POLITECNICO
MILANO 1863

School of Industrial and Information Engineering

Master of Science in Materials Engineering and Nanotechnology

Department of Chemistry, Materials and Chemical Engineering Giulio Natta

Raman Spectroscopy of *Cannabis Sativa* trichomes and selected Cannabinoids

Elisa Crocioni
N. matricola: 939928

Supervisor: Prof. Matteo Maria Saverio Tommasini

Co-supervisor: Prof. Nisha Rani Agarwal

Academic Year 2020-2021

Acknowledgements

Ringraziamenti

Innanzitutto, vorrei ringraziare il mio relatore, Professore Matteo Maria Saverio Tommasini, per avermi proposto questo progetto di tesi e per tutte le conoscenze che ha condiviso con me durante il suo sviluppo. Lo vorrei ringraziare per il materiale ed I consigli che ha condiviso con me durante lo sviluppo di questo progetto.

Un grande ringraziamento va alla Professoressa Nisha Rani Agarwal per avermi coinvolto nel suo Progetto e per avermi aiutato nella sua realizzazione. Non è stato possibile recarmi ad Oshawa data la situazione sanitaria mondiale, tuttavia è stata un'esperienza appagante e vorrei ringraziarla per la sua disponibilità e pazienza.

Un sincero grazie va alla mia famiglia per avermi supportato durante questi anni e per avermi sempre aiutato quando ne avevo bisogno. Vi sarò sempre molto grata per tutte le opportunità ed esperienze che mi hanno reso quella che sono oggi.

Un ringraziamento va ad Alberto che mi ha supportato quest'ultimo anno e che mi spinge ad essere la miglior versione di me stessa.

Un ringraziamento speciale va ai miei coinquilini di questi cinque anni. Mi avete aiutato a non sentirmi mai sola e non avrei potuto sperare di incontrare persone migliori con cui crescere e condividere fallimenti e successi.

Vorrei ringraziare tutti i miei colleghi dell'università. Grazie per aver condiviso con me questa esperienza e per avermi aiutato e dato consigli universitari e di vita.

Ultimi, ma non per importanza, vorrei ringraziare i miei amici di Roma. Anche se passiamo tanto tempo distanti, quando ci vediamo è come se non fossi mai andata via.

Abstract

Cannabis Sativa (hemp) and its species are an annual crop diffused all over the world and one of the oldest economic plants which provided humans with fibers for textiles, seeds for animal feed and an aromatic resin which has been used for recreational and medical purposed for more than 5000 years. Its several beneficial properties, such as the psychotropic, anti-inflammatory and analgesic effects, are due to the presence of specific chemical compounds called cannabinoids. THC, CBD, CBG, THCA, CBGA and CBDA are the most common cannabinoids, and the most interesting for pharmaceutical applications. Such cannabinoids can be found in secretion vesicles, called trichomes.

My thesis work aims to verify the possibility to use micro-Raman spectroscopy to obtain detailed chemical information on the distribution of cannabinoids in a trichome sample. Raman spectroscopy is a convenient, fast, and non-destructive technology. By specific and molecular structure-dependent vibrational fingerprints, it may provide precise information about the chemical composition of a selected sample. Therefore, it is interesting to apply Raman spectroscopy to the analysis of the *Cannabis* trichomes. Through the principal component analysis (PCA) of the micro-Raman mapping experiment on a trichome, it has been possible to extract both chemical and spatial information. This approach could find a wide interest in the administration of medical *Cannabis*, diagnostics, law-enforcement and for the identification of the optimal maturation of the *Cannabis* plant, so to select the best harvesting time.

Sommario

La *Cannabis* Sativa (canapa) e le sue specie sono un raccolto annuale diffuso in tutto il mondo e una delle piante più antiche utilizzate per la produzione di fibre per tessuti, semi per foraggio e una resina aromatica utilizzata per scopi ricreativi e medici da più di 5000 anni. Le sue proprietà vantaggiose, tra cui quella psicotropa, antinfiammatoria ed analgesica, sono dovute alla presenza di specifici composti chimici chiamati cannabinoidi. THC, CBD, CBG, THCA, CBGA e CBDA sono i cannabinoidi più comuni e i più interessanti per applicazioni farmaceutiche. Questi cannabinoidi sono contenuti all'interno di vescicole secretorie, chiamate tricomi.

Il mio progetto di tesi vuole dimostrare la possibilità di utilizzo della spettroscopia micro-Raman per ottenere informazioni chimiche dettagliate sulla distribuzione di cannabinoidi in un campione di tricoma. La spettroscopia Raman è una tecnica economica, veloce e non distruttiva. Da fingerprint specifici e dipendenti dalla struttura vibrazionale, può fornire informazioni precise sulla composizione chimica del campione selezionato. Dunque, è interessante applicare la spettroscopia Raman per analizzare i tricomi della *Cannabis*. Attraverso l'analisi delle componenti principali (PCA) dell'esperimento di mapping micro-Raman su un tricoma, è stato possibile estrarre sia informazioni chimiche che spaziali. Questo approccio potrebbe trovare ampio interesse nell'amministrazione medica della *Cannabis*, diagnostica, law-enforcement e per l'identificazione del momento di massima maturazione della pianta per selezionare il momento di raccolta.

Table of contents

List of figures

List of tables

correspondent normal mode are reported. The picture represents the correspondent normal mode. The color green represents the stretching of the bond, the blue the shrinking. _____ *120*

1. Introduction

This thesis project focuses on the Raman spectroscopy of *Cannabis* trichomes (the secretive vesicles present on the surface of the plant) and cannabinoids. *Cannabis Sativa,* one of the most abundant species of the *Cannabis* plant*,* grows in many regions of the world and has been used since 2900 B.C., at first for the production of textiles and ropes and then because of its pharmacological effects, both for medical and recreational purposes [1]. In 1800s *Cannabis* diffused in many countries all over the world but, within a century, the British and the US Pharmacopeia removed it from the listings. The reasons for this are the short shelf-life, the variation of the chemical composition of different plants, and therefore the different effects, and the introduction in the market of new drugs with more reliable effects. Moreover, the abuse and intoxication led to restrictions which prohibited the use of this plant in several countries [2]. The new scientific developments brought to an increase of interest on reviving *Cannabis* in the medical field. *Cannabis* has, in fact, several beneficial effects and the source of these effects can be identified in the *Cannabis* metabolites that accumulate in different regions of the plant. Amongst the most common metabolites, terpenes are known as the source of the odor typical of *Cannabis*, while cannabinoids have been studied because of their beneficial properties. The psychotropic effects of *Cannabis* are, in fact, due to the interaction of these chemical compounds with the receptors present in the brain. Cannabinoids have become an interesting topic of research for their anti-inflammatory and analgesic effects and have been recently used in the treatment of chronic pain, nausea and depression. The number of cannabinoid and non-cannabinoid compounds identified in *Cannabis* keeps increasing during the years. Today, more than 100 cannabinoids have been identified but the physiological properties are only known for some of them. The most relevant are THC, CBD, CBG and the correspondent cannabinoid acids (THCA, CBDA, CBGA). Such cannabinoids can be found in the resin produced by the secretive vesicles present on the surface of the plant, called trichomes [3].

Raman spectroscopy is a useful tool which can be used to get information on the chemical composition of a selected sample. Moreover, it is a non-destructive technique, which means that the sample is intact after testing. Raman spectroscopy can be used to analyze the trichomes of the *Cannabis* plant in detail, both from the point of view of its chemical composition and its spatial distribution. In fact, by analyzing the peaks of the Raman spectrum of trichomes as a function of the measurement point it is possible to identify the varying markers of specific cannabinoids present in the plant. These studies require the Raman spectra of each pure cannabinoid, especially those found in large quantities, in order to assign them to the experimental spectra obtained by analyzing the trichome sample. This approach is not conventional on samples with such a complex geometry as the trichomes and can be used to

detect the presence of specific cannabinoids and, eventually, it could be used to calculate their concentration in a sample.

The main goal of this thesis project is to employ a micro-Raman spectroscopy technique for the chemical characterization of cannabinoids in a sample of *Cannabis Sativa*. Besides being a non-destructive technique, it is fast, and the cheap analysis represent the main reasons for which this technique could be appealing compared to those currently used (*e.g.*, HPLC-MS). The use of Raman spectroscopy for the determination of the concentration of cannabinoids in a given sample could be exploited for the realization of sensors used to identify the moment in which the plant has reached its maximum maturation and therefore determine the best harvesting time. The development of analytic techniques for cannabinoids could also have large relevance in forensic applications and law enforcement.

This thesis work stems from an active collaboration with Professor Nisha Agarwal of the University of Ontario Institute of Technology (Oshawa, Canada). According to the new Canadian legislation on cannabinoids, in Canada it is possible to analyze samples with a concentration of THC (and other cannabinoids) higher than those which are considered legal in Italy. This thesis work also takes part in the recent exchange agreement between Politecnico di Milano and University of Ontario Institute of Technology.

This thesis project is divided into two sections. The first section has the goal to assign the peaks of the average spectrum obtained from the analysis of several points of a trichome sample, using as reference the peaks of pure cannabinoids reported in literature. The main reference paper for the Raman spectra of pure cannabinoids is called "*Raman-Based Differentiation of Hemp, Cannabidiol-Rich Hemp, and Cannabis*" by Lee Sanchez, David Batensperger, and Dmitry Kurouski, which has been published recently in the year 2020. This paper provides Raman spectra of some of the most abundant pure cannabinoids found in a trichome and detailed tables reporting the main peaks. By using principal component analysis, we could get spectroscopic information about the largest variations of the Raman signal across the trichome Raman mapping dataset and spatial information about the relative concentration of chemical species in the sample. This analysis was carried out both on background corrected data provided automatically by the Raman equipment software, and on the raw Raman data. By comparing the results, we could identify artifacts introduced by the software, which could make the spectra misleading or more complicated to read. Therefore, the spectra affected by artifacts were identified and removed, so to repeat the data analysis on a more reliable dataset. Finally, the analysis of the main principal components was done without removing the fluorescence background, which confirmed the validity of the previous analysis on the filtered dataset.

The second section is dedicated to the assignment of the main Raman peaks of selected cannabinoids (THC, THCA, CBD, CBDA, CBG, CBGA) to the normal modes determined by DFT calculations. The structure of each molecule was analyzed by using the software Avogadro and each Raman spectrum was simulated by using DFT. The main peaks of the simulated spectrum were compared with the main peaks of the experimental spectra, so that they could be assigned to the corresponding normal mode, whose animation was observed by using the software Molden [4]. This peaks assignment helps understanding the sensitivity of Raman spectroscopy to the molecular structure of different cannabinoids, which is important for

reliable sensing applications. Since similar molecules have similar spectra, it is necessary to determine to which normal mode the marker peaks are associated to, in order to possibly distinguish each cannabinoid. It is also important to understand how selective these peaks are compared to the chemical structure of each cannabinoid. In principle, we are interested in vibrational modes that are more delocalized and can be used in diagnostics to differentiate different chemical compounds. The goal of the following study is the one to give a more general overview on the meaning of the peaks and their relative intensity.

This thesis work represents the starting point for future studies and analysis on cannabinoids based on vibrational spectroscopy. *Cannabis* is a very complex plant, with important application perspectives, and exhaustive information on its chemical constituents is not yet available in the literature, which motivates the kind of spectroscopic investigations carried out here.

2. The pharmaceutical constituents of Cannabis

2.1 Cannabis Sativa and trichomes

Cannabis Sativa is a widely distributed plant, originated and firstly cultivated in Southern Asia and Egypt; it has been diffused all over the world because of its several interesting properties. Its use by humans goes back for 5000 years [1]. It is one of the oldest plants used as animal feeds, fibers for textiles and paper production, and aromatic resins containing compounds that have been widely studied for their recreational and medical purposes [1]. *Cannabis Sativa* has been considered for the cure of pain, nausea, depression and glaucoma besides for management of HIV symptoms and treatments for multiple sclerosis [5]. Natural and human selection have led to a wide variety of growth forms and chemical compositions, and new breeding techniques have been used for the production of cannabinoid rich species for medical use [6].

Figure 1. Example of medical Cannabis cultivars grown in the United Kingdom by GW Pharmaceuticals. The small inflorescence is a THC-rich cultivar containing only traces of CBD [2].

Cannabis is an annual crop plant grown from seeds when provided with open and sunny locations, light well-drained soil, water and nutrients. It can reach up to 5 m in height in a 4 to

8 months growing season, which goes from spring to autumn. *Cannabis* is dioecious since male and female flowers develop on separate plants. The gender of the plant is indistinguishable before flowering. The female plant produces flowers covered by resins glands (trichomes) that produce the psychoactive resin which contains specialized metabolites (cannabinoids and terpenes). Such metabolites have a wide pharmacological interest. While hemp crops, grown for seed or fibers production, contain both male and female plants, most medical *Cannabis* is produced with a different technique [7]. In fact, the metabolites are largely produced in the glandular trichomes of the female flowers while male flowers are typically not consumed for medical and recreational purposed because of the scarce number of glandular trichomes on their surface and consequently the lower number of cannabinoids and terpenes [8].

On the female plant flowers, three kinds of glandular trichomes have been identified: bulbous, sessile and stalked. They differ because of their morphology (figure 2).

1. Bulbous trichomes are the smallest and produce a limited selection of metabolites.

2. Sessile trichomes are made of a short stalk, which is bound to the epidermis, and a globose head. The head contains a multicellular disk of secretory cells, responsible for the metabolites production, and a subcuticular storage cavity, in which those metabolites are stored.

3. Stalked trichomes have a similar shape to the sessile trichomes. However, they are slightly larger, and the globose head is elevated hundreds of micrometers from the epidermis through a multicellular stalk.
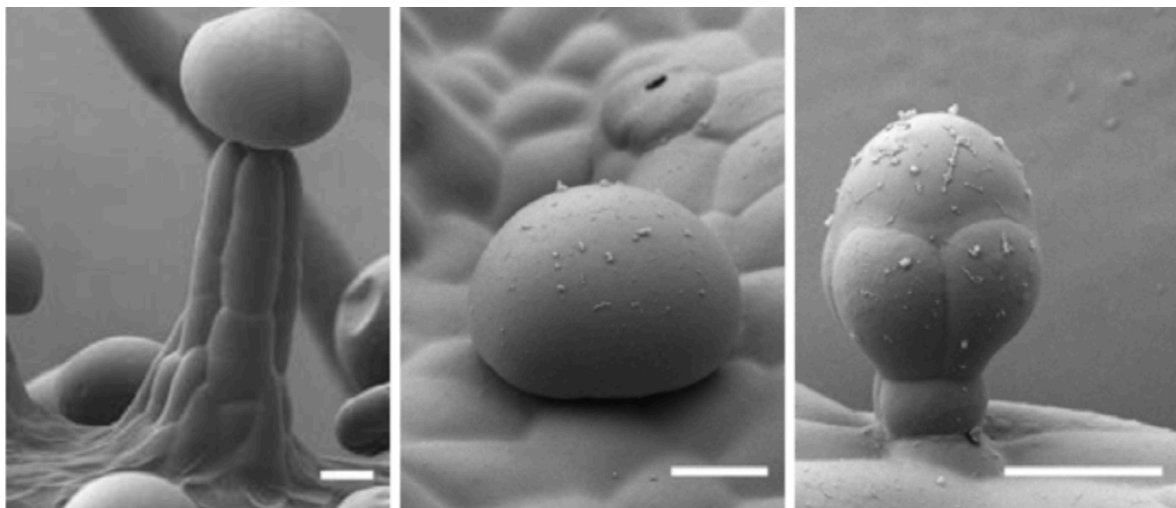
Figure 2. Cryo-SEM images of the three types of Cannabis glandular trichomes, classified as stalked (left), sessile (middle), and bulbous (right); scale bars 20 μm [4].

The anatomic and metabolic specialization of trichomes has been investigated for several species. In *Lamiaceae* and *Solanaceae* the glandular trichomes can be classified as peltate and capitate [9]. Peltate trichomes are similar to sessile ones since they are positioned on the organ surface. They have eight secretory disc cells which produce several metabolites, many of which are monoterpenes. Such metabolites accumulate in a subcuticular cavity. Capitate trichomes,

instead, are characterized by a small head, connected to the epidermis through a long stalk. They produce metabolites which are not stored in the subcultural cavity. In the same plant several kinds of capitated trichomes can be present and each one of them is responsible for the production of a specific metabolite, such as terpenes or acyl sugars [3].

The major trichomes that are found in *Cannabis* are the sessile and the stalked ones. Sessile and stalked trichomes are both characterized by a disk of secretory cells that develop an extracellular storage cavity. Their similarity has brought to some reports classifying these trichomes as one single group [10]. Another theory, instead, assumes that sessile trichomes represent a premature stage of stalked trichomes development [11]. The latter was verified by monitoring the concentration of stalked and sessile trichomes in the same plant over time [3]. Stalked trichomes were scarcely observed on the young plant, while the concentration of stalked ones was relevant in a more mature plant. The higher concentration of stalked trichomes on the mature plant indicates that the characteristics of these trichomes will influence and determine most the sensory and pharmacological properties of the Cannabis flower. The sensory properties are defined as taste, odor, irritation, texture and visual properties [12].
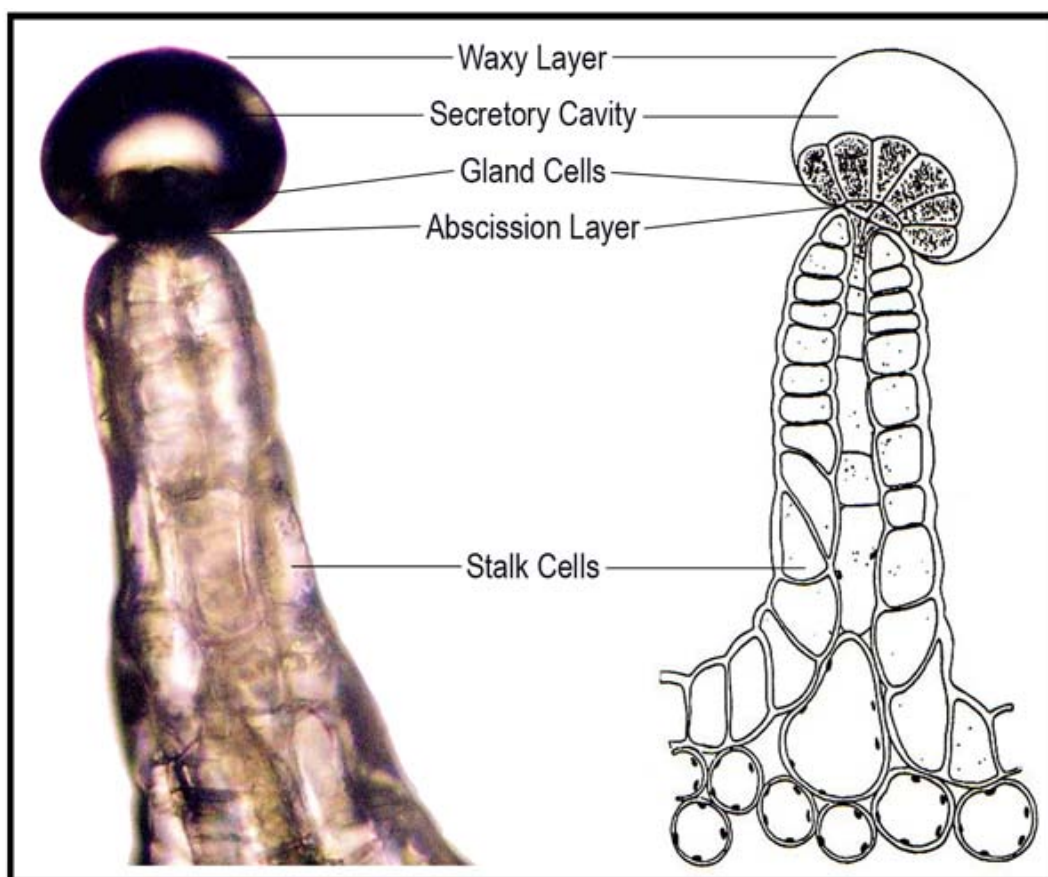
Figure 3. Microscope photograph and drawing of a Cannabis resin gland [2].

Specialization of various metabolites exists amongst different trichomes [3], but it has not been fully analyzed yet. Their relative contribution to the cannabinoids and terpenes profile is also unknown. However, it is known how the stalked trichomes produce a higher concentration of

cannabinoids when compared to the sessile ones. This may be due to the difference in the diameter of the head of the trichome.

The cluster of two dozens of glandular head cells produce a terpene rich resin with an high concentration of cannabinoids (more than 80%) and it collects such resin in vesicles covered by a thin membrane that surrounds the secretory head cell. The segregation of the resin protects the cannabinoids and terpenes from enzymatic and oxidative degradation. Below the vesicles lies an abscission layer that allows easy mechanical extraction of the resin (Figure 3)[3].

2.2 Non cannabinoid-type constituents

The number of chemical compounds isolated from *Cannabis Sativa* has been increasing in the last decades. 423 compounds were identified in 1980, today the number reaches 545 and hundreds of them are non-cannabinoid constituents, belonging to several different chemical classes.

Terpenoids are the primary aromatic constituents of *Cannabis*, responsible for the scent of its flowers and different flavour qualities. The interactions between cannabinoids and terpenes also determines the pharmaceutical properties attributed to different *Cannabis* strains and they enhance its therapeutic effect, especially in the case of aromatherapy [14]. They are synthesized starting from the isoprene units (C_5H_8). Isoprene (Figure 4) is the main unit of terpenes and it is characterized by 5 carbon atoms. Multiple isoprene units contribute to the formation of terpenes.
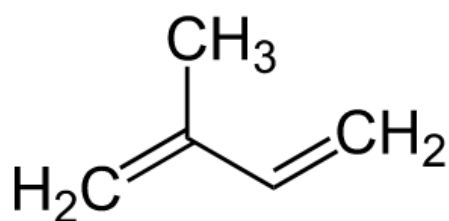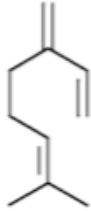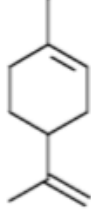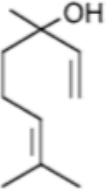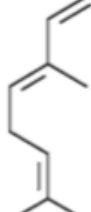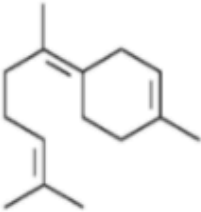
Figure 4. Chemical structure of isoprene [27].

Monoterpenes (C_{10} skeleton) and sesquiterpenes (C_{15} skeleton) are the most common species found in *Cannabis*. Hundreds of terpenes are found in *Cannabis*, but less than 40 have been identified. Table 1 contains some examples of the most common terpenoids, both monoterpenes and sesquiterpenes, found in *Cannabis Sativa*.

Compound	Class[a]	Structure
Myrcene	M	
Limonene	M	
Linalool	M	
trans-Ocimene	M	
delta-3-Carene	M	
trans-gamma-Bisabolene	S	
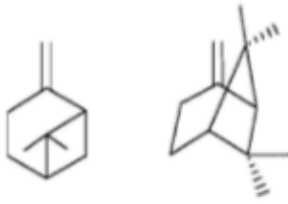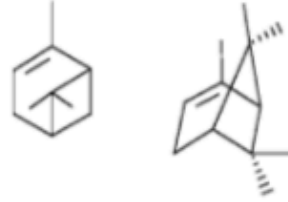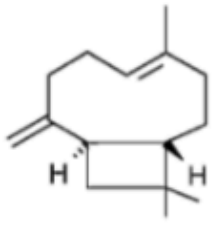
Compound	Class[a]	Structure
beta-Pinene	M	
alpha-Pinene	M	
beta-Caryophyllene	S	

Table 1. Names and chemical structures of some of the most common terpenoids found in Cannabis Sativa. M indicates monoterpenes and S sesquiterpenes [6].

Terpenes are produced trough biosynthetic pathways controlled by specific enzymes. It is possible to produce plants with several combinations of terpenes and with targeted medical use. The side chain substitution results in several different physiological responses. For example, some terpenoid compounds can stimulate the membrane of the pulmonary system, facilitating drug absorption. They are used in pulmonary medical products. This effect changes if together with terpenes also Cannabinoids are present [7]. Besides terpenes, other non-cannabinoid compounds, such as steroids, fatty acids, flavonoids, phenantrenes, spiroindans, nitrogenous compounds and xanthones have been identified as well but they are not as relevant for this thesis project [7].

2.3 Cannabinoids

104 cannabinoids have been isolated from *Cannabis Sativa*. The term cannabinoid refers to all the chemical substances isolated from C. Sativa which exhibit the typical C21 terpenophenolic skeleton, their derivative and transformation products. With the term phytocannabinoid we indicate the cannabinoids which derive directly from the plant. The 104 cannabinoids that have been identified in *C. Sativa* are classified in 11 subclasses: (-)-delta-9-trans-

tetrahydrocannabinol (Δ9-THC), (-)-delta-8-trans-tetrahydrocannabinol (Δ8-THC), cannabigerol (CBG), cannabichromene (CBC), cannabidiol (CBD), cannabinodiol (CBND), cannabielsoin (CBE), cannabicyclol (CBL), cannabinol (CBN), cannabitriol (CBT) and miscellaneous-type cannabinoids [6].

1. (-)-delta-9-trans-tetrahydrocannabinol (Δ^9-THC): nine THC-types of cannabinoids were identified with C_1 to C_5 side chains. The major precursor is THC acid A, while THC acid B is present in lower concentrations. The cannabinoid acid has no psychoactive properties. Through its decarboxylation the correspondent neutral cannabinoid is obtained (THC). The decarboxylation rate increases after harvesting, during storage or through heating and exposure to light. THC was first isolated in 1942, however, its structure was only assigned in 1964. It is known for its euphoriant, analgesic, anti-inflammatory, antioxidant and antiemetic pharmacological properties.
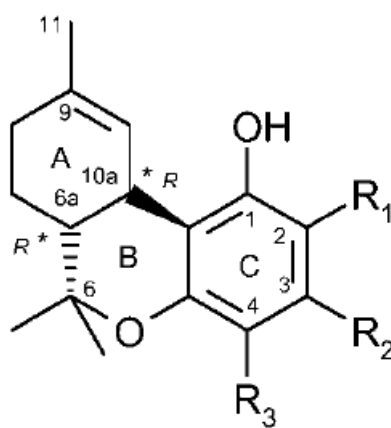
Figure 5. Δ9-THC general chemical structure. Δ9-THC: R1=H, R2=C5H11, R3=H; Δ9-THC acid A: R1=COOH, R2=C5H11, R3=H; Δ9-THC acid B: R1=H, R2=C5H11, R3=COOH [5].

2. (-)-delta-8-trans-tetrahydrocannabinol (Δ^8-THC): Δ^8-THC and its acid precursor are considered as artifacts of THC and THC acid and they are 20% less active.
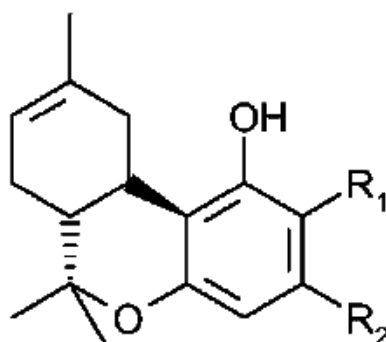
Figure 6. Δ8-THC general chemical structure. Δ8-THC: R1=H, R2=C5H11; Δ8-THC acid A: R1=COOH, R2=C5H11 [5].

3. cannabigerol (CBG): CBG was the first cannabinoid to be identified and the first one to be isolated from the *Cannabis* resin in pure form, while its cannabinoid acid (CBGA) was

demonstrated to be the first produced by the plant. CBG is produced trough decarboxylation of the correspondent cannabinoid acid.

Figure 7. CBG general chemical structure. CBG: R1=H, R2=C5H11, R3=H; CBGA: R1=COOH, R2=C5H11, R3=H [5].

4. cannabichromene (CBC): there are 5 types of CBC cannabinoids that have been identified so far.
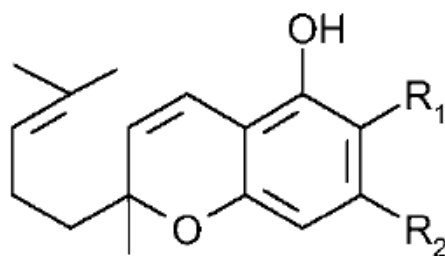
Figure 8. CBC chemical general structure. CBC: R1=H, R2=C5H11 [5].

5. cannabidiol (CBD): seven types of CBD-types cannabinoid were identified with C1 to C5 side chains. CBD and its corresponding cannabinoid acid (CBDA) are the most abundant in fiber-type *Cannabis* (hemp). CBD was first isolated in 1940, while CBDA, isolated in 1955, was the first discovered cannabinoid acid.
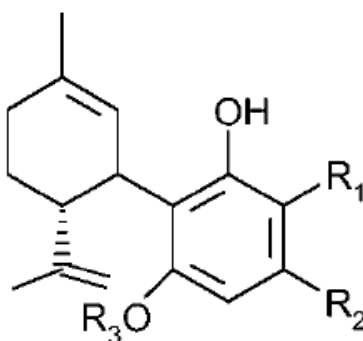
Figure 9. CBD general chemical structure. CBD: R1=H, R2=C5H11, R3=H; CBDA: R1=COOH, R2=C5H11, R3=H [5].

6. cannabinodiol (CBND): two CBND-type cannabinoids were discovered. They are considered to be artifacts of CBD.
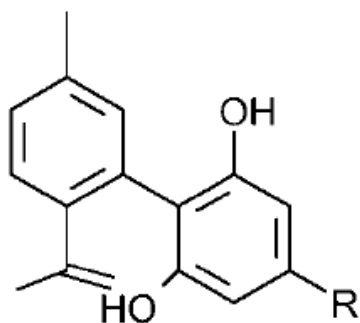
Figure 10. CBND general chemical structure. CBND: R=C3H7 [5].

7. cannabielsoin (CBE): There are five CBE-type cannabinoids, amongst which CBE and its acid precursors A and B. They are artifacts of CBD.
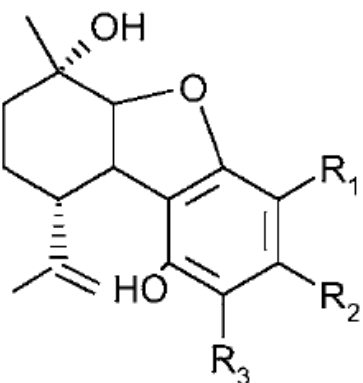
Figure 11. CBE general chemical structure. CBE: R1=H, R2=C5H11, R3=H; CBE acid A: R1=COOH, R2=C5H11, R3=H; CBE acid B: R1=H, R2=C5H11, R3=COOH [5].

8. cannabicyclol (CBL): known to be a heat generated artifact of CBC.
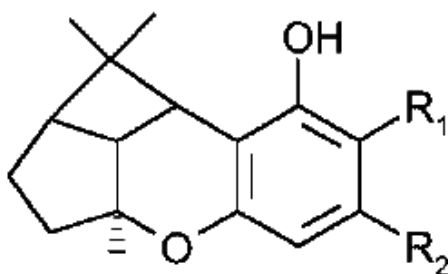
Figure 12. CBL general chemical structure. CBL: R1=H, R2=C5H11; CBLA: R1=COOH, R2=C5H11; CBLV: R1=H, R2=C3H7 [5].

9. cannabinol (CBN): six CBN-type cannabinoids were discovered. They are artifacts of THC and their structure was first identified in 1940.
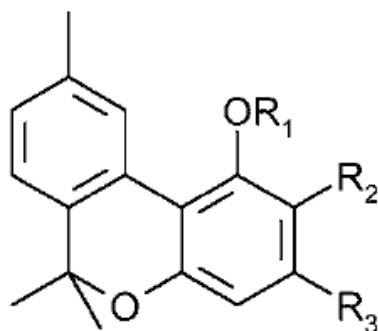
Figure 13. CBN general chemical structure. CBN: R1=H, R2=H, R3=C5H11 [5].

10. cannabitriol (CBT): there are nine CBT-type cannabinoids, characterized by an additional OH substitution. CBT itself exists in the form of both isomers and the racemate, whereas two isomers (9-a- and 9-b-hydroxy) of CBTV were identified. CBDA tetrahydrocannabitriol ester (ester at 9-hydroxy group) is the only reported ester of any naturally occurring cannabinoids.
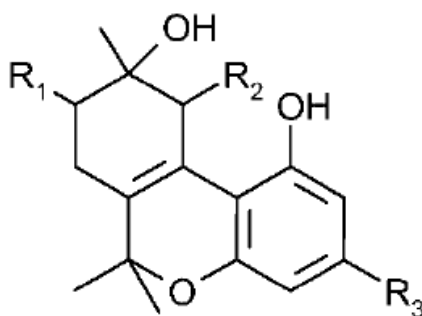
Figure 14. CBT general chemical structure. CBT: R1=H, R2=OH, R3=C5H11; CBTV: R1=H, R2=OH, R3=C3H7; CBTVE: R1=H, R2=OC2H5, R3=C3H7 [5].

11. Miscellaneous types: Eleven cannabinoids of various unusual structure were identified. They are characterized, for example, by a furan ring, carbonyl function or tetrahydroxy substitution [7].
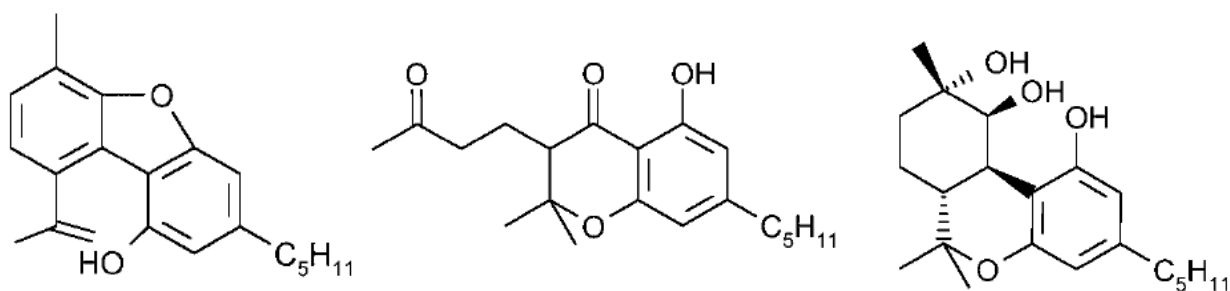
Figure 15. Dehydrocannabifuran, Cannabichromanon and Cannabiripsol chemical structures [5].

2.4 Accumulation of metabolites in trichomes

To analyze how metabolites are accumulated in the secretory cavity, Samuel J. Livingston and coworkers used multi photon imaging to analyze the internal structure of the differently shaped trichomes [3]. They developed a two-photon scanning fluorescence microscopy in which a red shifted pulsed laser penetrates the trichome's surface and activate the intrinsic fluorescence of the metabolites that follows a two-photon absorption process. The emitted light had a wavelength range of 420-530 nm, which was divided into two ranges:

1) 420-460 nm, false colored in teal.

2) 460-530 nm, false colored in red.

The superposition of the two fluorescence images allows to obtain high quality 3D images of sessile, stalked and bulbous trichomes. The fluorescence of such trichomes appeared to be different. Stalked trichomes show a large blue-shifted fluorescent region, false colored in teal, and some smaller round regions around it. Sessile trichomes show small droplets with red-shifted fluorescence, false colored in red. Finally, bulbous trichomes showed cells with blue-shifted fluorescence (teal) and a very small cavity with red-shifted fluorescence (red).

Livingstone et al. [3] concluded that the teal region represents points in the plant with the higher concentration of cannabinoids. This assumption is also supported by the fact that stalked trichomes, which show the highest blue shifted fluorescence, have a higher concentration of cannabinoids if compared to sessile and bulbous trichomes. This is due to the larger volume of the glandular head.

Transmission electron microscopy (TEM) was also used to further identify the arrangement and distribution of the metabolites in the storage cavity. TEM identified a large droplet in the storage cavity of stalked trichomes and many small droplets in the cavity of sessile trichomes. Therefore, by combining the results of the two imaging techniques it was possible to

distinguish the three kinds of trichomes not only for their shape but also for the spatial arrangement of their metabolites.
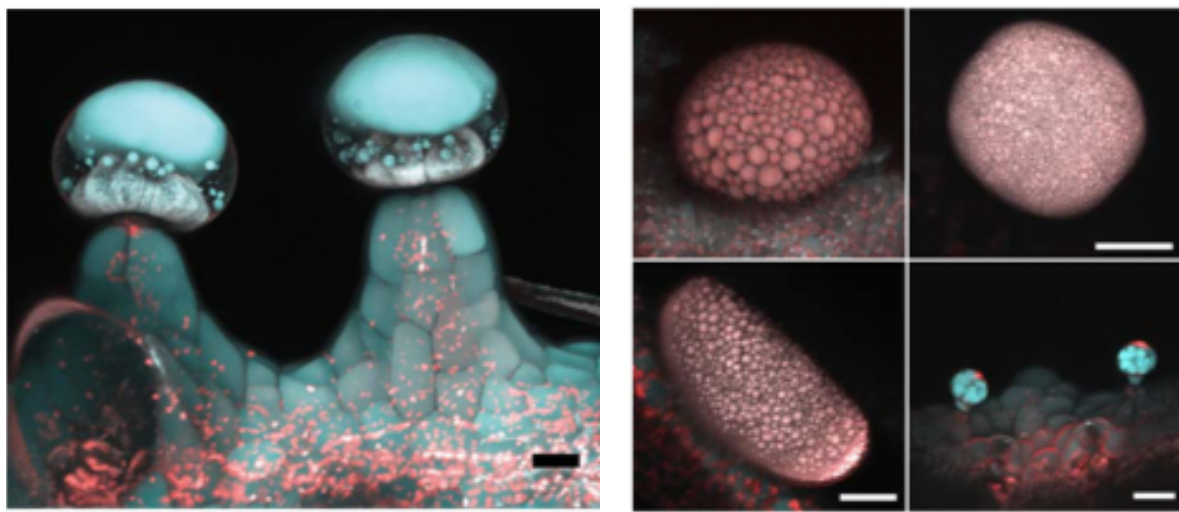
Figure 16. Glandular trichomes exhibit distinct intrinsic fluorescence and metabolite organizations.
The pictures are multi-photon microscopy images of the glandular trichomes on Cannabis plants, including female calyx stalked (on the left), calyx sessile, leaf sessile, male antherial sessile and calyx bulbous (on the right, from top to bottom and right to left). These pictures reveal distinct organization and intrinsic fluorescence of the metabolites stored in the storage cavity [3].

Another difference that was identified among the three kinds of trichomes is the number of secretory disk cells that are responsible for the secretion of metabolites. In bulbous trichomes the secretory cells could not be differentiated from pre-secretory trichomes. By using a staining method, it was instead possible to identify the number of secretory cells in sessile and stalked trichomes. Sessile trichomes contain exactly eight cells in the mature plant. Stalked trichomes are characterized by 12-16 cells. This larger range might be due to the larger dimension of the corresponding trichomes. These secretory cells have an average dimension of around 10 μm [7].
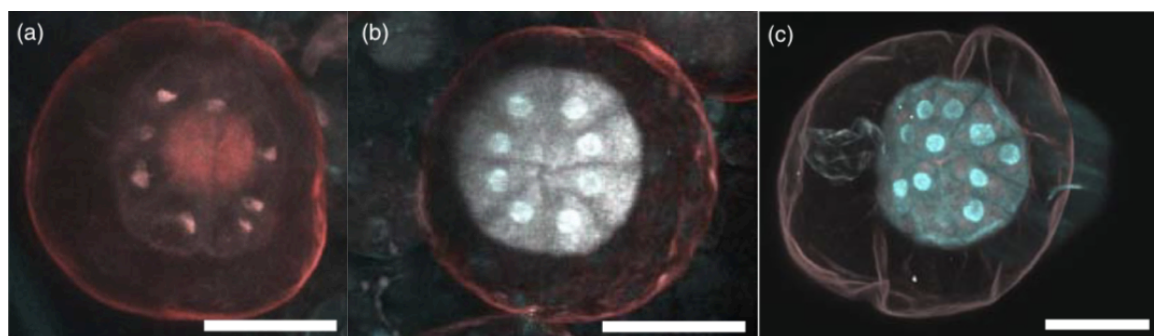
Figure 17. Stalked glandular trichomes on mature Cannabis flowers have a proliferation of cells in their secretory disc. The pictures are multi-photon microscopy images of disc cell for leaf sessile (a), calyx sessile (b) and calyx stalked (c) glandular trichomes

Two-photon scanning fluorescence microscopy was also used to verify the hypothesis that floral stalked trichomes develop from trichomes that appear to be sessile. Due to the different accumulation of metabolites in the three kinds of trichomes demonstrated with fluorescence, the expectation was that the intrinsic fluorescence and the droplets organization in the storage cavity would vary from red-shifted small droplets, which have been identified in sessile trichomes, to blue-shifted larger droplets, as seen in stalked trichomes. The intrinsic fluorescence properties were measured for both sessile and stalked trichomes to verify if there was a change in the size and fluorescence of the droplets between stalked and sessile trichomes. The analysis on the majority of sessile trichomes showed a fluorescence ratio which appeared to be really similar to the one of premature stalked trichomes. The calculation of the number of secretory cells was a further demonstration of the fact that the majority of the sessile trichomes are just a premature stage of the sessile ones. If sessile trichomes are premature stalked trichomes we would expect them to have a number of secretory cells higher than 8. If instead they are not, we would expect to see 8 secretory cells, number which characterizes the sessile trichomes of vegetative leaves which don't develop into stalked trichomes. The majority of the sessile cells appear to have more than 8 secretory disk cells, from which we can deduct that the majority of them will develop in stalked trichomes. However, a high number of premature trichomes showing 8 secretory cells was still present. Based on this analysis, Livingstone et al. [3] predicted that 86% of secretory trichomes with more than eight cells are premature stalked trichomes, while 89% of secretory trichomes with eight cells are sessile trichomes, like those of vegetative leaves. These data demonstrated that stalked glandular trichomes develop from premature stalked trichomes with greater than eight cells.

If the majority of sessile floral trichomes will develop in stalked trichomes, their biochemical composition should be more similar to the latter than to the sessile trichomes of vegetative leaves which don't develop in stalked trichomes. Sessile trichomes from leaves appeared to have a low concentration of monoterpenes if compared to sesquiterpenes. Similar results were obtained after analyzing sessile trichomes from anthers. Stalked trichomes, instead, appeared to be dominated by monoterpenes, contained in a percentage of 92 %, and a monoterpene sesquiterpene ratio of 12. Both trichomes types appeared to have a similar composition of cannabinoids. This further proved that sessile trichomes found on anthers and vegetative leaves, which do not develop in stalked trichomes, have a chemical profile which is completely different from the one of stalked trichomes [3].

The pictures obtained by Livingstone et al. [3] using cryo-SEM and TEM are of very high-quality. The reason why a micro-Raman mapping investigation of a trichome was carried out in this thesis work is not to obtain better images. Indeed, compared to the Raman signal, the fluorescence signal of different substances, which is represented in the pictures reported by Livingstone et al. [3], is less sensitive to the chemical composition. Hence, by analyzing the Raman spectra one hopes to obtain more detailed chemical information than that contained in the fluorescence spectral images. This is precisely the motivation at the basis of this thesis work.
This kind of analysis of the trichome was not attempted before with micro-Raman spectroscopy, and it is challenging because of the complex chemical composition of the sample, its intrinsic fluorescence, and the complicated geometry. Hence a proper data analysis approach is required to possibly extract the desired information from the Raman mapping.

3. Materials and methods

3.1 Experimental methods

To collect the data used for the principal component analysis of the micro-Raman mapping, a sample of *Cannabis Sativa* flower trichomes was analyzed. The sample was purchased from a *Cannabis* licensed distributor in Oshawa (ON, Canada) and was used as received. As described in Chapter 2.1, each flower of *C. Sativa*, especially those of the female plant, are covered in glandular trichomes. These are responsible for the secretion and production of metabolites. Among such metabolites, we are mainly interested in cannabinoids (because of their relevant applications) and we would like to assess their spatial distribution in the sample at the microscopic scale. For this reason, micro-Raman spectroscopy was used. The goal is to determine if Raman spectroscopy can be used to differentiate the vesicles in the trichomes (figure 18). As evidenced by Livingstone et al. [3] some regions of the trichome are characterized by a higher concentration of chemical species than others. Such regions can be identified as the secretory vesicles that are present, in different numbers, in the glandular head of trichomes.
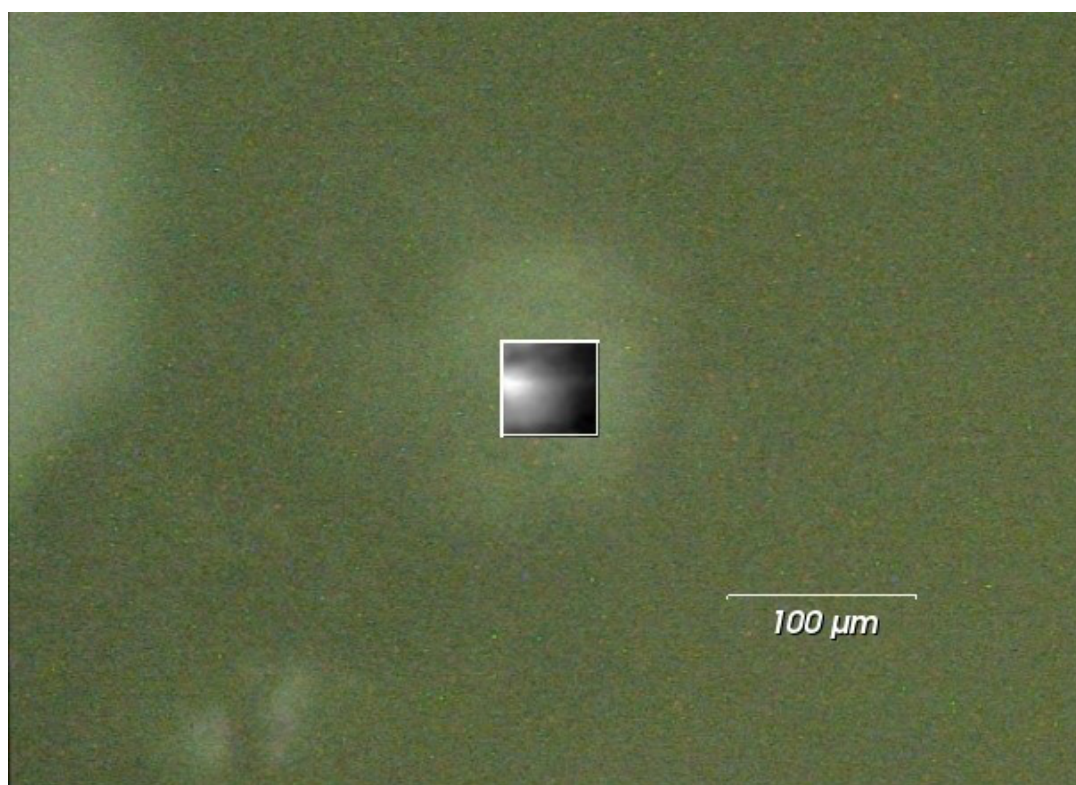
Figure 18. Bright field image of a trichome.

The experiments were carried out at the University of Ontario Institute of Technology (Oshawa, ON Canada), by using a Renishaw inVia Raman microscope. The inVia is a Raman spectrometer

coupled to a microscope. The microscope focuses the 532 nm laser source on specific areas of the sample, then it also collects the light backscattered from the surface of the sample and directs it through a Raman spectrometer. The 532 nm is a common wavelength used in Raman spectroscopy. Generally, the shorter the wavelength of the laser, the larger the Raman scattering cross section is, that also implies a higher signal in the measurements. However, at shorter wavelengths the background signal from fluorescence may increase as well. The light scattered by the sample is collected and transmitted to the spectrometer by fiber optic cables. A suitable filter is interposed to suppress the Rayleigh scattering. As a collateral effect of the edge filter used, also the anti-Stokes scattering is involuntarily filtered out. The Stokes scattered light, which is the only one which is not filtered out, passes through a dispersion element and it is then captured by a CCD detector, which provides a spectrum as a result of its reading process.

Figure 19. Picture of an inVia Raman spectrometer [32].

The area scanned during the Raman mapping of the trichome was spanning over a grid of 21x21 points, with 1 µm spacing. Each single point Raman experiment had a duration of 10 sec with 10 accumulations each.

Since Raman scattering, especially with such short wavelength, is competing with sizeable fluorescence background which can make the spectra difficult to read, the software included with the Raman spectrometer (Windows®-based Raman Environment, WiRE, from Renishaw) was used to automatically remove the fluorescence background. The software both controls the acquisition of Raman data and provides specific data processing and analysis options. Among them, WiRE can be used for the fluorescence removal. It is a completely automated software which removes the background noise and fluorescence consistently and repeatably. Removing the background fluorescence is important because it allows to have a flat spectrum which contains Raman information only. This allows clearer spectra which are simpler to analyze. However, as it will be evidenced in the data analysis Chapter, WiRE may also introduce artifacts that should be carefully taken into consideration to avoid misleading conclusions.

3.2 Raman Spectroscopy

Raman spectroscopy is a vibrational spectroscopy technique based on the inelastic scattering of light. It was first postulated by Smekal in 1923 but it was experimentally demonstrated only five years later by Krishnan and Raman, from whom the technique takes its name. Nowadays, it is a very popular spectroscopy tool used in several fields such as biology and medicine [15]. The original experiment was done by directing sunlight, using a telescope, into a sample (purified liquid or dust-free vapor). A second lens was positioned close to the sample, to detect the scattered radiations. A series of optical filters could determine the presence of scattered radiation with a different frequency from the one of the incident beam. This is the main feature of Raman spectroscopy.

When a photon interacts with the sample, it can be either absorbed, scattered or it can pass straight through it. If the energy of the incident photon corresponds to the energy gap between the ground state and the excited state of the molecule, the molecule can absorb the photon and it can be excited from the ground state to a higher energy excited state. This difference in energy can be measured with absorption spectroscopy, which determines the loss of intensity of the transmitted light beam as a function of the wavelength.

It is also possible for light to be scattered. In this case the energy of the photon does not have to match the difference in energy between the ground state and the excited state. For instance, the scattered photon can be observed by positioning a detector which collects light at an angle to the incident light beam. At room temperature, most of the molecules are present in the lowest energy vibrational level. When light interacts with the molecule, it distorts its electronic distribution, because the oscillating electromagnetic radiation induces a dipole due the molecule polarization. This forms a short-lived non-stationary state that is called the virtual state. This state is not stable; therefore, the photon is radiated again upon the de-excitation of the virtual state. Usually, a monochromatic laser is used in Raman spectroscopy as a light source [16]. One factor which limited the use of Raman spectroscopy when the technology was just created, was the very weak signal. Large intensities of monochromatic light are necessary to excite a detectable and intense signal. Therefore, monochromatic lasers are used [15]. When the laser beam interacts with the sample, it generates three different scattered components, with same (1), lower (2) or higher (3) frequency than the incident laser beam:

1. Rayleigh scattering: most of the photons are elastically scattered, so this process is the most intense. It does not involve any energy changes; therefore, the scattered photons will have the same wavelength, frequency and energy of the incident ones.

2. Stokes scattering: the Raman scattering process from the ground vibrational state m leads to absorption of energy by the molecule and promotion to a higher energy vibrational state n. It is characterized by an inelastic collision with the sample and the frequency of the scattered light is lower than the incident one.

3. Anti-Stokes scattering: due to thermal energy some molecules can be initially in an excited state n. Scattering from this state to the ground state m is called anti-Stokes scattering and it is associated to a photon energy increase of the scattered photon. It

is characterized by an inelastic collision with the sample and the frequency of the scattered light is higher than the incident beam [16].
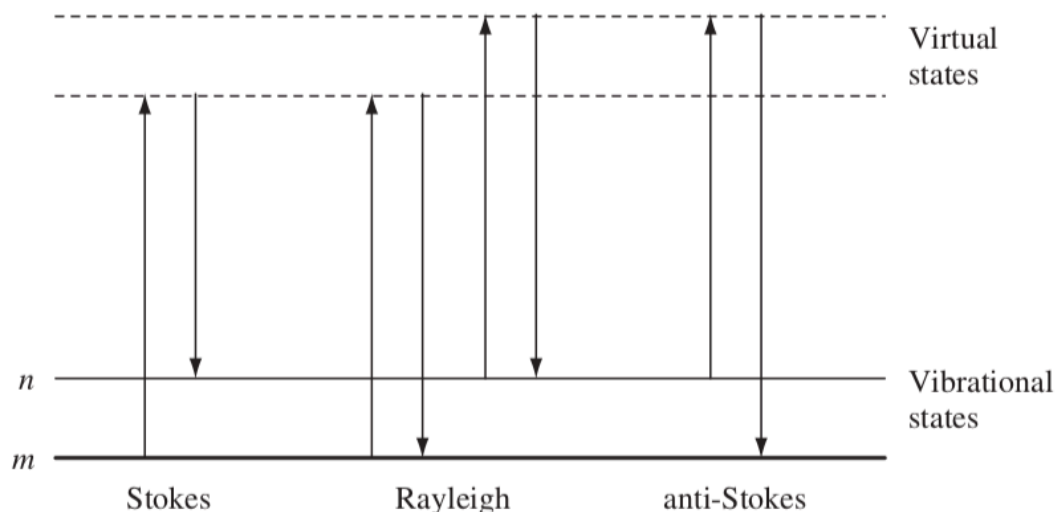
Figure 20. Energy levels scheme for a molecule. The ground vibrational state is indicated with the letter m, while n is the excited vibrational state [19].

The frequency shifts associated to the Stokes and anti-Stokes processes can be exploited to obtain information about the vibrational states of the sample and therefore determine its molecular composition with very high accuracy. Amongst the several applications of Raman spectroscopy in life sciences some examples are the quantification of biomolecules, and the hyperspectral molecular imaging of cells for medical diagnosis. Compared to other spectroscopic techniques, a large advantage of Raman spectroscopy is that it is non-destructive, and it does not require any sample preparation (usually). Therefore, Raman spectroscopy is ideal for applications that do not allow any alteration of the sample, like *in vivo* analysis [15].

When considering Raman scattering, it is possible to analyze its physics using two approaches: the classical and the quantum mechanical one. In the classical treatment, light is considered as an electromagnetic radiation, made by an oscillating electric field which can interact with the molecule through its time-dependent, oscillating, polarizability. Polarizability is defined as the ability of the electronic cloud of the molecule to deform because of the interaction with the electric field. In the quantum mechanical interpretation, instead, light is considered as a photon which interacts with the molecule and it is inelastically scattered.

It is convenient to start with the classical interpretation. A diatomic molecule can be considered as two masses (m_1 and m_2) connected by a spring, which has elastic constant K. m represent the atomic mass, whereas K is the bond stiffness.
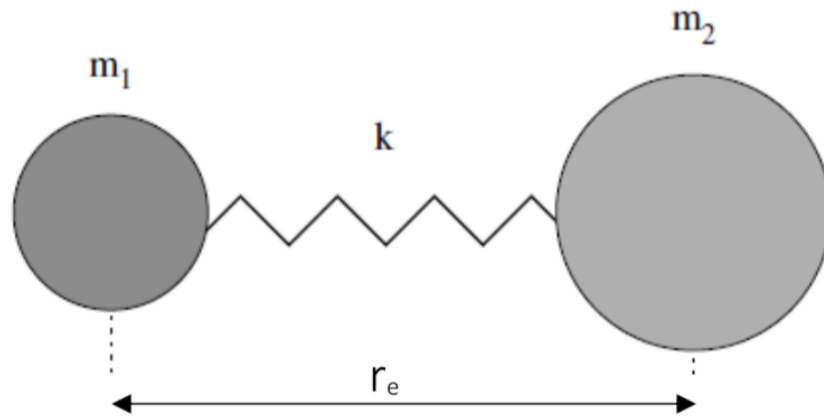
Figure 21. Model of a diatomic molecule. The atoms of masses m1 and m2, the chemical bond is represented by a spring with elastic constant k and equilibrium internuclear distance r_e [7].

Using this approach and the Hooke's law, the molecule's displacement can be calculated, and the associated equation of motion can be written as follows:

$$\frac{m_1 m_2}{m_1 + m_2}\left(\frac{d^2 x_1}{dt^2} + \frac{d^2 x_2}{dt^2}\right) = -K(x_1 + x_2)$$

By introducing the reduced mass ($\mu = m_1 m_2/(m_1+m_2)$) and the vibrational displacement $q = x_2-x_1$, associated to the bond stretching, it is possible to simplify the equation.

$$\mu \frac{d^2 q}{dt^2} = -Kq$$

Solving it for q we obtain:

$$q = q_0 \cos(2\pi v_m t)$$

v_m is the frequency of vibration of the molecule, defined as:

$$v_m = \frac{1}{2\pi}\sqrt{\frac{K}{\mu}}$$

From these results one can notice that the molecule vibrates by following a cosinusoidal pattern with a frequency proportional to the bond strength and inversely proportional to the reduced mass. Each molecule will have a characteristic vibrational frequency that depends both on the constituent atoms and the strength of the bond. By using Raman spectroscopy, it is possible to measure the vibrational frequencies from the frequency shift between the incident photon and the inelastically scattered photon.

As described above, when the laser beam interacts with the sample, it induces a dipole moment ($\Delta\mu$) given by the product between the polarizability (α), defined as the tendency of a molecule to be polarized, and the electric field I.

36

$$\Delta\mu = \alpha E$$

for a monochromatic laser of frequency ω_L, its electric field can be described as

$$E = E_0 \cos(\omega_L t)$$

By defining q as a generic vibrational coordinate:

$$q = q_0 \cos(\omega m t)$$

and by expanding the polarizability as a Taylor series in the q variable, one can obtain the following result:

$$\Delta\mu = \alpha_0 E_0 \cos(\omega_L t) + \frac{1}{2}\frac{\partial\alpha}{\partial q} q_0 E_0 [\cos(\omega_L - \omega_m)t + \cos(\omega_L + \omega_m)t]$$

From this expression one can see how the dipole moment oscillates with different frequencies, that are associated to the Rayleigh (ω_L), the Raman Stokes ($\omega_L - \omega_m$) and the Raman anti-Stokes ($\omega_L + \omega_m$) scattering processes. This means that the emitted radiation depends both on the frequency of the incident radiation (ω_L) and on the frequency of the molecular vibration (ω_m).

The intensity of the electromagnetic radiation is proportional to the square modulus of the electric field E. It is then possible to evaluate the Stokes and anti-Stokes Raman intensity:

$$I_{Stokes} \div (\omega_L - \omega_m)^4 \left(\frac{\partial\alpha}{\partial q}\right)^2 I_0(\omega_L)$$

$$I_{anti-Stokes} \div (\omega_L + \omega_m)^4 \left(\frac{\partial\alpha}{\partial q}\right)^2 I_0(\omega_L)$$

From such equations it is possible to determine the necessary condition for Raman scattering. To be Raman active, a displacement along a vibrational mode must produce a change in the molecular polarizability:

$$\left|\frac{\partial\alpha}{\partial q_i}\right| \neq 0$$

For non-linear molecules with N atoms, the number of vibrational modes is given by:

$$3N - 6$$

3N is the total degrees of freedom of the molecule. The reason why 6 is subtracted is because it considers the 3 translational and 3 rotational degrees of freedom. For linear molecules, there is one less rotational degree of freedom (the one around the molecular axis) and the number of vibrational modes is given by $3N - 5$.

The quantum mechanical approach to Raman spectroscopy provides further details and allows understanding the reason of the different intensity of Stokes and Anti-Stokes scattering. The

Raman effect occurs by the inelastic scattering of a photon when it interacts with a vibrating molecular bond. The Jablonski diagram can be used to describe how the photon excites the molecules in the virtual excited state:
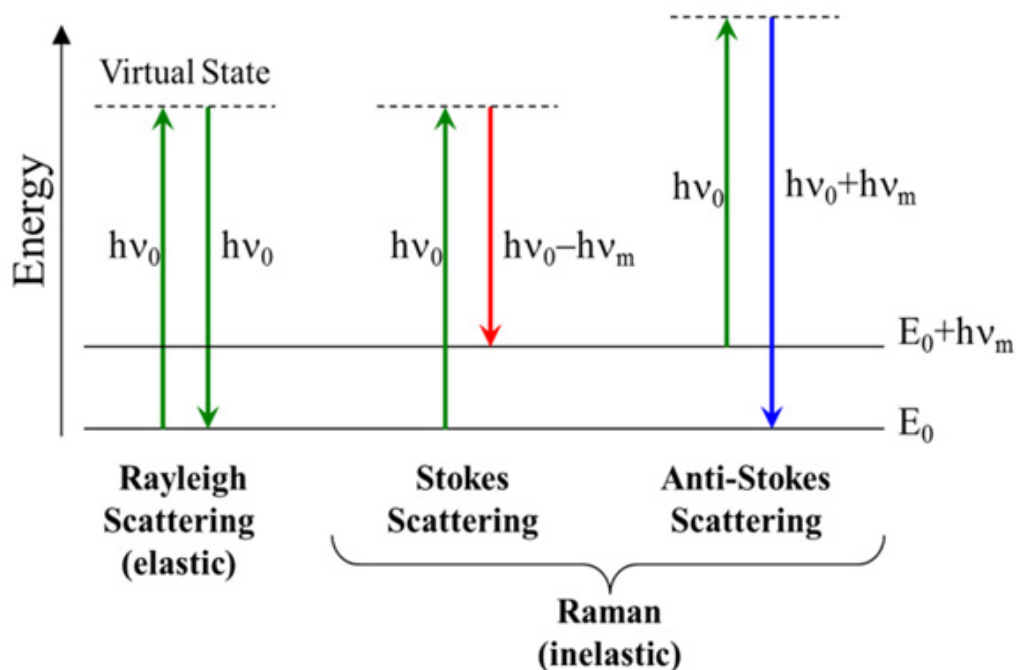
Figure 22. Representation of the Rayleigh, Stokes and anti-Stokes scattering [17].

After the molecule is excited into a higher energy state, the following three outcomes are possible:

1. the molecule can relax back to the ground state and emit a photon of equal energy to the incident one (Rayleigh scattering);

2. the molecule relaxes back to a vibrationally excited state of the electronic ground state and emits a photon with a lower energy than the one of the incident photon (Stokes shifted Raman scattering);

3. the molecule is initially in an excited vibrational state of the electronic ground state; the incident photon excites the molecule to the virtual excited state, then the molecule relaxes back to the ground state, and it emits a photon with higher energy than the incident one (anti-Stokes shifted Raman scattering).

The reason why the Stokes scattering is more probable is because most of the molecules are in the ground state at room temperature. Therefore, most Raman measurements consider the Stokes part of the Raman spectrum.

By further analysis it is possible to determine the power of the scattered light (P_s), which is given by the product between the intensity of the incident photons (I_0) and the Raman cross section (σ_R):

$$\sigma_R \propto \frac{1}{\lambda^4}$$

where λ is the wavelength of the incident photon.
Therefore,

$$P_s \propto \frac{I_0}{\lambda^4}$$

Hence, there is a linear relationship between the power of the scattered light and the intensity of the incident light, and a relationship between the power of the scattered light and the inverse of the wavelength to the fourth power. Therefore, to improve the Raman signal, it is evident that it is desirable to use a short excitation wavelength and a high-power excitation source. This applies provided that no sample degradation occurs, and no strong fluorescence background is excited by short-wavelength light sources.

3.3 Raman Spectra of pure Cannabinoids

D9-Tetrahydrocannabinolic acid (THCA), cannabidiolic acid (CBDA), cannabigerolic acid (CBGA), D9-tetrahydrocannabinol (THC), cannabidiol (CBD) and cannabigerol (CBG) were detected in trichomes at the flowering stages [8]. THCA, CBDA, and CBGA are present in high concentrations, while THC, CBD, and CBG were detected at lower concentrations. Their presence is due to the decarboxylation of the Cannabinoid acids into neutral cannabinoids in the living plant. The decarboxylation rate increases after harvesting, during storage or through heating and exposure to light. Other cannabinoids detected in trichomes, but in even lower concentrations, were Cannabichromene (CBC), cannabinol (CBN), which is an artifact of THC and can be an indicator of its degradation, and cannabichromenic acid (CBCA) [8]. THC, THCA, CBD, CBDA, CBG and CBGA were considered in our work. The main reason is that, besides being the ones present in higher concentrations in the plant, they are the most interesting for pharmaceutical applications because of their psychotropic, analgesic and anti-inflammatory properties, amongst others.

THC, whose term is commonly used to indicate (-)-delta-9-trans-tetrahydrocannabinol (Δ^9-THC), is the pharmacologically and toxicologically most relevant chemical constituent in *Cannabis*. It is responsible for most effects of natural *Cannabis* preparation. It acts trough the binding of the CB-1 receptor. It is known for its euphoriant, analgesic, anti-inflammatory, antioxidant and antiemetic pharmacological properties and its effects are dose dependent and can be enhanced or modulated by the presence of other chemical species. There are four isomers of THC, but only two exist in nature: the (-)-trans isomer and the (+)-trans isomer, which is less potent. THC contains one oxygen atom in the phenol moiety and one, which acts like a hydrogen bond acceptor, in an oxane ring. Differently from other cannabinoids, THC is an illicit drug because of its psychoactive effects. Its possession and consumption can lead to monetary fines and jail time. For these reasons, during the last decades many efforts have been

made to develop *Cannabis* varieties with a low concentration of THC (<0,3%), and high concentrations of CBD and CBG.
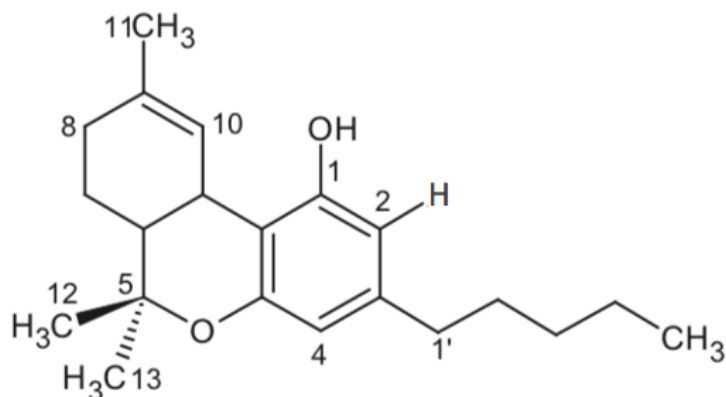
Figure 23. Δ9-Tetrahydrocannabinol chemical structure [25].

THCA is the chemical precursor of THC and it is the major constituent in *Cannabis* resin. It is found in high concentrations in *Cannabis* and it is slowly decarboxylated with heating and drying. For this reason, its structure differs from that of THC because of the presence of the carboxylic group (COOH).
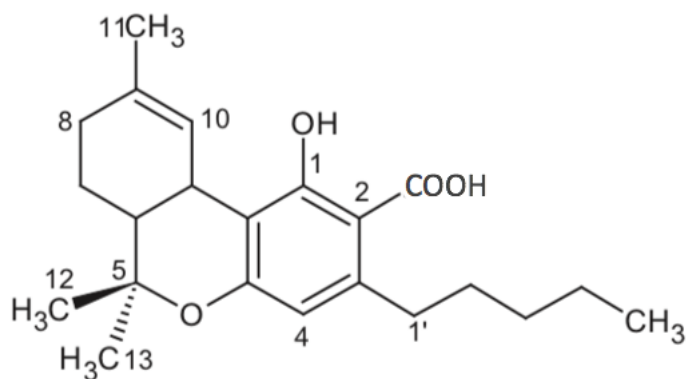
Figure 24. Δ9-Tetrahydrocannabinol acid chemical structure [26].

CBD is the next-best photocannabinoid after THC. Its structure differs from that of THC because of the two oxygens bound to the phenol rings and the presence of the terminal alkene. Moreover, the two rings, which are bound together with a single bond, are almost perpendicular to each other because of the steric endurance of the two rings. From the pharmacological point of view, differently from THC, CBD does not have any psychoactive effect. However, CBD can be used to relieve chronic pain, lower blood pressure, reduce inflammation, anxiety and depression, alleviate side effects related to cancer treatment, such as nausea and vomiting, and mitigate symptoms of neurological disorders, such as reduction of muscle spasticity in people with multiple sclerosis [7].
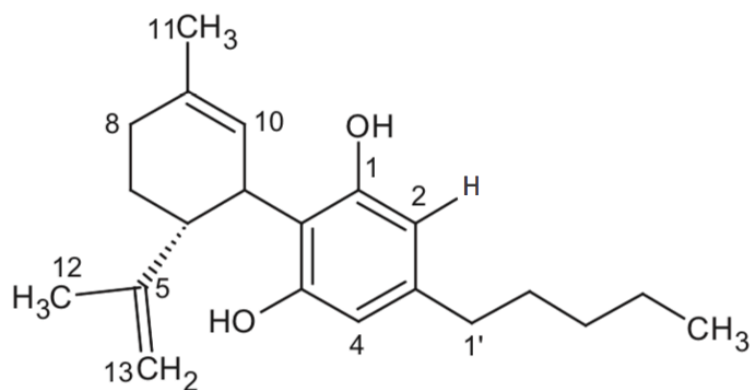
Figure 25. CBD chemical structure [28].

Similar to THCA, CBDA is the chemical precursor of CBD. Its structure differs from the one of CBD because of the presence of the carboxylic group (COOH); CBDA slowly decarboxylates with heating and drying, turning to CBD.
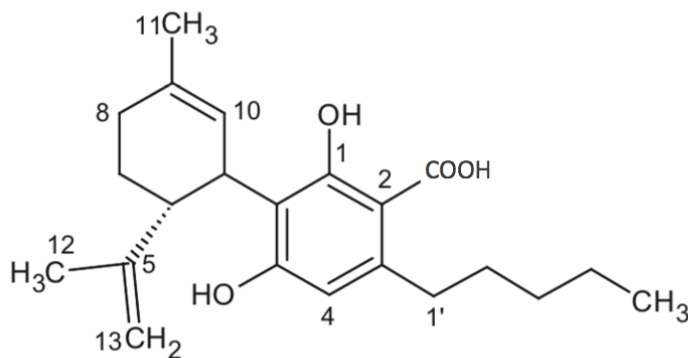
Figure 26. CBDA chemical structure [29].

CBG is a minor cannabinoid. During the plant growth the majority of CBG is turned into other cannabinoids, especially THC and CBD, and only a small concentration of CBG (< 1%) remains in the plant. CBG has a large spectrum of pharmacological activity. It is known to kill or decelerate bacterial growth, to treat glaucoma, reduce inflammation, and inhibit tumor cell growth [8]. Its precursor is CBGA which produces CBG trough decarboxylation [6].
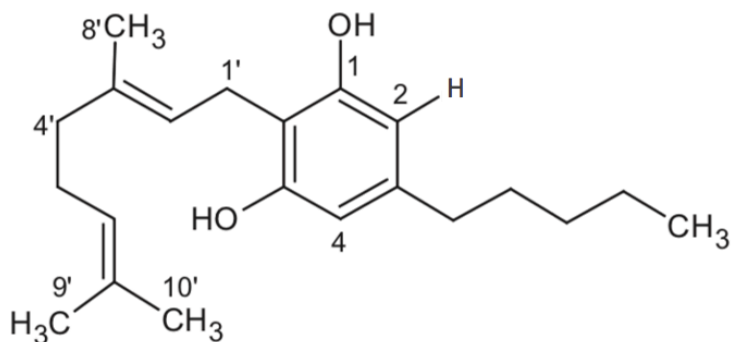
Figure 27. CBG chemical structure [30].

41
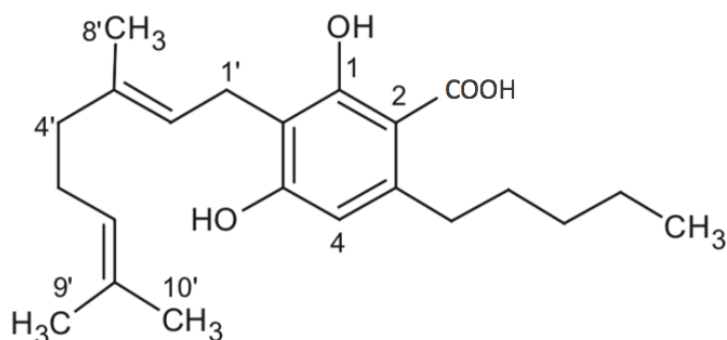
Figure 28. CBGA chemical structure [31].

Raman spectroscopy has been recently used to identify the Raman spectrum of the pure Cannabinoids [1]. Raman spectroscopy turns out to be a very powerful tool in the analysis of such chemical compounds. It can be applied both on prepared samples and directly on the plant flower and it can be used to analyze the chemical and structural characteristics of the sample.

The Raman spectra of the pure cannabinoids which has been used as reference data in this thesis work were obtained by Sanchez et al. [1]. Their goal was to demonstrate that *Cannabis*/hemp diagnostics can be possible with a non-destructive and non-invasive spectroscopy on fresh plant materials. *Cannabis* refers to hemp plants that are rich in THCA. In [1] it is shown that it is possible to record the Raman spectra of hemp and the *Cannabis* variety "TS" using a handheld device. The difference between the two species is in the concentration of cannabinoids:

1. TS: THCA: 4.05%, THC: 0.04%, CBDA: 0%, CBD: 0%, CBGA: 0.23%, CBG: 0%.

2. Hemp: THCA: 0%, THC: 0%, CBDA: 1.08%, CBD: 0.07%, CBGA: 0%, CBG: 0%.

By using such Raman spectrometers, the measurements were directly made through the vile. This approach circumvents the problems associated to the extraction of cannabinoids and/or their evaporation (cannabinoids are volatile) but doesn't provide data in the low wavenumber region of the Raman spectrum. With this technique it was possible to recognize *Cannabis* from hemp with 100% accuracy and to recognize different *Cannabis* varieties with an accuracy of 97%. The Raman spectra for CBD-rich hemp (T5-005 and TS006) was obtained. T5-005 and TS006 are characterized by the following concentrations of cannabinoids:

1. T5−005: THCA: 0.09%, THC: 0%, CBDA: 1.68%, CBD: 0.64%, CBGA: 0.1%, CBG: 0.02%.

2. TS006: THCA: 0.1%, THC: 0%, CBDA: 2.27%, CBD: 0.67%, CBGA: 0.05%, CBG: 0%;

The Raman spectra show vibrational bands that are assigned to cellulose, carotenoids, lignin and xylan. These same bands are present in the spectra of hemp and *Cannabis* (TS) as well. However, there are relevant differences among the intensity of these peaks in the three plants. For T5-005 and TS006, the peak at 1525 cm^{-1} is the one which shows and highest intensity. It is associated to carotenoid. The intensity of this same peak is lower in hemp and *Cannabis*. Similar considerations can be done for the peaks at 1155-1228 cm^{-1} which are assigned to cellulose and xylan. The different intensity of the peaks suggests a difference in the relative concentration of such molecules in hemp*, Cannabis*, and CBD-rich plants, which could be interesting for their differentiation.
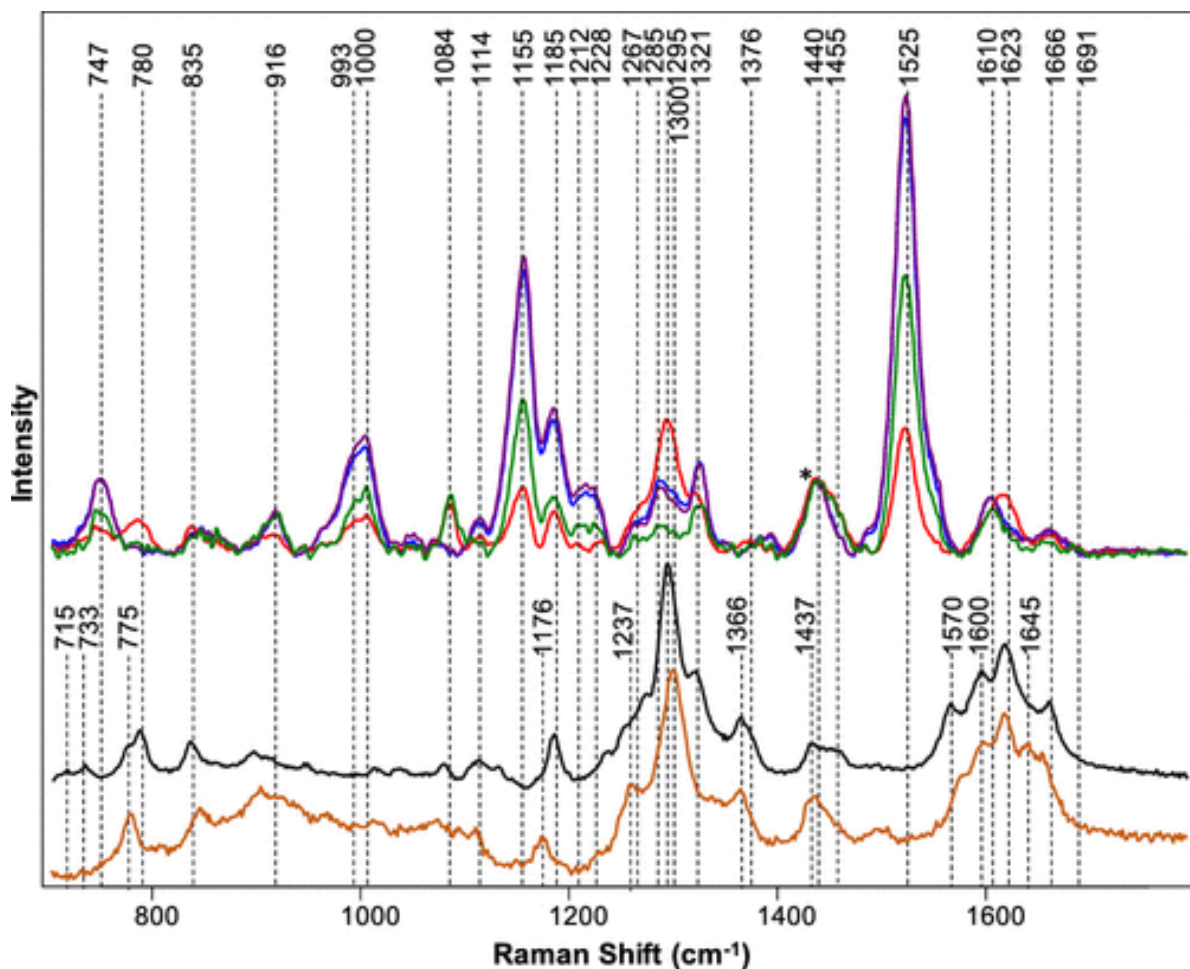
Figure 29. On the top: Raman spectra collected from T5−005 (purple) and TS006 (blue), hemp (green) and TS (red). On the bottom: Raman spectra of THCA (black) and CBDA (orange) [1].

band	vibrational mode	assignment
780	TBA	cannabinoids[12]
835	TBA	cannabinoids[12]
916	ν(C−O−C) in plane, symmetric	cellulose, lignin[15]
993−1000	ν3 (C−CH3 stretching) and phenylalanine	carotenoids, protein[16,17]
1084	ν(C−O)+ν(C−C)+δ(C−O−H)	carbohydrates[18]
1114	νsym(C−O−C), C−OH bending	cellulose[19,20]
1155	νasym(C−C) ring breathing	carbohydrates, cellulose[15]
1185	ν(C−O−H) next to aromatic ring+σ(CH)	xylan[21,22]
1212−1228	δ(C−C−H)	aliphatic,[23] xylan[21]
1267	C−O stretching (aromatic)	lignin[24]
1285	δ(C−C−H)	aliphatic[23]
1295	TBA	cannabinoids[12]
1321	δCH2 bending vibration	cellulose, lignin[15]
1376	δCH2 bending vibration	aliphatic[23]
1440	δ(CH2) + δ(CH3)	aliphatic[23]
1455	δCH2 bending vibration	aliphatic[23]
1525	—C=C— (in plane)	carotenoids[25,26]
1610	ν(C−C) aromatic ring +σ(CH)	lignin[27,28]
1623−1666	aromatic	cannabinoids[12,29]

Figure 30. Vibrational Bands and their assignments for Hemp, Cannabis species and THCA [1].

Substantial differences in the peaks associated to THCA were also observed among hemp, *Cannabis* and CBD-rich plants. TS has bands at 780, 1295, 1623 and 1666 cm^{-1} which can be associated to THCA. T5-005 and TS006 have the same peaks of TS, besides those at 780 cm^{-1} and 1623 cm^{-1}, which are not present. The peak at 1666 cm^{-1} has the same intensity as in TS. The difference between the peaks at 780 cm^{-1} and 1623 cm^{-1} can be used to differentiate between *Cannabis* and CBD-rich hemp. Besides this, we can also notice a change in intensity of the vibrational bands in the 1295-1300 cm^{-1} range. Their intensity is higher in T5-005 and TS006 than in TS. This difference can be used as a further indication of the difference between CBD-rich hemp and industrial hemp.

The Raman spectra of selected pure cannabinoids were also reported in [1]. The sampled analyzed were produced trough solvent evaporation (i.e., by leaving the uncapped vials for several hours under a fume hood). Next, the residual cannabinol oils were redissolved in ethanol and deposited onto clean microscope coverslips and dried under room temperature. The spectra were obtained by using a home built inverted microscope and a 488 nm laser. THCA and CBDA exhibit very similar spectra (figure 31). However, THCA has a vibrational band centered at 1295 cm^{-1}, which has been found to be 5 cm^{-1} red-shifted in the spectrum of CBDA. Other differences are the peak at 1185 cm^{-1} and the doublet at 775 and 780 cm^{-1} in the spectrum of THCA, which are missing in the spectrum of CBDA. The only peak that can be used to differentiate the spectra of CBDA and THCA is the one at 780 cm^{-1}. The intensity of the peak

at 1623 cm^{-1} can be used to distinguish *Cannabis* from CBD-rich hemp. Indeed, the first contains a concentration of THCA of 4.05%, so the peak at 1623 cm^{-1} is present and intense. CBD-rich hemp, instead, has a lower concentration of THCA (1.68 % for T5-005 ad 2.27% for TS006) and that peak is not as visible. Because of the difference in the concentration of THCA and CBDA in the two plants, the intensity of the corresponding peak will be different, and this can be exploited for their recognition. In [ref] it was also important to determine how different are the spectra of each cannabinoid, and how they differ from the corresponding acids (i.e., THC vs. TCHA, CBD vs. CBDA, etc.).
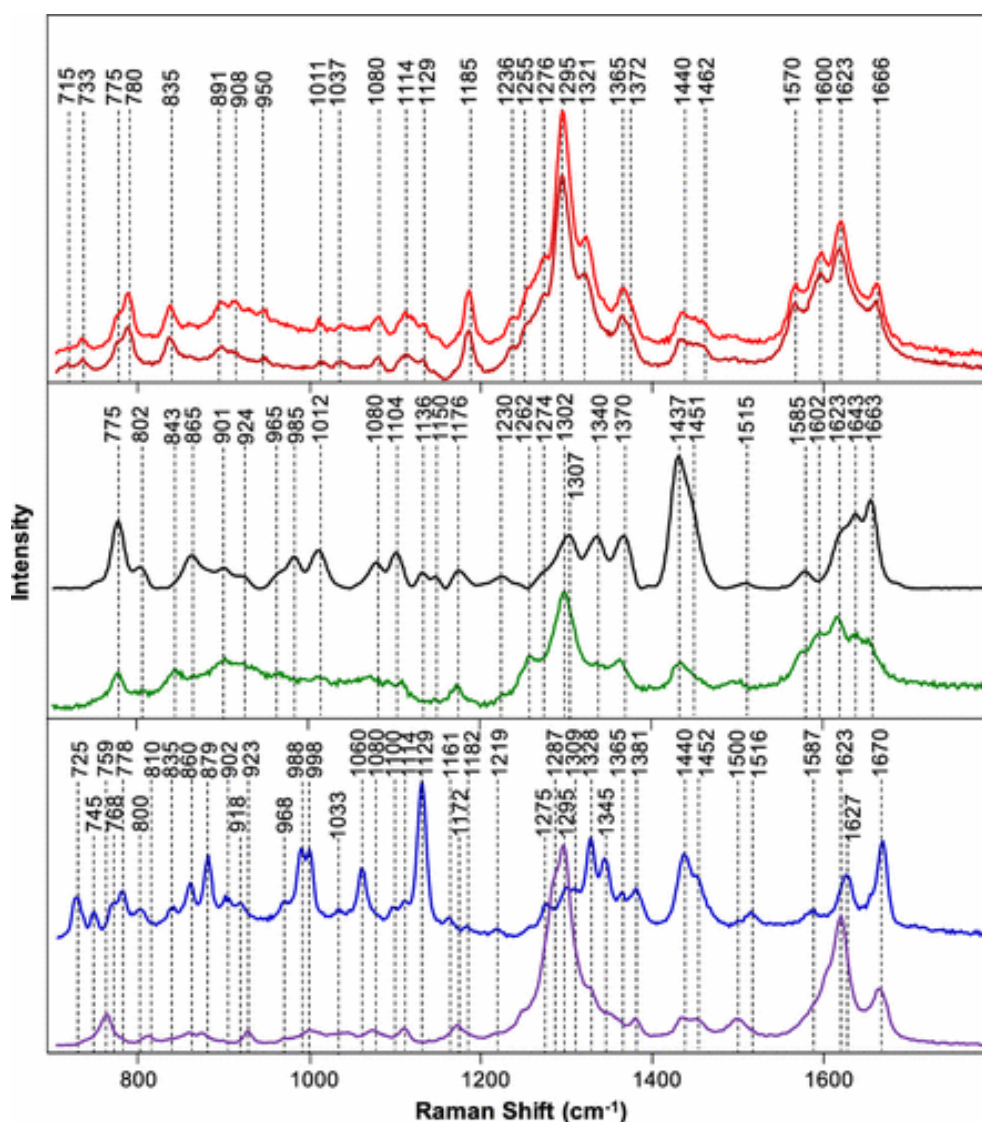
Figure 31. Raman spectra of THC (red), THCA (maroon), CBD (black), CBDA (green), CBG (blue), and CBGA (violet) [1].

THC and THCA have a nearly identical Raman spectra, just like CBD and CBDA. Both spectra exhibit vibrational bands at 775, 802, 901, 1176, 1370, 1437, 1585, 1623, 1643, and 1663 cm^{-1}. However, CBD had peaks at 865, 985, 1080, 1104, 1136, 1150, 1274, and 1515 cm^{-1} that were not observed in the spectra of CBDA. Instead, CBDA has peaks at 843, 1262, and 1602 cm^{-1} that were not observed in the spectrum of CBD. The peak centered at 1307 in the spectrum of CBD is also 5 cm^{-1} blue-shifted in the spectrum of CBDA.

At difference from the other cannabinoids and their acid precursors, CBGA and CBG exhibit a greater spectroscopic difference. Even if nearly all vibrational bands were present in both spectra, their relative intensities are different. For instance, in CBGA the most intense peaks are those at 1287-1295 cm^{-1}. In CBG the most intense peak is the one at 1129 cm^{-1} which is not present in the spectrum of CBGA. The bands at 725, 768, 879, 998, and 1060 cm^{-1} are unique for CBG/CBGA and are not present in the Raman spectra of THC/THCA and CBD/ CBDA. Therefore, the latter bands can be used for identification of this cannabinoid in CBG/CBGA-rich hemp, even if their concentration in the plant is relatively small [1].

compound	vibrational bands, cm^{-1}
THC	733, 775, 780, 835, 891, 908, 950, 1011, 1037, 1080, 1114, 1129, 1185, 1236, 1255, 1276, 1295, 1321, 1365, 1372, 1440, 1462, 1570, 1600, 1623, 1666
THCA	715, 733, 775, 780, 835, 891, 908, 950, 1011, 1037, 1080, 1114, 1129, 1185, 1236, 1255, 1276, 1295, 1321, 1365, 1372, 1440, 1462, 1570, 1600, 1623, 1666
CBD	775, 802, 865, 901, 924, 965, 985, 1012, 1080, 1104, 1136, 1150, 1176, 1230, 1262, 1274, 1302, 1340, 1370, 1437, 1451, 1515, 1585, 1623, 1643, 1663
CBDA	775, 802, 843, 901, 1176, 1262, 1307, 1370, 1437, 1451, 1585, 1602, 1623, 1643, 1663
CBG	725, 745, 768, 778, 800, 810, 835, 860, 879, 902, 918, 968, 988, 998, 1033, 1060, 1100, 1114, 1129, 1161, 1182, 1219, 1275, 1295, 1309, 1328, 1345, 1365, 1381, 1440, 1452, 1516, 1587, 1623, 1627, 1670
CBGA	759, 810, 860, 879, 923, 998, 1080, 1114, 1172, 1287, 1295, 1381, 1440, 1452, 1500, 1623, 1670

Figure 32. Vibrational Bands Observed in the Raman Spectra of THC, THCA, CBD, CBDA, CBG, and CBGA [1].

3.4 Elements of multivariate statistics: principal component analysis of spectral data

Multivariate statistics deals with the analysis of data, called "observations", which are described by at least two components, called "variables". This approach is commonly used in several fields, like chemistry, physics, medicine, and engineering. An example of multivariate data can be identified when talking about spectroscopy. The spectrum of a sample, in fact, is the result of the observation of several single measurements of the intensity of light, one for each discrete wavelength (λ_i) considered in the spectrum. The recorded vector of intensities represents the set of variables associated to the given observation. Therefore, in general, in a group of multivariate data, for each observation (o) we collect the values of a certain number of variables. We can represent this as a matrix (Table 2), in which each line (X_o) represents an

observation, and the columns represent the different variables. This matrix can be indicated as X_{ov}, in which the o indicates that the observations are represented by the rows, and the variables (v) are associated to the columns. With this notation the transposed matrix is indicated as $(\boldsymbol{X}_{ov})^t = \boldsymbol{X}_{vo}$.

	v_1	v_2	v_3	v_4	v_5	v_6
o_1	0.29	0.15	0.67	0.57	0.97	0.53
o_2	0.37	0.19	0.45	0.37	0.07	0.59
o_3	0.98	0.82	0.13	0.62	0.54	0.37
o_4	0.73	0.29	0.50	0.43	0.64	0.76
o_5	0.82	0.75	0.23	0.11	0.58	0.12
o_6	0.53	0.99	0.64	0.04	0.16	0.63
o_7	0.21	0.23	0.94	0.38	0.61	0.75
o_8	0.57	0.68	0.29	0.90	0.77	0.33

Table 2. Group of multivariate data represented as a matrix. The lines, o, represent the observations and the columns, v, represent the variables [17].

Therefore, each observation appears to be multidimensional. This is the reason why such kind of data are at the center of multivariate statistics. One of the fundamental methods of multivariate statistics is the Principal Component Analysis (PCA). It was introduced by Harold Hotelling in the thirties [18], but it became of common use only in the seventies. To introduce PCA one can extend the concept of variance to multivariate data. It is known that, when considering two independent variables (x_i, y_i), the derived quantity $q = q(x,y)$ has a variance given by the squared weighted sum of the variances:

$$\sigma_q^{\;2} = \left(\frac{\partial q}{\partial x}\right)^2 \sigma_x^{\;2} + \left(\frac{\partial q}{\partial y}\right)^2 \sigma_y^{\;2}$$

This is valid only in the hypothesis that the two variables, x and y, are completely independent one from the other. In the case of multivariate data, it is not necessarily true that the measurements of the variables (x_i, y_i) are independent from each other. For example, considering a spectrum with several peaks due to different substances present in the sample, the measurement of the intensity of light around a peak will increase or decrease depending on the quantity of the chemical compounds which produces that peak in the spectrum. We cannot say that those variables are independent one another, especially when we consider two peaks originated by the same compound.

It can be demonstrated that for non-independent quantities the variance is not given only by the sum of the variances of the single quantities, but it is necessary to add an additional term σ_{xy}, which can be positive or negative. This term changes this expression in a significant way. The term σ_{xy} is called covariance of the variables x and y.

It is possible to introduce PCA through an eigenvalue problem that involves the covariance matrix of the dataset. The definition of the covariance Σ_{xy} between two variables x and y observed in $N=N_0$ independent measurements is the following:

$$\Sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})$$

If instead of two variables we measure three variables (say: x, y, z) we need to introduce two further covariances among the variables:

$$\Sigma_{xz} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(z_i - \bar{z})$$

$$\Sigma_{yz} = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})(z_i - \bar{z})$$

From their definition it is immediate to see the symmetry of the covariance matrix, i.e., Σ_{xy} = Σ_{yx}. When the two variables coincide, the following matrix element is obtained:

$$\Sigma_{xx} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x}) = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

The result coincides with the definition of the variance of the x variable. Finally, we can collect all this information in a matrix Σ, which can be defined as the covariance matrix or (more precisely) the variance-covariance matrix. The latter name indicates that along the diagonal of the matrix we can find the variances of each single variable:

$$\Sigma = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} & \Sigma_{xz} \\ \Sigma_{yx} & \Sigma_{yy} & \Sigma_{yz} \\ \Sigma_{zx} & \Sigma_{zy} & \Sigma_{zz} \end{pmatrix}$$

We can use the same approach if we examine a system with more than three variables, in which, for each observed data (I = 1:N), we measure the value of N_v variables. We begin by computing the average of the λ-variable:

$$\bar{\lambda} = \frac{1}{N_o}\sum_{i=1}^{N_0}X_{i\lambda}$$

By using such definition, it is possible to define the covariance for two given variables (λ, μ) as follows:

$$\Sigma_{\lambda\mu} = \frac{1}{N_0-1}\sum_{i=1}^{N_0}(X_{i\lambda} - \bar{\lambda})(X_{i\mu} - \bar{\mu})$$

Because of the relevance of the covariance matrix, it is important to analyze in detail its definition. This allows to express the correspondent matrix with algebraic notation, and allows

to work, more conveniently, with MATLAB. We therefore introduce a matrix (χ) that represents the centered with respect to the average values of the variables:

$$X_{i\lambda} = X_{i\lambda} - \bar{\lambda}$$

It is now possible to introduce in matrix form the covariance matrix between variables:

$$\Sigma_{vv} = \frac{1}{N_0 - 1} \chi_{vo} \chi_{ov} = \frac{1}{N_0 - 1} \chi^t \chi$$

To facilitate the interaction with Matlab, it is also convenient to introduce the functional notation:

$$\Sigma_{vv} = Cov(X_{ov}) = \frac{1}{N_0 - 1} [\chi_{ov} - \langle \chi_{ov} \rangle]^t [\chi_{ov} - \langle \chi_{ov} \rangle]$$

This approach uses $\langle X_{ov} \rangle$ which is the operator mean of the matrix, for which the average is carried out with respect to the first index of the matrix. It produces a row vector that runs over the variable indexes:

$$(\langle X_{ov} \rangle)_\lambda = \frac{1}{N_0} \sum_i X_{i\lambda}$$

The last two equations define covariance and mean in a way which is equivalent to the functions cov( ) and mean( ) in Matlab. The variance-covariance matrix is a symmetric matrix. If we define the matrix A as follows

$$A = \frac{\chi}{\sqrt{(N_0 - 1)}}$$

It is possible to define the covariance matrix Σ as the product between A and its transposed:

$$\Sigma = A^t A \qquad (1)$$

By evaluating the transposed matrix of Σ, we easily reach the following conclusion:

$$\Sigma^t = (A^t A)^t = A^t (A^t)^t = A^t A = \Sigma$$

which further demonstrates the symmetry of the variance-covariance matrix. Another relevant property of the variance-covariance matrix is that it is positive-defined, which means that its eigenvalues are ≥ 0. This can be demonstrated starting from equation (1). A semi defined positive matrix is such that given a generic vector x, it is true that:

$$x^t M x \geq 0$$

In this specific case, using the definition of equation 1, we can write

$$x^t \Sigma x = x^t A^t A x = (Ax)^t (Ax) = |Ax|^2$$

We end up obtaining the squared modulus of a generic vector Ax. This value is ≥ 0.

The goal of PCA is to introduce a linear transformation which allows to go from the indexes of the variables (v) to new set of indexes (s), which will be associated to the principal components. The idea is to substitute to the set of variables which were instrumental in the measurements (e.g. the light intensities sampled along the wavenumber axis) a new set of variables that are potentially "better" for two reasons:

1. they are mutually independent (orthogonal) and statistically independent (i.e. their covariance is zero)
2. their number is reduced compared to that of the original variables.

This set of principal components (associated to the indexes s) allows to describe the behavior of the experimental observations with a reduced set of independent quantities. To do so, from a technical point of view, one applies the spectral decomposition to the covariance matrix. This is possible since the matrix is symmetrical. Hence, one can try to cast the spectral decomposition as the product of a matrix and its transposed, like in eq n. The identification of the structure like A^tA gives the definition of the dataset in the new exemplificative variables, the principal components. The requirement of independence of the principal components is met only if we associate each component to the eigenvectors of the covariance matrix. In fact, different eigenvectors are orthogonal to each other, and they provide the coefficients of the linear transformation which allow to go from the indexes v (experimental variables) to a certain principal component with index s.

Let us proceed by starting with the eigenvalue equation for the covariance matrix Σ, in which we associate to each eigenvalue/eigenvector pair the index of a principal component s:

$$\Sigma_{vv}L_{vs} = L_{vs}\sigma_{ss}$$

σ_{ss} is the diagonal matrix of the eigenvalues of Σ_{vv}, L_{vs} is the matrix of the eigenvectors of Σ_{vv}. Because of the symmetry of Σ, the eigenvector matrix L_{vs} is an orthogonal matrix, so the following relations are valid:

$$L_{vs}L_{sv} = 1_{vv}$$

$$L_{sv}L_{vs} = 1_{ss}$$

By multiplying the latter equation by L_{sv} and by exploiting the orthogonality relation $L_{sv} L_{vs} = 1_{ss}$ we obtain the spectral decomposition of the covariance-variance matrix:

$$L_{sv}\Sigma_{vv}L_{vs} = \sigma_{ss} \ (2)$$

By introducing in eq. (2) the definition of Σ:

$$\Sigma_{vv} = \frac{\chi_{vo}\chi_{ov}}{N_0-1},$$

and by rewriting the full matrix product on the right-hand side as the product between a new matrix S and its transposed, it is possible to finally write:

$$\sigma_{ss} = \left[\frac{1}{\sqrt{N_0 - 1}}L_{sv}\chi_{vo}\right]\left[\chi_{ov}L_{vs}\frac{1}{\sqrt{N_0 - 1}}\right] = S_{so}S_{os} = S^t S$$

The rows of the S matrix (S = S_{os}, the so-called scores matrix), defines the observations (o) in terms of the principal components (s):

$$S_{os} \equiv \left[\frac{1}{\sqrt{N_0 - 1}}\chi_0 L_{vs}\right]$$

The diagonalization of the covariance matrix Σ provides the eigenvalues associated to the principal components, which are usually represented in a screeplot. This graph provides the decreasing trend of the principal variances (σ_{ss}) as a function of the s-index of the principal components. From this graph we can judge the relative importance of the principal components in describing the total variance of the dataset, which is given by the trace of Σ (notably, $Tr(\Sigma) = Tr(\sigma)$).

The matrix L_{vs} of the eigenvalues of the covariance matrix Σ_{vv} is called the loadings matrix. It represents the relation between the variables (v) and the principal components (s). In the case in which the dataset is made by a collection of spectra, the representation of the loadings of a principal component in function of the variables is conceptually similar to a spectrum. The positive and negative peaks in the representation of the loadings shows in which spectral regions we observe the largest variations (increasing/decreasing) of the spectral signal moving across the dataset. For the interpretation of the PCA analysis of spectroscopic data, it is very useful to represent the loadings, and in particular the average data and its variation along the relevant principal components. This can be thought as the extension of the concept of interval of confidence, very usual in the description of univariate data, i.e., <x> ± σ. In the case of multivariate data, we have a higher number of ways by which the variation of the data around its average may happen. Some "modes" (i.e., principal components) have larger principal variance than others and this representation allows, together with the screeplot, to judge the relative importance of the principal components in producing relevant spectral chenges across the dataset. Given a certain component n, we can consider the representation of the data $\langle X \rangle \pm \sigma^{1/2} L^t$ as if it was a spectrum to confront with the average one $\langle X \rangle$. The transposition of the column vector L_s relative to the s-component is necessary to be able to sum thus column vector to the tow vector that expresses the average spectrum. The graphic representation given by plot($\bar{v}$, $\langle X \rangle \pm \sigma_{ss}^{1/2} L^t$) – or by plot($\lambda$, $\langle X \rangle \pm \sigma_{ss}^{1/2} L^t$) clearly shows in which region of the spectrum each principal component has a variation and of which entity. This kind of graphical representations will be used in the Chapter dedicated to the data analysis of the Raman mapping data collected on a *Cannabis* trichome.

The S matrix is called the scores. The 2D Cartesian plot of the first two columns of the matrix S_{os} represents the position of the observations of the dataset as a function of the orthogonal reference system given by the principal components s_1 and s_2. This graph is called scatterplot, and it is useful to observe whether the multivariate observations tend to group (forming clusters) or are merely disperse.

51

3.5 DFT and Normal Modes

The cannabinoids that have been considered so far in the PCA of the Raman mapping data are six (THC, THCA, CBD, CBDA, CBG, CBGA) and their structures have been obtained from PubChem [19]. Given the sizeable number of selected cannabinoids and their conformational freedom, to keep the complexity of the problem under control in this first approach to the vibrational structure of cannabinoids, we have decided to assume the conformer reported in PubChem [19] to be the most stable one, with no further analysis. However, each cannabinoid has several conformers as they are relatively flexible molecules. Their conformations are important, and are worth further investigations, because the pharmacological properties of cannabinoids arise after their interaction with the CB1 receptor present in the brain. Different cannabinoids in different conformations can interact differently with this receptor, for two main reasons [15]:

1. The CB1 receptor has different conformations in different areas of the brain which are responsible for several functions. Because of their conformation, each ligand can only bind to a specific area and will exhibit specific properties, which are not shown by other cannabinoids. For example, THC can bind to CB1 in a region of the brain, which is responsible for its psychoactive effects, while CBD cannot bind to it, but can bind to a region that stimulates appetite [15].

2. CB1 has the same structure in different regions of the brain, therefore all cannabinoids can bind to the same receptor. However, after binding they cause structural changes on a protein which are specific for each cannabinoid, this will determine different signals to the cells and different effects [20].

The input structure for each cannabinoid has been taken from available information published on PubChem [19] and it has been optimized by using the DFT method B3LYP/6-31G(d,p) – a popular method for the early screening of the chemical structure and associated vibrational structure. We verified that the input PubChem conformations of the selected cannabinoids are reasonable minima since the DFT calculations provide optimized structures that are very similar to those on PubChem [19]. Thereafter, the assignment of the main Raman peaks of the cannabinoids was done based on the comparison between the simulated and the experimental spectra available in the literature [1]. By using the software Molden [4], I could inspect the animation of the relevant vibrational modes for each peak of interest, which has been instrumental for the mode assignment and description.

All the atoms in a molecule are always in motion. This is explained by the Heisenberg's uncertainty principle:

$$\Delta p \Delta x \geq \frac{\hbar}{2}$$

which states that the greater is the certainty of the measurement of x (or p), the higher is the uncertainty of p (or x). Hence, if Δp approaches zero and can therefore be determined with infinite precision, there will be an infinite uncertainty of Δx which must approach ∞. This is the case of the free particle when V(x) = 0. Therefore, we can say that molecules are in continuous

motion, otherwise – at rest condition – the position and momentum should be both known accurately. There exist three kinds of motions: translation, rotation and vibration. Molecules with two atoms are characterized by a single vibrational motion (bond stretching), while more complex molecules, such as cannabinoids, possess a manifold of more complex vibrations, also known as normal modes. Normal modes are used to describe the vibrational modes of molecules. A molecule with N atoms, and therefore N nuclei, is characterized by 3N degrees of freedom used to describe its position in the 3D space. In the limit of small displacements, suitable linear combinations of such coordinates can be classified as follows:

1. three translational degrees of freedom that describe the rigid translation of the molecule and its center of mass;
2. three rotational degrees of freedom (which become two for of a linear molecule); such degrees of freedom describe the rotational motion of the molecule in space as a rigid body;
3. 3N-6 (3N-5 in a linear molecule) vibrational degrees of freedom that describe the internal vibrational motions of the molecule.

Starting from the definition of kinetic energy and potential, it is possible to write the vibrational Hamiltonian in terms of independent coordinates:

$$\hat{H} = \hat{T} + \hat{V} = \sum_j \left(-\frac{\hbar^2}{2} \frac{\partial^2}{\partial q_j^2} + \frac{1}{2} \omega_j^2 q_j^2 \right) = \sum_j \hat{H}_{vib}(q_j)$$

Since the Hamiltonian is a sum of independent Hamiltonians (one for each normal mode), the wavefunction is the product of the individual wavefunctions and the energy the sum of the independent energies:

$$E_{vib} = \sum_{j=1}^{N_{vin}} h v_j \left(v_j + \frac{1}{2} \right)$$

where v_j= 0,1,2,3…

As a conclusion we can consider each vibrational mode as a harmonic oscillator approximation where N_{vib} represents the number of harmonic oscillators corresponding to the number of vibrational modes in the molecule. When the molecule is at the ground state (v_j = 0), also known as the zero-point energy, it has an energy of:

$$E = \frac{1}{2} h v_j$$

When a photon with energy equal to

$$E = hv$$

is absorbed, since v_j becomes 1, the molecule will reach a vibrationally excited state of energy:

$$E_{vib} = \frac{3}{2} h v_j$$

This approach is valid for low vibrational quantum numbers, as it does not consider the dissociation of the molecule that might happen once the molecule absorbs an amount of energy sufficient to break the vibrating bond (the Morse potential could amend this issue by considering also the anharmonicity).

There are several kinds of collective vibrations that can be observed in a molecule; for convenience, some recurrent nuclear displacement patterns have been given a name in the literature:

1. Stretching: change in length of the bond, for example C-C or C-H.

2. Bending (or scissoring): variation of the angle between two bonds, such as the HCH angle.

3. Rocking: oscillation of the HCH group on its own plane. The bond length does not change but the bond angle does.

4. Wagging: oscillation of the HCH group orthogonally to its plane in the same direction. The bond length does not change but the bond angle does.

5. Twisting: oscillation of the HCH triangle group orthogonally to its plane. Differently from wagging, in the case of twisting, half of the triangle oscillates in one direction and the other half in the opposite direction. The bond length does not change but the bond angle does.

6. Out of plane: change in the angle between a bond and the plane formed by the rest of the molecule.

4. Principal Component Analysis of the micro-Raman mapping of a Cannabis trichome

4.1 Analysis of the average spectrum and assignment of the main peaks

By using micro-Raman spectroscopy, it was possible to analyze the trichome of a *Cannabis* sample (see Chapter 3.1) over a 21x21 spatial grid (for a total of 441 individual Raman spectra). All the 441 Raman spectra were used to obtain an average spectrum. Before using PCA to obtain relevant chemical information, the necessary assignment of the most relevant peaks of the average spectrum was done. By comparing the reference spectra available from the literature for selected cannabinoids [1] with the average spectrum, it was then possible to assign some of the peaks to a single cannabinoid, while others were assigned to a combination of cannabinoids that, together, contribute to the intensity of the same peak.
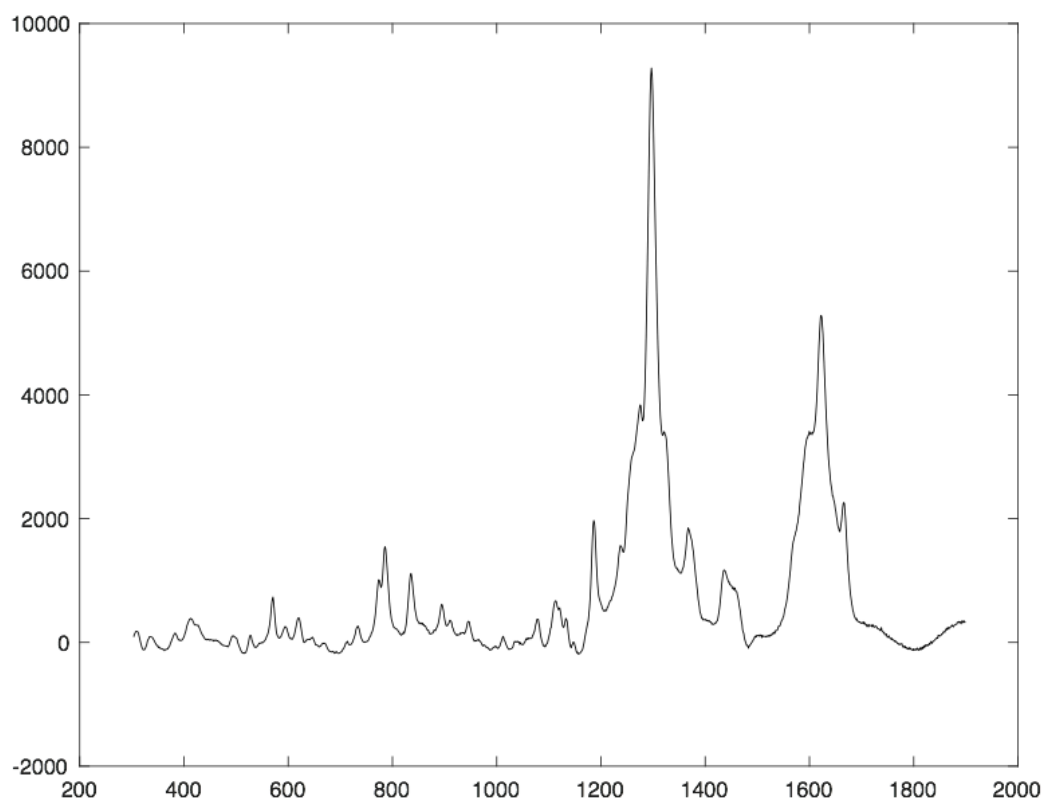
Figure 33. Average spectrum obtained from Raman spectroscopy of several points of a flower trichome of a sample of commercial Cannabis.

Unfortunately, due to the device used in the literature [1] to obtain the spectra for the pure cannabinoids, no information is available for peaks with Raman shifts lower than 700 cm^{-1}. This could represent a limitation since many clear peaks of relevant intensity appear in the 300 −

700 cm^{-1} range of the average spectrum. At higher Raman shifts, between 2900 and 3000 cm^{-1} instead we see the manifold of the peaks associated to the C-H stretching vibrations. These are not much useful for chemical recognition since they do not provide distinct molecular-structure dependent fingerprints for any of the compounds here considered (i.e., they could be associated to all of these cannabinoids).

Not all of the peaks with a Raman shift between 700 and 1800 cm^{-1} were assigned, for several reasons. First, for time constraints we could consider just a selection of cannabinoids, but *Cannabis* is a quite complex plant that contains several metabolites. Furthermore, there is not yet much literature that reports the Raman spectra of pure cannabinoids. This happens because the experimental analysis is difficult on such volatile compounds (that are also not straightforward to obtain in pure forms from *Cannabis* extracts). Indeed, the light beam of the Raman setup could accelerate the evaporation of the sample, thus impeding the Raman measurement. This issue was observed on THC samples by Prof. N.R. Agarwal in Oshawa. Furthermore, besides cannabinoids, some of the peaks might be associated to the waxy layer that characterizes the trichomes of the *Cannabis* plant. A further issue is the similarity between the Raman spectra of each cannabinoid and the correspondent cannabinoid acid. For instance, THC and THCA have similar Raman spectra. For this reason, it is quite complicated to assign some peaks uniquely to one single cannabinoid. Indeed, by inspecting the average spectrum (Figure 33), one clearly recognizes the peak at 1297 cm^{-1} as the most intense one. By looking at the reference data, one verifies that this peak is present in the Raman spectra of both pure THC and THCA. Hence, we can conclude that both THC and THCA contribute to the intensity of the same peak. In some other cases, the same peak could not be associated only to a cannabinoid and the corresponding acid, but to several different metabolites or even to all of them.

The analysis of the peaks of the average spectrum was done by taking into consideration the limitations discussed above and the limited Raman data on cannabinoids currently available in the literature. Among the peaks identified in the average spectrum some were shoulders or were characterized by a low intensity and were therefore neglected in the following stage of analysis (PCA). However, they have still been reported in the following list and indicated in parenthesis and in the graph using a dotted line. Based on their intensity (reported in instrumental units) the peaks were classified as very strong, strong, medium, weak, and very weak:

1. I > 6000   very strong (vs);
2. 3000 < I < 6000   strong (s);
3. 1000 < I < 3000   medium (m);
4. 100 < I < 1000   weak (w);
5. I < 100   very weak (vw).

Comparing the peaks of the average spectra and the reference data of pure cannabinoids it was possible to assign several peaks to a single cannabinoid.

1. Peaks associated to THCA (cm^{-1}) represented in maroon: 714 vw.
2. Peaks associated to CBD (cm^{-1}) represented in green: 925 w, 966 vw, (986), 1012 vw, (1103), 1515 vw.
3. Peak associated to CBDA (cm^{-1}) represented in dark green: 1603 s.
4. Peaks associated to CBG (cm^{-1}) represented in pink: (745), 988 vw, 1060 vw, (1218), (1346), 1517 w.

5. Peaks associated to CBGA (cm^{-1}) represented in purple: 1500 w.

The following peaks could be assigned to more than one cannabinoid:

1. THC and THCA (cm^{-1}) represented in red: 733 w, 1038 vw, 1186 m, 1237 m, 1297 vs, 1321 s, 1367 m, 1600 w, 1666 s.
2. CBD and CBDA (cm^{-1}) represented in yellow: 1436 m, (1451).
3. CBG and CBGA (cm^{-1}) represented in grey: (809), 860 w, 998 vw.
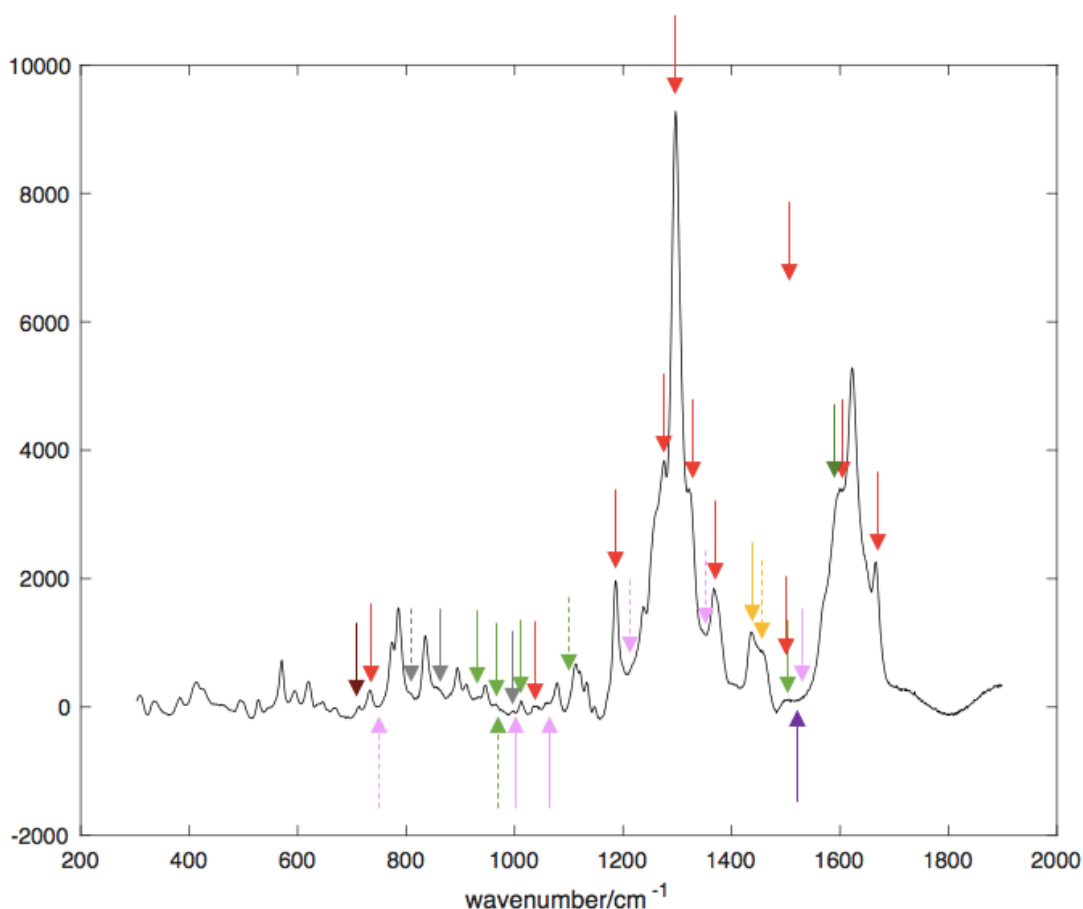
Figure 34. Representation of the peaks in the average spectrum associated to THCA (red), CBD (green), CBDA (dark green), CBG (pink), CBGA (purple), THC and THCA (red), CBD and CBDA (yellow), CBG and CBGA (grey). The shoulders are represented using dotted arrows.

Some peaks in the average spectrum have very high intensity. However, unfortunately, they cannot be associated to one type of cannabinoid, but to a larger variety of them:
- 775 cm^{-1}, assigned to THC, THCA, CBD, CBDA;
- 835 cm^{-1}, assigned to THC, THCA, CBG;
- 1080 cm^{-1}, assigned to THC, THCA, CBD, CBGA;
- 1113 cm^{-1}, assigned to THC, THCA, CBG, CBGA;
- 1276 cm^{-1}, assigned to THC, THCA, CBG;

57

- 1623 cm-1, assigned to THC, THCA, CBD, CBDA, CBG, CBGA.
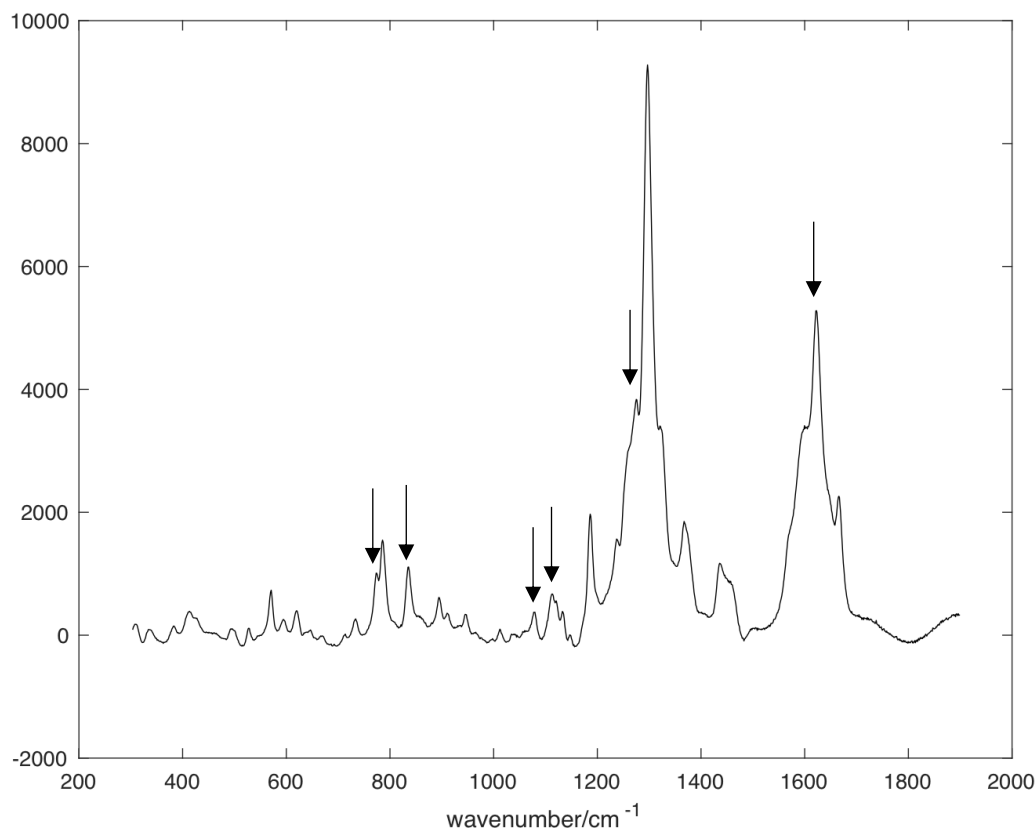
Figure 35. Representation of the peaks in the average spectrum associated to several different cannabinoids.

4.2 PC1 and PC2

Once the peaks of the average spectrum have been assigned to selected cannabinoids it is possible to proceed with the PCA and analyze the loadings of the principal components by considering the chemical information encoded in specific positions of the wavenumber axis (i.e., the marker bands of the cannabinoids listed above). As in our case the dataset for PCA is made of the collection of 21x21 spectra (measured during the Raman mapping), the representation of the loadings of a principal component as a function of the variables is conceptually like a spectrum. The positive (negative) peaks in the representation of the loading show in which spectral regions we observe the increase (decrease) of the signal, with respect to its average, across the dataset. For analysis of spectroscopic nature, it is therefore useful to represent directly the average spectrum and its variation along the selected principal components. Hence, by considering the assignment of the main peaks identified on the average spectrum, and the changes brought by a given loading, one can retrieve spatial information about the spatial variation of the concentration of the considered chemical species. This will be possible by reporting the scores of a given principal component directly on the 21x21 grid of the Raman mapping experiment.

The first step of PCA requires the identification of the relevant principal components. To do so, after running the PCA scripts within Matlab [21], one can obtain and inspect the screeplot

58

reported in figure 36 in the logarithmic scale. The screeplot is a graph that represents the eigenvalues of the principal components (principal variance σ_s) in decreasing order as a function of the principal component index s. From this plot it is immediate to identify the relative importance of the principal components in describing the total variance of the data across the multivariate dataset. By looking at the screeplot (figure 36) we can assume that the most relevant principal components are the PC1 and PC2. Such two components show a higher variance than the others that will be therefore neglected.
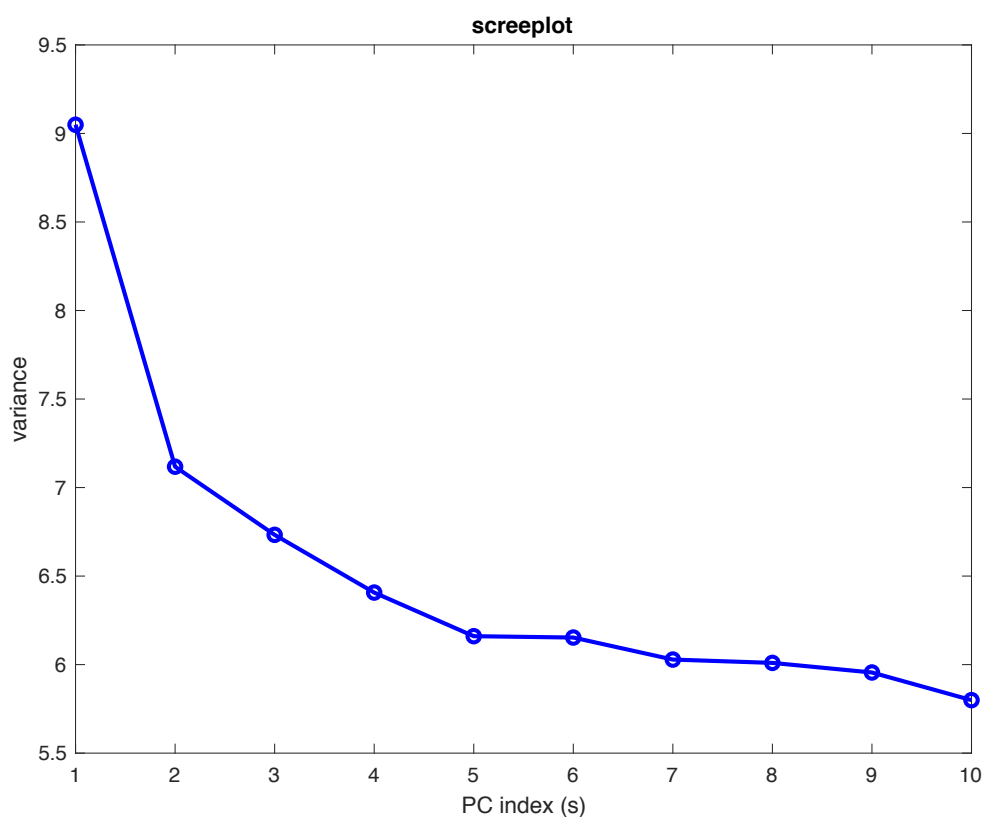
Figure 36. Screeplot of the Principal Components in the logarithmic scale on the y axis: variance of the dataset as a function of the PC index (s).

To check the chemical meaning of the most relevant PCs, we can look at the representation of each loading as a function of the variables (wavenumber axis, see figure 37). The loadings along PC1 (blue) and PC2 (red) are characterized by clear peaks that can be assumed to contain relevant chemical information. However, PC3 (yellow) has mostly an undulatory behavior and does not show clear peaks. Most likely, PC3 is the signature of the automatic fluorescence background correction introduced by the instrument software (WiRE) – this point will be further investigated later on by running the PCA on the raw (uncorrected) dataset. PC4 (purple) is a noisy signal, and we cannot obtain neither spatial nor spectroscopical information out of it. There is only one visible very sharp peak. This is just a spike originated by the interaction of cosmic rays with the CCD detector during the Raman mapping experiment. Therefore, it is not associated to any cannabinoid and cannot provide chemical information about the trichome.

59
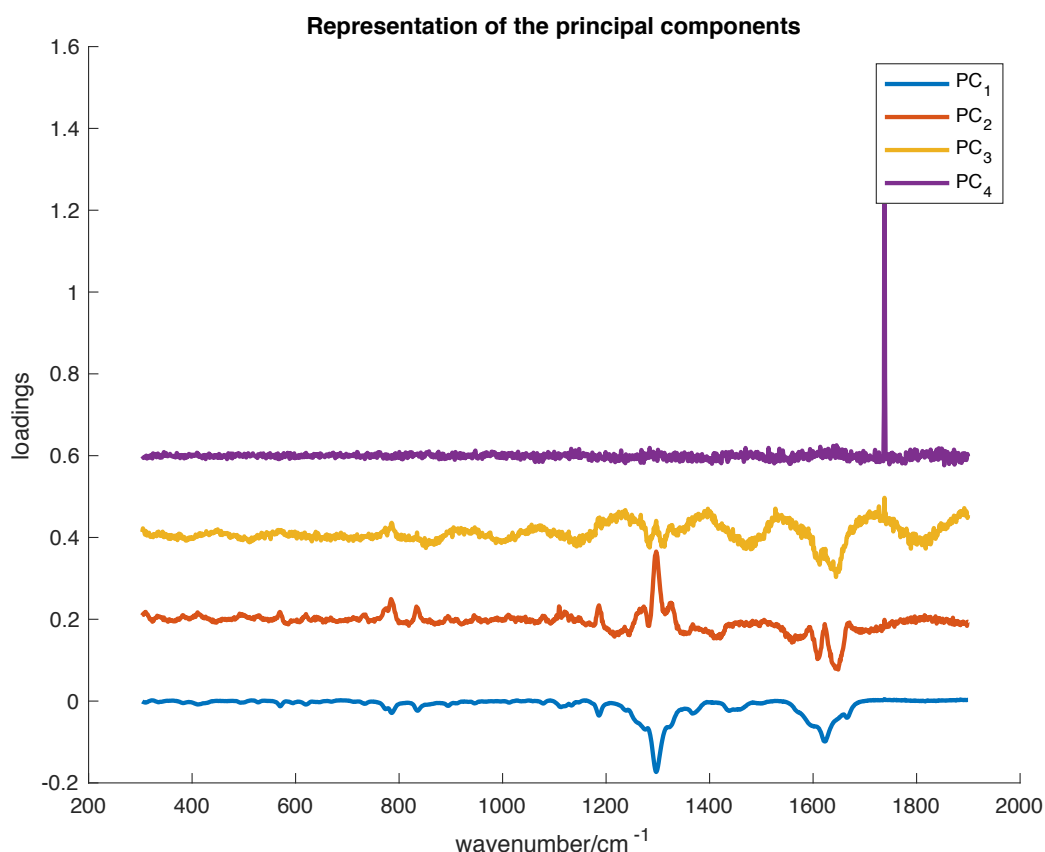
Figure 37. Representation of the principal components: the loading along the first (blue) and second (red) principal component show clear peaks, the loading along the third (yellow) shows an undulatory behavior, the loading along the fourth (purple) is noise only.

The scoremaps can also be obtained with the Matlab PCA scripts [21]. Each pixel in the scoremap represents a single measurement point in the Raman mapping experiment on the trichome, characterized by its (x, y) spatial coordinates. Different colors, or more precisely different shades of the same color, are used to identify the value of the score (of a given PC) for each measurement in the Raman mapping. Therefore, the scoremap can be used to assess how much the Raman spectrum associated to each point in the map varies (along a given PC) with respect to the average Raman spectrum of the whole map. By looking at the scoremaps along PC1 and PC2 we see an evident lack of homogeneity in the shades of the map: this is an indication of chemical information associated to local changes of concentration of chemical species.

In the first scoremap, we can identify the presence of a dark spot around coordinates (-20, 0) which becomes brighter moving along the x axis. In the second scoremap, we can identify three brighter regions, whose dimension is about 10 micrometers. To provide an explanation for this change in shade in the scoremaps, it is important to analyze the loadings of the associated PCs and inspect how their peaks vary if compared to the average spectrum.
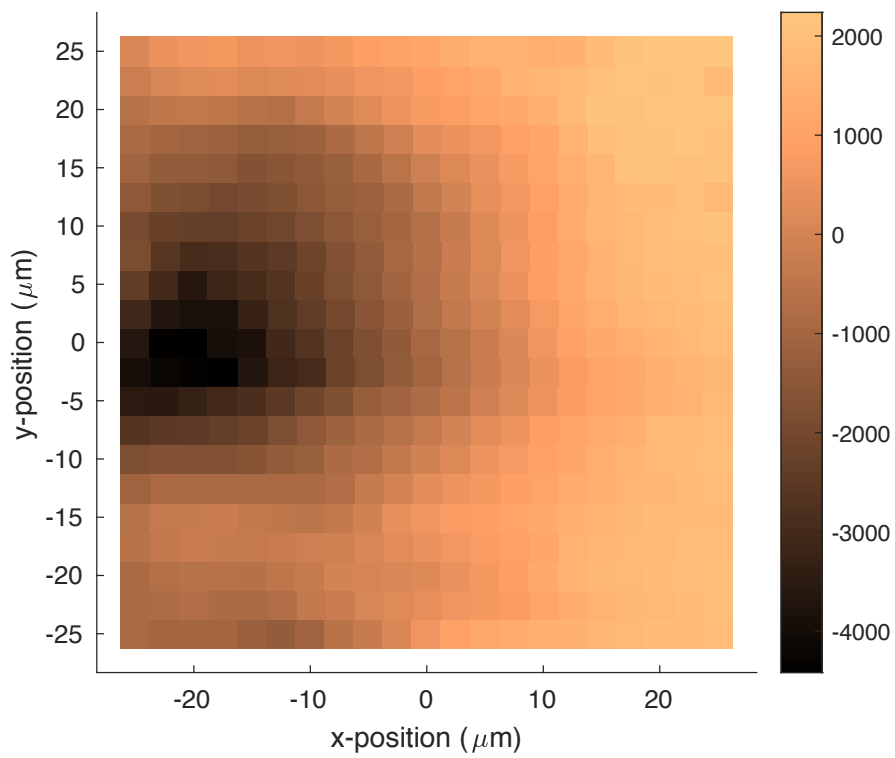
Figure 38. Scoremap of PC1.
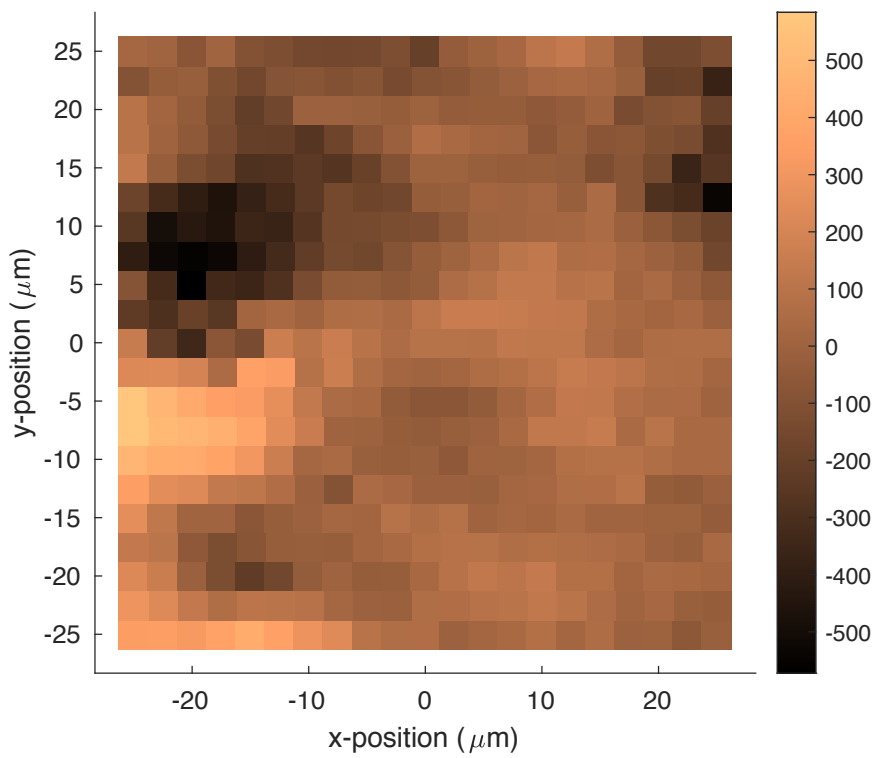
Figure 39. Scoremap of PC2.

The loadings associated to the first two PCs are shown in figure 40 and 41 below.
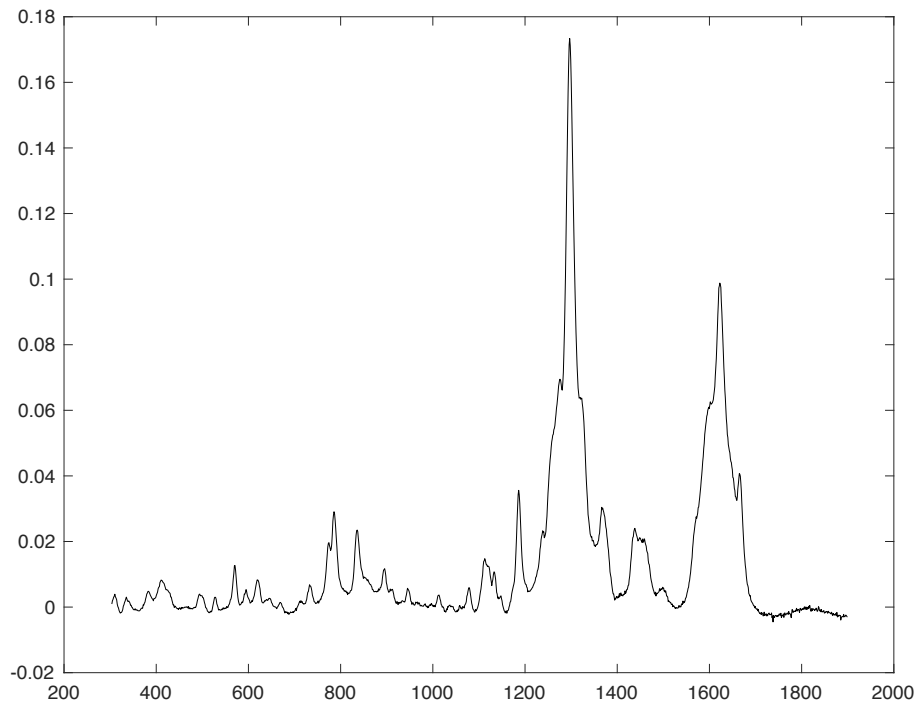
Figure 40. Loading along PC1.
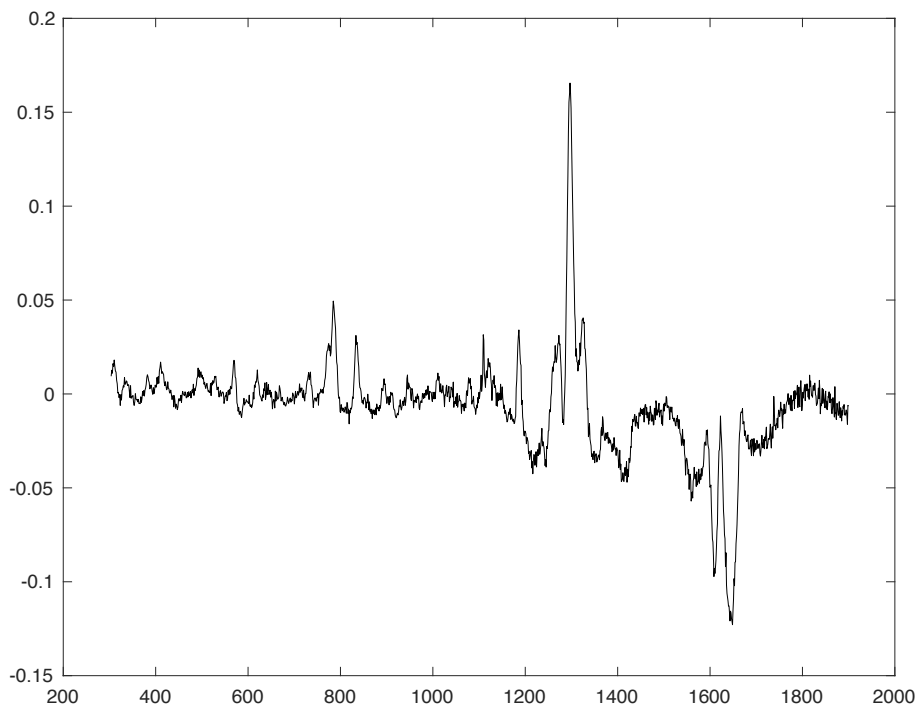
Figure 41. Loading along PC2.

The loading along the first principal component is very similar to the average spectrum. This is even clearer after normalizing and superposing the two graphs (Figure 42) that show overall the same trend. We can find some small differences in the region between 1000 and 1700 cm^{-1}. For instance, the minor peak at 1700 cm^{-1}, associated to the carbonyl group, is not seen in the loading along PC1. However, such peak could also be an artifact due to the undulations of the spectrum caused by the fluorescence background removal algorithm (see above). Hence, because above a certain value of intensity, the two curves practically coincide, we can assume that the loading along PC1 mainly correlates with the overall strength of the Raman signal, i.e., the varying focusing conditions in the mapping experiment on the curved bulb of the trichome surface. Therefore, this loading does not contain any chemical information on the changes of relative composition of the sample from point to point.
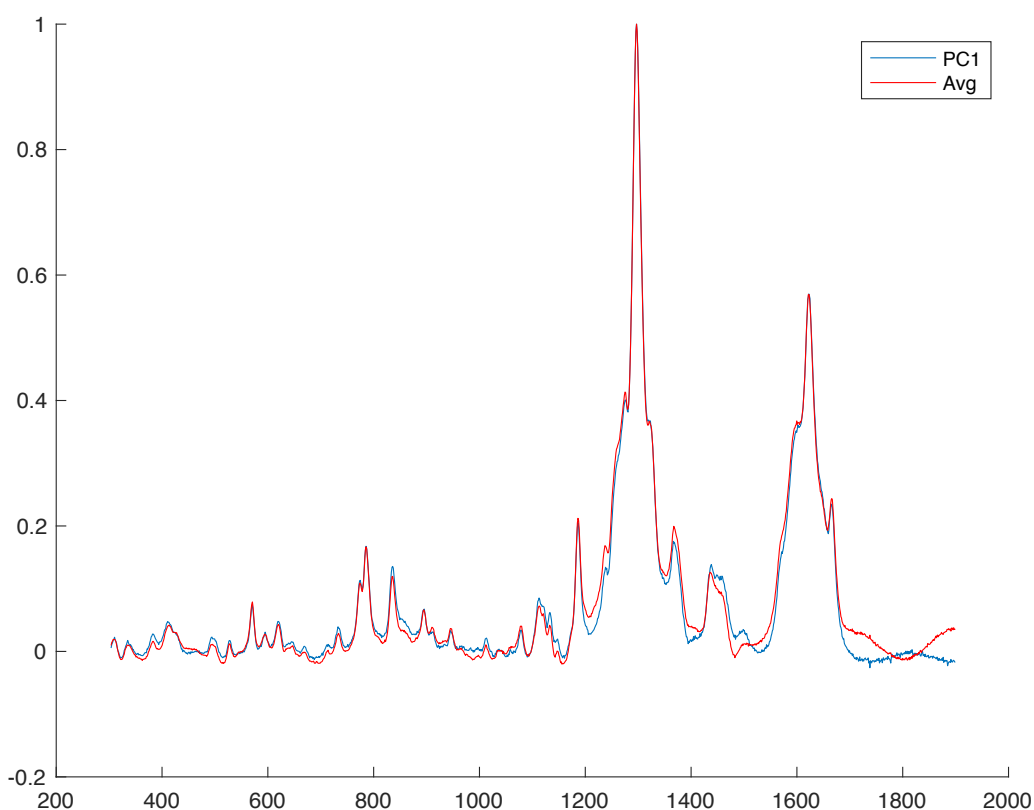
Figure 42. Superposition of the normalized average spectrum and loading along the first principal component.

The loading along PC2, instead, appears to be very different from the average spectrum (Figure 43), with peaks of opposite sign that may contain relevant chemical information, after proper data interpretation. As we did for the loading along PC1, it is possible to normalize and superimpose the curves of the loading and of the average spectrum on the same graph, to better understand the chemical information one may derive from it (Figure 43).
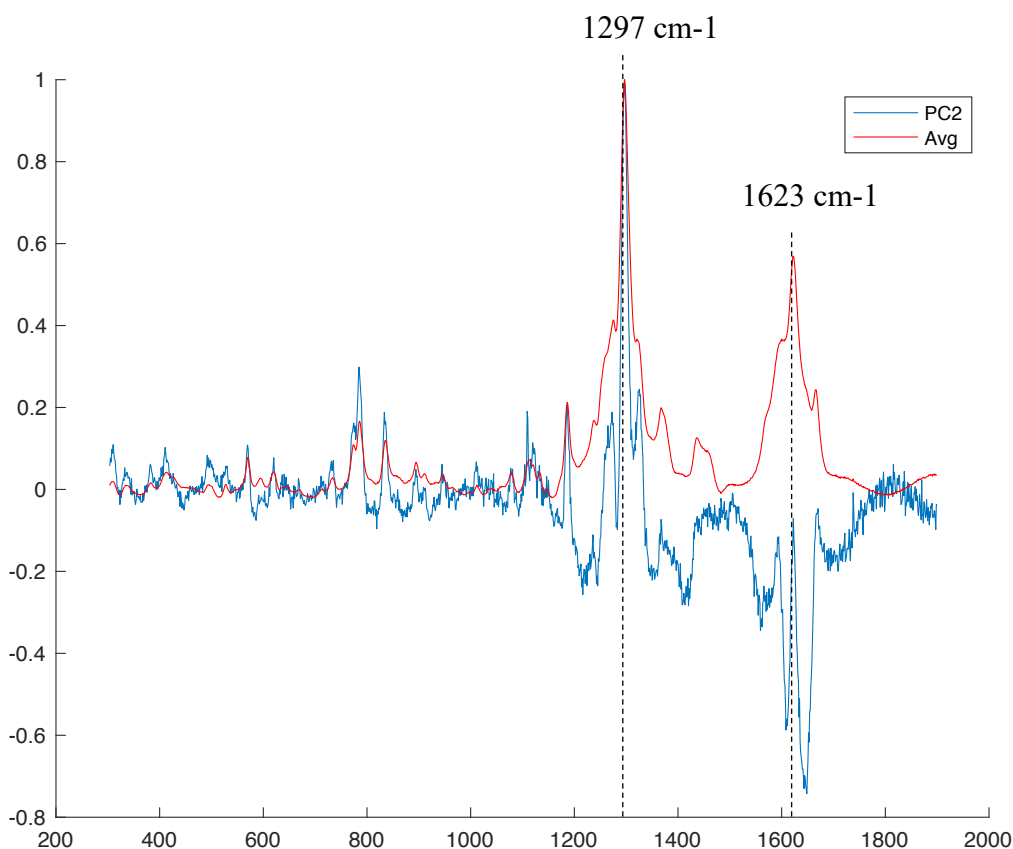
Figure 43. Superposition of the normalized average spectrum and the loading along the second principal component.

Two intense peaks which we can identify in the loading along the second principal component are those at 1297 and 1623 cm $^{-1}$. In the analysis of the average spectrum (see Section 4.1 above) the peak at 1297 cm^{-1} was associated to THC and THCA whereas the one at 1623 cm $^{-1}$ was associated to all the cannabinoids analysed. There are other peaks in the graph of the loading of PC2, and the plot is fairly complex, therefore it is really important to clearly see how the peaks of the Raman spectrum may change from the average value while moving along the PC2 coordinate. Based on Figure 44, while we expect the peak at 1297 cm^{-1} to increase in intensity as the score along PC2 increases, the variation of the peak at 1623 cm^{-1} is not only a change in intensity of the peaks, but a thinning of the peak which becomes highly selective. This feature of the loading along PC2 could seem to be an artifact; however, it can be shown to be reliable by analysing the spectral changes in the 1600 cm^{-1} region as one increases the score along the second principal component. To show this behaviour, we can represent the average spectrum plus the loading (L_2) multiplied by increasing numbers (from 0 to s_{max}). This helps us to assess how the spectrum, and specific peaks, change by moving along the second principal component. The complete graph obtained is reported below. The blue curve represents the spectrum corresponding to s_{max}, the black lines represent how the peaks change moving along the second principal component (from 0 to s_{max}), the red curve represents the average spectrum (corresponding to s = 0).

64
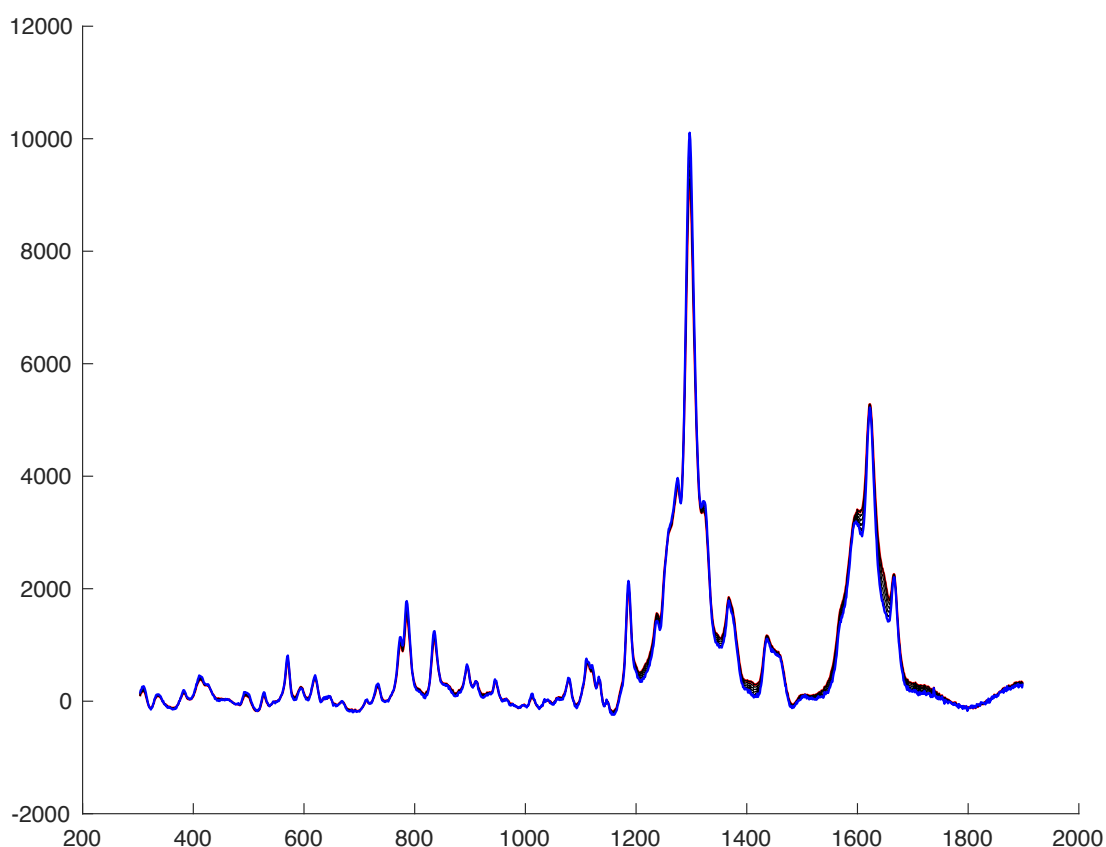
Figure 44.Representation of how the average spectrum change moving along the second principal component. The blue curve represents the final spectrum, the black lines represent how the peaks change moving along the second principal component, the red curve represents the average spectrum.

To better see the behaviour of the peaks at 1297 and 1623 cm $^{-1}$ it is possible to zoom in and represent the graph only for specific sections of Raman shift. First, we can select the part of the spectrum with Raman shift between 1150-1400 cm^{-1} (Figure 45) which includes the most intense peak at 1297 cm^{-1}. It is noticeable how, moving along the second principal component (from the red to the blue line), the intensity of the peak increases.

In the second range of interest, the one of 1500 - 1750 cm^{-1} (figure 46), we can see how the peak at 1623 cm^{-1} changes as well. In this case, there is not a relevant change of the intensity of the peak as it happens for the one at 1297 cm^{-1}. In fact, we can see that the intensity of the peak tends to remain the same. The main difference is in the width of this peak. By moving along PC2 (towards the blue line) the peak becomes sharper.
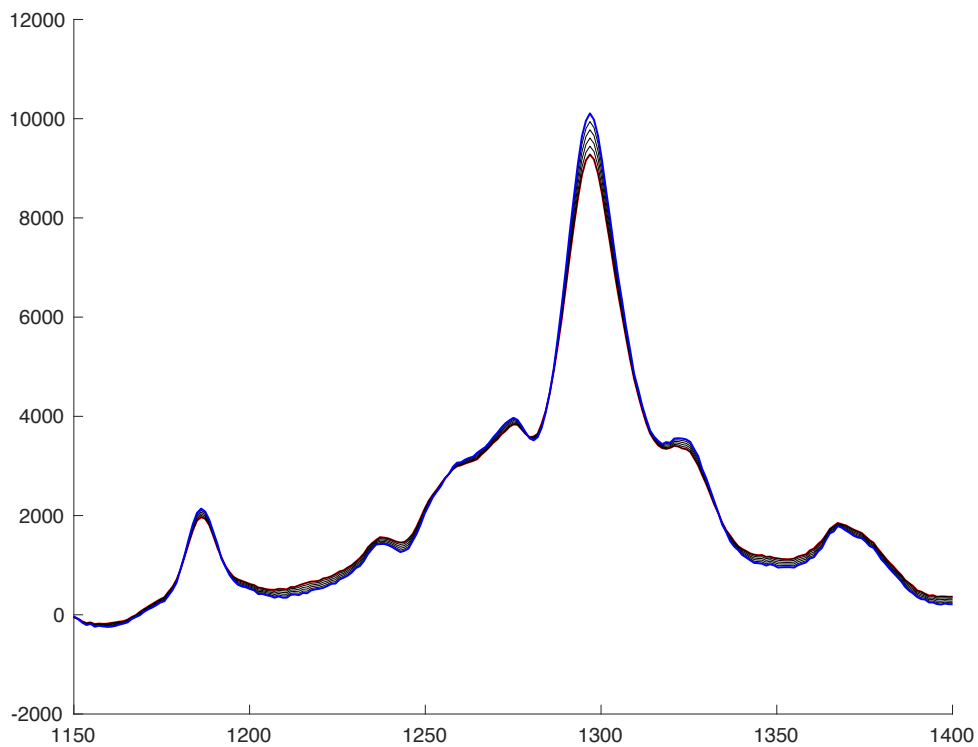
Figure 45. Representation of how the average spectrum change moving along the second principal component in the region between 1150 and 1400 cm^{-1}. The blue curve represents the final spectrum, the black lines represent how the peaks change moving along the second principal component, the red curve represents the average spectrum.
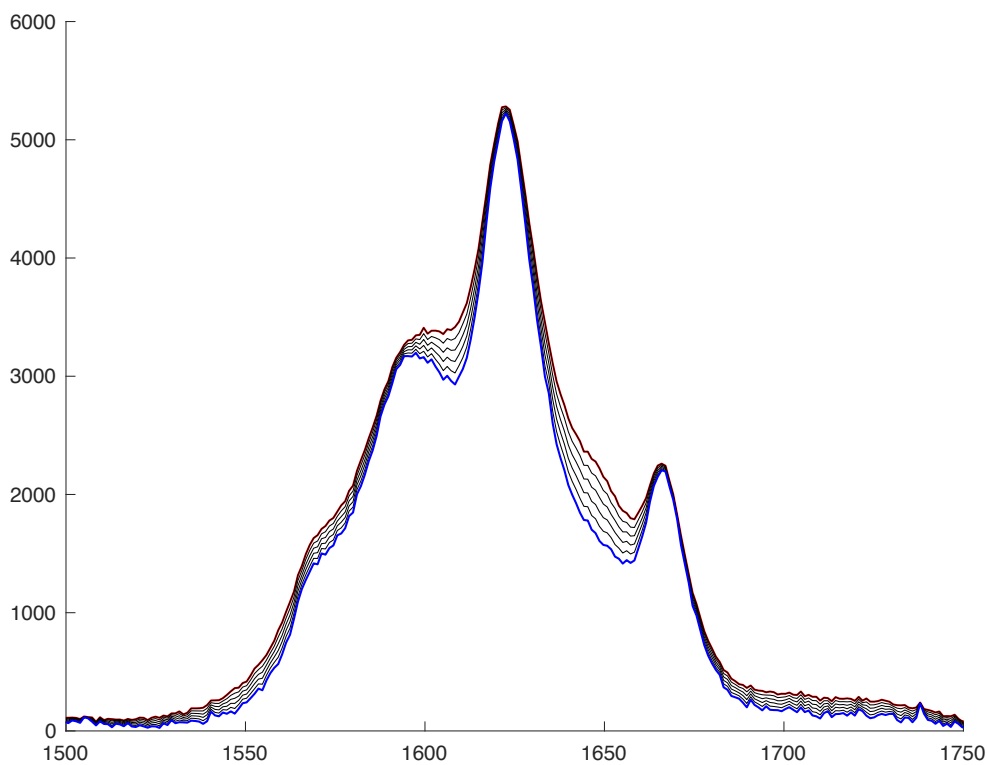
Figure 46. Representation of how the average spectrum change moving along the second principal component in the region between 1500 and 1750 cm-1. The blue curve represents the final spectrum, the black lines represent how the peaks change moving along the second principal component, the red curve represents the average spectrum.

Similar considerations can be done for the loading along the first principal component. Selecting both the first (1150-1400 cm^{-1}) and second range (1500 1750 cm^{-1}). In Figure 47 it is possible to see how, by moving along the first principal component, the spectrum changes over all the wavenumber range, in a more uniform way if compared to the changes occurring along the second principal component (Figure 45).
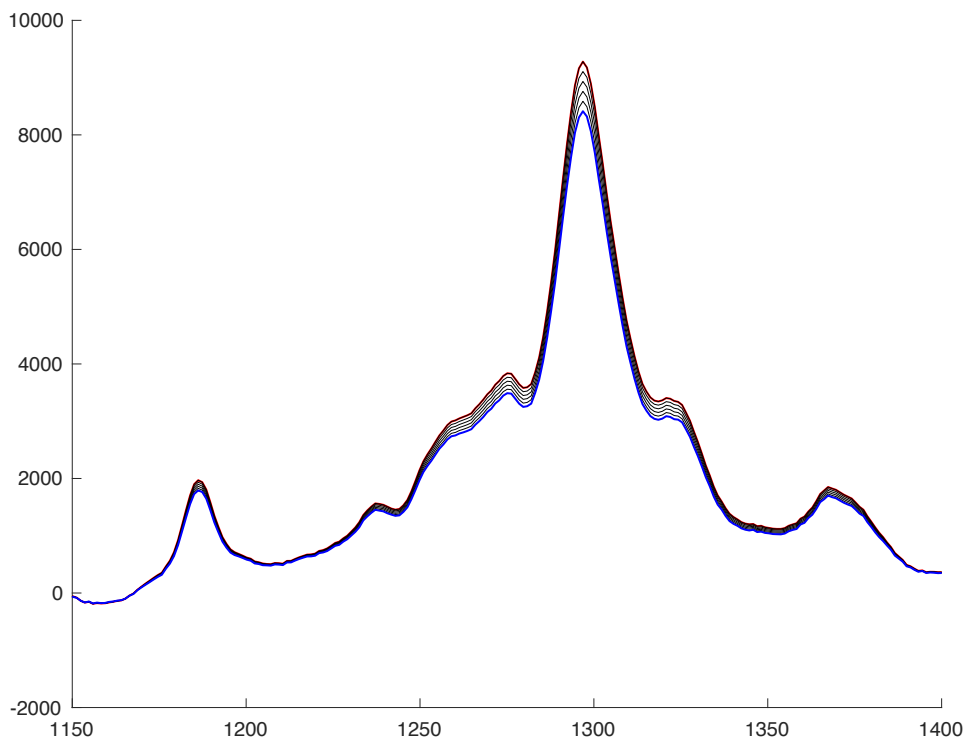
Figure 47. Representation of how the average spectrum change moving along the first principal component in the region between 1150 and 1400 cm^{-1}. The blue curve represents the final spectrum, the black lines represent how the peaks change moving along the first principal component and the red curve represents the average spectrum.
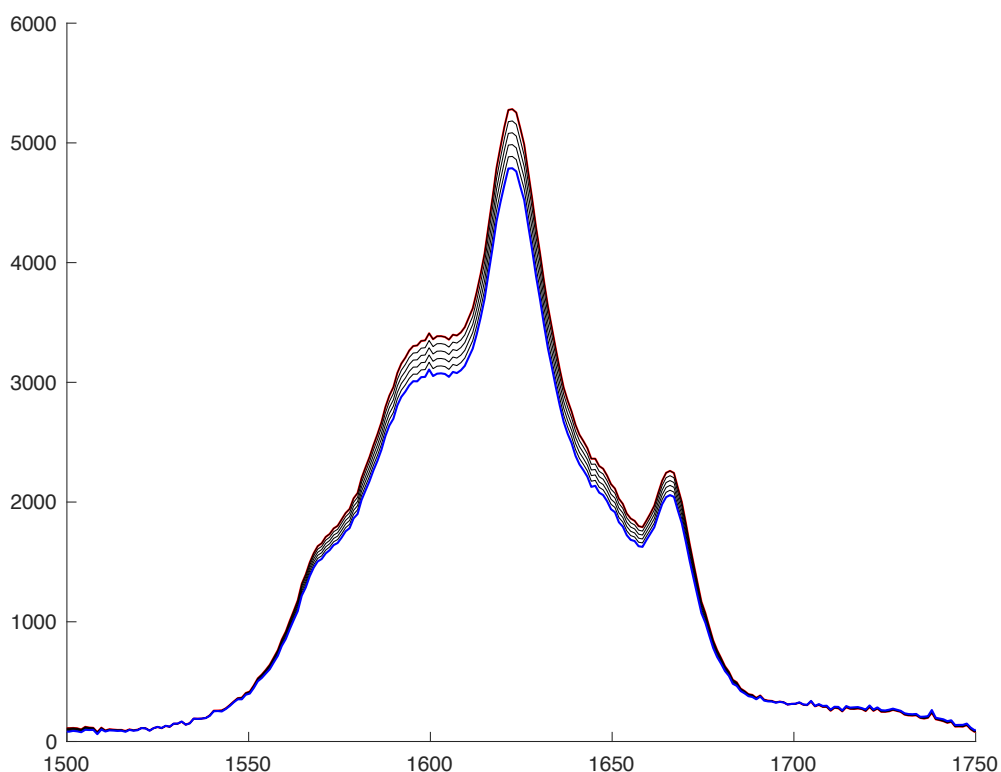
Figure 48. Representation of how the average spectrum change moving along the first principal component in the region between 1500 and 1750 cm^{-1}. The blue curve represents the final spectrum, the black lines represent how the peaks change moving along the first principal component and the red curve represents the average spectrum.

Once we have identified the relevant features in the PC2 loading and we have observed how the Raman spectra evolve moving along PC2, we can obtain chemical information by considering the assignment of the major peaks of the spectrum to selected cannabinoids (see Section 4.1 above). We can thus determine which chemical compounds are responsible for the change of intensity of these peaks along the selected principal component. Since the peak at 1297 cm^{-1} is associated to THC and THCA, and the one at 1623 cm^{-1} is associated, besides THC and THCA, also to other cannabinoids, we can expect that when the intensity ratio between THC/THCA and the other compounds changes, the peak at 1297 cm^{-1}, which is only associated to THC and THCA, will increase in intensity. We should not notice a large change in intensity for the peak at 1623 cm^{-1}: this happens because all the chemical compounds analysed so far contribute to that peak. Therefore, I can conclude that when the second principal component is high there is high concentration of THC and THCA.
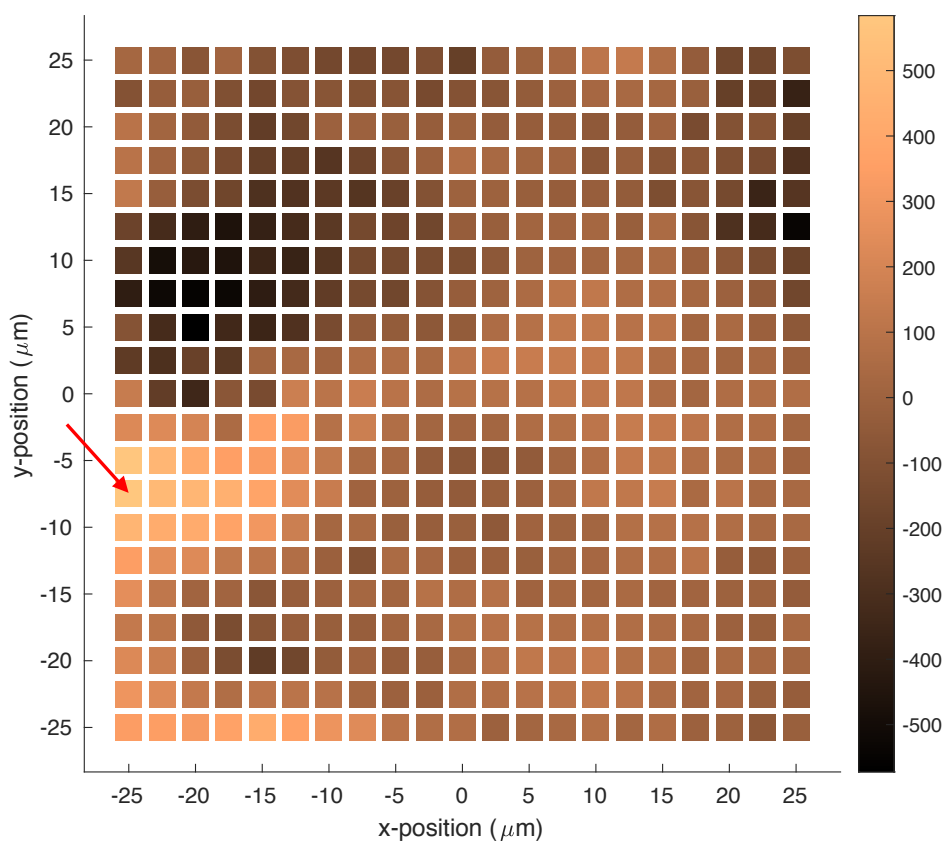
Figure 49. Scoremap of the second principal component. The red arrow indicates the coordinate selected for the analysis of the single spectrum.

By PCA it is then possible to get qualitative information about the relative concentration of selected cannabinoids in different regions of the sample. However, it is not straightforward to quantify how much this increase in concentration corresponds to. To determine the exact change in concentration, a dedicated calibration would be necessary, with *ad hoc* prepared samples – i.e., with known cannabinoid concentrations.

Looking now at the score map 2 (figure 49), we can say that the bright spots are associated to a point with a high concentration of THC and THCA, while the dark spots correspond to a point with a low concentration.

By using Matlab, it is possible to extract the spectrum associated to each point in the scoremap, (given the set of coordinates). This can help us double check the considerations done insofar. In fact, we expect that in the spectra corresponding to the bright pixel we have high intensity of the peaks associated to THC and THCA. We expect the peak at 1297 cm^{-1} to have high intensity, since it is only associated to THC and THCA, and a secondary increase of intensity of the peak at 1623 cm^{-1} (as discussed above, the intensity of the 1623 cm^{-1} peak is not only due to THC and THCA, but also to other cannabinoids, which makes less direct the interpretation of this marker band). For instance, we can extract the spectrum of coordinates (-25,-7.5), which corresponds to a bright spot in scoremap 2, with expected high concentrations of THC and THCA.

69

The representation of the spectrum selected from the scoremap (Figure 50), compared with the average spectrum, allows to quickly identify the rise of intensity of the relevant peaks. Indeed, we see that the peaks at 1186 and 1297 cm^{-1} (associated to THC and THCA only) have a higher intensity if compared to the average spectrum (red spectrum). Furthermore, also the peak at 1623 cm^{-1} has a higher intensity, and this may also be due to the increase in concentration of THC and THCA, but not exclusively (see above).
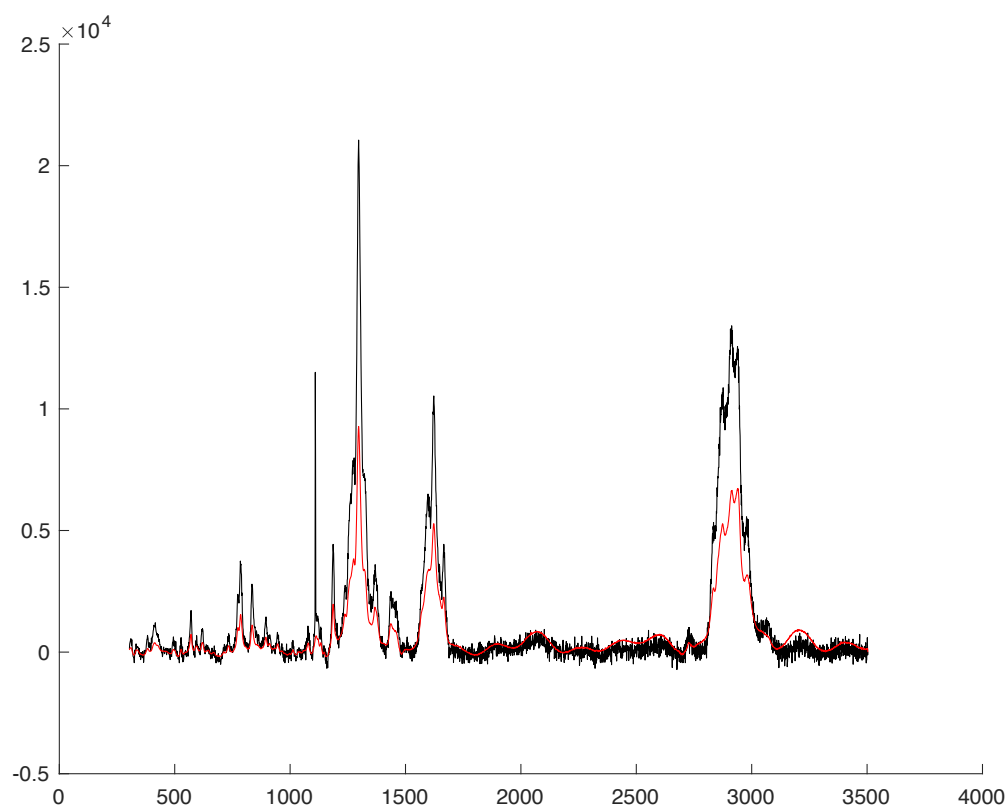
Figure 50. Superposition of the spectrum of the coordinates (-25,-7.5) in black and the average spectrum (red).

From the scoremap of PC2 is it possible to identify brighter regions that have a dimension of about 8-10 μm, and that according to the analysis done so far, represent points with higher THC/THCA concentration. We can compare the score map obtained with the dimension of the vesicle in the *Cannabis Sativa* plant to see why THC is accumulated in that specific region of space. The dimension of trichomes in a living plant shows a diameter that ranges from about 76–84 μm, with a mean of 80.1 μm [7]. After 25 days, the mean shrinkage in diameter was reported to be 13.7 %; after one year, the mean reported shrinkage was 24.5 %. The mean shrinkage of fresh material subjected to oven drying for 24 h at 103 °C was 39.9 % [22]. We can assume that each trichome contains on the average 8 vesicles [3], so for a representative trichome size of 65 micron we can deduce the dimension of the vesicle between 6 and 10 μm. The length of the bright region in scoremap 2 is 8 μm, that is compatible with the size of a vesicle [3]. We may conclude that the bright regions, which are characterized by a higher accumulation of THC and THCA, represent the vesicles of the trichomes.

4.3 Investigation of WiRES software performance for the automatic fluorescence background correction

As discussed before, the spectrometer used to analyze the trichome sample is coupled with a software used to automatically correct the baseline and filter the fluorescence background. This software could be responsible for the creation of artifacts such as a background modulation which makes the spectra confusing and complex to analyze. This could be a limitation of the instrument worth investigating to avoid misleading data interpretation. To obtain a clearer point of view of the situation, one can manually remove from the dataset those observations that are presumably harder to correct by the WiRES software because the Raman signal is too low compared to the background. Such datapoints are expected to correspond to the peripheral regions of the mapping that are out of focus – as discussed in presenting the interpretation of the first principal component. By using Matlab we can extract the spectrum for each specific coordinate and plot it together with the average spectrum for comparison. There are some locations in the mapping where the spectrum appears markedly different from the average. To identify those points that can be considered valid, we can monitor the intensity of the strong peak at 1297 cm^{-1} to make sure that it is well above the modulated background signal. When the 1287 cm^{-1} peak is lower than half the modulation, we can assume that the background artifact is too high, and that the spectrum is not reliable enough to get relevant chemical information. By scrutinizing the single spectra for selected coordinates, I have realized that when moving towards higher x, for fixed values of y, the peaks intensity decreases, until in some cases there is so much noise that the spectrum is not like the average spectrum anymore.

Figure 51. Superposition of the spectrum corresponding to the coordinates (-20; 0) (black) with the average spectrum (red).

71

For instance, let us select the point at coordinates (-20; 0), which corresponds to a black pixel in the scoremap of PC2 (Figure 39). Figure 51 represents the spectrum corresponding to that coordinate (black) and it compares it with the average spectrum (red). As we can notice, the peaks are very clear and resemble much the average spectrum. We can notice that the peaks of the region between 1000-1500 cm^{-1} are well defined, with not much background modulation, and the intensity of the most intense peak (1297 cm^{-1}) is also well defined. Hence, in this case we can assume that the spectrum of coordinates (-20; 0) is not too noisy and it contributes to the creation of a reliable average spectrum.

Instead, if we select a different coordinate, towards higher values of x, e.g., (25; 0), we can see that the spectrum will appear completely different from the average spectrum. In this case, the peaks are not clearly identifiable anymore (Figure 52). In the region between 1000-1500 cm^{-1}, which in the previous case was characterized by clear and sharp peaks, there is now a strong background modulation. We are not able to identify clear peaks anymore and the most intense peak (1297 cm^{-1}) does not exceed more than one time the intensity of the modulation. Hence, we conclude that the spectrum in correspondence of this coordinate is too noisy, not reliable and it will contribute to the dataset by adding a modulation to the background, thus making the data interpretation more complex.
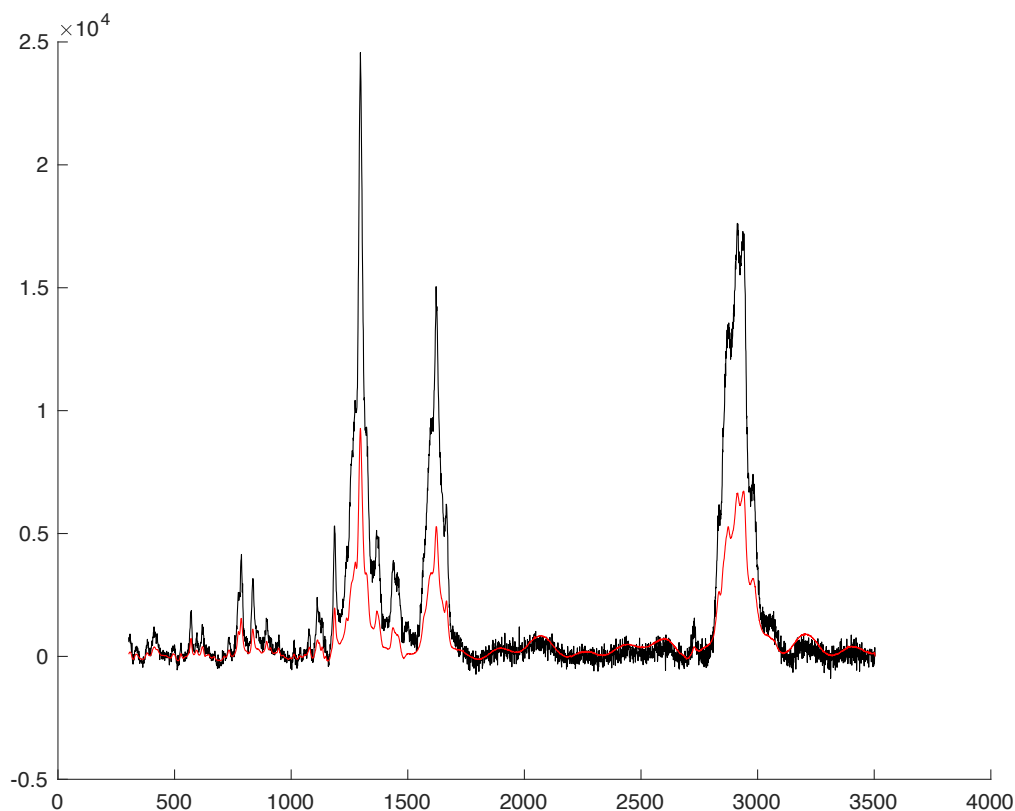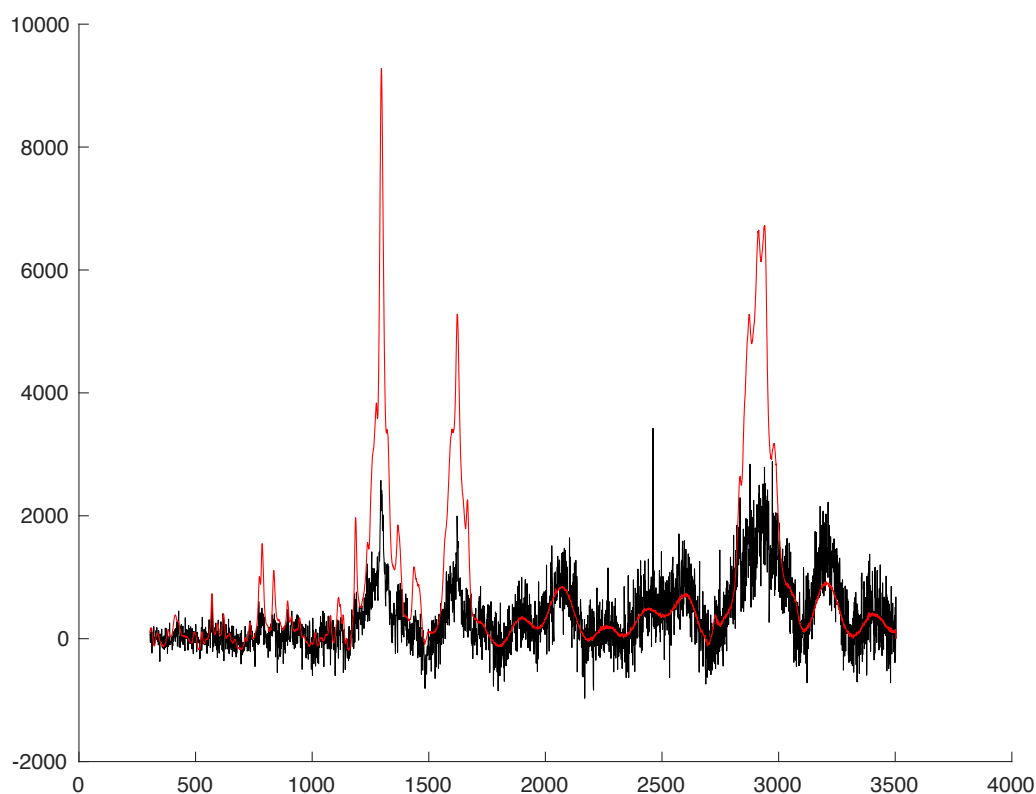

Figure 52. Superposition of the spectrum corresponding to the coordinates (25; 0) (black) with the average spectrum (red).

By following the same approach discussed above for different coordinates we can identify all those coordinates in the mapping that should be excluded from our PCA calculation to obtain

72

a better-quality average spectrum and loadings from which we can expect to obtain clearer chemical information. This identification of the outliers was done just by observing the spectra corresponding to different coordinates and assuming that if the intensity of the band of largest intensity (1297 cm^{-1}) is not higher than the intensity of the modulation, the signal is too noisy and must be excluded. The same analysis can also be done using a mathematical approach and defining a ratio signal/noise. However, the latter approach would require a more complex definition and Matlab programming. Instead, by looking at the scoremap of the first principal component, it is straightforward to see which color of the pixels corresponds to the noisy spectra that we want to exclude. To remove the coordinates associated to such noisy spectra we can use Matlab straightforwardly. We can select a specific threshold associated to the score of the pixel which are related to the noisy spectra and, selecting that threshold in Matlbab, we can reject all datapoints in the mapping for which the score along PC1 is less than that threshold. In this way one obtains a filtered dataset on which a new PCA calculation will provide a scoremap containing only the points with good focusing conditions during the Raman mapping.

The first threshold we can select is -1500 which, by looking at the score map 1, we can see that corresponds to the brighter pixels. By selecting this threshold, it is possible to obtain the loadings reported in figure 57 which are associated to the first four principal components.
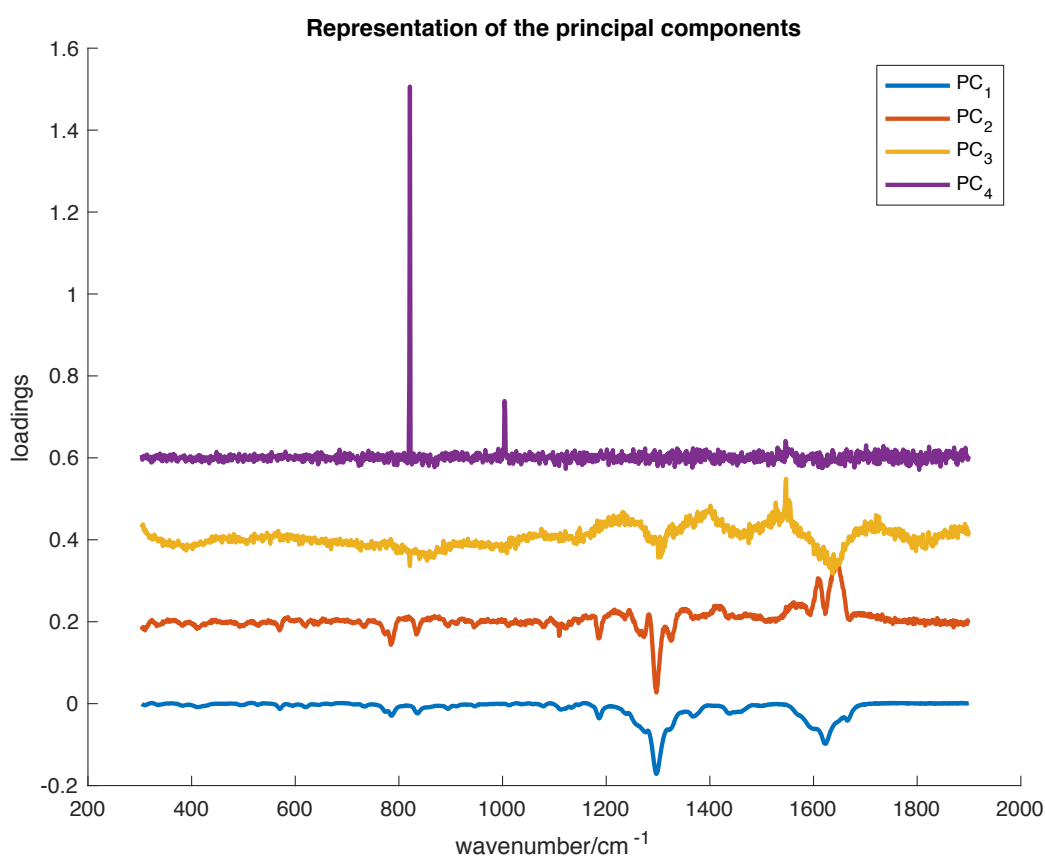
Figure 53. Loadings along PC1, PC2, PC3 and PC4 of the filtered dataset.

By examining the loadings along the principal components, we can describe them as follows:

1. Loading along PC1: it is very similar to the average spectrum and shows overall the same trend. It mainly correlates to the strength of the Raman signal and it does not contain any chemical information on the change of relative composition of the sample point to point.

2. Loading along PC2: it appears to be very different from the average spectrum, with peaks of opposite sign. It can be used to obtain chemical information considering, as done before, the assignment of the main peaks of the spectrum to selected cannabinoids.

3. Loading along PC3: it has mostly an undulatory behavior and does not show clear peaks. Most likely, PC3 is the signature of the automatic fluorescence background correction introduced by the instrument software (WiRE)

4. Loading along PC4: as expected, the loading along PC4 is a noisy signal, and we cannot obtain neither spatial nor spectroscopical information out of it. There is only one visible very sharp peak. This is just a spike originated by the interaction of cosmic rays with the CCD detector during the Raman mapping experiment. Therefore, it is not associated to any cannabinoid and cannot provide chemical information about the trichome.

Such assessment of the PCs is further supported by looking at the related scoremaps. The score maps associated to PC1 (figure 54) and PC2 (figure 55) show a high spatial variation of the score. For PC1 there are some dark (bright) regions that can be used to identify which points are better (worse) on focus. As for PC2, we retrieve a similar situation as in the first PCA of the complete dataset, with the scoremap highlighting the regions with relative higher concentration of THC/THCA in the sample.
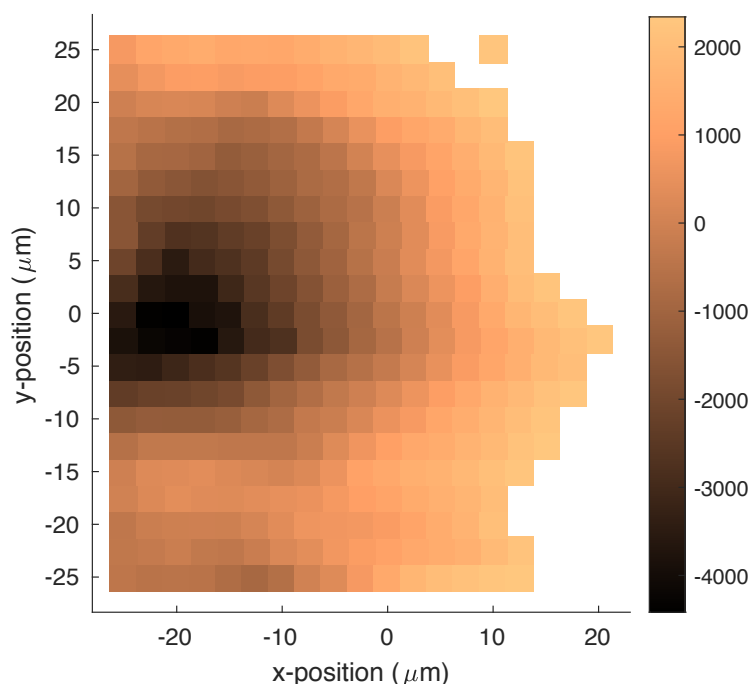
Figure 54. Scoremap one of the filtered dataset.
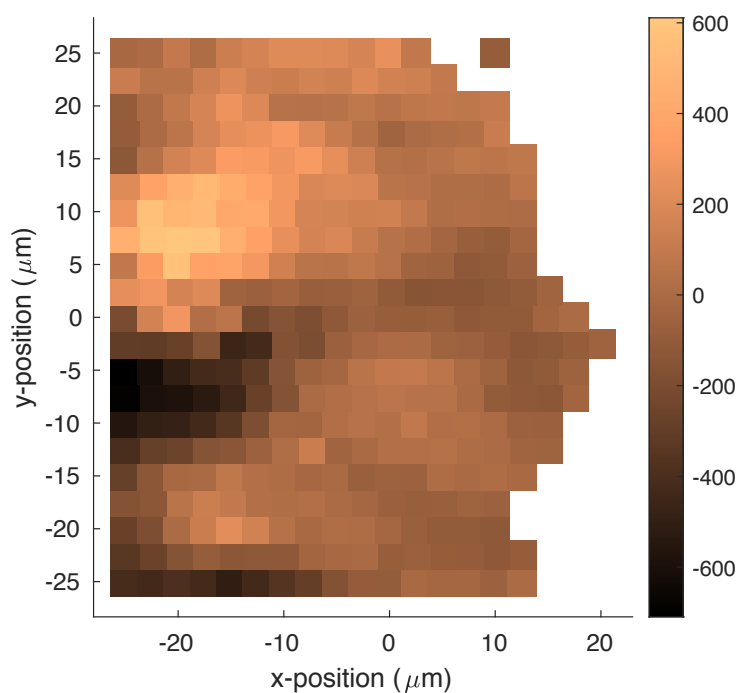
Figure 55. Scoremap two of the filtered dataset.

Instead, the PC3 scoremap (figure 56), does not show a clear and well-defined structure. It is a rather noisy map where one cannot see a gradual modification of the score from point to point; the map is overall scattered and confused. This matches with the conclusions reached by observing the loading along the third principal component, which presents an undulatory behavior due to artifacts produced by the WiRE software in the procedure for the automatic removal of the fluorescence background.
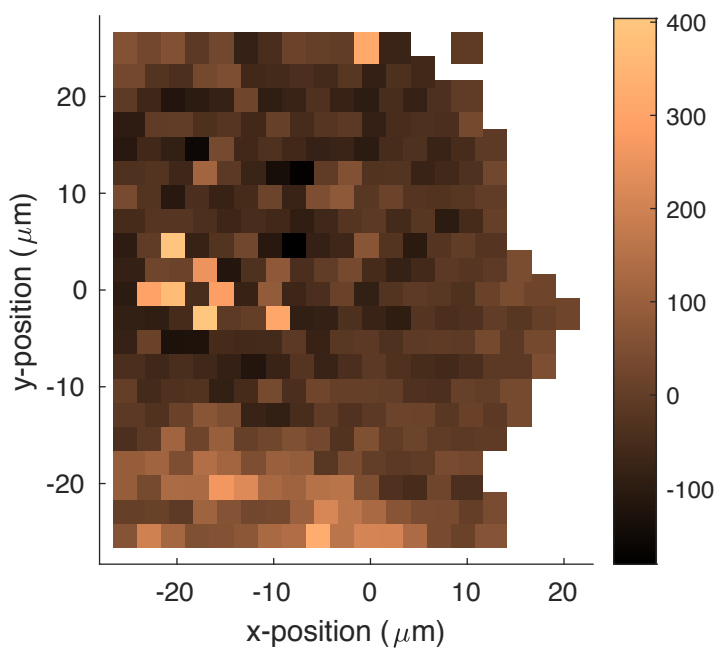
Figure 56. Scoremap three of the filtered dataset.
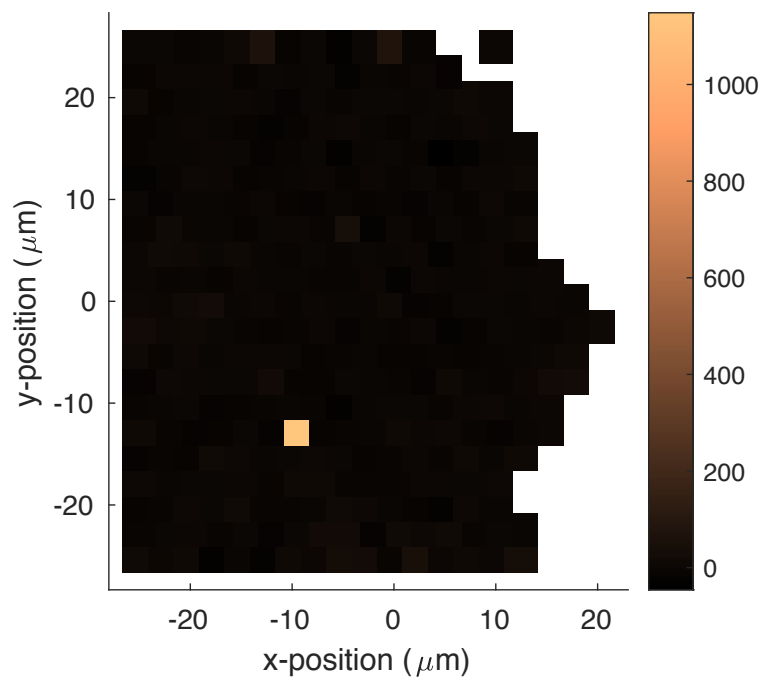
Figure 57. Scoremap four of the filtered dataset.

A similar consideration can be done for the scoremap of the fourth principal component (figure 57). Such scoremap is completely dark, we see no variation of the score nor of the signal. No spatial nor spectroscopic information can be obtained from this principal component.

The relevance of the principal components can be also assessed by looking at the screeplot (Figure 58). It is clear that the third principal component still has a high statistic relevance. This is not ideal since the corresponding loading is interpreted as artifact and it is still desirable to reduce its effects on the PCA results.
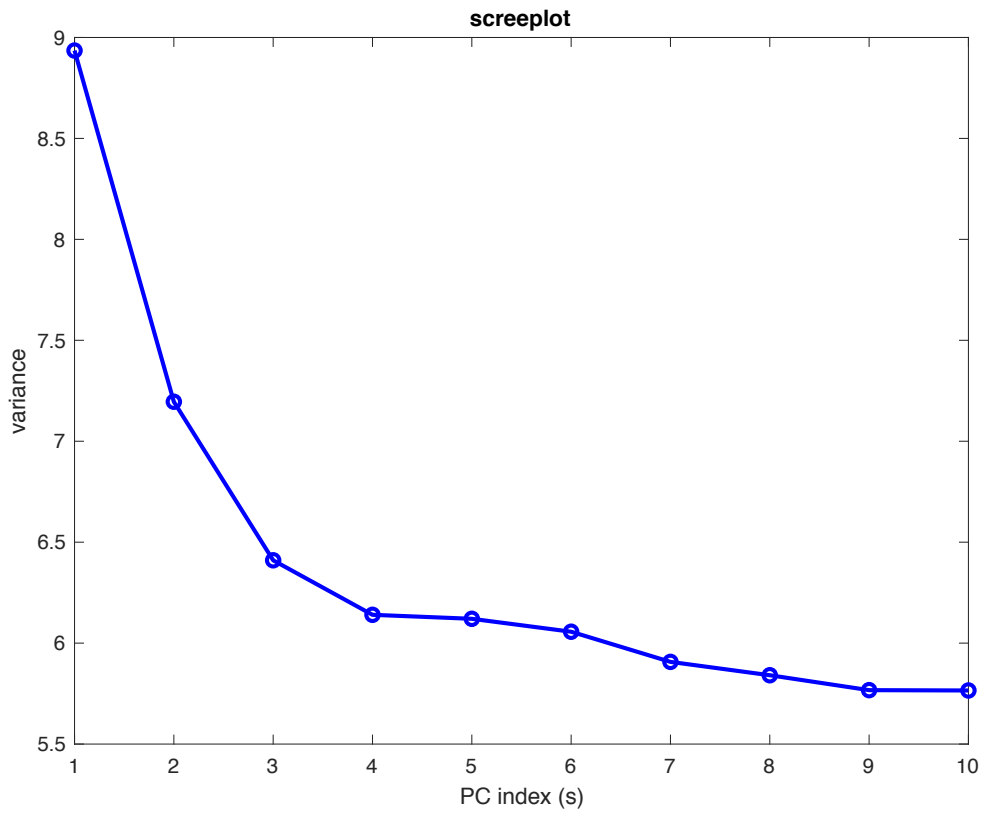
Figure 58. Screeplot of the filtered dataset.
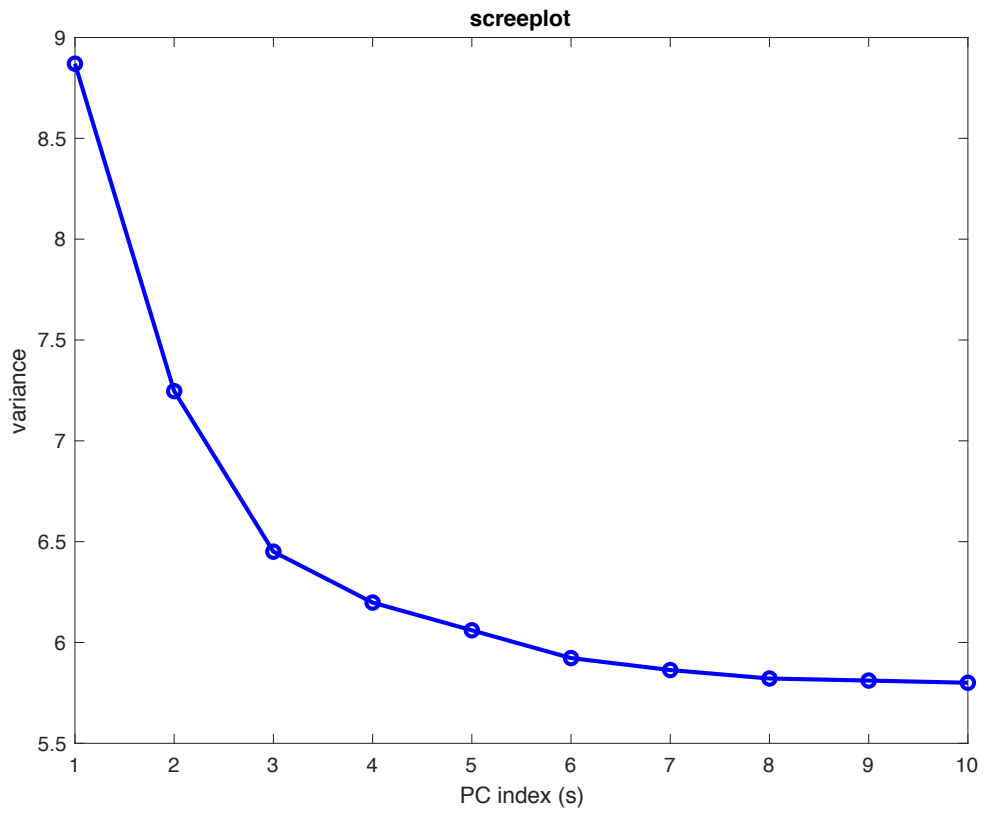
Figure 59. Screeplot of the filtered dataset.

Therefore, to attempt to solve this problem a new more restrictive threshold was selected, at -1000. However, by looking at the updated screeplot (Figure 59), the third principal component still appears to have a statistic relevance which is too high.

The loadings along the first four principal components (figure 60) have a very similar trend to the previous case (Figure 53). Therefore, a similar reasoning to what has been done in the first case (-1500) can be repeated.
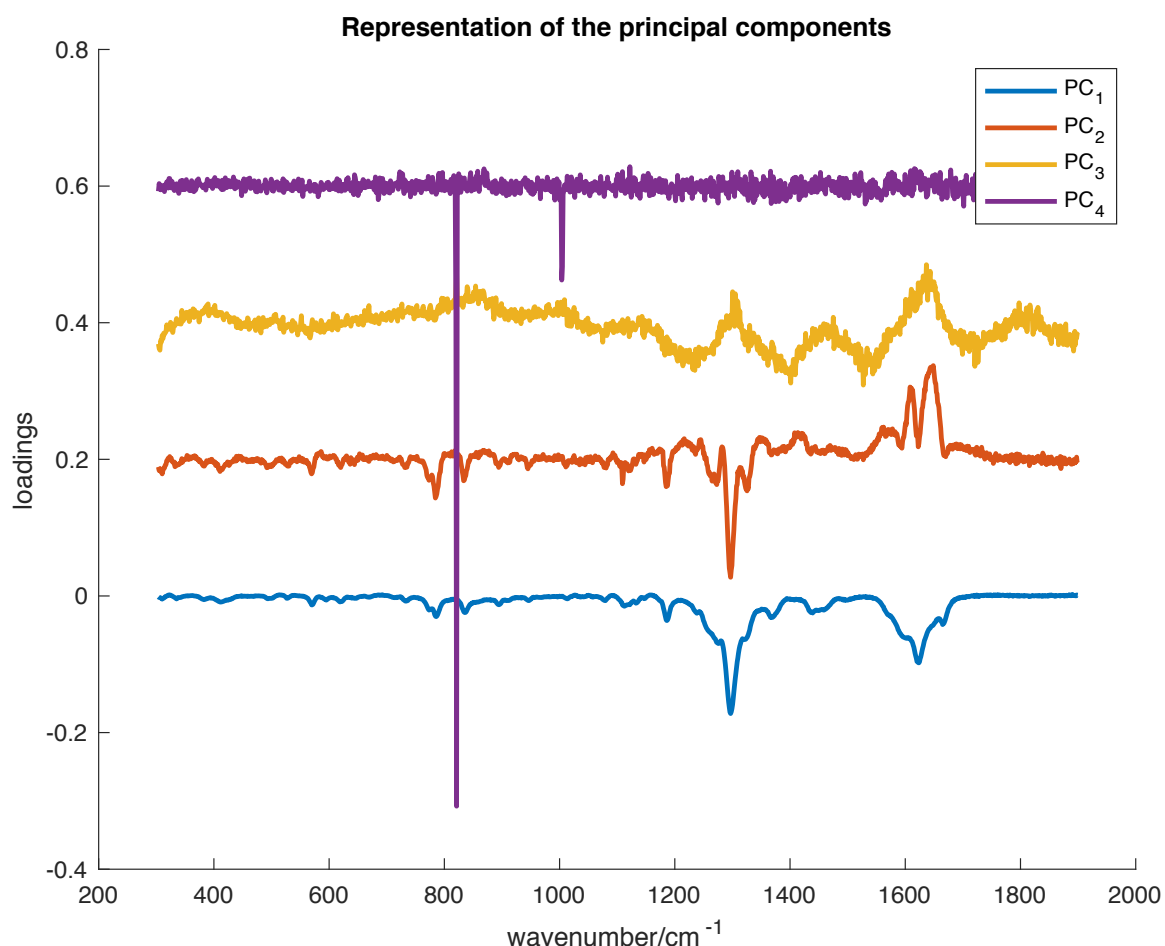
Figure 60. Loadings along PC1, PC2, PC3 and PC4 of the filtered dataset.

Similar conclusions can be drawn observing the different score maps. Scoremap 1 (figure 61) and 2 (figure 62) show a large score variation, while the one correspondent to PC4 (figure 64) is completely bright. Since it is completely flat, we cannot expect any relevant spatial information. Finally, the score map correspondent to PC3 (figure 63) has some pixels of different scores, but they are only confused fluctuations of the signal, symbol of a noisy map.
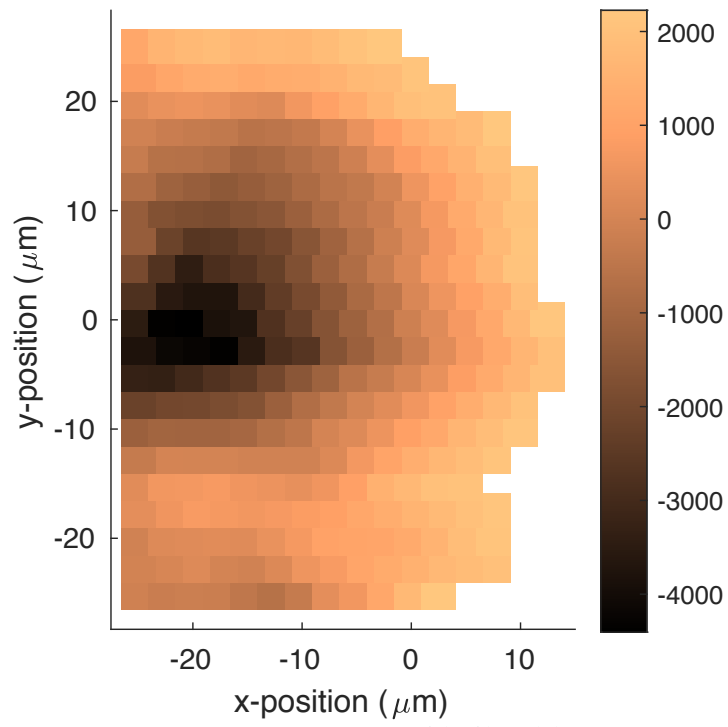
Figure 61. Scoremap one of the filtered dataset.
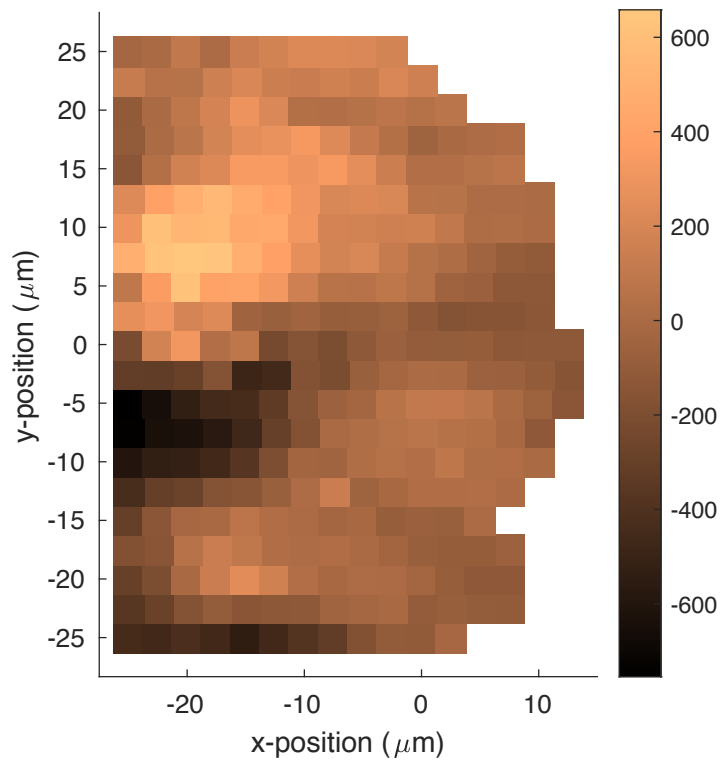
Figure 62. Scoremap two of the filtered dataset.
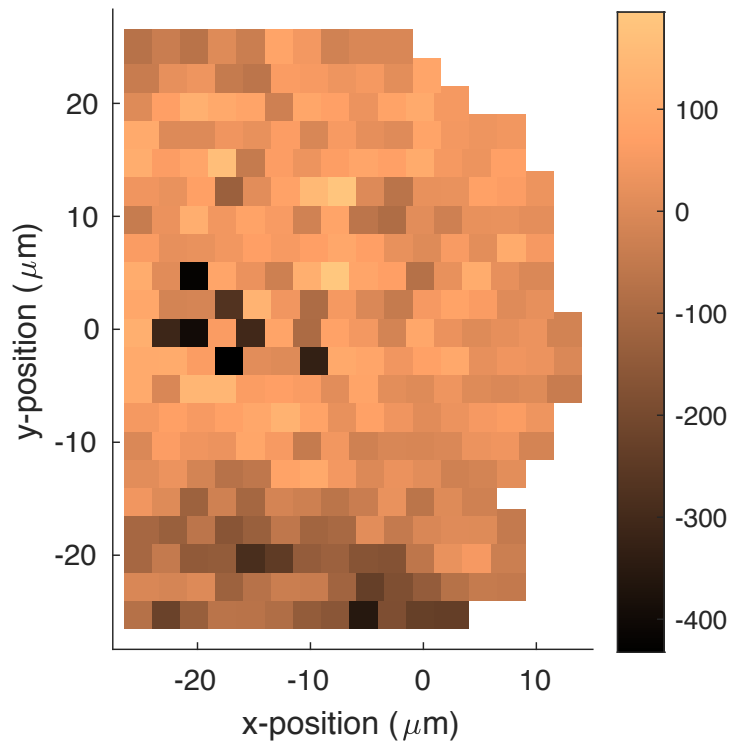
Figure 63. Scoremap three of the filtered dataset.
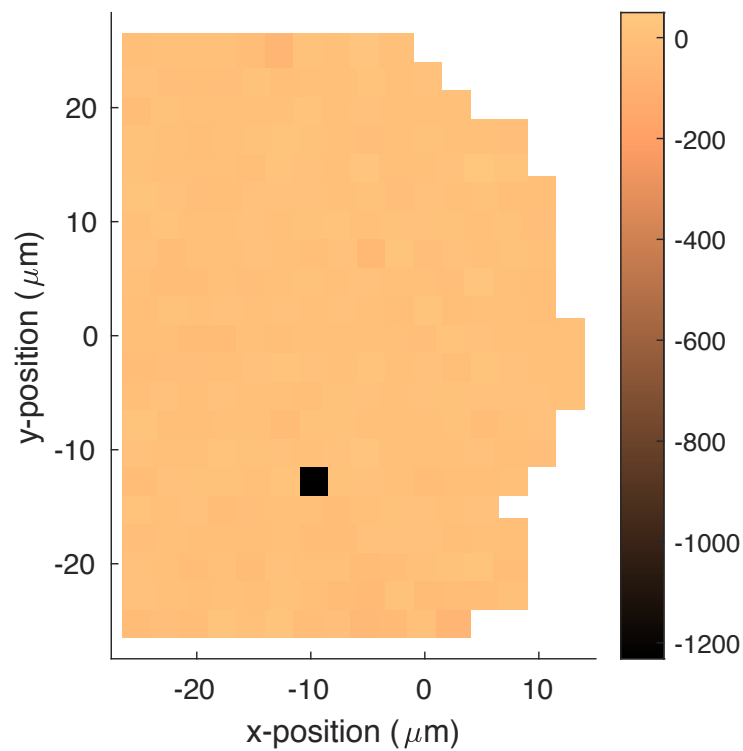
Figure 64. Scoremap four of the filtered dataset.

Therefore, a further restriction of the threshold was applied, by selecting a value of +500. What we expect is a change of the loading of PC3, which we expect to lose its undulatory behavior.

By looking at the loading along PC1, PC2 and PC4 we see no change, and the same considerations done before still hold. The loading along PC3, instead, appears to be different. It does not have the same trend it had before, and it lost some of the undulatory behavior which was characterizing it (figure 65).

Thereafter, by looking at the score maps, we can see that the first and the second PC show fluctuations which contain relevant information. As before, PC1 (figure 66) is related to the focusing conditions in the Raman mapping; PC2 (figure 67) instead gives information on the compositional difference across the sample. From the considerations done in the previous paragraphs, we can assume that the bright regions are associated to a higher concentration and accumulation of THC and THCA, while the darker ones are associated to a lower concentration. Such bright regions that can be clearly identified in the score map of PC2, correlate with the dimension of the secretory vesicles in which cannabinoids and more in general metabolites are secreted within the trichomes [3].

The third principal component (figure 68) does not show a clear structure, but only random fluctuations. We cannot identify a clear modification of the signal from point to point, and it is a rather noisy map. Besides the third principal component, we can also conclude that PC4 (figure 69) is just noise and does not provide useful chemical/spatial information.
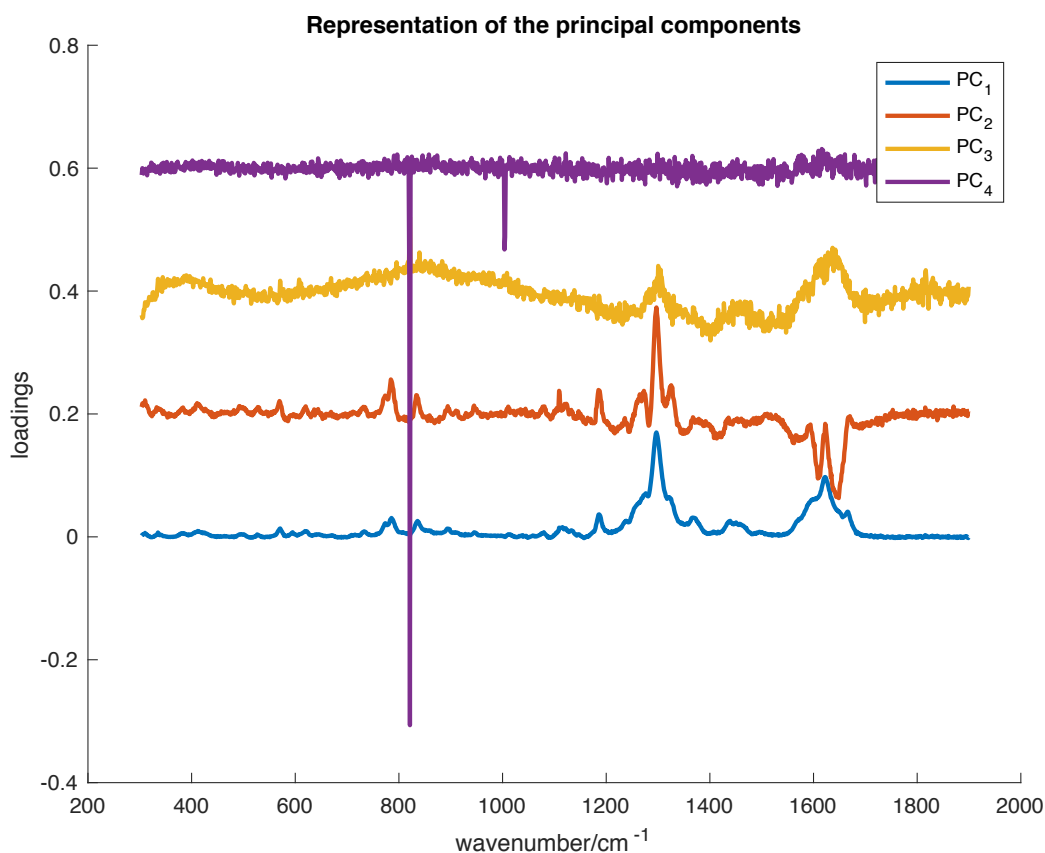
Figure 65. Loadings along PC1, PC2, PC3 and PC4 of the filtered dataset.
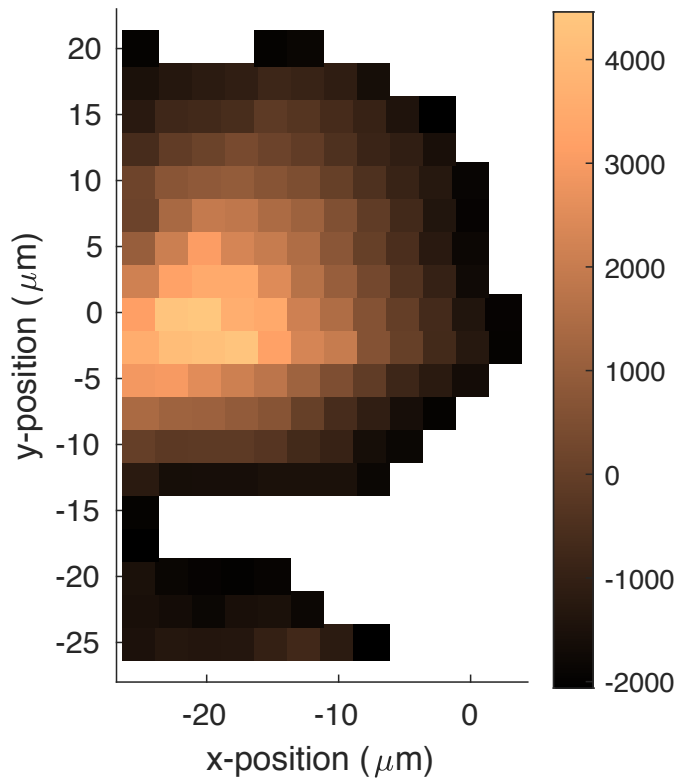
Figure 66. Scoremap one of the filtered dataset.
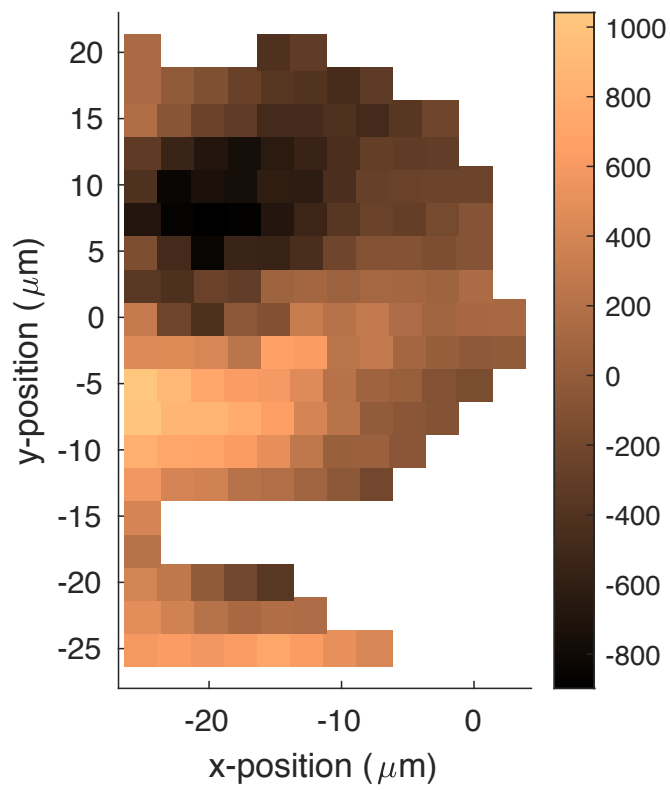
Figure 67. Score map two of the filtered dataset.
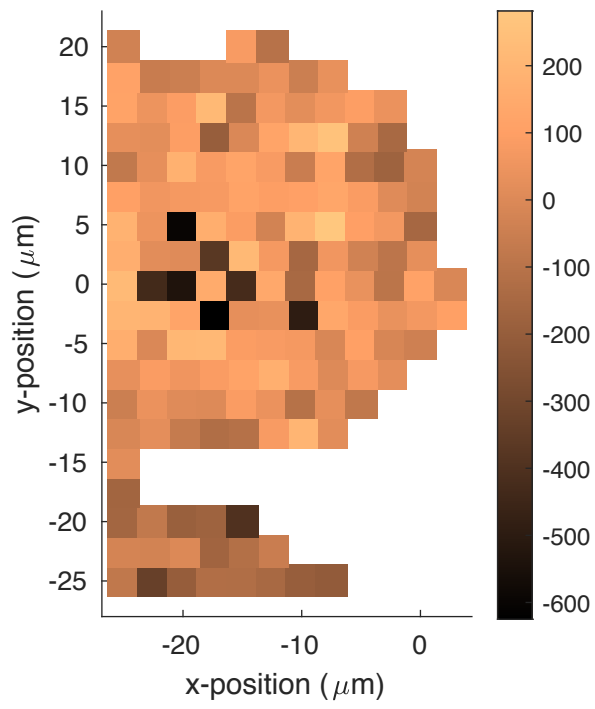
Figure 68. Scoremap three of the filtered dataset.
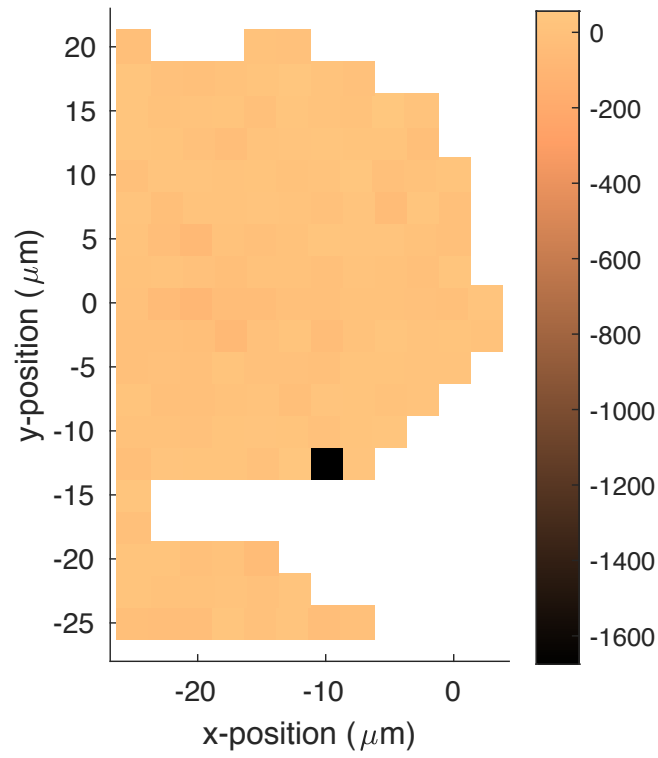
Figure 69. Scoremap four of the filtered dataset.

Figure 70 represents the screeplot of the newly filtered data. The relative importance of PC3 is still sizeable, however PC3 may be neglected, as it does not seem to provide any relevant information. PC3 may have some spectroscopic information in the loading vector, but this does not find a justification in the spatial analysis of the associated scoremap. Therefore, we do not consider it any further.
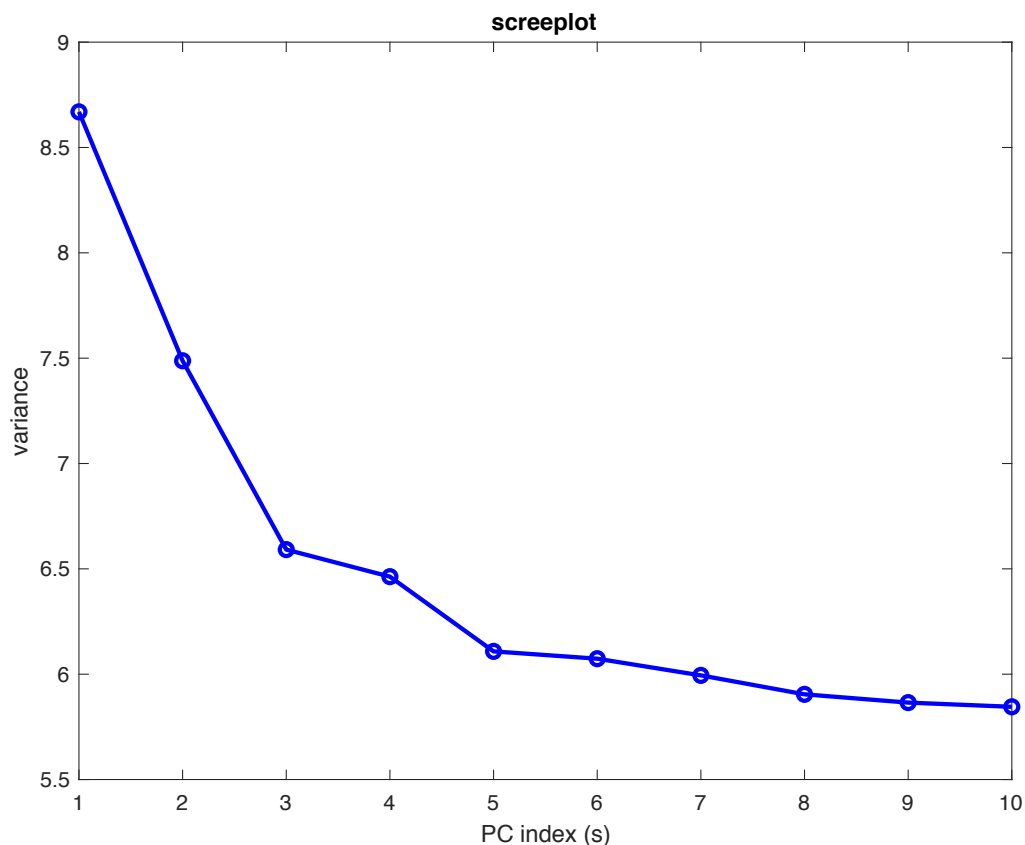
Figure 70. Screeplot of the filtered data.

4.4 PCA of the Raman mapping of the trichome including the fluorescence background

The average spectrum of the trichomes and the loadings along the principal components have been analyzed in their raw form, to try to obtain a clearer assessment of the Raman mapping data, without the background artifacts introduced by the WiRE correction algorithm. This was possible by merely applying the Matlab scripts developed for PCA [17] directly on the raw mapping data collected on the trichome.

The screeplot of such PCA is reported in Figure 71, which shows that the weight of the first principal component is dominant. By comparison with respect to the previous PCA on the fluorescence-corrected data – where such dominant PC1 was absent – we expect that this component should be related with the fluorescence background.

In Figure 72 the variance of the principal components, excluded PC1, are represented to better understand the relevance of the other principal components. Finally, by plotting the curve on a semi-logarithmic scale (Figure 73), it is easier to see how quickly the intensity of the principal components decreases. For this reason, the loadings along the first four principal components were analyzed. The components starting from PC5 have been neglected since their variance is very low compared to the previous PCs.
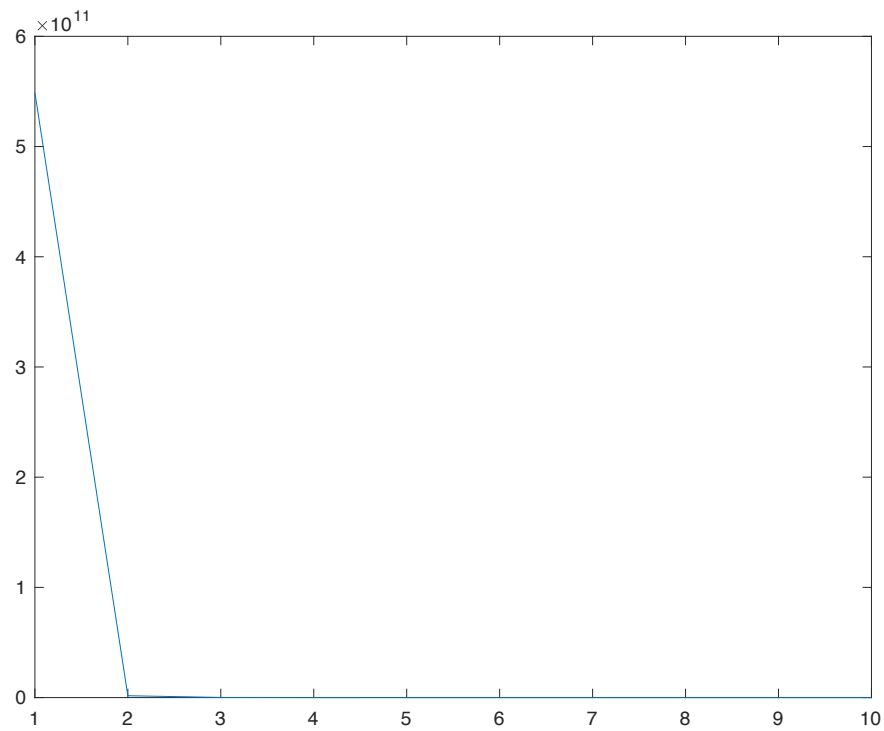
Figure 71. Screeplot of the principal components.
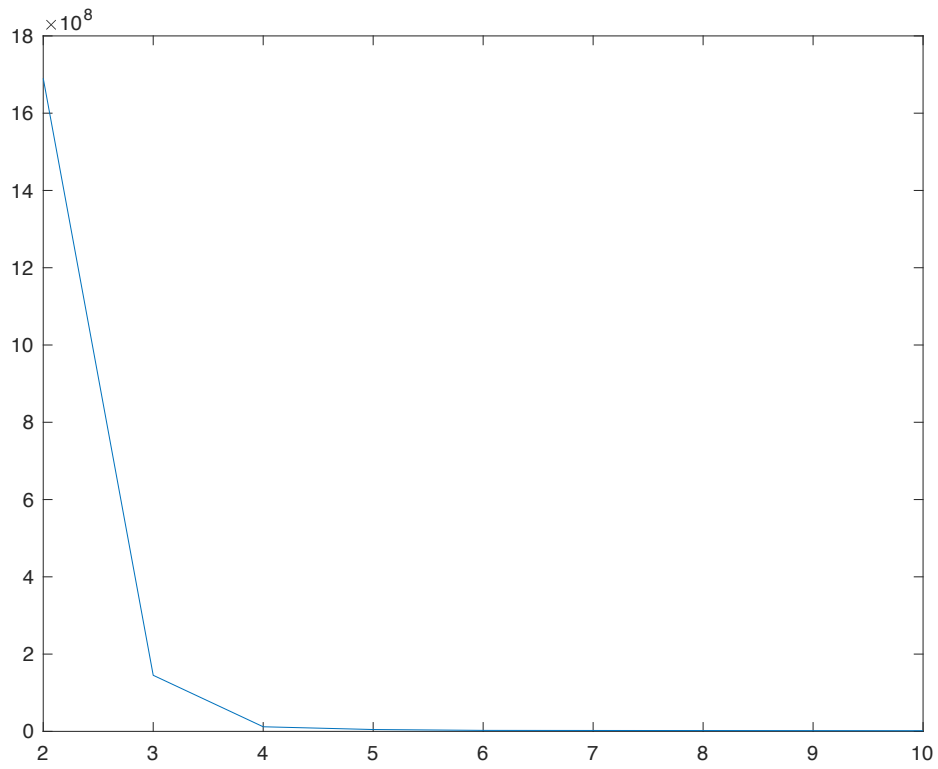
Figure 72. Score plot of the principal components starting from PC2.
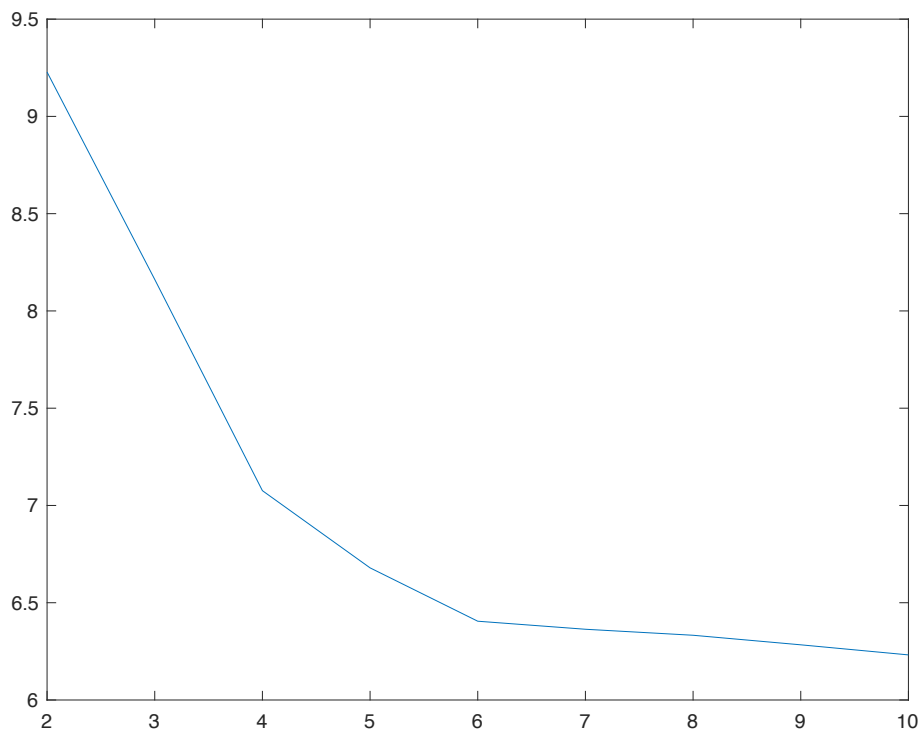
Figure 73. Score plot of the principal components starting from PC2 in the logarithmic scale.

Once the most relevant components have been selected, the correspondent loadings can be analyzed. The loadings along PC1 (black line) and PC2 (red line) are superimposed in Figure 74. The two curves were properly normalized and multiplied for a selected number so that the most intense Raman peak (1297 cm^{-1}) could have a similar intensity. This makes it easier to confront the two curves. The PC1 loadings PC1 clearly display an intense fluorescence background, which is almost missing in the loadings of PC2. This can be easily concluded since the black curve appears to be tilted and the maximum does not have a relevant intensity if compared to the difference between the maximum and the minimum of the curve. Instead, the fluorescence background in the loadings along the second principal component is negligible, for the same reason. In PC1 and PC2 it is not possible to fully decouple the fluorescence and Raman contributions, as PC1 and PC2 are both characterized by a fluorescence and Raman component. However, while in PC1 it is the fluorescence component to be more intense, in PC2 it is the Raman contribution to be dominant. Hence, with a little degree of approximation, when considering the associated scoremaps we may assume that PC1 describes the areas of the trichome that are more fluorescent, whereas PC2 indicates the areas that are more Raman active. Since the fluorescence signal is less strongly related to the chemical structure of the compound than the Raman signal, one may expect to get chemical information out of PC2 (and possibly the higher PCs).
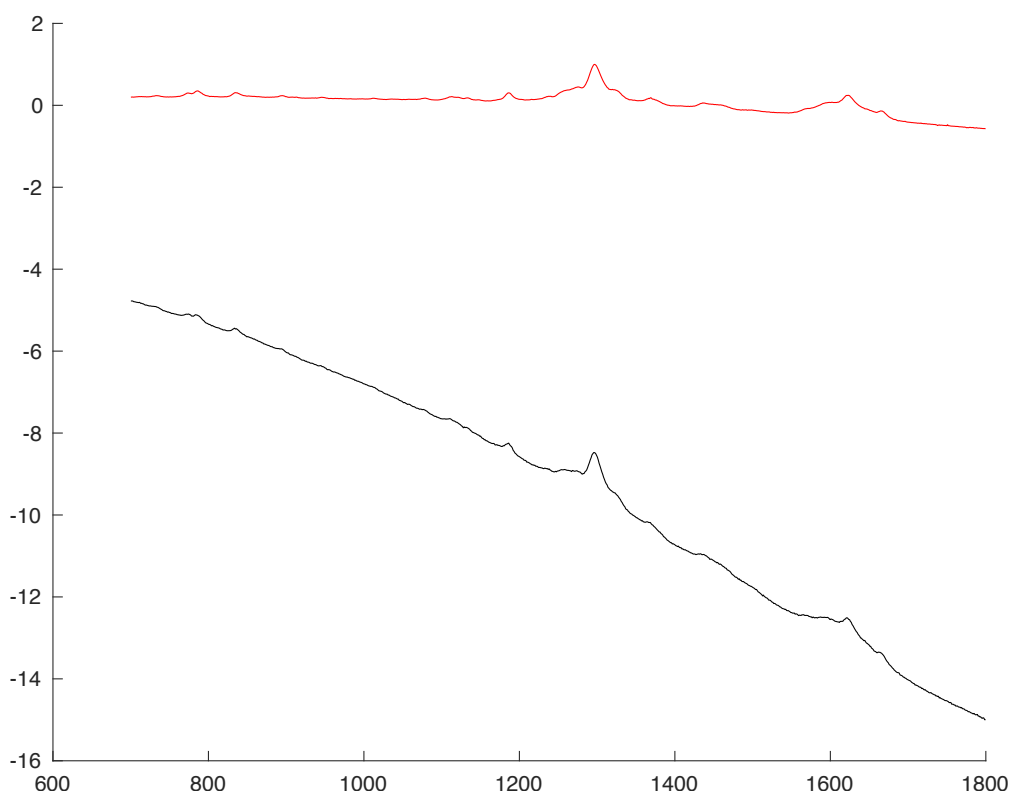
Figure 74. Loadings along PC1 (black) and PC2 (red).

The score map of the first principal component (figure 75) does not provide many details. There are strong signals or weak signals in different regions, but there is a low contrast and modulation in a wide central part of the map. The low modulation present in coordinates like (0, -20) or (0, -25) is due to the fact that PC1 does not contain just information on fluorescence but also on Raman. However, the PC2 score map (figure 76) is easier to read and characterized by higher modulation in the central part of the grid. This tells us where we have a good Raman signal, however it does not provide yet information about compositional changes from point to point.
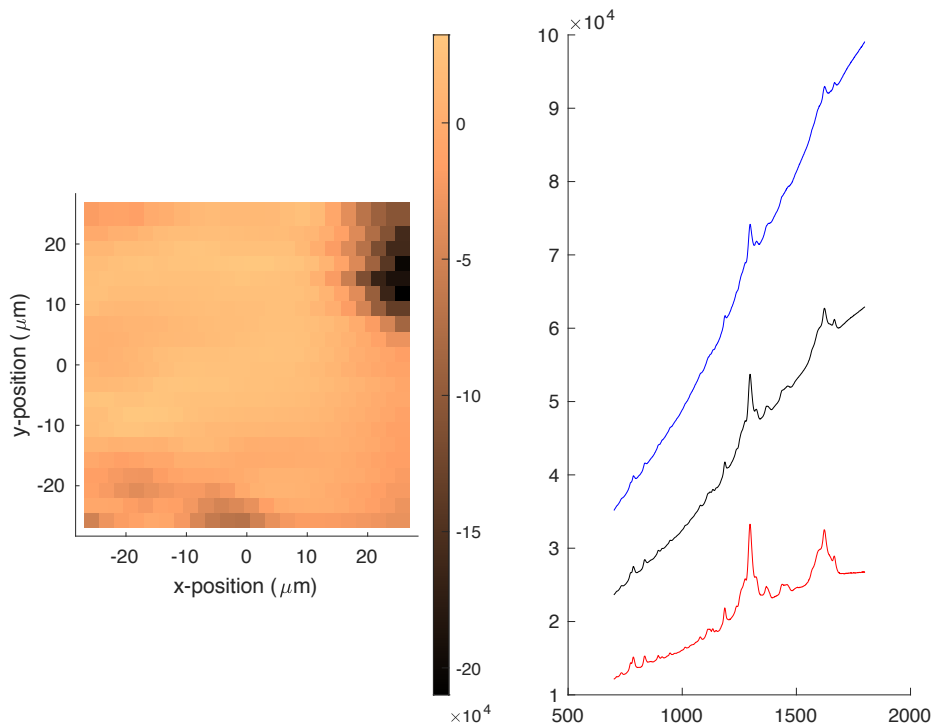
Figure 75. Score map of PC1.

88
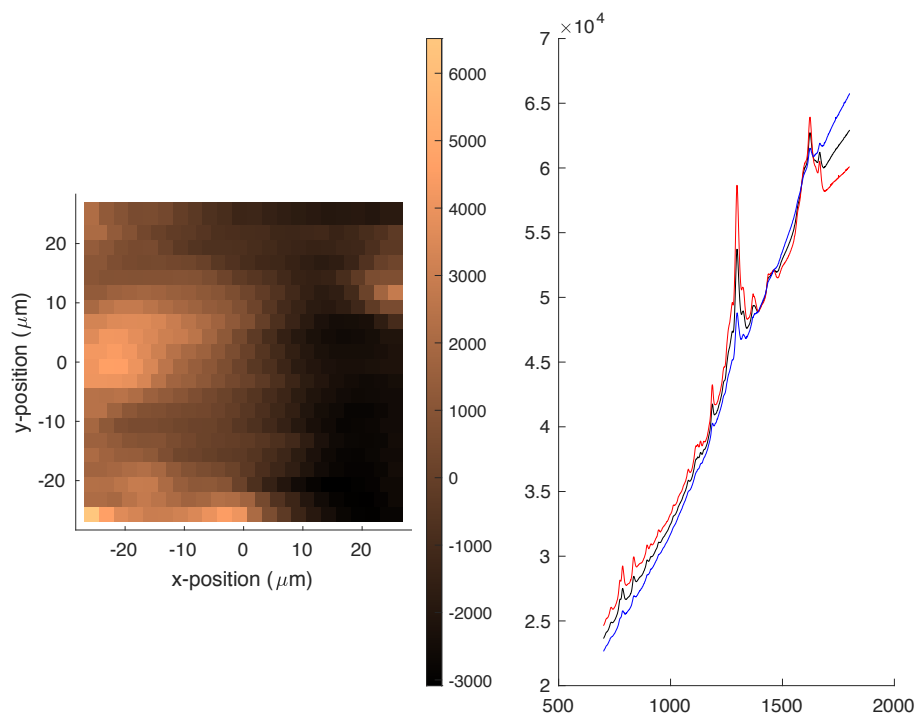
Figure 76. Score map of PC2.

The loadings along PC2 and PC3 are really similar and their intensity ratios are similar as well. We can superpose them in the same graph to compare them easier (figure 77). PC2 is represented in red, PC3 in blue. The differences that can be noticed are two:

1. in the region around 1600 cm^{-1} the blue curve shows somewhat richer sub-band details, if compared to the red one;

2. the two curves are characterized by a different fluorescence background. The red curve has a fluorescence background that tends to move downwards, the blue upwards.

Since the peaks of PC2 and PC3 are similar, we cannot easily identify chemical differences associated to the two components. The major difference between the two is the relative

weight of the fluorescence background
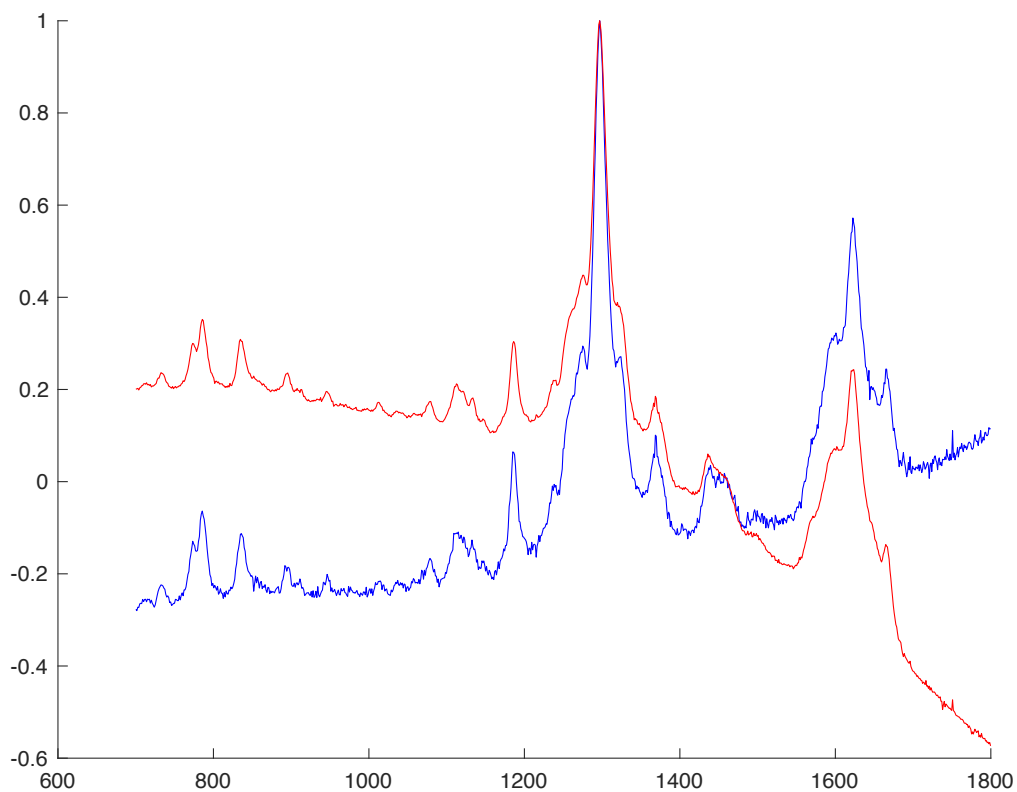
Figure 77. Superposition of the loadings along PC2 and PC3.

The loading along the fourth principal component (figure 78), instead, contains relevant chemical information and it is fairly similar to the second principal component that was examined on the florescence-corrected dataset. The peak at 1297 cm^{-1} is associated to THC and THCA, whereas the one at 1623 cm^{-1} is associated to all the cannabinoids.
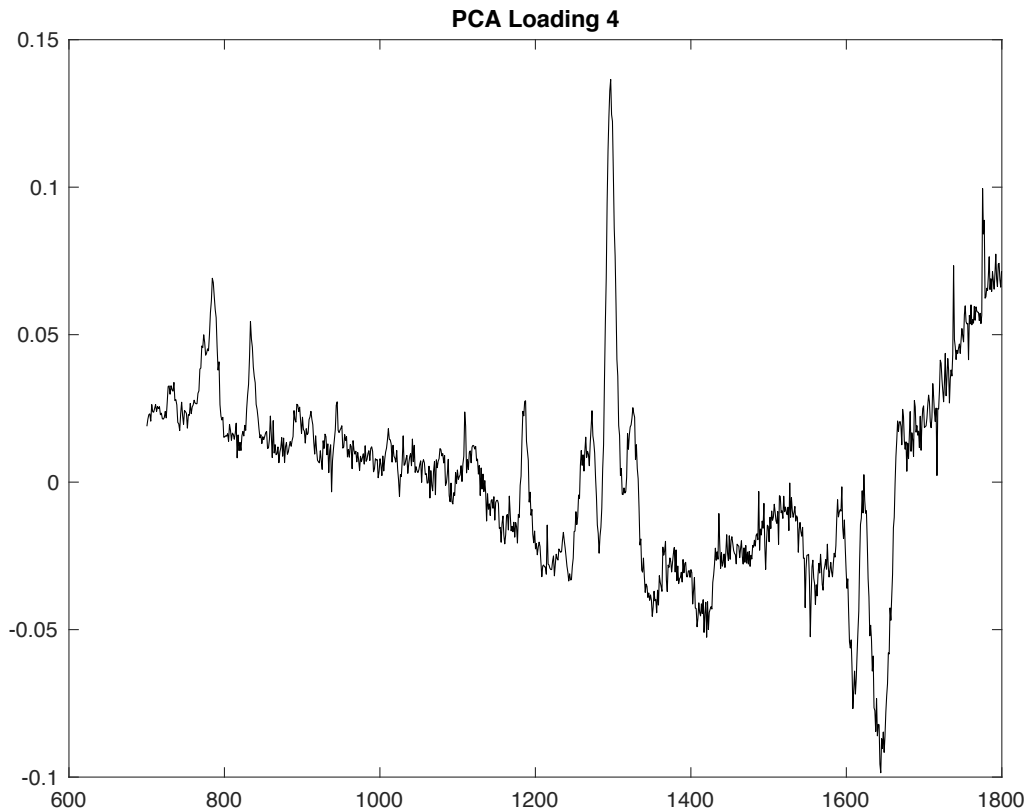
Figure 78. Loading along PC4.

This means that the bright pixels in the score map (figure 79) are associated to higher concentration of THC and THCA. Looking at the score map is it also possible to see three circular bright regions. These correspond to the vesicle of the trichomes. We can in fact determine a diameter of around 10 μm, which matches the expected vesicle size based on literature data [3]. The dark pixels at the top of the score map (y=25) are associated to out of focus points. The spectra associated to those coordinates are quite noisy. We can identify the darkest point in the score map of PC4, which corresponds to the coordinates (-22.5,10). By using Matlab it is possible to extract the spectra for that point in the map and plot it as in figure 80. The distinct peaks which can be identified in the spectrum are associated to THC and THCA. This confirms that the loading along the fourth principal component gives us chemical information about the concentration of THC and THCA in the trichome.

If we select a coordinate which corresponds to a bright spot, for example (-25, -10), we can extract the spectrum using Matlab. This spectrum is reported in figure 81. The main peaks are the ones associated to THC and THCA.
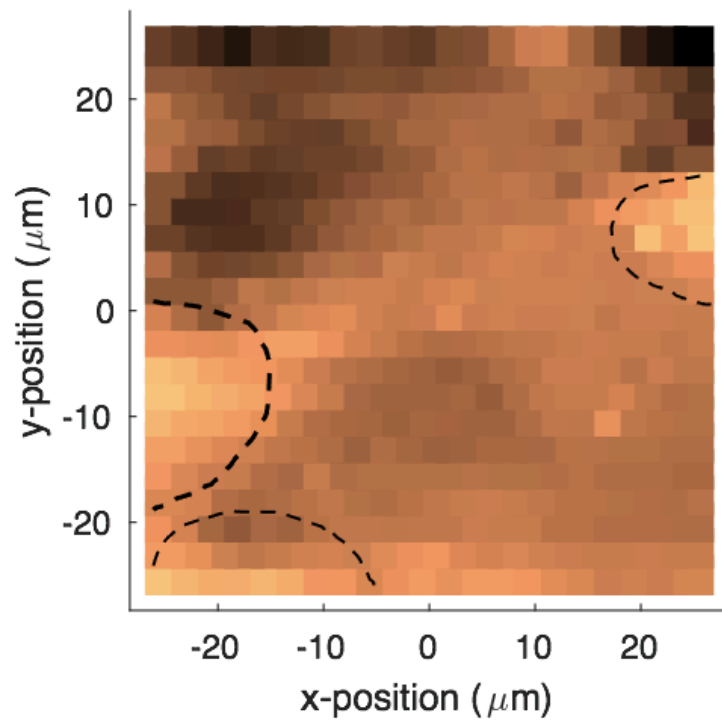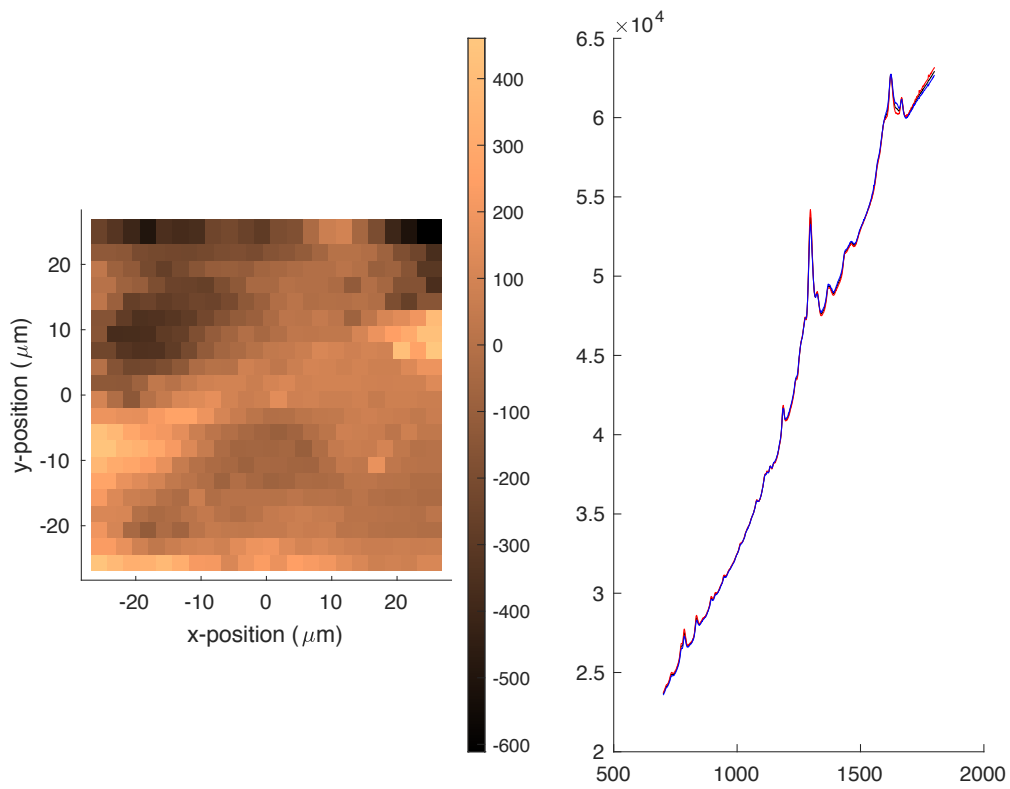
Figure 79. Scoremap four (on top). On the bottom, scoremap four with dotted lines that show the three bright reagions which can be identified as the trichomes.
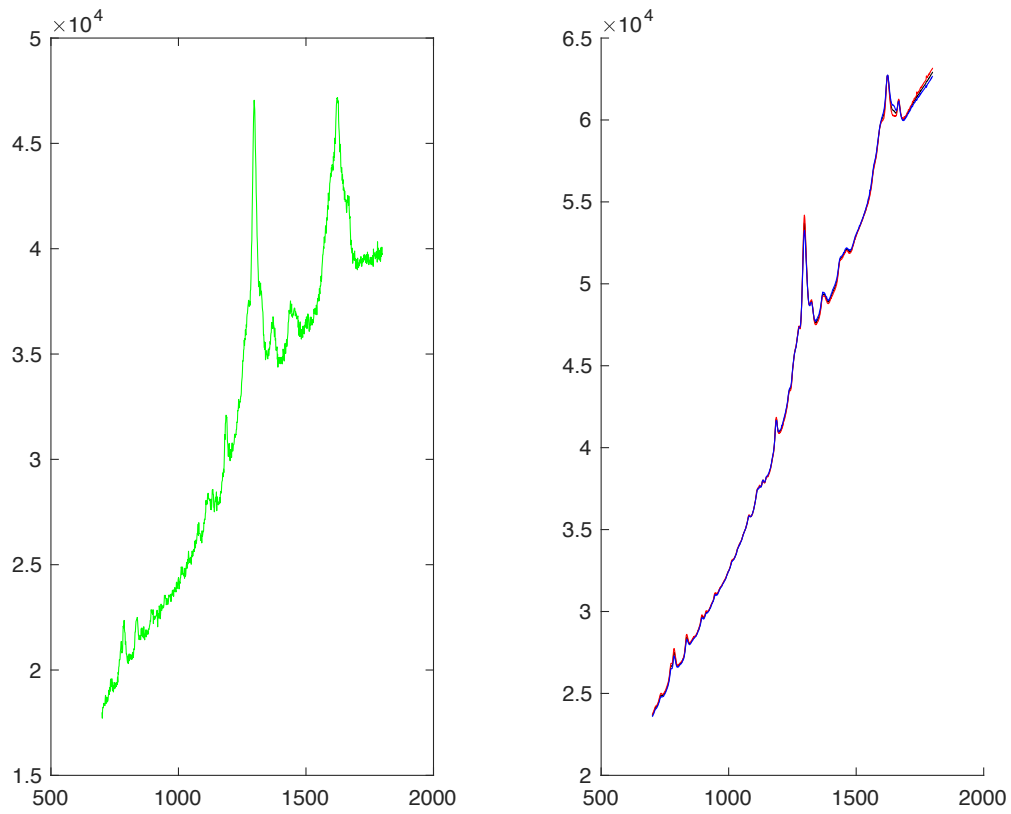
Figure 80. Extraction of the spectrum of coordinates (-22.5,10), which corresponds to a dark pixel.
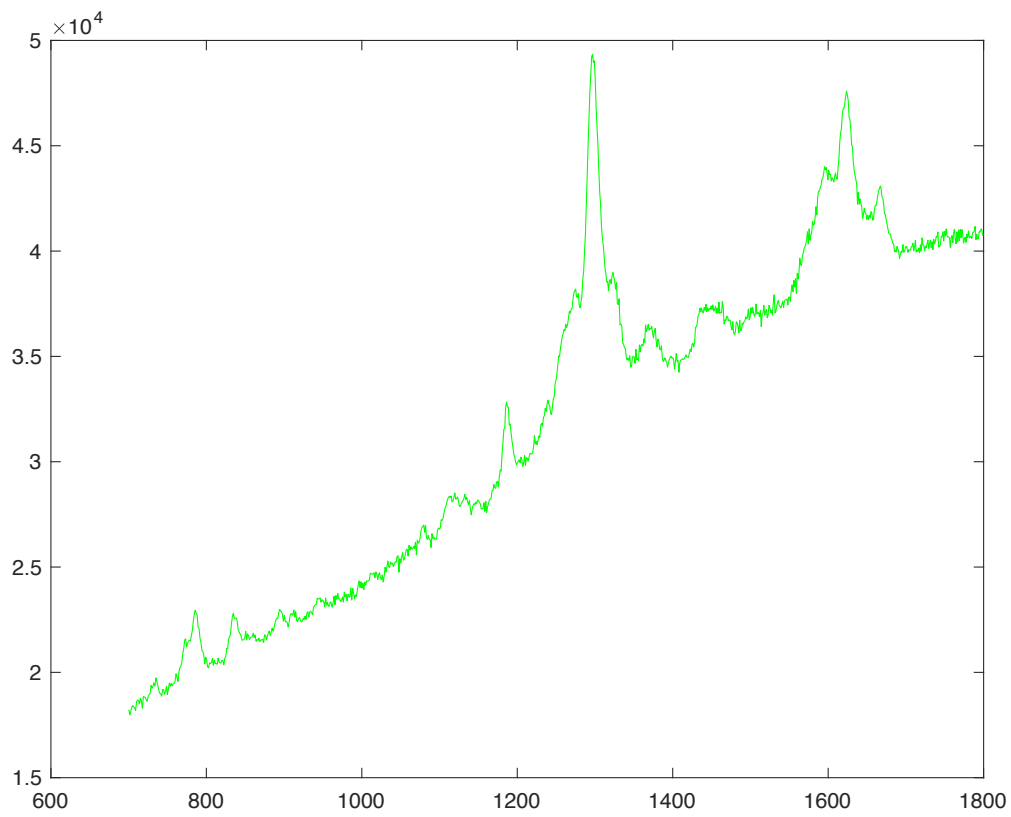
Figure 81. Extraction of the spectrum of coordinates (-25,-10), which corresponds to a bright pixel.

93

5. Interpretation of the peaks of cannabinoids

5.1 Structure of cannabinoids and their main Raman markers

In the previous chapter we saw that the Raman mapping of a trichome shows spatial variations of the signal associated to specific chemical species. The aim of this Chapter is to better understand the nature of such Raman signals, with the attempt to assign them to the normal modes of the chemical compounds we are interested in. The cannabinoids considered are reported below, together with their stoichiometry, a description of their structure and their CAS Number, a unique numerical identifier assigned by the Chemical Abstract Service (CAS) to all the chemicals described in scientific literature. For convenience, the Table of the CAS Numbers of the cannabinoids here considered is also reported below.

NAME	ABBREVIATION	CAS
Tetrahydrocannabinol	THC	1972-08-3
Tetrahydrocannabinolic acid	THCA	23978-85-0
Cannabidiol	CBD	13956-29-1
Cannabidiolic acid	CBDA	1244-58-2
Cannabigerol	CBG	25654-31-3
Cannabigerolic acid	CBGA	25555-57-1

Table 3. Table containing the name of the selected cannabinoids, the abbreviations and the correspondent CAS.

All the chemical structures have been obtained from PubChem [19] using their CAS, they have been optimized uwith the force field MMFF94 and their geometries have been investigated with the software Avogadro [23], focusing on bond lengths and selected dihedral angles. B3LYP/6-31G(d,p) DFT calculations have been carried out later to determine equilibrium structures, normal modes and Raman intensities. All the structures obtained from DFT calculations have been compared against the input structures obtained from PubChem [19].

Tetrahydrocannabinol (THC), $C_{21}H_{30}O_2$, CAS 1972-08-3. THC has a molecular weight of 314,46 g/mol. Its structure is characterized by three rings: phenol (A), pyrane (B) and cyclohexene (C), which are responsible for the rigidity of a large part of the molecule. However, the presence of the alkyl chain is brings flexibility and a potentially large number of conformers with specific relative probabilities. There are four isomers of THC, but only two exist in nature: (-)-trans isomer and (+)-trans isomer, which is less potent. THC contains one oxygen atom in the phenol moiety and one, which acts like a hydrogen bond acceptor, in an oxane ring (B). There are two functional groups, the phenol and the ether. The phenol group is characterized the OH group (O2-H23) while the ether is characterized by the oxygen (O1) atom bound to two R groups on both sides (carbons in this case).
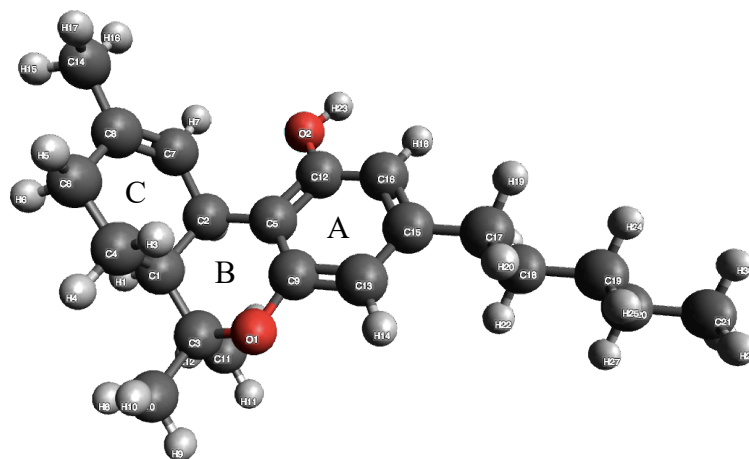
Figure 82. Chemical structure of THC. The rings are phenol (A), pyrane (B) and cyclohexene (C).

By using programs such as the software Avogadro, it is possible to collect more information on the structure of THC. For example, it is possible to determine the dihedral angles and the bond lengths. Table 4 reports some of the bond lengths determined using Avogadro on the PubChem structures. The initial atom and final atom are numbered according to the picture above and the length is calculated up to the third rounded decimal. In the third column the bond lengths of the PubChem structure of THC are reported. To verify that the structure of THC optimized using DFT is actually comparable to the one reported in PubChem, the bond lengths of that structure were also calculated using Avogadro and reported in the fourth column. Since the bond lengths depend on the theoretical method used to model the molecular structure (and we have no information about PubChem's method of choice), it is acceptable that the bond lengths are slightly different.

INITIAL ATOM	FINAL ATOM	LENGTH (Å)	LENGHT (Å) OF THE OPTIMIZED STRUCTURE
H12	C11	1.095	1.095
H13	C11	1.094	1.092
H2	C2	1.099	1.101
C11	H11	1.097	1.094
C11	C3	1.544	1.538
H23	O2	0.971	0.966
H26	C20	1.096	1.098
O2	C12	1.369	1.373
H7	C7	1.088	1.084
H1	C1	1.100	1.099
H21	C18	1.097	1.098
C2	C7	1.518	1.520
C2	C5	1.523	1.534
C2	C1	1.554	1.553
H27	C20	1.096	1.098
H29	C21	1.095	1.095
C12	C5	1.405	1.408
C12	C16	1.400	1.395
C7	C8	1.344	1.338

C20	C21	1.520	1.532
C20	C19	1.528	1.534
H18	C16	1.087	1.089
C5	C9	1.403	1.406
C3	C1	1.578	1.547
C3	C10	1.541	1.531
C3	O1	1.436	1.451
C1	C4	1.548	1.540
H22	C18	1.097	1.098
H8	C10	1.095	1.094
C16	C15	1.397	1.400
H16	C14	1.095	1.093
C21	H30	1.095	1.096
C21	H28	1.095	1.096
C18	C19	1.530	1.533
C18	C17	1.530	1.542
H9	C10	1.097	1.094
H15	C14	1.095	1.098
C9	O1	1.370	1.368
C9	C13	1.400	1.401
C10	H10	1.095	1.091
C8	C14	1.501	1.507
C8	C6	1.500	1.509
C14	H17	1.095	1.098
C19	H24	1.097	1.100
C19	H25	1.097	1.100
C15	C13	1.396	1.392
C15	C17	1.507	1.514
C13	H14	1.088	1.085
H6	C6	1.097	1.102
C17	H19	1.097	1.098
C17	H20	1.097	1.097
C4	C6	1.528	1.531
C4	H4	1.095	1.093
C4	H3	1.097	1.096
C6	H5	1.097	1.099

Table 4. Table containing some relevant bond lengths of THC and the lengths of the structure optimized using DFT. The bond lengths were calculated using the software Avogadro.

Similarly, (Table 5) it is also possible to determine the values of the dihedral angles both for the structure taken from PubChem and the one optimized by DFT (third column).

AVOGADRO ATOMIC LABELS	ANGLE	ANGLE OF THE OPTIMIZED STRUCTURE
H1-C1-C4-H3	180° (-)	177° (-)
H1-C1-C4-H4	62°	65.7°
H1-C1-C2-H2	42.2° (-)	40.8° (-)
H4-C4-C6-H6	43.9° (-)	46.3° (-)
H3-C4-C6-H6	159° (-)	163° (-)
H4-C4-C6-H5	72.5°	69.1°
H3-C4-C6-H5	42.7° (-)	47.6° (-)
H2-C2-C7-H7	46.5° (-)	49.1° (-)

96

As for the case of the bond lengths, these values are quite similar but not identical. Anyway, for our purposes it is enough to verify a good similarity between PubChem and DFT structures.

Tetrahydrocannabinolic acid (THCA), $C_{22}H_{30}O_4$, CAS 23978-85-0. THCA has a structure very similar to that of THC. As for THC, its structure is characterized by three rings: phenol (A), pyrane (B) and cyclohexene (C), which are responsible for the rigidity of a part of the molecule. The only difference is that H16 in THC is substituted by a carboxylic group (COOH) which can be removed in the form of CO2 by heating. It is possible to describe THCA as a diterpenoid 6a,7,8,10a-tetrahydro-6H-benzo[c]chromene substituted at position 1 by one hydroxy group, by methyl groups in positions 6 and 9, and by a penthyl group in position 3.
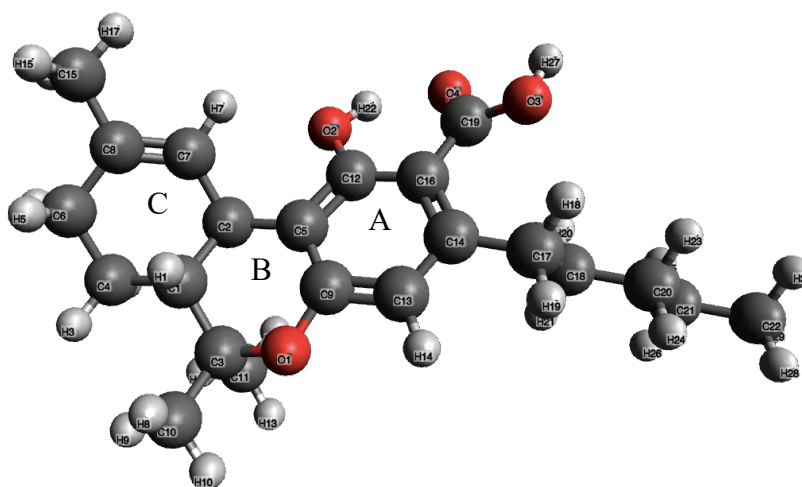
Figure 83. Chemical structure of THCA. The three rings are phenol (A), pyrane (B) and cyclohexene (C).

We report in the Tables 6 and 7 below the bond lengths and the dihedral angles of THCA.

INITIAL ATOM	FINAL ATOM	LENGTH (Å)	LENGHT (Å) OF THE OPTIMIZED STRUCTURE
H11	C11	1.094	1.094
H12	C11	1.094	1.093
H2	C2	1.099	1.103
H13	C11	1.097	1.094
C11	C3	1.537	1.534
H4	C4	1.097	1.098
H25	C21	1.096	1.098
H22	O2	0.982	0.997
H26	C21	1.096	1.098
H29	C22	1.095	1.095
O4	C19	1.221	1.242
O2	C12	1.378	1.342
C21	C22	1.520	1.532
C21	C20	1.529	1.534
H20	C18	1.096	1.095
H21	C18	1.098	1.098

H6	C6	1.096	1.099
C2	C7	1.520	1.522
C2	C5	1.520	1.523
C2	C1	1.549	1.544
C4	H3	1.094	1.095
C4	C6	1.534	1.539
C4	C1	1.536	1.534
C22	H30	1.095	1.096
C22	H28	1.095	1.096
C12	C5	1.410	1.409
C12	C16	1.420	1.435
H16	C15	1.095	1.098
C19	C16	1.487	1.495
C19	O3	1.340	1.351
C7	H7	1.086	1.083
C7	C8	1.348	1.340
C5	C9	1.398	1.398
C18	C20	1.530	1.533
C18	C17	1.532	1.542
C3	C1	1.539	1.538
C3	C10	1.540	1.529
C3	O1	1.442	1.463
C6	C8	1.509	1.514
C6	H5	1.098	1.101
C16	C14	1.409	1.432
C8	C15	1.502	1.507
H9	C10	1.094	1.093
C1	H1	1.104	1.100
H27	O3	0.980	0.972
C20	H23	1.097	1.099
C20	H24	1.097	1.100
C15	H17	1.095	1.093
C15	H15	1.095	1.098
C9	O1	1.369	1.356
C9	C13	1.398	1.407
C10	H10	1.097	1.093
C10	H8	1.096	1.094
C14	C13	1.405	1.382
C14	C17	1.514	1.518
C13	H14	1.090	1.085
C17	H18	1.094	1.091
C17	H19	1.099	1.096

Table 6. Table containing some relevant bond lengths of THCA and the lengths of the structure optimized using DFT. The bond lengths were calculated using the software Avogadro.

AVOGADRO ATOMIC LABELS	ANGLE	ANGLE OF THE OPTIMIZED STRUCTURE
H1-C1-C4-H3	68.2° (-)	71.4° (-)
H1-C1-C4-H4	173.3°	170.7°
H1-C1-C2-H2	174.2° (-)	173.9° (-)
H4-C4-C6-H6	38.4°	41.2°
H3-C4-C6-H6	77.6° (-)	75.7° (-)
H4-C4-C6-H5	154.5°	156.3°

H3-C4-C6-H5	38.5°	39.4°
H2-C2-C7-H7	88.2° (-)	88.0° (-)

Table 7. Table containing some relevant dihedral angles of THCA and the angles of the structure optimized using DFT. The angles were calculated using the software Avogadro.

Cannabidiol (CBD). $C_{21}H_{30}O_2$. CAS 13956-29-1. CBD has the same molecular weight of THC (314.46 g/mol) and their structures are similar as well. The phenol ring (A) and cyclohexene (C) are clearly identifiable. The only difference is in the pyran ring (B). This ring is closed in THC. while it is open in CBD and a hydroxyl group (O2-H25) is present instead of the oxygen atom. Such differences in the structure are responsible for the different pharmacological effects of CBD compared to THC. The $C(sp^2)H$ bond at the C=C is indicated with the letter "a" for simplicity in the description of the normal modes.
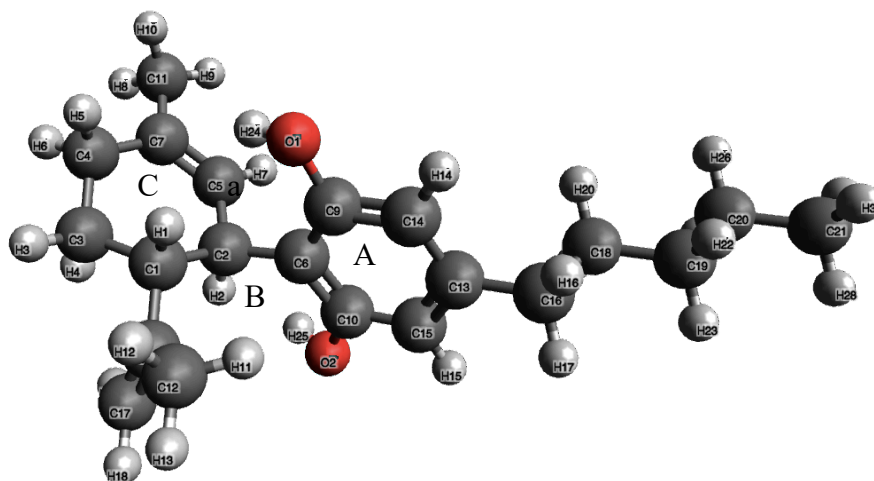
Figure 84. Chemical structure of CBD.

INITIAL ATOM	FINAL ATOM	LENGTH (Å)	LENGHT (Å) OF THE OPTIMIZED STRUCTURE
H12	C12	1.095	1.098
H14	C14	1.088	1.086
O1	H24	0.970	0.973
O1	C9	1.369	1.370
H16	C16	1.097	1.097
H5	C4	1.097	1.101
H11	C12	1.094	1.096
C12	H13	1.095	1.093
C12	C8	1.506	1.507
H1	C1	1.098	1.098
H22	C19	1.097	1.100
H30	C21	1.095	1.096
C14	C9	1.398	1.397

C14	C13	1.396	1.397
H3	C3	1.972	1.096
C9	C6	1.409	1.411
C16	C13	1.507	1.514
C16	H17	1.097	1.097
C16	C18	1.530	1.542
C4	C3	1.530	1.533
C4	H6	1.097	1.098
C4	C7	1.508	1.514
H10	C11	1.095	1.098
C1	C8	1.530	1.527
C1	C3	1.542	1.541
C1	C2	1.566	1.559
H20	C18	1.097	1.098
C8	C17	1.345	1.342
C19	C18	1.530	1.533
C19	H23	1.097	1.100
C19	C20	1.529	1.534
C3	H4	1.098	1.097
C13	C15	1.396	1.397
H26	C20	1.096	1.098
C21	C20	1.520	1.532
C21	H29	1.095	1.095
C21	H28	1.095	1.096
C18	H21	1.097	1,098
C20	H27	1,096	1,098
C7	C11	1,502	1,506
C7	C5	1,345	1,344
C6	C2	1,529	1,528
C6	C10	1,408	1,411
C11	H8	1,095	1,098
C11	H9	1,095	1,093
C17	H18	1,086	1,086
C17	H19	1,085	1,087
C2	C5	1,520	1,520
C2	H2	1,096	1,099
C5	H7	1,090	1,090
C15	C10	1,398	1,396
C15	H15	1,088	1,085
C10	O2	1,369	1,374
O2	H25	0,970	0,971

Table 8. Table containing some relevant bond lengths of CBD and the lengths of the structure optimized using DFT. The bond lengths were calculated using the software Avogadro.

AVOGADRO ATOMIC LABELS	ANGLE	ANGLE OF THE OPTIMIZED STRUCTURE
H1-C1-C3-H3	62.0° (-)	67.1° (-)
H1-C1-C3-H4	179.8° (-)	176.3°
H1-C1-C2-H2	169.3°	168.9° (-)
H4-C3-C4-H6	48.2°	48.3°
H4-C3-C4-H5	164.8°	164.0°
O1-C9-C14-H14	0.4°	1.1° (-)
H1-C1-C2-H2	169.3°	168.9°

H1-C1-C3-H4	179.8° (-)	176.3°
H1-C1-C3-H3	62° (-)	67.1° (-)

Cannabidiolic acid (CBDA), $C_{22}H_{30}O_4$, CAS 1244-58-2. CBDA has a molecular weight of 358,5 g/mol. It is a dihydroxybenzoic acid, an olivetolic acid in which the hydrogen in position 3 is substituted by a limonene group. It is the carboxylic acid of CBD, characterized by a carboxylic group bound to the benzene ring, replacing the H14 of CBD. The phenol ring (A) and cyclohexene (B) are clearly identifiable.
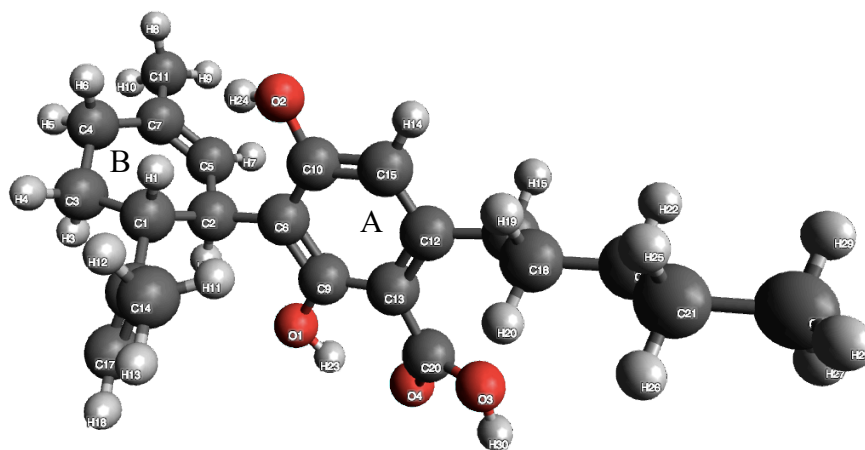
Figure 85. Chemical structure of CBDA.

AVOGADRO ATOMIC LABELS	ANGLE	ANGLE OF THE OPTIMIZED STRUCTURE
H1-C1-C3-H3	179.5°	175°
H1-C1-C3-H4	62.7° (-)	68.9° (-)
H1-C1-C2-H2	170.4°	171°
H4-C3-C4-H6	163.3°	46°
H4-C3-C4-H5	68.3° (-)	69.7° (-)
H1-C1-C2-H2	169.3°	171°
H1-C1-C3-H4	170.4°	68.9° (-)
H1-C1-C3-H3	179.5°	175°

Cannabigerol (CBG), $C_{21}H_{32}O_2$, CAS 25654-31-3: it is a minor cannabinoid present in *C. Sativa.* It is present in low concentration in the plant, while the majority is converted into other cannabinoids, especially THC and CBD.
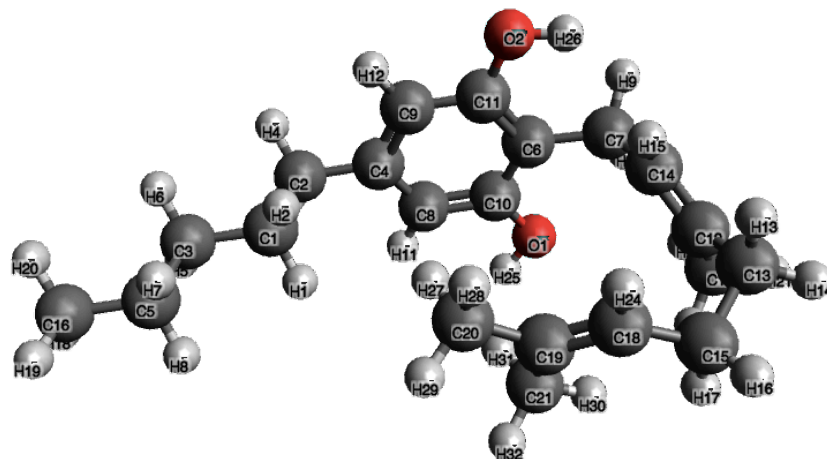
Figure 86. Chemical structure of CBG.

AVOGADRO ATOMIC LABELS	ANGLE	ANGLE OF THE OPTIMIZED STRUCTURE
H1-C1-C3-H6	179.9°	179.6° (-)
H1-C1-C3-H5	62.8° (-)	64.0 (-)
H2-C1-C3-H6	63.2°	64.4°
H2-C1-C3-H5	179.5° (-)	179.9°
H1-C1-C2-H4	179.7°	179.6° (-)
H1-C1-C2-H3	63.1°	64.6°
H2-C1-C2-H3	179.8° (-)	179.9°
H2-C1-C2-H4	63.2° (-)	64.3° (-)
H10-C7-C14-H15	169.0°	177.3° (-)
H9-C7-C14-H15	55.1°	67.0°

Table 11. Table containing some relevant dihedral angles of CBG and the angles of the structure optimized using DFT. The angles were calculated using the software Avogadro.

Cannabigerolic acid (CBGA), $C_{22}H_{32}O_4$, CAS 25555-57-1: its structure is very similar to that of CBG. The main difference is the presence of one carboxylic group which substitutes the H12 atom of CBG.

Figure 87. Table containing some relevant dihedral angles of CBG and the angles of the structure optimized using DFT. The angles were calculated using the software Avogadro.

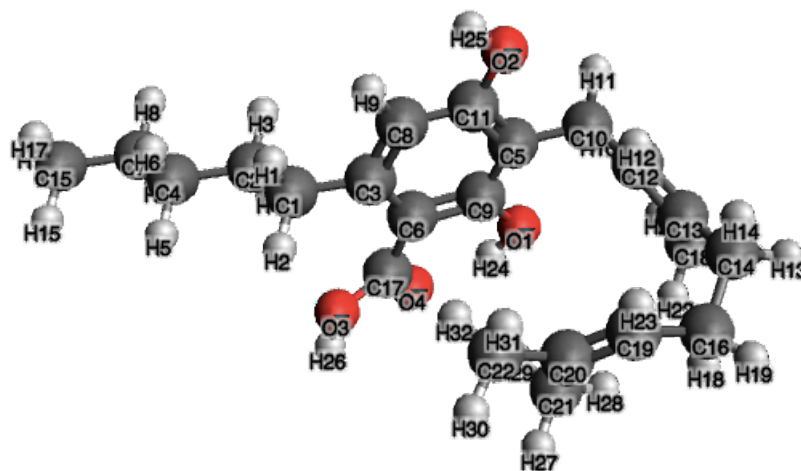| AVOGADRO ATOMIC LABELS	ANGLE	ANGLE OF THE OPTIMIZED STRUCTURE
H1-C1-C2-H3	63.0°	62.2°
H1-C1-C2-H4	179.9° (-)	178.2°
H2-C1-C2-H4	65.9° (-)	65.7° (-)
H2-C1-C2-H3	176.9°	178.2°
H3-C2-C4-H6	62.9° (-)	65.1° (-)
H3-C2-C4-H5	179.8°	179.3°
H4-C2-C4-H5	63.4°	62.7°
H4-C2-C4-H6	179.3° (-)	178.3°

Table 12. Table containing some relevant dihedral angles of CBGA and the angles of the structure optimized using DFT. The angles were calculated using the software Avogadro.

The analysis of the previous cannabinoids demonstrated how the bond length appears to be the similar for the non-optimized and optimized structures. Since the force field used is different, we expect the values not to be identical. This study has been done for THC, THCA and CBD and was not repeated for the other cannabinoids. Therefore, we can just assume the similarity of the bond length. The calculation of the dihedral angles was instead done also for the other cannabinoids. As expected, the values of bond length calculated for all the molecules are similar but not identical. However. this is not a problem since different force fields have been used. As for the case of the bond lengths, the values of the dihedral angles appear to be similar but not identical. Anyway. it is sufficient to assume the similarity between the two structures.

The Figures 88-93 below show the comparison between the 3D structures of the selected cannabinoids taken from PubChem and the optimized structures determined by DFT. It is

evident also by visual inspection DFT and PubChem structures are fairly close, as the numerical Tables above demonstrate in detail.

Figure 88. Chemical structure of THC. On the left there is the structure reported in PubChem, on the right, the one optimized with the MMFF94 force field.

Figure 89. Chemical structure of THCA. On the left there is the structure reported in PubChem, on the right, the one optimized with the MMFF94 force field.

Figure 90. Chemical structure of CBD. On the left there is the structure reported in PubChem, on the right, the one optimized with the MMFF94 force field.

Figure 91. Chemical structure of CBDA. On the left there is the structure reported in PubChem, on the right, the one optimized with the MMFF94 force field.
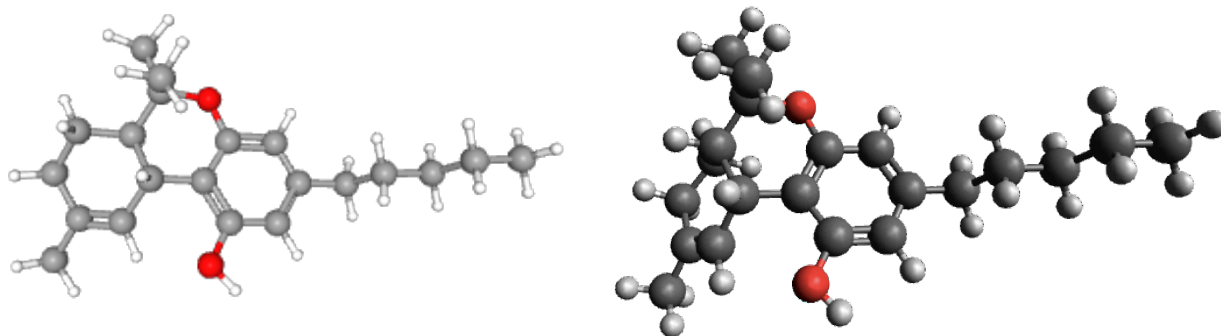

Figure 92. Chemical structure of CBG. On the left there is the structure reported in PubChem, on the right, the one optimized with the MMFF94 force field.
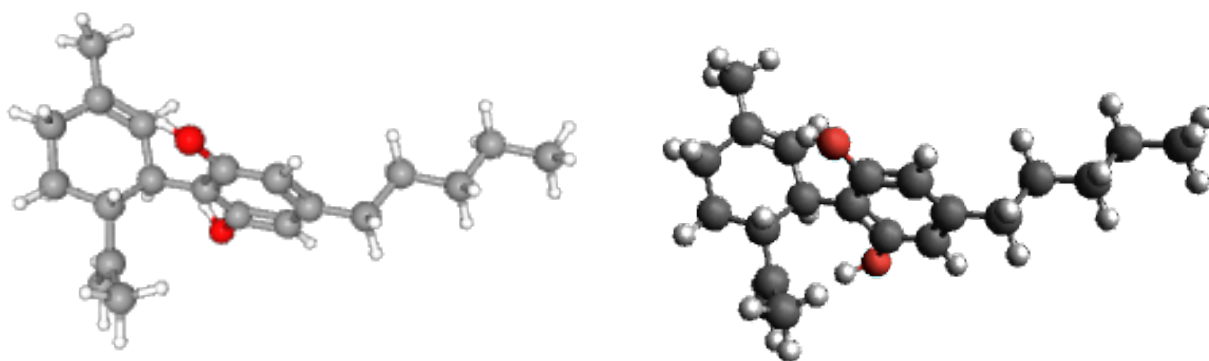

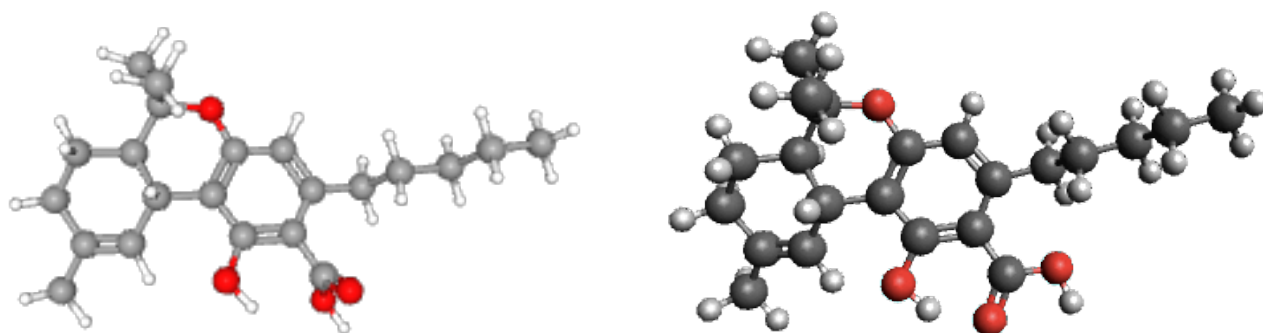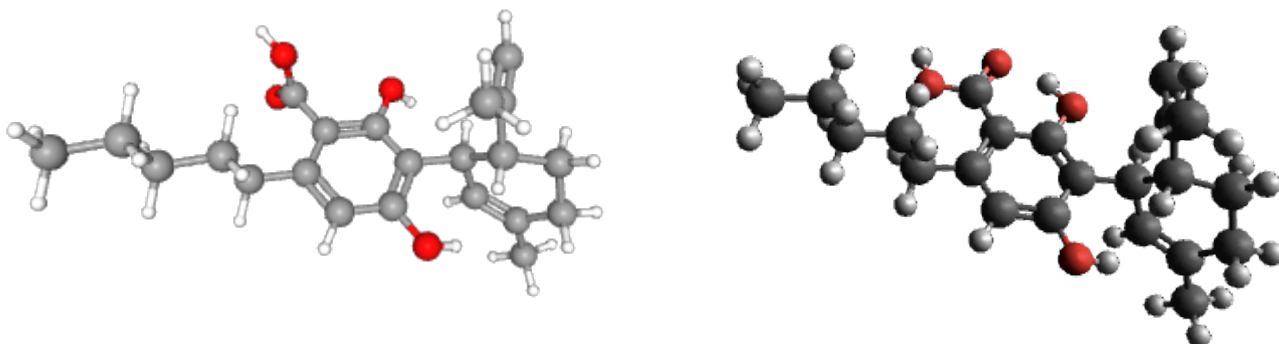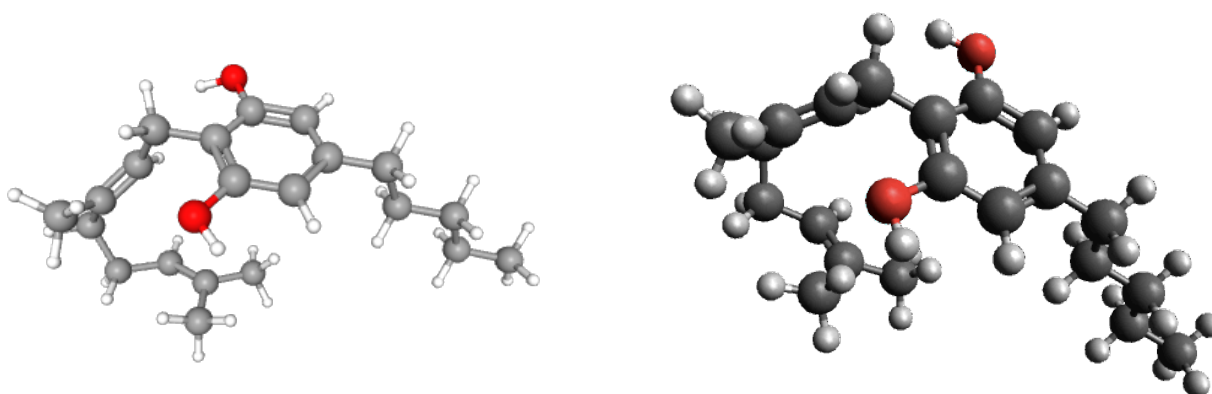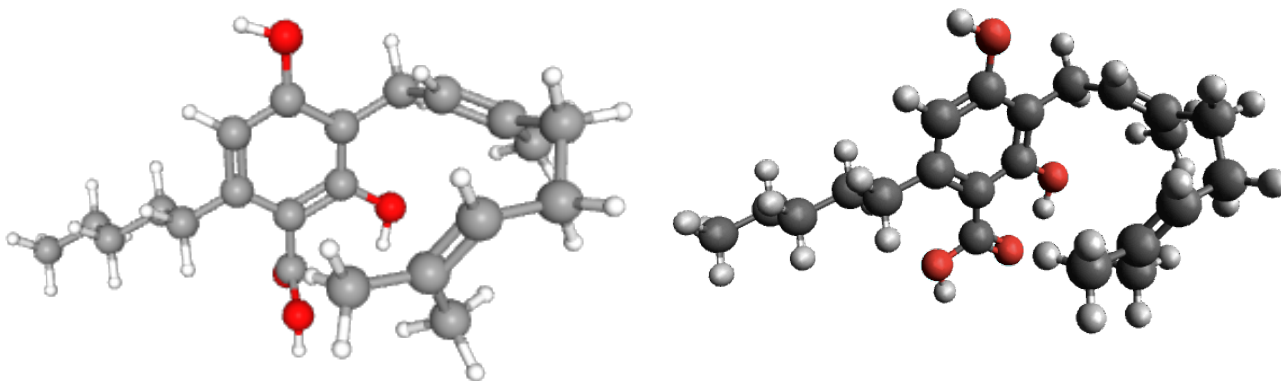*Figure 93. Chemical structure of CBGA. On the left there is the structure reported in PubChem, on the right, the one optimized with the MMFF94 force field.*

5.2 Raman peaks assignment and normal modes of the selected cannabinoids

By using a set of programs already present in the laboratory where I have carried out my thesis work, it is possible to obtain from DFT calculations the simulated spectra for each cannabinoid here considered. Since, compared with experiments, the DFT method B3LYP/6-31G(d,p) tends to increase the value of frequency, we can multiply the wavenumber axis for a correction factor < 1. Usually, the choice of 0.98 is optimal in the wavenumber range $400 - 2000$ cm^{-1}. This allows to transform the graph of the simulated spectrum and facilitate the comparison with available experimental data. Before opening the spectra on MATLAB, it is important to decide to neglect the region of the spectrum corresponding to the C-H stretching, which is located at around 3000 cm^{-1}. In fact, as also seen in the first chapter, the C-H stretching region is not a fingerprint from any of the molecule considered. Furthermore, plain DFT methods within the harmonic approximation (such as the present approach) are not very reliable in this region of the vibrational spectrum. The peaks below 700 cm^{-1} have been also neglected. The reason for this choice is that there is no available experimental information in the literature about the most relevant peaks of each cannabinoid below that threshold.

By using Matlab, it is possible to plot the Raman spectrum simulated by DFT for each cannabinoid and it is possible to compare the main peaks to the data collected in literature. One should not expect that the simulated spectra exactly match those in the literature. This is mainly due to the fact that DFT neglects the intermolecular forces and the analysis has been done on an isolated molecule within the harmonic approximation. However, we expect the trend of the DFT spectrum to be similar to the experimental one, and we thus expect to be able to assign at least the most intense peaks reported in the literature. This peak assignment procedure will allow us to understand how to use the spectroscopic data for chemical sensing purposes. Since similar molecules have similar spectra, it is necessary to assess to which normal modes the peaks are associated, to be able to reliably distinguish each cannabinoid. It is also important to understand how selective such Raman peaks can be compared to the chemical structure of each cannabinoid. We are interested to those vibrational modes that are delocalized over the cannabinoid molecule and can be expected to be distinctive features useful to differentiate among the different compounds.

The goal of the following study is to provide a general overview of the relevant Raman peaks of selected cannabinoids. This study could be an important starting point for further developments employing Raman spectroscopy to investigate cannabinoid-containing samples. This analysis done for each single molecule can be very complicated because most of the molecular structures of cannabinoids in condensed phase are unknown (X-ray diffraction experiments are difficult on such volatile compounds, moreover not always they provide crystals phases but rather amorphous phases). Therefore, we selected a more feasible and reductional approach: we decided to focus on the most intense peaks, and we tried to associate the most intense theoretical peaks to the closer experimental values. Therefore, within Matlab, the simulated spectrum (after wavenumber scaling – see above) was represented together with vertical lines representing the wavenumbers reported by the literature [1] in correspondence of which relevant peaks in the experimental spectra could be detected. Such experimental positions are marked by using different colors, depending on the relative intensity of the peaks in the experimental spectra:

1. Red for the most intense peaks;
2. Yellow for the peaks with medium intensity;
3. Grey for the ones with low intensity.

The following approach was used to assign the experimental peaks to DFT modes. For each cannabinoid the graph which contains both the simulated spectrum, and the experimental peaks was obtained using Matlab. A table was then prepared, containing:
- the experimental peaks reported in literature,
- the scaled (by 0.98) theoretical wavenumber (that correspond to the position of the peaks of the black curve in the graph),
- the theoretical wavenumber,
- the theoretical intensity,
- the description of the associated normal mode.

By confronting the trend of the simulated Raman spectra with the position of the peaks reported by the literature we were able to associate some of the literature peaks with the DFT data. Of course, this was not possible for each peak, for several reasons. For instance, the simulated spectrum deals with the isolated molecule, whereas in the experiments the molecules experience intermolecular interactions with each other. This is the reason why some peaks that are very strong in the simulated spectrum could not be present in the experimental one. Moreover, sometimes the calculation cannot predict correctly the coupling of different vibrations on interacting groups, which leads to issues in estimating reliably the relative Raman intensities.

The results from the DFT calculations carried out on the selected cannabinoids were analyzed by using the software Molden [4] that shows the animation of the specific vibrational modes associated to each peak. The modes were then named and described in the table together with a picture that represents them.

The rest of this paragraph reports the simulated spectrum and the table containing the assignment of the peaks for each cannabinoid and the normal modes associated to it. The last part of the paragraph will be dedicated to the description of such information, including the normal modes.

The first cannabinoid to be considered is THC. The list of experimental peaks of THC is: 733, 775, 780, 835, 891, 908, 950, 1011, 1037, 1080, 1114, 1129, 1185, 1236, 1255, 1276, 1295, 1321, 1365, 1372, 1440, 1462, 1570, 1600, 1623, 1666 cm^{-1}. By looking at the experimental Raman spectrum of THC [1], it is possible to divide it in three equivalent regions and classify the intensity of such peaks as follows:
- Strong: 1295 cm^{-1}
- Medium 780, 835, 891, 908, 1185, 1255, 1276, 1321, 1365, 1372, 1623, 1666 cm^{-1}
- Weak 733, 775, 950, 1011, 1037, 1080, 1114, 1129, 1236, 1440, 1462, 1570, 1600 cm^{-1}.

The simulated DFT spectrum (Figure 94) and the assignment table for THC (Table 13) are reported below.
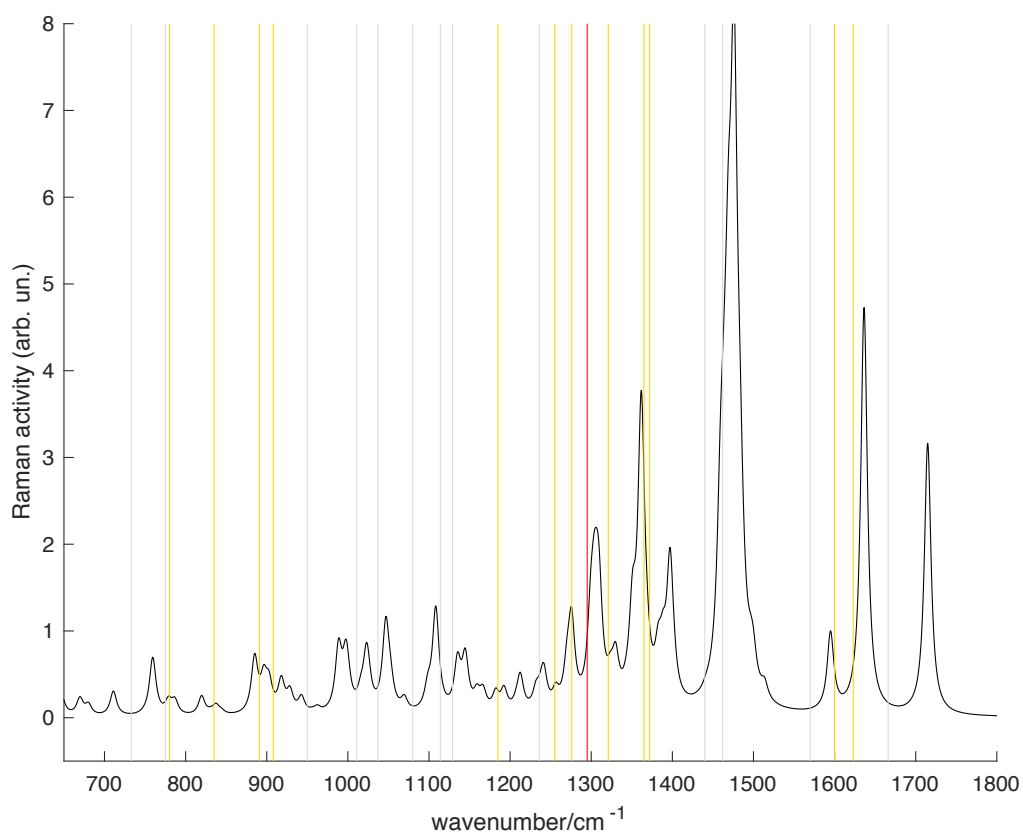
Figure 94. Simulated spectrum of THC. The vertical lines represent the wavenumbers corresponding to experimental peaks. Strong peaks are represented in red, medium in yellow and weak in grey.

	Experimental data	Scaled theoretical data	Theoretical data	Intensity	Normal mode	Representation of the normal mode
1	1109	1114	1131	weak	alkyl chain vibration.	
2	1276	1275	1302	medium	CH bending (CH bound to the cyclohexene).	
3	1295	1306	1326	strong	CH_2 twisting, CH and OH bending.	
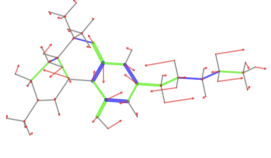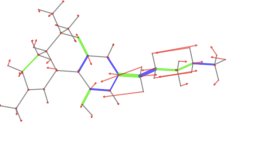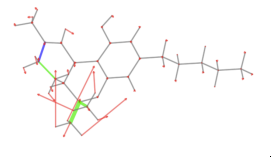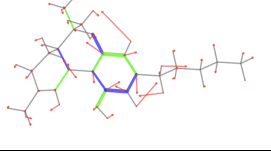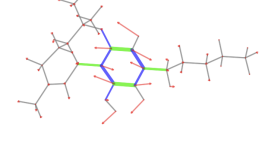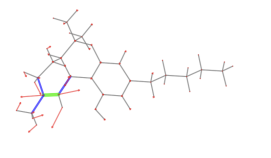
4	1321	1330	1379	medium	aromatic ring stretching; CH_2 wagging.	
5	1365	1362	1389	medium	CH_2 wagging in the alkyl chain; aromatic ring vibration.	
6	1372	1397	1431	medium	CH_3 vibration.	
7	1462	1476	1471	weak	stretching of the aromatic ring.	
8	1623	1636	1670	medium	stretching of the aromatic ring.	
9	1666	1715	1750	medium	C=C stretching of the cyclohexene.	

Table 13. Table containing the assignment of the peaks of the experimental spectrum and the simulated one for THC. For each experimental peak the scaled theoretical value, the theoretical data, the intensity and the correspondent normal mode are reported. The picture represents the correspondent normal mode. The color green represents the stretching of the bond, the blue the shrinking.

By looking at the experimental Raman spectrum of THCA [1], it is possible to divide it in three regions and classify the intensity of the peaks as follows:

- Strong: 1295 cm^{-1}
- Medium 780, 835, 891, 908, 1185, 1255, 1276, 1321, 1365, 1372, 1570, 1600, 1623, 1666 cm^{-1}
- Weak 715, 733, 775, 950, 1011, 1037, 1080, 1114, 1129, 1236, 1440, 1462 cm^{-1}.

The simulated DFT spectrum (Figure 95) and the assignment table for THCA (Table 14) are reported below.
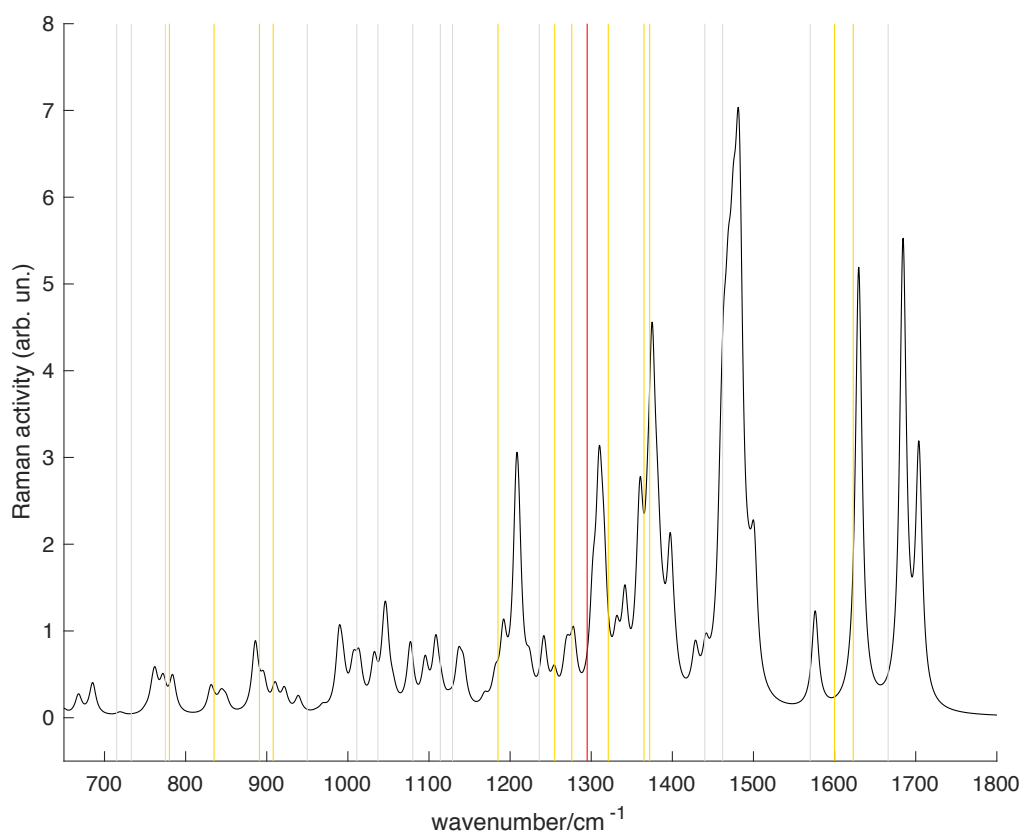
Figure 95. Simulated spectrum of THCA. The vertical lines represent the wavenumbers corresponding to experimental peaks. Strong peaks are represented in red, medium in yellow and weak in grey.

	Experimental data	Scaled theoretical data	Theoretical data	Intensity	Normal mode	Representation of the normal mode
1	891	886	904	medium	alkyl chain vibration.	
2	1185	1208	1232	medium	CH$_2$ twisting in alkyl chain; aromatic ring vibration; bending of CH bound to the aromatic ring; bending of OH in the	

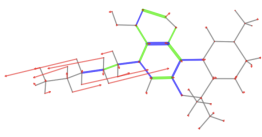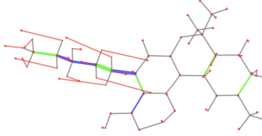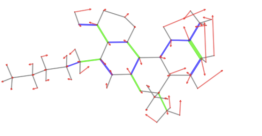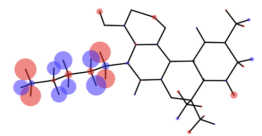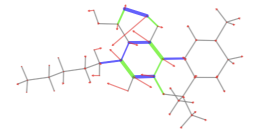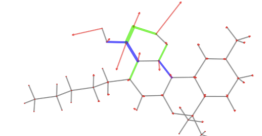|						
					COOH group	
3	1295	1310	1337	strong	asymmetric twisting of the CH$_2$ in the alkyl chain and OH vibration	
4	1365	1360	1388	medium	CH$_2$ wagging in the alkyl chain; OH bending.	
5	1372	1375	1403	medium	symmetric twisting of the CH$_2$ in the ring and CH vibration	
6	1462	1481	1511	weak	CH$_2$ wagging in the cyclohexane.	
7	1623	1630	1663	medium	stretching of the aromatic ring; OH bending.	
8	1666	1685	1719	weak	OH bending; C=O contraction in the COOH group	

Table 14. Table containing the assignment of the peaks of the experimental spectrum and the simulated one for THCA. For each experimental peak the scaled theoretical value, the theoretical data, the intensity and the correspondent normal mode are reported. The picture represents the correspondent normal mode. The color green represents the stretching of the bond, the blue the shrinking. The blue and red circles are used to indicate a movement which is perpendicular to the plane of the molecule.

The picture of mode 6, which corresponds to the experimental peak of THCA at 1462 cm^{-1}, appears different from the others. While for the other peaks the bonds are represented in green or blue (for stretching/shrinking), in this picture there are blue and red circles. This happens because for this mode the movement of the H of the CH$_2$ groups in the alkyl chain is perpendicular to the orientation of the molecule selected by the software that produces the

plot of the modes. The animation provided by Molden in this case was very helpful to reliably assign the mode.

Looking at the experimental Raman spectrum of CBD, it is possible to divide it in three regions and classify the intensity of the peaks in:

- Strong: 1437, 1451, 1643, 1663 cm^{-1}
- Medium 775, 865, 985, 1012, 1104, 1302, 1340, 1370, 1623 cm^{-1}
- Weak: 802, 901, 924, 965, 1080, 1136, 1150, 1176, 1230, 1262, 1274, 1515, 1585 cm^{-1}

The simulated DFT spectrum (Figure 96) and the assignment table for CBD (Table 15) are reported below.
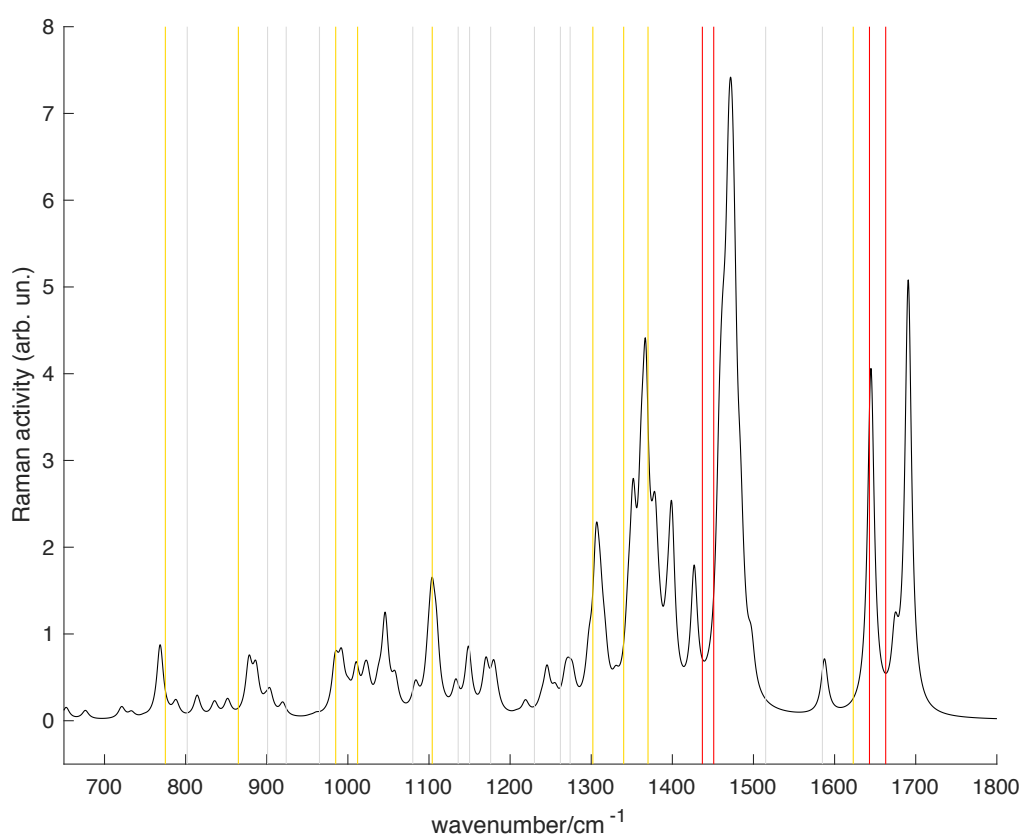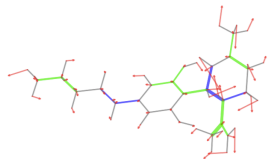
Figure 96. Simulated spectrum of CBD. The vertical lines represent the wavenumbers corresponding to experimental peaks. Strong peaks are represented in red, medium in yellow and weak in grey.

	Experimental data	Scaled theoretical data	Theoretical data	Intensity	Normal mode	Representation of the normal mode
1	775	768	784	medium	cyclohexene ring deformation; CH_2 twisting.	

2	865	878	896	medium	out of plane bending of the a bond at the C=C.	
3	901	902	922	weak	cyclohexene deformation ; CH2 rocking.	
4	924	920	939	weak	CH2 wagging; cyclohexane deformation .	
5	1012	1010	1031	medium	CH2 wagging; cyclohexene deformation .	
6	1104	1104	1126	medium	CC stretching, CH bending at the cyclohexene .	
7	1302	1306	1333	medium	CH2 twisting of the alkyl chain.	
8	1340	1352	1379	medium	wagging of CH2 in alkyl chain, OH bending, aromatic ring deformation .	
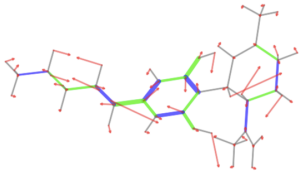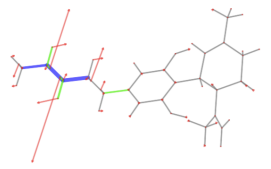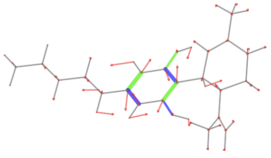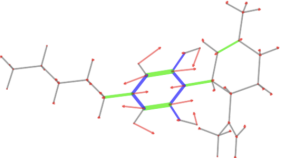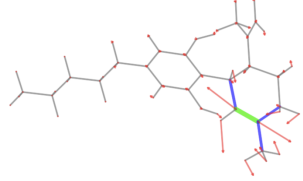
9	1370	1367	1395	medium	CH bending, OH bending, aromatic ring deformation.	
10	1437	1472	1500	strong	CH2 scissoring in the alkyl chain.	
11	1585	1588	1620	weak	aromatic ring stretching.	
12	1623	1645	1678	strong	aromatic ring stretching.	
13	1663	1690	1725	strong	C=C stretching of the cyclohexene and contraction of two C-C single bonds.	

Table 15. Table containing the assignment of the peaks of the experimental spectrum and the simulated one for CBD. For each experimental peak the scaled theoretical value, the theoretical data, the intensity and the correspondent normal mode are reported. The picture represents the correspondent normal mode. The color green represents the stretching of the bond, the blue the shrinking.

By looking at the experimental Raman spectrum of CBDA [1], it is possible to divide it in three regions and classify the intensity of the peaks in:

- Strong: 1307, 1623 cm^{-1}
- Medium 775, 843, 901, 1262, 1370, 1437, 1451, 1585, 1602, 1643, 1663 cm^{-1}
- Weak: 802, 1176 cm^{-1}

The simulated DFT spectrum (Figure 97) and the assignment table for CBDA (Table 16) are reported below.

114
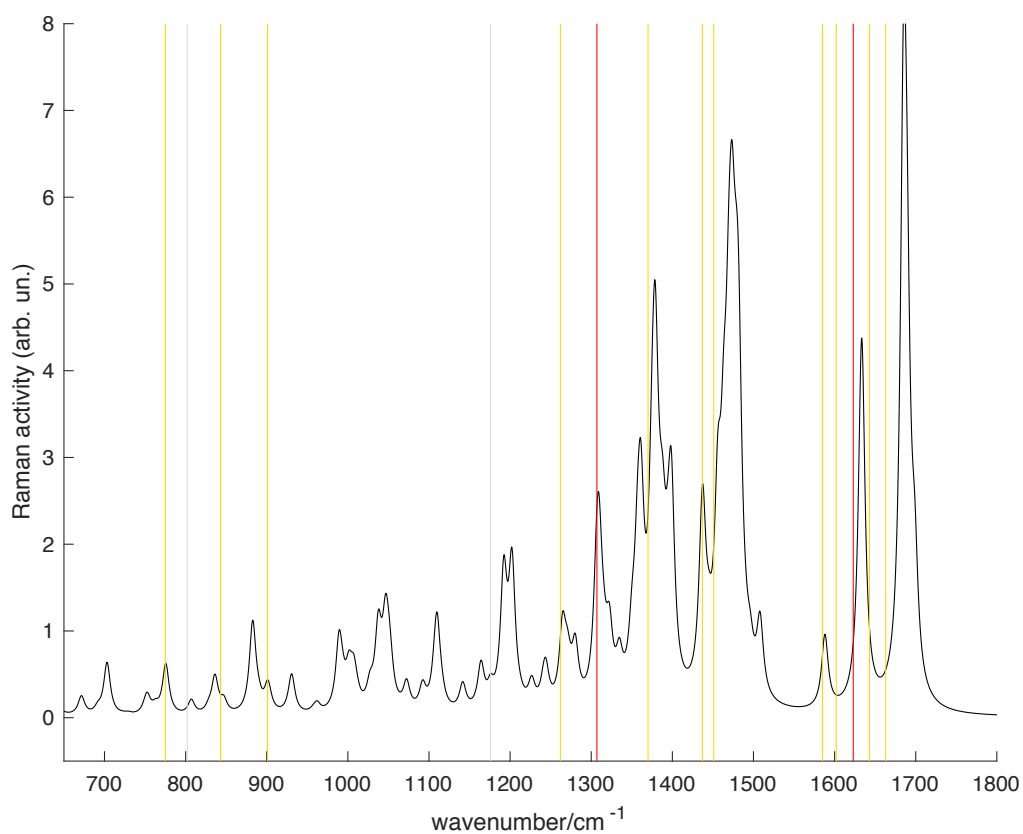
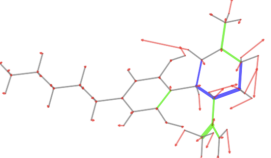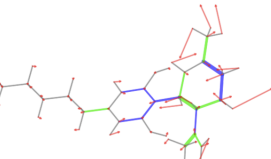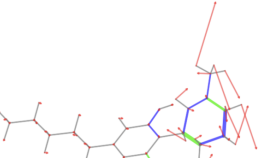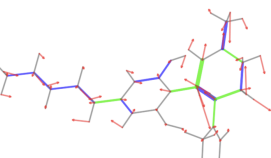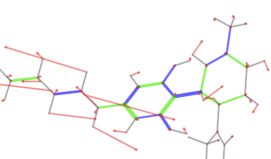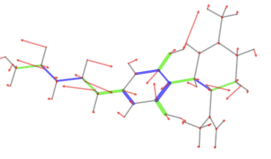*Figure 97. Simulated spectrum of CBDA. The vertical lines represent the wavenumbers corresponding to experimental peaks. Strong peaks are represented in red, medium in yellow and weak in grey.*

	Experimental data	Scaled theoretical data	Theoretical data	Intensity	Normal mode	Representation of the normal mode
1	775	775	791	medium	cyclohexene ring deformation; CH2 rocking.	
2	901	901	920	medium	cyclohexene ring deformation; CH2 wagging.	
3	1262	1266	1291	medium	wagging of CH bound to the cyclohexane; aromatic	
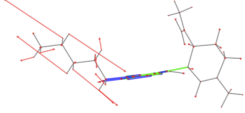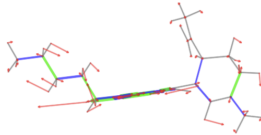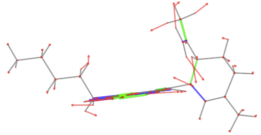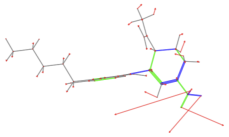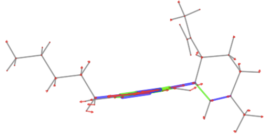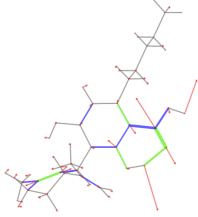
					ring deformation; OH bending.	
4	1307	1310	1337	strong	twisting of CH2 of the alkyl chain.	
5	1370	1379	1407	medium	aromatic ring deformation; OH bending in the carboxylic group.	
6	1437	1437	1466	medium	vibration of the aromatic ring and OH vibration	
7	1451	1473	1504	medium	CH3 vibration; aromatic ring vibration; OH bending.	
8	1623	1633	1667	strong	aromatic ring vibration; OH bending.	
9	1663	1686	1720	medium	OH bending; contraction of the C=O bond in the COOH group.	

Table 16. Table containing the assignment of the peaks of the experimental spectrum and the simulated one for CBDA. For each experimental peak the scaled theoretical value, the theoretical data, the intensity and the correspondent normal mode are reported. The picture represents the correspondent normal mode. The color green represents the stretching of the bond, the blue the shrinking

By looking at the experimental Raman spectrum of CBG [1], it is possible to divide it in three regions and classify the intensity of the peaks as follows:

- Strong: 1129cm^{-1}
- Medium: 879, 988, 998, 1060, 1309, 1328, 1345, 1440, 1452, 1623, 1627, 1670 cm^{-1}
- Weak: 725, 745, 768, 778, 800, 810, 835, 860, 902, 918, 968, 1033, 1100, 1114, 1161, 1182, 1219, 1275, 1295, 1365, 1381, 1516, 1587 cm^{-1}

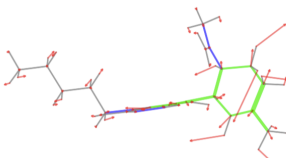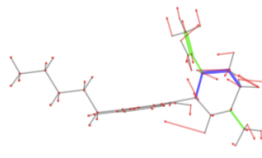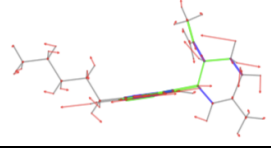The simulated DFT spectrum (Figure 98) and the assignment table for CBG (Table 17) are reported below.
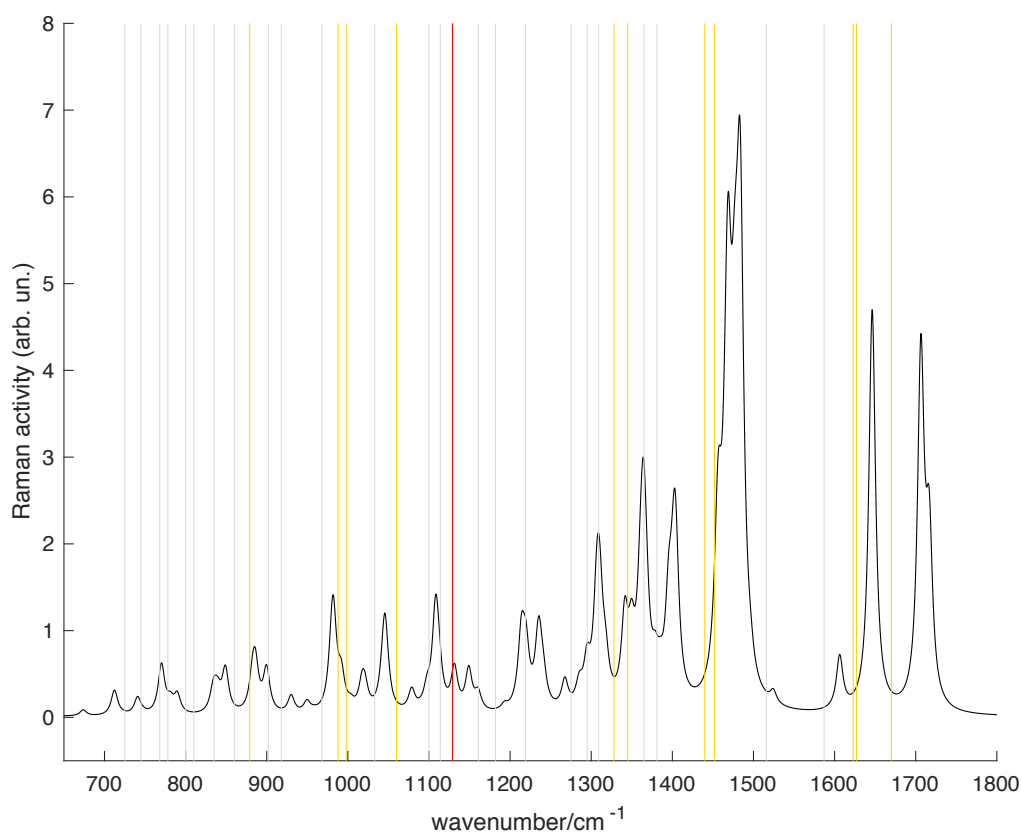
Figure 98. Simulated spectrum of CBG. The vertical lines represent the wavenumbers corresponding to experimental peaks. Strong peaks are represented in red, medium in yellow and weak in grey.

	Experimental data	Scaled theoretical data	Theoretical data	Intensity	Normal mode	Representation of the normal mode
1	879	885	904	medium	CH3 vibration in the alkyl chain.	

117

2	988	982	1002	medium	C-C stretching.	
3	1129	1109	1131	strong	alkyl chain vibration, CH2 wagging.	
4	1328	1364	1392	medium	CH2 wagging; aromatic ring oscillation; OH bending.	
5	1440	1469	1499	medium	CH2 scissoring.	
6	1452	1483	1513	medium	CH3 vibration, CH2 scissoring.	
7	1627	1646	1680	medium	stretching of the aromatic ring, OH bending.	
8	1670	1706	1741	medium	expansion of the C=C double bond in the lateral chain.	

Table 17. Table containing the assignment of the peaks of the experimental spectrum and the simulated one for CBG. For each experimental peak the scaled theoretical value, the theoretical data, the intensity and the correspondent normal mode are reported. The picture represents the correspondent normal mode. The color green represents the stretching of the bond, the blue the shrinking.

By looking at the experimental Raman spectrum of CBGA [1], it is possible to divide it in three regions and classify the intensity of the peaks as follows:

- Strong: 1287, 1295, 1623 cm^{-1}
- Medium 1670 cm^{-1}
- Weak: 759, 810, 860, 879, 923, 998, 1080, 1114, 1172, 1381, 1440, 1452, 1500 cm^{-1}

The simulated DFT spectrum (Figure 99) and the assignment table for CBGA (Table 18) are reported below.
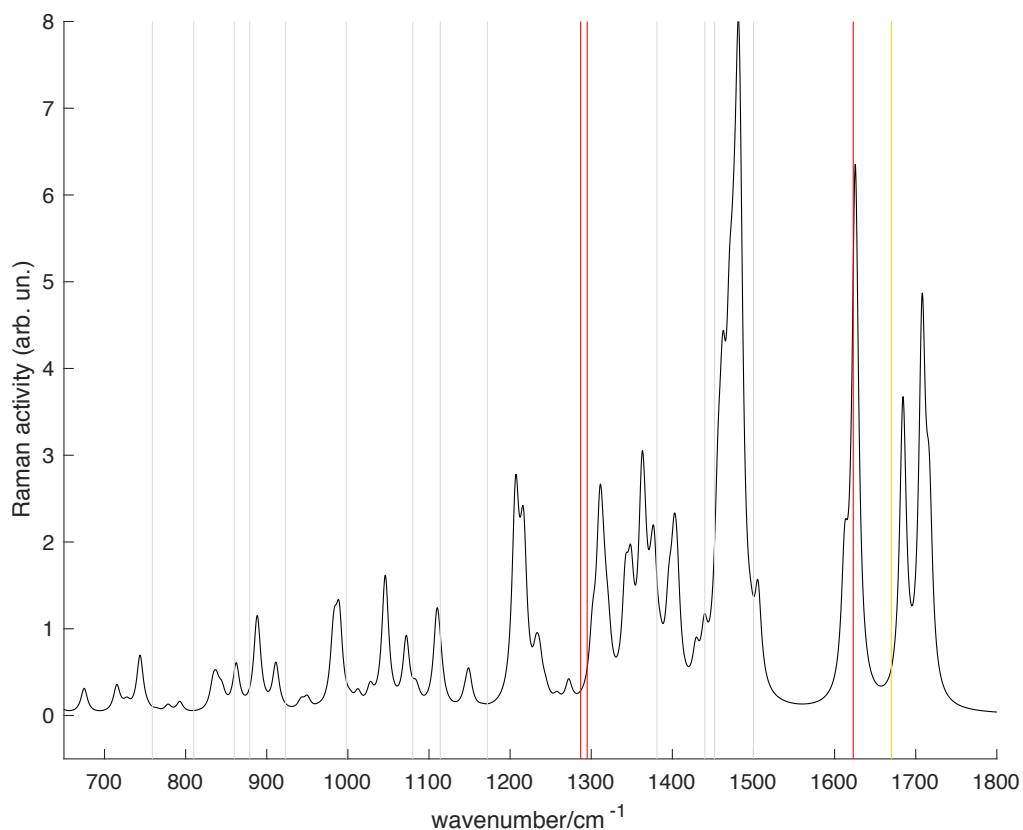
Figure 99. Simulated spectrum of CBGA. The vertical lines represent the wavenumbers corresponding to experimental peaks. Strong peaks are represented in red, medium in yellow and weak in grey.

	Experimental data	Scaled theoretical data	Theoretical data	Intensity	Normal mode	Representation of the normal mode
1	1114	1109	1132	weak	alkyl chain vibration.	
2	1295	1312	1338	strong	CH2 twisting.	

119
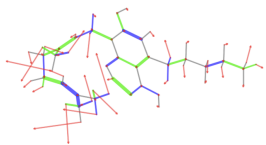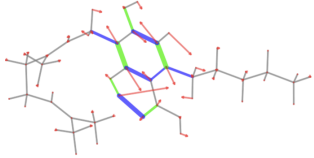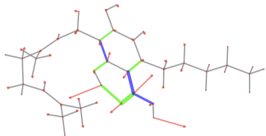
3	1500	1481	1511	weak	CH2 scissoring; CH3 vibration.	
4	1623	1625	1659	strong	stretching of the aromatic ring, OH bending.	
5	1670	1685	1719	medium	OH bending; contraction of C=O in the COOH group	

Table 18. Table containing the assignment of the peaks of the experimental spectrum and the simulated one for CBGA. For each experimental peak the scaled theoretical value, the theoretical data, the intensity, and the correspondent normal mode are reported. The picture represents the correspondent normal mode. The color green represents the stretching of the bond, the blue the shrinking.

5.3 Concluding remarks on the Raman peaks assignment

After the normal modes analysis of the previous paragraph, one first remarkable observation emerges about THC. The most intense peak of the simulated spectrum of THC (computed at 1476 cm^{-1}) is associated to a weak peak in the experimental spectrum (1462 cm^{-1}). This peak is assigned to the CH$_2$ bending, hence we can say that the CH$_2$ bending is weak in the experimental analysis of THC, since the corresponding peak is weak. This is unexpected from the theoretical (DFT) point of view. The discrepancy may be due to several reasons, such as for instance the interaction between molecules or the effect of the conformation. The latter is associated to the change in conformation of the alkyl chain that could facilitate the packing of the condensed phase and could justify this effect. To verify it, it would be necessary to analyze a crystalline sample of THC, which is not so easy since THC is volatile. In fact, no studies have been published on this topic yet. Remarkably, the peak associated to CH$_2$ stretching is very strong in alkanes. Consistently it also shows up a a strong peak in the spectrum simulated by DFT. To verify that the CH$_2$ bending peak is normally intense in the Raman spectroscopy of alkanes, we selected alkanes of similar molecular weight as the alkyl chain of THC and we assessed their Raman spectra (taken from Chemical Book [13]). For instance, the Raman spectrum of cyclohexane is reported in Figure 100.

CYCLOHEXANE

C_6H_{12} RM-01-00018 : 4880A.100M.LIQUID

2938	41
2923	37
2686	13
2853	53
1446	12
1267	12
1029	15
803	95

Figure 100. Raman soectrum of cyclohexane [13].

The peak associated to the CH_2 scissoring (also named bending), is not the most intense but it is clearly visible at wavenumber close to 1450 cm^{-1} (red arrow). It can be considered as a peak of medium intensity. We can also look at the Raman spectrum of hexane for further confirmation (Figure 101).

HEXANE

C_6H_{14} RM-01-00880 : 4880A.200M.LIQUID

2963	60
2939	76
2917	65
2901	59
2875	96
2863	72
1040	11
892	15
871	10

Figure 101. Raman spectrum of hexane [24].

In this case, the peaks are in general broader, and this is due to the higher flexibility of the hexane molecule. In fact, cyclohexane is a more rigid molecule, it only has two conformations (chair and boat), while hexane has several, being more flexible. Notably, the peak associated to the CH_2 bending of hexane is the most intense in the region of the medium IR.

In the experimental spectrum of THC this peak does not appear. The same issue can be observed in THCA, CBGA and CBDA. In the case of THCA, the stronger peak in the simulated spectrum at 1481 cm^{-1} can be associated to the experimental peak at 1461 cm^{-1}. While the first one is extremely strong, the experimental one is weak.

Another phenomenon that can be observed in the assignment of the peaks of cannabinoids is the change of relative intensity of the theoretical *vs.* experimental peaks (other than the CH_2 bending). In the experimental spectrum of THC we can see a strong peak at 1295 cm^{-1} and two medium peaks close to it at 1276 cm^{-1} and 1321 cm^{-1}. In the simulated spectrum the peak at 1295 cm^{-1} is associated to the one at 1306 cm^{-1}, which is stronger that the two peaks at 1275 cm^{-1} and 1330 cm^{-1} (which are associated to those at 1276 cm^{-1} and 1321 cm^{-1}) but it is not the stronger peak in the whole spectrum. The experimental peaks at 1365 cm^{-1} and 1371 cm^{-1} can be associated to 1362 cm^{-1} and 1397 cm^{-1} in the simulated spectrum of THC. The peak at 1362 cm^{-1} is more intense than the one at 1306 cm^{-1}, even if the corresponding peaks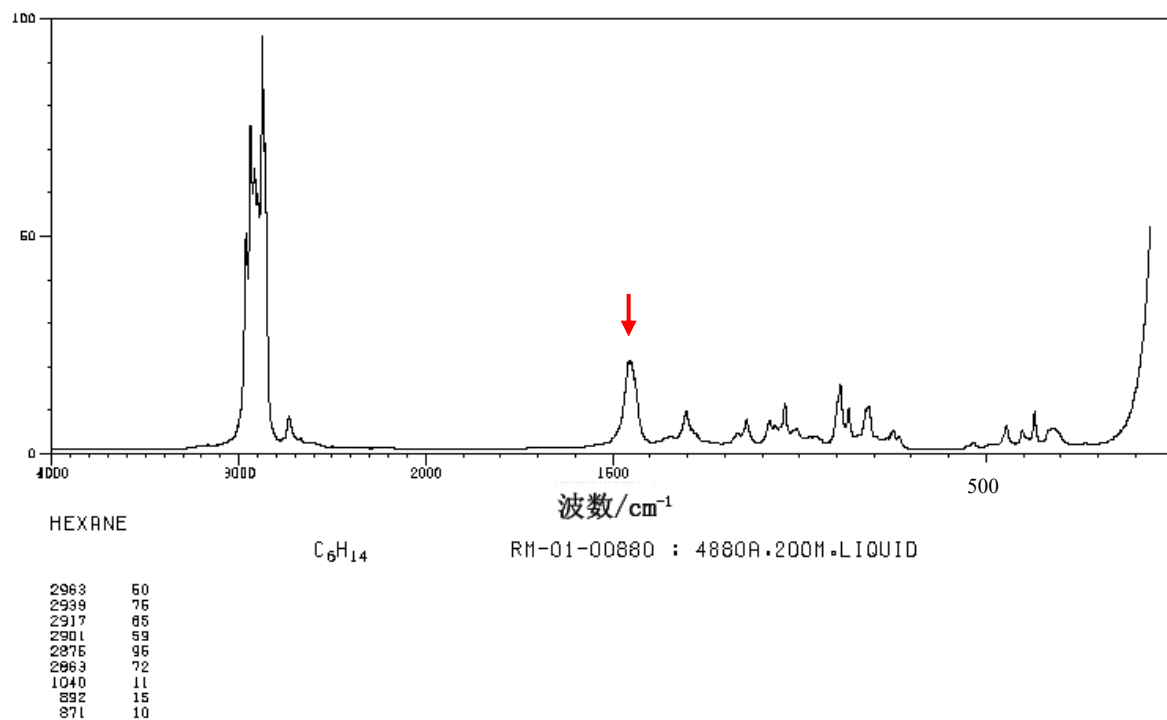 in the experimental spectrum have opposite relative intensities. The difference in the relative intensities may be due to intermolecular interactions with the surroundings of the molecule.

For CBD, the experimental peaks at 1663 cm^{-1} and 1623 cm^{-1} can be associated to the two strong peaks visible at 1645 cm^{-1} and 1690 cm^{-1}, which are the two most intense peak in the region above 1600 cm^{-1}. While the peak at 1663 cm^{-1} is the most intense experimental one, the peak at 1623 cm^{-1} is not. Between them we can see in the experimental spectrum a peak at 1643 cm^{-1}, which is stronger in the experimental spectrum but weaker in the calculation. This happens because the DFT calculation does not reproduce well the coupling of the two vibrations. This phenomenon can be also observed in CBGA. The peak at 1481 cm^{-1} in the simulated spectrum is very strong and it is associated to the experimental peak at 1500 cm^{-1}, which is instead very weak.

Interestingly, for THC the peak located at 1046 cm^{-1} in the simulated spectrum does not show up in the experimental spectrum. This peak is assigned to the OH bending. It appears in the simulated spectrum since it considers an isolated molecule, while in reality the THC molecules interact one another through the OH groups that may form hydrogen bonds that affect significantly the position of the OH bending vibration. Therefore, the peak at 1046 cm^{-1} does not appear in the experimental spectrum. The same phenomenum can be observed in THCA, where the peak at 1046 cm^{-1} in the simulated spectrum does not appear in the experiment. Similarly, the peak at 1047 cm^{-1} in the simulated spectrum of CBD does not appear in the experimental spectrum, and the same happens in CBGA. All such peaks are associated to the OH bending of the free molecule. Therefore, in the simulated spectrum, which represents an isolated molecule, this peak is clearly visible, and it is relatively strong. However, in the experimental spectrum, this peak is not present, since the OH groups of different molecules interact one with each other or with water molecules from the ambient. This observation

indicates clearly that the molecules are aggregated, and also shows how the analysis and assignment of the peak helps us to understand some characteristics of the sample.

6. Conclusions and perspectives

The interest in *Cannabis Sativa* and in the cannabinoids stored in its trichomes is continuously growing. Cannabinoids have several properties, such as analgesic, anti-inflammatory, antioxidant, which are responsible for their fast diffusion and their use in several fields, for example the pharmaceutical one. The methods which are currently used to differentiate amongst cannabinoids in a given sample are expensive, time consuming and complex. Therefore, a new technique, which is faster, cheaper and non-destructive must be identified. Raman spectroscopy could be the solution to this issue, since it has been used for years for the chemical characterization of chemical compounds. The main goal of this thesis is the one to use micro-Raman spectroscopy to obtain relevant and detailed chemical information on the presence of specific cannabinoids in a given trichome sample. The main problematic associated to Raman spectroscopy is that, in order to obtain a precise calibration, reference spectra of selected chemical compounds at different testing conditions are needed. However, there is a lack of vibrational information on cannabinoids, not much literature is available on the topic and the the intrinsic fluorescence, the complex chemical composition and the complex geometry of the sample makes the analysis very complex. An analysis of this kind was never done on such a complex geometry and this thesis project wants to provide an initial demonstration of the use of this technology.

More than one hundred cannabinoids are currently known, but only the most relevant ones, for concentration and properties, were considered in our work (THC, THCA, CBD, CBDA, CBG, CBGA). A sample of *Cannabis Sativa* flower trichome was analyzed using Raman spectroscopy and principal component analysis. It was possible to obtain scoremaps that show three brighter regions of dimensions of about 8 μm. This diameter correlates with the dimension of the secretory vesicles present in the trichome. Moreover, the analysis of the second principal component, which is characterized by positive and negative peaks, demonstrated that when the score along the second principal component is high there is a high concentration of THC and THCA. Since in the score map we can notice the presence of three brighter regions, we can say that they are associated to coordinates with a higher second principal component, and therefore a higher concentration of THC and THCA. We can therefore see how, using Raman spectroscopy and principal component analysis, it is possible to observe that THC and THCA accumulate in the trichomes of the *Cannabis* flower.

After demonstrating that micro-Raman spectroscopy and PCA can be used to extract spectroscopic and spatial information on the selected trichome sample, the nature of the Raman signals observed was analyzed. The goal was to assign by Density Functional Theory calculations the main peaks observed in the average spectrum of the trichome to specific normal modes of the relevant cannabinoids. For instance, we could observe that the mode associated to the main peak in the spectrum of THC (1295 cm^{-1}) is delocalized in the cannabinoid molecule, and is described as a CH_2 twisting, CH and OH bending, while the same peak of THCA is described as asymmetric twisting of the CH_2 in the alkyl chain and OH vibration. No similar peaks have been identified in the other cannabinoids here considered. This allows us to differentiate amongst different compounds and it is important for a reliable interpretation of the Raman data, especially in a scenario where the intensity of specific Raman markers is taken as a proxy of the amount of cannabinoids in the sample.

This thesis provided a general analysis of the relevant Raman peaks for each cannabinoid. However, literature is still scarce, and cannabinoids are very complex molecules with several different conformers. This thesis could be a starting point for future developments on the use of Raman spectroscopy to investigate cannabinoid-containing samples. Besides their use in the pharmaceutical field, the development of this technique could lead to the application of Raman spectroscopy in forensics and to a non-destructive analysis of the plant, and the identification of the best harvesting time.

Scientific acknowledgements

References

[1] Sánchez, L., D. Baltensperger and D. Kurouski. "Raman-Based Differentiation of Hemp, Canabdiol-Rich Hemp and *Cannabis*." Analytical chemistry (2020), 10.1021/acs.analchem.0c00828, 2020.

[2] Avinash, Priya Ranjan. (2020) "Update of Cannabis and its medical use".

[3] Livingston, S. J., Quilichini, T. D., Booth, J. K., Wong, D., Rensing, K. H., Laflamme-Yonkman, J., Castellarin, S. D., Bohlmann, J., Page, J. E., & Samuels, A. L. (2020). "*Cannabis* glandular trichomes alter morphology and metabolite content during flower maturation". The Plant journal : for cell and molecular biology, 101(1), 37–56. https://doi.org/10.1111/tpj.14516

[4] Gijs Schaftenaar. Molden [Internet]. (2018). Available from: https://www3.cmbi.umcn.nl/molden/

[5] WHO Expert Committee on Drug Dependence (2018). "Critical Review - Delta-9-tetrahyrocannabinol".

[6] R. G. Pertwee, "Handbook of *Cannabis*", (2018), Aberdeen, UK: Oxford University Press.

[7] P. Mahmoud A. ElSohly, "Marijuana and the Cannabinoids", (2007), Oxford, MS: SERIES EDITOR.

[8] 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. https://doi.org/10.1016/j.phytochem.2012.11.001

[9] Simmons, Aaron & Gurr, Geoff. (2005). "Trichomes of Lycopersicon species and their hybrids: Effects on pests and natural enemies". Agricultural and Forest Entomology. 7. 265 - 276. 10.1111/j.1461-9555.2005.00271.x.

[10] Dayanandan, P. & Kaufman, Peter. (1976). "Trichomes of *Cannabis Sativa L. (Cannabaceae)*". American Journal of Botany - AMER J BOT. 63. 10.2307/2441821.

[11] Hammond, C.T., & Mahlberg, P. (1973). "Morphology of glandular hairs of *Cannabis* Sativa from scanning electron microscopy". American Journal of Botany, 60, 524-528.

[12] Shepard, Glenn. (2002). "Nature's Madison Avenue: Sensory cues as mnemonic devices in the transmission of medicinal plant knowledge among the Matsigenka and Yora of Peru".

[13] Chemical Book,
https://www.chemicalbook.com/ProductList_en.aspx?kwd=cyclohexane. [Online].

[14] Sommano, Sarana R., Chuda Chittasupho, Warintorn Ruksiriwanich, and Pensak
Jantrawut. (2020). "The *Cannabis* Terpenes" Molecules 25, no. 24: 5792.
https://doi.org/10.3390/molecules25245792

[15] Dustin W. Shipp, Faris Sinjab, and Ioan Notingher, "Raman spectroscopy: techniques
and applications in the life sciences," (2017) Adv. Opt. Photon. 9, 315-428

[16] Smith, B.C., & Dent, G. (2019). "Modern Raman Spectroscopy: A practical approach". 1:
1-21, 3: 71-92. John Wiley & Sons, Ltd.

[17] Tommasini, M.M.S. (2017). "Lecture notes of the course: Physical Properties of
Molecular Materials", 2017.

[18] Jolliffe, Ian & Cadima, Jorge. (2016). "Principal component analysis: A review and recent
developments". Philosophical Transactions of the Royal Society A: Mathematical,
Physical and Engineering Sciences. 374. 20150202. 10.1098/rsta.2015.0202.

[19] Sunghwan Kim, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li,
Benjamin A Shoemaker, Paul A Thiessen, Bo Yu, Leonid Zaslavsky, Jian Zhang, Evan E
Bolton, PubChem in 2021: new data content and improved web interfaces, Nucleic
Acids Research, Volume 49, Issue D1, 8 January 2021, Pages D1388–
D1395, https://doi.org/10.1093/nar/gkaa971

[20] D. J. a. H. FE, "Conformational Study of Cannabinoid Docking to Cannabinoid Receptor 1
(CB1) via Linear and Nonlinear Circular Dichroism". (2016). Journal of Physical Chemistry
& Biophysics.

[21] Tommasini, M.M.S., & Zanchi, C. (2020). "Lecture notes of the course: Sperimentazione
e modellazione delle proprietà dei materiali".

[22] E. N. S. Small, "Size matters: evolution of large drug-secreting resin glands in elite
pharmaceutical strains of *Cannabis Sativa* (marijuana)". (2016). Genet Resour Crop
Evol63, 349–359 . https://doi.org/10.1007/s10722-015-0254-2.

[23] Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., & Hutchison, G.
R. (2012). Avogadro: an advanced semantic chemical editor, visualization, and analysis
platform. Journal of cheminformatics, 4(1), 17. https://doi.org/10.1186/1758-2946-4-17

[24] Chemical Book, https://www.chemicalbook.com/ProductList_en.aspx?kwd=hexane.
[Online].

[25] Wikipedia, L'enciclopedia libera (2021), "Tetrahydrocannabinol", https://en.wikipedia.org/wiki/Tetrahydrocannabinol

[26] Wikipedia, L'enciclopedia libera (2021), "Tetrahydrocannabinolic acid, https://en.wikipedia.org/wiki/Isoprene

[27] Wikipedia, L'enciclopedia libera (2021), "Isoprene" https://en.wikipedia.org/wiki/Isoprene

[28] Wikipedia, L'enciclopedia libera (2021), "Cannabidiol" https://en.wikipedia.org/wiki/Cannabidiol

[29] Wikipedia, L'enciclopedia libera (2021), "Cannabidiolic acid" https://en.wikipedia.org/wiki/Cannabidiolic_acid

[30] Wikipedia, L'enciclopedia libera (2021), "Cannabigerol" https://en.wikipedia.org/wiki/Cannabigerol

[31] Wikipedia, L'enciclopedia libera (2021), "Cannabigerolic acid" https://en.wikipedia.org/wiki/Cannabigerolic_acid

[32] Renishaw: apply innovation. https://www.renishaw.com/en/invia-confocal-raman-microscope--6260. [Online].