x.

Park, S. K., Seo, J. B., & Lee, M. Y. (2012). Proteomic profiling of hempseed proteins from Cheungsam. [Research Support, Non-U.S. Gov't]. *Biochimica et biophysica acta, 1824*(2), 374-382, doi:10.1016/j.bbapap.2011.10.005.

Pe´rez-Garcı´a, A., Pereira, S., Pissarra, J., Garcı´a Gutie´rrez, A., Cazorla, F. M., Salema, R., et al. (1998). Cytosolic localization in tomato mesophyll cells of a novel glutamine synthetase induced in response to bacterial infection or phosphinothricin treatment. *Planta, 206*, 426-434.

Petri, G., Oroszlan, P., & Fridvalszky, L. (1988). Histochemical Detection of Hemp Trichomes and Their Correlation with the Content. *Acta Biologica Hungarica, 39*(1), 59-74.

Petro, D. J., & Ellenberger, C. (1981). Treatment of Human Spasticity with Delta-9-Tetrahydrocannabinol. *Journal of Clinical Pharmacology, 21*(8-9), S413-S416.

Plumb, R. S., Johnson, K. A., Rainville, P., Smith, B. W., Wilson, I. D., Castro-Perez, J. M., et al. (2006). UPLIC/MSE; a new approach for generating molecular fragment information for biomarker structure elucidation. *Rapid Communications in Mass Spectrometry, 20*(13), 1989-1994, doi:10.1002/rcm.2550.

Qiu, Y. Q., Lu, X., Pang, T., Zhu, S. K., Kong, H. W., & Xu, G. W. (2007). Study of traditional Chinese medicine volatile oils from different geographical origins by comprehensive two-dimensional gas chromatography-time-of-flight mass spectrometry (GC x GC-TOFMS) in combination with multivariate analysis. *Journal of Pharmaceutical and Biomedical Analysis, 43*(5), 1721-1727, doi:DOI 10.1016/j.jpba.2007.01.013.

Radwan, M. M., ElSohly, M. A., Slade, D., Ahmed, S. A., Khan, I. A., & Ross, S. A. (2009). Biologically Active Cannabinoids from High-Potency Cannabis sativa. *Journal of Natural Products, 72*(5), 906-911, doi:Doi 10.1021/Np900067k.

Radwan, M. M., ElSohly, M. A., Slade, D., Ahmed, S. A., Wilson, L., El-Alfy, A. T., et al. (2008a). Non-cannabinoid constituents from a high potency Cannabis sativa variety. *Phytochemistry, 69*(14), 2627-2633, doi:DOI 10.1016/j.phytochem.2008.07.010.

Radwan, M. M., Ross, S. A., Slade, D., Ahmed, S. A., Zulfiqar, F., & ElSohly, M. A. (2008b). Isolation and characterization of new cannabis constituents from a high potency variety. *Planta medica, 74*(3), 327-327.

Raharjo, T. J., Widjaja, I., Roytrakul, S., & Verpoorte, R. (2004). Comparative proteomics of Cannabis sativa plant tissues. [Comparative Study

Research Support, Non-U.S. Gov't]. *Journal of biomolecular techniques : JBT, 15*(2), 97-106.

Rasmussen, B., Cloarec, O., Tang, H. R., Staerk, D., & Jaroszewski, J. W. (2006). Multivariate analysis of integrated and full-resolution H-1-NMR spectral data from complex pharmaceutical preparations: St. John's wort. *Planta medica, 72*(6), 556-563, doi:DOI 10.1055/s-2006-931567.

Ratcliffe, R. G., Roscher, A., & Shachar-Hill, Y. (2001). Plant NMR spectroscopy. *Progress in Nuclear Magnetic Resonance Spectroscopy, 39*(4), 267-300.

Ratcliffe, R. G., & Shachar-Hill, Y. (2005). Revealing metabolic phenotypes in plants: inputs from NMR analysis. *Biological Reviews, 80*(1), 27-43, doi:Doi 10.1017/S1464793104006530.

Raven, E. L. (2000). Peroxidase-Catalyzed Oxidation of Ascorbate Structural, Spectroscopic and Mechanistic Correlations in Ascorbate Peroxidase. *Subcellular Biochemistry, 35*, 317-349.

Razdan, R. K., Puttick, A. J., Zitko, B. A., & Handrick, G. R. (1972). Hashish .6. Conversion of (-)-Delta1(6)-Tetrahydrocannabinol to (-)-Delta1(7)-Tetrahydrocannabinol - Stability of (-)-Delta1- and "(-)-Delta1(6)-Tetrahydrocannabinols. *Experientia, 28*(2), 121-&.

Rhee, S. G., Kang, S. W., Chang, T. S., Jeong, W., & Kim, K. (2001). Peroxiredoxin, a Novel Family of Peroxidases. *IUBMB Life, 52*(1), 35-41.

Roberts, J. D. (1959). *Nuclear Magnetic Resonance Applications to Organic Chemistry*. NEW YORK, TORONTO, LONDON: McGRAW-HILL BOOK COMPANY, INC.

Roberts, J. K. M. (2000). NMR adventures in the metabolic labyrinth within plants. *Trends in Plant Science, 5*(1), 30-34.

Romero, G. Q., Souza, J. C., & Vasconcellos-Neto, J. (2008). Anti-Herbivore Protection by Mutualistic Spiders and the Role of Plant Glandular Trichomes. *Ecology, 89*(11), 3105-3115, doi:Doi 10.1890/08-0267.1.

Ross, S. A., ElSohly, M. A., Sultana, G. N. N., Mehmedic, Z., Hossain, C. F., & Chandra, S. (2005). Flavonoid glycosides and cannabinoids from the pollen of Cannabis sativa L. *Phytochemical Analysis, 16*(1), 45-48, doi:Doi 10.1002/Pca.809.

Rufty, T. W., Israel, D. W., Volk, R. J., Qiu, J., & Sa, T. (1993). Phosphate regulation of nitrate assimilation in soybean. *Journal of Experimental Botany, 44*(879-891).

Rufty, T. W., MacKown, C. T., & Israel, D. W. (1990). Phosphorus stress effects on the assimilation of nitrate. *Plant Physiology, 94*, 328–333.

Sallan, S. E., Zinberg, N. E., & Frei, E. (1975). Antiemetic Effect of Delta-9-Tetrahydrocannabinol in Patients Receiving Cancer Chemotherapy. *New England Journal of Medicine, 293*(16), 795-797.

Sanchez-Sampedro, A., Kim, H. K., Choi, Y. H., Verpoorte, R., & Corchete, P. (2007). Metabolomic alterations in elicitor treated Silybum marianum suspension cultures monitored by nuclear magnetic resonance spectroscopy. *Journal of Biotechnology, 130*(2), 133-142, doi:DOI 10.1016/j.jbiotec.2007.03.007.

Scalbert, A., Brennan, L., Fiehn, O., Hankemeier, T., Kristal, B. S., van Ommen, B., et al. (2009). Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics, 5*(4), 435-458, doi:DOI 10.1007/s11306-009-0168-0.

Scarpari, L. M., Meinhardt, L. W., Mazzafera, P., Pomella, A. W. V., Schiavinato, M. A., Cascardo, J. C. M., et al. (2005). Biochemical changes during the development of witches' broom: the most important disease of cocoa in Brazil caused by Crinipellis perniciosa. *Journal of Experimental*

Botany, 56, 865-877.

Schad, M., Lipton, M. S., Giavalisco, P., Smith, R. D., & Kehr, J. (2005a). Evaluation of two-dimensional electrophoresis and liquid chromatography tandem mass spectrometry for tissue-specific protein profiling of laser-microdissected plant samples (vol 26, pg 2729, 2005). *Electrophoresis, 26*(17), 3406-3406.

Schad, M., Mungur, R., Fiehn, O., & Kehr, J. (2005b). Metabolic profiling of laser microdissected vascular bundles of Arabidopsis thaliana. *Plant Methods, 1*(2), 1-10, doi:10.1186/1746-4811-1-2.

Schilmiller, A. L., Last, R. L., & Pichersky, E. (2008). Harnessing plant trichome biochemistry for the production of useful compounds. *Plant Journal, 54*(4), 702-711, doi:DOI 10.1111/j.1365-313X.2008.03432.x.

Schneider, B., & Holscher, D. (2007). Laser microdissection and cryogenic nuclear magnetic resonance spectroscopy: an alliance for cell type-specific metabolite profiling. *Planta, 225*(3), 763-770, doi:DOI 10.1007/s00425-006-0404-z.

Schneider, B., & Hölscher, D. (2007). Laser microdissection and cryogenic nuclear magnetic resonance spectroscopy: an alliance for cell type-speciWc metabolite profiling. *Planta, 225*, 763-770, doi:10.1007/s00425-006-0404-z.

Scholz, M., Gatzek, S., Sterling, A., Fiehn, O., & Selbig, J. (2004). Metabolite fingerprinting: detecting biological features by independent component analysis. *Bioinformatics, 20*(15), 2447-2454, doi:10.1093/bioinformatics/bth270.

Schripsema, J. (2010). Application of NMR in Plant Metabolomics: Techniques, Problems and Prospects. *Phytochemical Analysis, 21*(1), 14-21, doi:Doi 10.1002/Pca.1185.

Schroder, H. F. (1996). Polar organic pollutants from textile industries in the wastewater treatment process - Biochemical and physico-chemical elimination and degradation monitoring by LC-MS, FIA-MS and MS-MS. *Trac-Trends in Analytical Chemistry, 15*(8), 349-362.

Schultes, R. E., Klein, W. M., Plowman, T., & Lockwood, T. E. (1974). Cannabis: an example of taxonomic neglect. *Harvard University Botanical Museum Leaflets, 23*, 337–367.

Schwab, B., Folkers, U., Ilgenfritz, H., & Hulskamp, M. (2000). Trichome morphogenesis in Arabidopsis. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences, 355*(1399), 879-883.

Segelman, A. B., Segelman, F. P., Star, A. E., Wagner, H., & Seligmann, O. (1978). Structure of 2 C-Diglycosylflavones from Cannabis-Sativa. *Phytochemistry, 17*(4), 824-826.

Simmons, A. T., & Gurr, G. M. (2005). Trichomes of Lycopersicon species and their hybrids: effects on pests and natural enemies. *Agricultural and Forest Entomology, 7*(4), 265-276, doi:DOI 10.1111/j.1461-9555.2005.00271.x.

Simpson, P. J., Tantitadapitak , C., Reed, A. M., Mather, O. C., Bunce, C. M., White, S. A., et al. (2009). Characterization of two novel aldo-keto reductases from Arabidopsis: expression patterns, broad substrate specificity, and an open active-site structure suggest a role in toxicant metabolism following stress. *Journal of Molecular Biology, 392*(2), 465-480.

Sirikantaramas, S., Morimoto, S., Shoyama, Y., Ishikawa, Y., Wada, Y., Shoyama, Y., et al. (2004). The gene controlling marijuana psychoactivity - Molecular cloning and heterologous expression of Delta(1)-tetrahydrocannabinolic acid synthase from Cannabis sativa L. *Journal of Biological Chemistry, 279*(38), 39767-39774, doi:DOI 10.1074/jbc.M403693200.

Sirikantaramas, S., Taura, F., Tanaka, Y., Ishikawa, Y., Morimoto, S., & Shoyama, Y. (2005). Tetrahydrocannabinolic acid synthase, the enzyme controlling marijuana psychoactivity, is secreted into the storage cavity of the glandular trichomes. *Plant and Cell Physiology, 46*(9), 1578-1582.

Small E., & A., C. (1976). A practical and natural taxonomy for Cannabis. *Taxon 25*(4), 405–435.

Spraul, M., Neidig, P., Klauck, U., Kessler, P., Holmes, E., Nicholson, J. K., et al. (1994). Automatic Reduction of Nmr Spectroscopic Data for Statistical and Pattern-Recognition Classification of Samples. *Journal of Pharmaceutical and Biomedical Analysis, 12*(10), 1215-1225.

Staab, J. M., O'Connell, T. M., & Gomez, S. M. (2010). Enhancing metabolomic data analysis with Progressive Consensus Alignment of NMR Spectra (PCANS). *Bmc Bioinformatics, 11*, -, doi:Artn 123

Doi 10.1186/1471-2105-11-123.

Stout, J. M., Boubakir, Z., Ambrose, S. J., Purves, R. W., & Page, J. E. (2012a). The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in Cannabis sativa trichomes. *The plant Journal : for cell and molecular biology, 71*(3), 353-365.

Stout, J. M., Boubakir, Z., Ambrose, S. J., Purves, R. W., & Page, J. E. (2012b). The hexanoyl-CoA precursor for cannabinoid biosynthesis is formed by an acyl-activating enzyme in Cannabis sativa trichomes. *Plant Journal, 71*(3), 353-365, doi:DOI 10.1111/j.1365-313X.2012.04949.x.

Suhartono, L., Van Iren, F., de Winter, W., Roytrakul, S., Choi, Y. H., & Verpoorte, R. (2005). Metabolic comparison of cryopreserved and normal cells from Tabernaemontana divaricata suspension cultures. *Plant Cell Tissue and Organ Culture, 83*(1), 59-66, doi:DOI 10.1007/s11240-005-3869-8.

Sumner, L. W., Mendes, P., & Dixon, R. A. (2003). Plant metabolomics: large-scale phytochemistry in the functional genomics era. *Phytochemistry, 62*(6), 817-836, doi:10.1016/S0031-9422(02)00708-2.

Szczepanowski, R., Bekel, T., Goesmann, A., Krause, L., Krömeke, H., Kaiser, O., et al. (2008). Insight into the plasmid metagenome of wastewater treatment plant bacteria showing reduced susceptibility to antimicrobial drugs analysed by the 454-pyrosequencing technology. *136*, 54-64.

Tashkin, D. P., Reiss, S., Shapiro, B. J., Calvarese, B., Olsen, J. L., & Lodge, J. W. (1977). Bronchial Effects of Aerosolized Delta-9-Tetrahydrocannabinol in Healthy and Asthmatic Subjects. *American Review of Respiratory Disease, 115*(1), 57-65.

Tauch, A., Trost, E., Bekel, T., Goesmann, A., Ludewig, U., & Puhler, A. (2006). Ultrafast De Novo Sequencing of Corynebacterium urealyticum Using the Genome Sequencer 20 System. *Journal of Biotechnology, 136*, 22-30.

Taura, F., Dono, E., Sirikantaramas, S., Yoshimura, K., Shoyama, Y., & Morimoto, S. (2007a). Production of Delta(1)-tetrahydrocannabinolic acid by the biosynthetic enzyme secreted from transgenic Pichia pastoris. *Biochemical and Biophysical Research Communications, 361*, 675-680.

Taura, F., Morimoto, S., & Shoyama, Y. (1996). Purification and characterization of cannabidiolic-acid synthase from Cannabis sativa L. Biochemical analysis of a novel enzyme that catalyzes the oxidocyclization of cannabigerolic acid to cannabidiolic acid. *Journal of Biological Chemistry, 271*(29), 17411-17416.

Taura, F., Morimoto, S., Shoyama, Y., & Mechoulam, R. (1995). First Direct Evidence for the Mechanism of Delta(1)-Tetrahydricannabinolic Acid Biosynthesis. *Journal of the American Chemical Society, 117*(38), 9766-9767.

Taura, F., Sirikantaramas, S., Shoyama, Y., Shoyama, Y., & Morimoto, S. (2007b). Phytocannabinoids in Cannabis sativa: Recent studies on Biosynthetic enzymes. *Chemistry & Biodiversity, 4*(8), 1649-1663.

Taura, F., Sirikantaramas, S., Shoyama, Y., Yoshikai, K., Shoyama, Y., & Morimoto, S. (2007c). Cannabidiolic-acid synthase, the chemotype-determining enzyme in the fiber-type Cannabis sativa. *Febs Letters, 581*(16), 2929-2934, doi:DOI 10.1016/j.febslet.2007.05.043.

Taura, F., Tanaka, S., Taguchi, C., Fukamizu, T., Tanaka, H., Shoyama, Y., et al. (2009). Characterization of olivetol synthase, a polyketide synthase putatively involved in cannabinoid biosynthetic pathway. *Febs Letters, 583*(12), 2061-2066, doi:DOI 10.1016/j.febslet.2009.05.024.

Tian, C., Tsuboi, Y., Chikayama, E., Sekiyama, Y., Kuromori, T., Shinozaki, K., et al. (2006). Advances in hetero-nuclear NMR-based metabolomics approach in plant systems. *Plant and Cell Physiology, 47*, S25-S25.

Tolstikov, V. V., & Fiehn, O. (2002). Analysis of highly polar compounds of plant origin: Combination of hydrophilic interaction chromatography and electrospray ion trap mass spectrometry. *Analytical Biochemistry, 301*(2), 298-307, doi:10.1006/abio.2001.5513.

Turner, C. E., Elsohly, M. A., & Boeren, E. G. (1980). Constituents of Cannabis-Sativa L .17. A Review of the Natural Constituents. *Journal of Natural Products, 43*(2), 169-234, doi:Doi 10.1021/Np50008a001.

Turner, J. C., Hemphill, J. K., & Mahlberg, P. G. (1978). Quantitative-Determination of Cannabinoids in Individual Glandular Trichomes of Cannabis-Sativa L (Cannabaceae). *American Journal of Botany, 65*(10), 1103-1106.

Turner, J. C., Hemphill, J. K., & Mahlberg, P. G. (1980). Trichomes and Cannabinoid Content of Developing Leaves and Bracts of Cannabis-Sativa L (Cannabaceae). *American Journal of Botany, 67*(10), 1397-1406.

Turner, J. C., Hemphill, J. K., & Mahlberg, P. G. (1981). Interrelationships of Glandular Trichomes and Cannabinoid Content .1. Developing Pistillate Bracts of Cannabis-Sativa L (Cannabaceae). *Bulletin on Narcotics, 33*(2), 59-69.

United-Nation (2009). World Population Prospects, The 2008 Revision, Highlights (D. o. E. a. S. A. (DESA), Trans.). New York: Department of Economic and Social Affairs, United Nations.

van Bakel, H., Stout, J. M., Cote, A. G., Tallon, C. M., Sharpe, A. G., Hughes, T. R., et al. (2011). The draft genome and transcriptome of Cannabis sativa. *Genome Biol, 12*(10), R102, doi:gb-2011-12-10-r102 [pii]
10.1186/gb-2011-12-10-r102.

Van Dam, N. M., & Hare, J. D. (1998). Biological activity of Datura wrightii glandular trichome exudate against Manduca sexta larvae. *Journal of Chemical Ecology, 24*(9), 1529-1549, doi:Doi 10.1023/A:1020963817685.

Van der Kooy, F., Verpoorte, R., & Meyer, J. J. M. (2008). Metabolomic quality control of claimed anti-malarial Artemisia afra herbal remedy and A. afra and A. annua plant extracts. *South African Journal of Botany, 74*(2), 186-189, doi:DOI 10.1016/j.sajb.2007.10.004.

Vanhoenacker, G., Van Rompaey, P., De Keukeleire, D., & Sandra, P. (2002). Chemotaxonomic features associated with flavonoids of cannabinoid-free cannabis (Cannabis sativa subsp sativa L.) in relation to hops (Humulus lupulus L.). *Natural Product Letters, 16*(1), 57-63, doi:Doi 10.1080/1057563029001/4863.

Vaudel, M., Burkhart, J. M., Breiter, D., Zahedi, R. P., Sickmann, A., & Martens, L. (2012). A Complex Standard for Protein Identification, Designed by Evolution. *Journal of Proteome Research, 11*(10), 5065-5071, doi:Doi 10.1021/Pr300055q.

Verpoorte, R., Choi, Y. H., & Choi, H. K. (2006). Metabolomics: Novel tool for studying complex biological systems. *Chemical Senses, 31*(8), E67-E67.

Verpoorte, R., Choi, Y. H., & Kim, H. K. (2008). Metabolomics: back to basic. *7*, 525-537, doi:10.1007/s11101-008-9091-7.

Vines, G. (2004). Herbal harvests with a future: towards sustainable sources for medicinal plants Plantlife International.

Volfe, Z., Dvilansky, A., & Nathan, I. (1985). Cannabinoids Block Release of Serotonin from Platelets Induced by Plasma from Migraine Patients. *International Journal of Clinical Pharmacology Research, 5*(4), 243-246.

Wagner, C., Sefkow, M., & Kopka, J. (2003). Construction and application of a mass spectral and retention time index database generated from plant GC/EI-TOF-MS metabolite profiles. *Phytochemistry, 62*(6), 887-900, doi:10.1016/S0031-9422(02)00703-3.

Wagner, G. J. (1991). Secreting Glandular Trichomes - More Than Just Hairs. *Plant Physiology, 96*(3), 675-679, doi:Doi 10.1104/Pp.96.3.675.

Wagner, G. J., Wang, E., & Shepherd, R. W. (2004). New approaches for studying and exploiting an old protuberance, the plant trichome. *Annals of Botany, 93*(1), 3-11, doi:Doi 10.1093/Aob/Mch011.

Wang, M., Lamers, R. J. A. N., Korthout, H. A. A. J., van Nesselrooij, J. H. J., Witkamp, R. F., van der Heijden, R., et al. (2005). Metabolomics in the context of systems biology: Bridging traditional Chinese medicine and molecular pharmacology. *Phytotherapy Research, 19*(3), 173-182, doi:Doi 10.1002/Ptr.1624.

Ward, J. L., Baker, J. M., & Beale, M. H. (2007). Recent applications of NMR spectroscopy in plant metabolomics. *Febs Journal, 274*(5), 1126-1131, doi:DOI 10.1111/j.1742-4658.2007.05675.x.

Weber, J., Schley, M., Casutt, M., Gerber, H., Schuepfer, G., Rukwied, R., et al. (2009, Clinical Study Tetrahydrocannabinol (Delta 9-THC) Treatment in Chronic Central Neuropathic Pain and Fibromyalgia Patients: Results of aMulticenter Survey. *Anesthesiology Research and Practice*, doi:10.1155/2009/827290.

Weissbach, H., Etienne, F., Hoshi, Heinemann, S. H., Lowther, W. T., Matthews, B., et al. (2002). Peptide Methionine Sulfoxide Reductase: Structure, Mechanism of Action, and Biological Function. *397*(2).

Weljie, A. M., Newton, J., Mercier, P., Carlson, E., & Slupsky, C. M. (2006). Targeted profiling: Quantitative analysis of H-1 NMR metabolomics data. *Analytical Chemistry, 78*(13), 4430-4442, doi:Doi 10.1021/Ac060209g.

Wells, W. W., & Xu, D. P. (1994). Dehydroascorbate reduction. *Journal of Bioenergetics and Biomembranes, 26*(4), 369-377.

Westerhuis, J. A., Hoefsloot, H. C. J., Smit, S., Vis, D. J., Smilde, A. K., van Velzen, E. J. J., et al. (2008). Assessment of PLSDA cross validation. *Metabolomics, 4*(1), 81-89, doi:DOI 10.1007/s11306-007-0099-6.

Westerhuis, J. A., van Velzen, E. J. J., Hoefsloot, H. C. J., & Smilde, A. K. (2010). Multivariate paired data analysis: multilevel PLSDA versus OPLSDA. *Metabolomics, 6*(1), 119-128, doi:DOI 10.1007/s11306-009-0185-z.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems, 2*(1-3), 37-52.

Wu, B., Yan, S. K., Lin, Z. Y., Wang, Q., Yang, Y., Yang, G. J., et al. (2008). Metabonomic study on ageing: NMR-based investigation into rat urinary metabolites and the effect of the total flavone of Epimedium. *Molecular Biosystems, 4*(8), 855-861, doi:Doi 10.1039/B800923f.

Wu, T., Wang, Y., & Guo, D. (2012). Investigation of Glandular Trichome Proteins in Artemisia annnua L. Using comparative Proteomics. *PLoS One, 7*(8), e41822.

Xiang, Z., Wang, X., Cai, X., & Zeng, S. (2011). Metabolomics Study on Quality Control and Discrimination of Three Curcuma Species based on Gas Chromatography-Mass Spectrometry. *Phytochemical Analysis*, doi:01.1002/pca.1296.

Xie, D. Y., Jackson, L. A., Cooper, J. D., Ferreira, D., & Paiva, N. L. (2004). Molecular and biochemical analysis of two cDNA clones encoding dihydroflavonol 4-reductase from Medicago truncatula. *Plant Physiology, 134*, 979-994.

Xie, G. X., Ni, Y., Su, M. M., Zhang, Y. Y., Zhao, A. H., Gao, X. F., et al. (2008). Application of ultra-performance LC-TOF MS metabolite profiling techniques to the analysis of medicinal Panax herbs. *Metabolomics, 4*(3), 248-260, doi:DOI 10.1007/s11306-008-0115-5.

Yadav, S. K., Singla-Pareek, S. L., Reddy, M. K., & Sopory, S. K. (2005). Transgenic tobacco plants overexpressing glyoxalase enzymes resist an increase in methylglyoxal and maintain higher reduced glutathione levels under salinity stress. *FEBS Letter, 579*(27), 6265-6271.

Yang, S. Y., Kim, H. K., Lefeber, A. W. M., Erkelens, C., Angelova, N., Choi, Y. H., et al. (2006). Application of two-dimensional nuclear magnetic resonance spectroscopy to quality control of ginseng commercial products. *Planta medica, 72*(4), 364-369, doi:DOI 10.1055/s-2005-916240.

Yerger, E. H., Grazzini, R. A., Hesk, D., Coxfoster, D. L., Craig, R., & Mumma, R. O. (1992). A Rapid Method for Isolating Glandular Trichomes. *Plant Physiology, 99*(1), 1-7.

Yuliana, N. D., Khatib, A., Choi, Y. H., & Verpoorte, R. (2011). Metabolomics for Bioactivity Assessment of Natural Products. *Phytotherapy Research, 25*(2), 157-169, doi:Doi 10.1002/Ptr.3258.

Zhang, P., Wen, P. W., Wan, S. B., Wang, W., Pan, Q. H., Zhan, J. C., et al. (2008). Molecular cloning of dihydroflavonol 4-reductase gene from grape berry and preparation of an anti-DFR polyclonal antibody. *Vitis, 47*(3), 141–145.

Zhang, X. Y., Wu, H. F., Liao, P. Q., Li, X. J., Ni, J. Z., & Pei, F. K. (2006). NMR-based metabonomic study on the subacute toxicity of aristolochic acid in rats. *Food and Chemical Toxicology, 44*(7), 1006-1014, doi:DOI 10.1016/j.fct.2005.12.004.

Zhao, F. J., Hawkesford, M. J., Warrilow, A., G.,S., McGrath, S. P., & Clarkson, D. T. (1996). Responses of two wheat varieties to sulphur addition and diagnosis of sulphur deficiency. *Plant and Soil, 181*(317-327).

Zhao, J., Yu, Q. T., Li, P., Zhou, P., Zhang, Y. J., & Wang, W. (2008). Determination of nine active components in Radix Hedysari and Radix Astragali using capillary HPLC with diode array detection and MS detection. *Journal of Separation Science, 31*(2), 255-261, doi:DOI 10.1002/jssc.200700379.

Zhi, H. J., Qin, X. M., Sun, H. F., Zhang, L. Z., Guo, X. Q., & Li, Z. Y. (2012). Metabolic Fingerprinting of Tussilago farfara L. Using 1H-NMR Spectroscopy and Multivariate Data Analysis. *Phytochemical Analysis, 23*(5), 492-501, doi:Doi 10.1002/Pca.2346.

Zhou, Y., Lu, L., Li, Z., Gao, X., Tian, J., Zhang, L., et al. (2011). Antidepressant-like effects of the fractions of Xiaoyaosan on rat model of chronic unpredictable mild stress. *Journal of Ethnopharmacology*, doi:10.1016/j.jep.2011.05.016.

Zuardi, A. W. (2006). History of cannabis as a medicine: a review. *Revista Brasileira De Psiquiatria, 28*(2), 153-157.

Curriculum Vitae

Personal data

Name	Nizar Happyana
Born	1st January, 1984
Nationality	Indonesian
Family status	Married (1 child)

Education

Since 2011	Doctoral thesis at the Laboratory of Technical Biochemistry, TU Dortmund, under the supervision of Prof. Dr. Oliver Kayser. Dortmund, Germany
2009-2011	Doctoral thesis at the Laboratory of Pharmaceutical Biology, Groningen University, under the supervision of Prof. Dr. Oliver Kayser. Groningen, the Netherlands
2006-2007	Master in Organic Chemestry (M.Sc.) at the Laboratory of Natural Product Chemistry, Bandung Institute of Technology, under the supervision of Prof. Dr. Euis Holisotan Hakim. Bandung, Indonesia
2002-2005	Bachelor in Chemistry (B.Sc.) at the Laboratory of Natural Product Chemistry, Bandung Institute of Technology, under the supervision of Prof. Dr. Euis Holisotan Hakim. Bandung, Indonesia

Work

Since 2010	Lecturer at Department of Chemistry, Bandung Institute of Technology Bandung, Indonesia

Publications

Peer Reviewed Articles

Happyana, N., Agnolet, S., Muntendam, R.,Van Dam, A., Schneider, B., Kayser, O., Analysis ofbcannabinoids in laser microdissected trichomes of medicinnal Cannabis sativa using LCMS andbcryogenic NMR, Phytochemistry, 2012, 87: 51-59.

Muntendam, R., **Happyana, N.**, Erkelens, T., Bruining, F. & Kayser, O. Time dependentbmetabolomics and transcriptional analysis of cannabinoid biosynthesis in Cannabis sativa var. Bedrobionol and Bediol grown under standardized condition and with genetic homogeneity, Online International J Med Plants Res. 2012,1(2): 31-40.

Book Chapter

Happyana, N., Muntendam, R., Kayser, O., Metabolomics as bioanalytical tool for characterization of medicinal plants and their phytomedical preparations, in: Pharmaceutical Biotechnology 2nd ed. (eds. Kayser, O., Warzecha, H.), 2012, pp. 527-552.

Oral Presentation

Happyana, N., Kayser, O., Monitoring Metabolites Production and Cannabinoids Analysis invMedicinal Cannabis Trichomes during Flowering Period by 1H NMR-Based Metabolomics, 61st International Congress and Annual Meeting of the Society for Medicinal Plant and Natural Product Research (GA), 1st – 5th of September, 2013, Münster.

Proceeding article

Happyana, N., Kayser, O., 1H NMR-Based metabolomics differentiation and real time PCR Analysis of medicinal Cannabis organs, International Horticulture Congress, 18th-22nd of August, 2014, Brisbane, Australia, in press.

Poster Presentation

Happyana, N., Agnolet, S., Muntendam, R., Van Dam, A., Schneider, B., Kayser, O., Cannabinoids Analysis of Laser-Microdissected Trichomes of Cannabis sativa L. by LC-MS and Cryogenic NMR, International Conference for natural Product Research (ICNPR), July 28th-August 1st, 2012, New York, United State of America.

Happyana, N., Muntendam, R., Van Dam, A., Kayser, O., Metabolic Profiling in the Specific cells of Capitate-Stalked Trichomes of Medicinal Cannabis, New Biotrend to Smarter Drugs, 8^{th}-9^{th} of December, 2011, Dortmund, Germany.

Pamplaniyil K., **Happyana, N.**, Kayser, O., Identification, isolation and functional characterization of terpene transferases in Cannabis sativa, New Biotrend to Smarter Drugs, 8^{th}-9^{th} of December, 2011, Dortmund, Germany.

Summary

Cannabis sativa L. is a medicinal plant which has a huge potency to treat a wide range of diseases. Many patients medically consume *Cannabis* for relieving suffering from various diseases, including multiple sclerosis, cancer, and HIV/AIDS. Cannabinoids, with THC as the leading compound, are valuable compounds that are mostly responsible for the therapeutic properties of this plant. However, the biosynthetic pathways of cannabinoids in particular and secondary metabolites in general, in this plant are still not fully understood, even though its complete genome sequencing has been published.

Cannabis trichomes have a very important role for the plant, which is as the main site for synthesizing and storing cannabinoids and other secondary metabolites. This makes the trichomes become an ideal model for studying the biosynthesis of secondary metabolites of *Cannabis*. Therefore, in this thesis we focused our works on this organ. Metabolomics, proteomics, and transcriptomics were applied in this research as analytical approaches in order to study secondary metabolites, proteins, and genes that are synthesized in the trichomes. Furthermore, the function of specific trichomes and their individual parts in the cannabinoid biosynthesis was investigated.

^{1}H NMR-based metabolomics differentiation together with RT-PCR analysis was performed in order to study metabolite profiles and cannabinoid biosynthesis in the different *Cannabis* organs (trichomes, flowers, and leaves). This combination method provided more information of cannabinoids study in *Cannabis* with revealing expression levels of cannabinoid genes, metabolite profiles, and discriminant compounds in the different organs. Furthermore, ^{1}H NMR-based metabolomics has been successfully applied for monitoring the production of metabolites, especially cannabinoids in the *Cannabis* trichomes during the last weeks of flowering period. Moreover, this method is simple, reproducible and can be coupled with ^{1}H NMR quantification.

Proteomic analysis of *Cannabis* trichomes was conducted using nano-LC-MS/MS method. As the results, house-keeping proteins participating in primary metabolism, energy production, protein synthesis, and other metabolism have been successfully identified in the trichomes. Besides that, identification of a set proteins involved in the disease/defense responses supported the functions of trichomes in protecting the plant against biotic and abiotic threats. Furthermore, many enzymes corresponding in the biosynthesis of secondary metabolites, including cannabinoids, flavonoids, and terpenoids were recorded. This finding confirmed the

function of *Cannabis* trichomes as the main site of secondary metabolite production. Although there is no flavonoid reported from the trichomes, however the identification of enzymes related to its biosynthesis indicated that this compound might be present in this organ.

Interestingly, identification of enzymes involved in the biosynthesis of cannabinoids, terpenoids, and flavonoids in the proteomic work has been also confirmed by the detection of their putative transcripts in the cDNA library of *Cannabis* trichomes. This cDNA library was constructed using 454 GS FLX pyrosequencing system and analyzed with SAMS for GO annotation and KEGG pathway assignments. Furthermore, putative transcripts of enzymes participating in the biosynthetic pathways of alkaloids were identified in the cDNA library as well. This finding probably might be a signature for the existence of alkaloids in the *Cannabis* trichomes. All of these have shown that cDNA library of *Cannabis* trichomes have been provided successfully valuable information for elucidating the biosynthetic pathways of secondary metabolites in this plant.

In this research, laser microdissection (LMD) have been used for collecting specific cells of the glandular hairs as a secreting plant organ, particularly in the intact capitate-sessile and capitate-stalked trichomes as well as in the heads and stems of the latter. Cannabinoids in the microdissected samples were analyzed by means of LCMS and cryogenic NMR. The quantitative analysis revealed that the concentration of cannabinoids in the capitate-stalked trichomes is higher than that in the capitate-sessile trichomes. Interestingly, cannabinoids were detected not only in the head of capitate-stalked trichomes but also in its stem part. This finding suggest that cannabinoid biosynthesis is not only limited to the expected head cells, but also the stems of *Cannabis* capitate-stalked trichomes might play a role in cannabinoids production.

Rangkuman

Cannabis sativa L. merupakan tanaman obat yang memiliki potensi besar untuk mengobati berbagai macam penyakit. Sekarang ini banyak pasien mengkonsumsi tanaman ini secara medis untuk meringankan rasa sakit dari berbagai penyakit yang dideritanya, seperti multipel sklerosis, kanker, dan HIV/AIDS. Cannabinoid, dengan THC sebagai senyawa yang paling berkhasiat, merupakan senyawa-senyawa yang biasanya bertanggung jawab untuk sifat-sifat terapeutik tanaman ini. Namun jalur-jalur biosintesis metabolit sekunder umumnya dan cannabinoid khususnya, pada tanaman ini masih belum sepenuhnya dipahami, meskipun sekuensing genom lengkap *Cannabis* sudah dipublikasikan.

Trikoma memiliki peranan yang sangat penting bagi tanaman ini, yaitu sebagai tempat utama untuk mensintesis dan menyimpan cannabinoid dan metabolit sekunder lainnya. Hal ini membuat trikoma menjadi model yang ideal untuk mempelajari biosintesis metabolit sekunder pada *Cannabis*. Sehingga pada disertasi ini, kami menitikberatkan penelitiannya pada trikoma *Cannabis*. Pada penelitian ini, metabolomik, proteomik, dan transkriptomik telah digunakan sebagai metode untuk mempelajari metabolit sekunder, protein, dan gen yang disintesis di organ ini. Selain itu, kami juga meneliti fungsi setiap jenis trikoma *Cannabis* dan bagian-bagian tertentunya di dalam biosintesis cannabinoid.

Metabolomik berbasis ^{1}H NMR bersama dengan metode RT-PCR telah digunakan untuk mempelajari profil metabolit dan biosintesis cannabinoid pada trikoma, bunga, dan daun *Cannabis*. Gabungan metode ini berhasil menyediakan informasi lebih lengkap dalam mempelajari biosintesis cannabinoids pada tanaman ini. Kombinasi keduanya dapat mengungkap tingkat ekspresi gen cannabinoid, profil metabolit, dan senyawa-senyawa yang membedakan organ-organ *Cannabis*. Metabolomik berbasis ^{1}H NMR telah berhasil digunakan juga untuk mengamati produksi metabolit, terutama cannabinoid, di dalam trikoma *Cannabis* pada minggu-minggu terakhir periode berbunga.

Analisis proteomik trikoma *Cannabis* telah dilakukankan dengan menggunakan metode nano-LC-MS/MS. Analisis ini berhasil mengidentifikasi protein-protein yang berpartisipasi dalam metabolisme primer, produksi energi, sintesis protein, respon terhadap penyakit/pertahanan, dan metabolisme lainnya. Pengidentifikasian protein-protein yang terlibat dalam respon terhadap penyakit/pertahanan mendukung fungsi trikoma dalam melindungi tanaman terhadap ancaman biotik dan abiotik. Selain itu, banyak enzim yang terlibat dalam biosintesis metabolit sekunder, termasuk cannabinoids, flavonoid, dan terpenoid berhasil diidentifikasi. Hasil ini

mengkonfirmasi fungsi trikoma *Cannabis* sebagai tempat utama untuk memproduksi metabolit sekunder. Meskipun tidak ada flavonoid yang telah dilaporkan dari trikoma, namun pengidentifikasian enzim-enzim yang berkaitan dengan jalur biosintesis senyawa ini mengindikasikan senyawa ini mungkin disintesis juga di dalam organ ini.

Menariknya, transkrip putatif dari enzim-enzim yang terlibat dalam biosintesis cannabinoids, terpenoid, dan flavonoid telah berhasil diidentifikasi di dalam *cDNA library* trikoma *Cannabis*. Oleh karena itu, hal ini mengkonfirmasi keberadaan enzim-enzim tersebut di dalam trikoma. *cDNA library* ini dibuat dengan menggunakan *454 pyrosequencing GS FLX system* dan transkrip-transkrip uniknya dianalisis dengan SAMS. Selain itu, transkrip putatif dari enzim-enzim yang terlibat dalam biosintesis alkaloid berhasil dideteksi, sehingga ini mungkin dapat dijadikan pertanda keberadaan alkaloid di dalam trikoma *Cannabis*. Dengan demikian, semua ini memperlihatkan bahwa *cDNA library* berhasil menyediakan informasi berharga untuk mempelajari jalur-jalur biosintesis metabolit sekunder pada *Cannabis*.

Pada penelitian ini, laser *microdissection* telah digunakan untuk mengisolasi trikoma *capitate-sessile*, trikoma *capitate-stalked*, serta bagian batang dan kepala trikoma *capitate-stalked*. Cannabinoid di dalam sampel-sampel trikoma tersebut dianalisis dengan LCMS dan *cryogenic* NMR. Analisis kuantitatif menunjukkan bahwa konsentrasi cannabinoid di trikoma *capitate-stalked* lebih tinggi dibandingkan di trikoma *capitate-sessile*. Menariknya, cannabinoids terdeteksi bukan hanya di bagian kepala trikoma *capitate-stalked*, tetapi juga di bagian batangnya. Hasil ini menunjukkan bahwa biosintesis cannabinoid bukan hanya terbatas pada sel-sel bagian kepala trikoma *capitate-stalked* seperti yang telah dilaporkan. Selain itu, hal ini juga memperlihatkan bahwa bagian batang dari trikoma *capitate-stalked* mungkin memiliki peranan dalam biosintesis cannabinoid.