UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE CHIMICHE

SCUOLA DI DOTTORATO DI RICERCA IN SCIENZE MOLECOLARI

INDIRIZZO SCIENZE CHIMICHE

CICLO XXVIII

NMR based Metabolomics in Food Chemistry

Direttore della Scuola: Ch.mo Prof. Antonino Polimeno

Coordinatore di Indirizzo: Ch.mo Prof. Antonino Polimeno

Supervisore: Ch.mo Prof. Stefano Mammi

Dottorando: Jalal Uddin

Dedicated To:

My Parents

Who Always Pray for me

and

My Supervisor

Who Always Encourages me

Summary

Metabolomics is defined as the systematic analysis of hundreds or thousands of small metabolites present in a living system. It has emerge as an important field of study along with other, already established 'omics' sciences, *i.e.*, genomics, proteomics and transcriptomics. Metabolomics is well established in the field of medicine, drug toxicity and disease diagnosis. Among the existing analytical techniques, NMR is a fast, reproducible and non-destructive technique to construct an informative snapshot of the metabolites under certain conditions. NMR data give metabolic signature information of the samples when it is combined with data preprocessing and chemometric tools, such as multivariate statistical techniques. NMR-based metabolomics is still expanding in the field of the food chemistry. In this context, this Ph.D. thesis is focused on two major aspects, which show applications of NMR-based metabolomics in food chemistry.

1. Many nutraceutical products possess powerful antioxidant activity as demonstrated in many chemical *in vitro* tests and in several *in vivo* trials. Nevertheless, the mechanism of their activity is not completely studied in detail. Due to their poor bioavailability and fast metabolism, studies on the *in vivo* antioxidant effects are still needed. We performed longitudinal experiments on Sprague Dawley (SD) rats using two commonly available nutraceutical antioxidant products, namely, Curcumin (chapter 2) and Resveratrol (chapter 3). The effects of different doses of orally administered standardized antioxidant extracts in healthy rats were investigated by untargeted metabolomic analysis based on LC-MS and NMR spectrometry. The experiments were carried out over different periods of time for different antioxidants. Changes in the urinary metabolome were evaluated by monitoring the 24-hour urine composition by ^{1}H-NMR and HPLC-MS. The two different approaches were able to detect variations in the urinary levels of antioxidant markers, leading to the observation of different metabolites thus proving the complementarity of these two analytical techniques for metabolomic purposes. Analytical tools such as MS and NMR spectroscopy in combination with chemometrics can profile the impact of time, stress, nutritional status, and environmental perturbations on hundreds of metabolites simultaneously. This results in complex, massive data sets that must be analyzed through a careful statistical protocol. Our strategy included data preprocessing, data analysis and validation of statistical models. After several data processing steps, principal component analysis (PCA)

and partial least-squares (PLS) were used to identify urine biomarkers. The PLS models were validated by permutation tests and critically important variables were validated through univariate analysis.

2. The second part of this thesis (chapters 4 and 5) describe the use of NMR-based metabolomics as a fast, convenient, and effective tool for origin discrimination and biomarker discovery in food analysis. Traditionally, the determination of the floral origin of honey is made from palynological analysis. The method is based on the identification of pollen by microscopic inspection. However, melissopalynological analysis needs expertise and also it is not a very reliable technique for the discrimination of botanical origin of some types of honey. Also, honey regulation in the EU (Codex Alimentarius Commission 2001; European Commission 2002) emphasizes that the botanical and geographical origins of the product must be printed on the label in order to avoid the fraud such as adulteration with industrial sugar, selling product under false name or mixing the honey of different floral origin. Therefore, there is need to establish a method to discriminate honey from different origin. The aim of this work was to develop an NMR-based metabolomic approach that used multivariate statistical analysis to discriminate the botanical (chapter 4) and entomological (chapter 5) origin of different types of honey. Multivariate statistical analysis helped us to identify the most relevant signals to differentiate honey botanically and entomologically. The obtained data sets were useful in the search of markers responsible for the discrimination of different honey samples from different botanical species and produced by different bee species.

Sommario

La metabolomica è definita come l'analisi sistematica di centinaia o migliaia di piccoli metaboliti presenti in un sistema vivente. È emerso come un importante campo di studio insieme ad altre, già affermate scienze "omiche", vale a dire, genomica, proteomica e trascrittomica. La metabolomica è ben consolidata nel campo della medicina, nello studio della tossicità di farmaci e nella diagnostica. Tra le tecniche analitiche esistenti, NMR è veloce, riproducibile e non distruttiva, utile per fornire una fotografia informativa sui metaboliti in determinate condizioni. Dati NMR forniscono informazioni metaboliche che caratterizzano i campioni quando combinati con una pre-elaborazione dei dati e con strumenti chemiometrici, come le tecniche di statistica multivariata. La metabolomica basata sull'NMR è ancora in espansione nel campo della chimica degli alimenti. In questo contesto, questa tesi di Dottorato si concentra su due aspetti principali, che mostrano applicazioni della metabolomica basata sull'NMR in chimica degli alimenti.

1. Molti prodotti nutraceutici possiedono potente attività antiossidante, come dimostrato in molti test chimici *in vitro* e in diverse prove *in vivo*. Tuttavia, il meccanismo della loro attività non è completamente studiato in dettaglio. A causa della loro scarsa biodisponibilità e metabolismo veloce, sono ancora necessari studi *in vivo* sugli effetti antiossidanti. Abbiamo condotto esperimenti longitudinali su ratti Sprague Dawley (SD) utilizzando due prodotti antiossidanti nutraceutici comunemente disponibili, vale a dire, curcumina (capitolo 2) e resveratrolo (capitolo 3). Gli effetti di diverse dosi di estratti antiossidanti standardizzati somministrati per via orale nei ratti sani sono stati studiati mediante analisi metabolomica non mirata (*untargeted*) basata su LC-MS e spettrometria NMR. Gli esperimenti sono stati eseguiti lungo diversi periodi di tempo per diversi antiossidanti. Le variazioni del metaboloma urinario sono state valutate attraverso il monitoraggio della composizione delle urine di 24 ore usando ^{1}H-NMR e HPLC-MS. I due differenti approcci sono stati in grado di rilevare le variazioni dei livelli urinari di marcatori antiossidanti, portando all'osservazione di diversi metaboliti e dimostrando così la complementarità di queste due tecniche analitiche per scopi metabolomici. Strumenti di analisi come la spettroscopia NMR e MS in combinazione con chemiometria possono delineare l'impatto del tempo, dello stress, dello stato nutrizionale, e di perturbazioni ambientali su centinaia di metaboliti contemporaneamente. Ciò comporta complessi enormi set di dati che

devono essere analizzati mediante un protocollo statistico accurato. La nostra strategia ha compreso una pre-elaborazione dei dati, l'analisi dei dati e la validazione dei modelli statistici. Dopo varie fasi di elaborazione di dati, l'analisi delle componenti principali (PCA) e l'analisi dei minimi quadrati parziali (PLS) sono state utilizzate per identificare i biomarcatori urinari. I modelli PLS sono stati convalidati dai test di permutazione e le variabili di importanza critica sono stati convalidati attraverso analisi univariata.

2. La seconda parte di questa tesi (capitoli 4 e 5) descrivono l'uso di metabolomica basata su NMR come strumento veloce, conveniente ed efficace per la discriminazione di origine e la scoperta di biomarcatori in analisi degli alimenti. Tradizionalmente, la determinazione dell'origine floreale del miele è condotta mediante analisi palinologica. Il metodo si basa sulla individuazione di polline mediante ispezione microscopica. Tuttavia, l'analisi melissopalinologica richiede perizia ed inoltre non è una tecnica molto affidabile per la discriminazione di origine botanica di alcuni titpi di miele. Inoltre, la regolamentazione del miele nell'Unione Europea (Codex Alimentarius 2001; Commissione Europea 2002) sottolinea che le origini botaniche e geografiche del prodotto devono essere stampate sull'etichetta, per evitare frodi, come l'adulterazione con zucchero industriale, vendita di prodotti sotto falso nome o aggiunte di miele di diversa origine floreale. Pertanto, vi è la necessità di stabilire un metodo per distinguere miele di diverse origini. Lo scopo di questo lavoro è stato quello di sviluppare un approccio metabolomico basato sull'NMR che ha utilizzato l'analisi statistica multivariata per discriminare l'origine botanica (capitolo 4) ed entomologica (capitolo 5) di diversi tipi di miele. statistica multivariata ci ha aiutato ad identificare i segnali più importanti per differenziare il miele sia dal punto di vista botanico che entomologico. I set di dati ottenuti sono stati utili nella ricerca di marcatori responsabili della discriminazione dei diversi campioni di miele di diverse specie botaniche e prodotti da diverse specie di api.

Acknowledgment

First of all, countless thanks to 'Almighty Allah', worthy of all praise, Most gracious and Most Merciful, Who guides me in difficulties and gives me courage to complete my work.

I am very grateful to my supervisor, Prof. Stefano Mammi for his precious attention and guidance, continuous support and encouragement throughout my PhD research work. This thesis wouldn't complete without his support and help. He was very patient with my bad writing and always spent time to check my yearly progress reports with great interest and high enthusiasm. By observing his personality, hard work and dedication, I learnt the key to success. Sir, this doesn't cease here, this relationship of student and teacher would go further and you will always have a respect and special place in my heart. He is one of the best mentor and calm listener I have ever met.

I wish to express my sincere thanks to Dr. Elisabetta Schievano (Betty) for her cooperation and valuable suggestions to complete my thesis. You were always being very patient about my all weird suggestions and discussions. The experience that I have learnt from you I would never forgive and would like to use in future. Thanks you for sharing your knowledge with me and I always consider you as my co-supervisor.

It has been pleasure to work with some amazing and talented peoples during my PhD stay in the lab. I would like to say thanks with bottom of my heart to Emmanuel, Claudia and Valentina for all their moral and technical support. Specially, Claudia who tolerate me a lot and help me to rescue many time from ordering chemicals, washing vials and NMR tube.

A very special thanks to the Dr. Stefano Dall'Acqua and his group (Gregorio and Irene) to perform all the Mass Spectrometry experiment and sharing their data with me. Without your help and support this thesis wouldn't have been possible. Particularly, Gregorio who always been supportive and quicker to share the data and information. Thank you Dr. Stefano for letting me introduce to the state-of-art mass spectrometry based metabolomics.

I would also not forgot the support and care shown by Ileana from NMR section. She teaches me in beginning how to use the NMR spectrometer. A special thanks to Dr. Massimo Bellanda for sharing his knowledge on the theory of NMR.

Outline of the thesis

This PhD thesis is divided into five chapters.

Chapter 1 discusses the concepts and terms used in the field of metabolomics**.** Also, it emphasizes the basic principles of NMR and its application in food metabolomics with some literature examples. Finally, the chapter ends with the definition and role of chemometrics in metabolomics and some specific aspects of chemometrics in NMR-based metabolomics.

Chapter 2 deals with the application of NMR-based metabolomics to assess the *in vivo* effect of nutraceuticals containing antioxidants on rat urine. Specifically, curcumin was used in a high dose in a longitudinal experiment in which urine was collected over a seven week period. An untargeted MS and NMR-based metabolomic approach was used to identify markers of oxidative stress in rat urine.

Chapter 3 shows the continuation of our study on the extract of *Polygonum cuspidatum*, a source of resveratrol, an antioxidant also found in red wines or in other foods of the Mediterranean diet. The aim of the work was to analyze the *in vivo* effect of resveratrol on rat urine, again using NMR and Mass spectrometry-based metabolomics.

Chapter 4 covers the application of NMR-based metabolomics in the field of food quality assurance; specifically, we used honey as a food matrix. The aim of the work was to use NMR as an analytical tool to differentiate citrus honey from clementine honey, two products deriving from very similar plants.

Chapter 5 describes the role of NMR-based metabolomics in combination of chemometrics to discriminate the entomological origin of Equador honey. It also dicusses the authentication of genuine honey and its differentiation from fake honey with a simple liquid-liquid chloroform extraction.

Chapter 6 briefly summarizes the finding of the research describes in this thesis.

List of Figures

X

List of Tables

List of Abbreviations

^{1}H-NMR	Proton Nuclear Magnetic Spectroscopy
2D-NMR	Two Dimensional Nuclear Magnetic Spectroscopy
2-HAF	2-Hydroxyacetyl-furan
2-OG	2-Oxogluturic acid
5-HMF	5-Hydroxymethyl furfural
8-*OHdG*	8-hydroxy-2′-deoxyguanosine
ACD	Advanced Chemistry Development
ACTH	Adrenocorticotropic hormone
BPI	Base Peak Intensity
CDCl$_3$	Deutrated Chloroform
CE	Capillary Electrophoresis
CEASA	Comitato Etico per la Sperimentazione Animale – University of Padova ethical committee
CHCl$_3$	Chloroform
CV	Cross Validation
DFI/LC-MS/MS	Direct Flow Injection/Liquid Chromatography-Tandem Mass Spectrometry
DHR	Dihydroresveratrol
DMG	Dimethylglycine
EDTA	Ethylenediaminetetraacetic acid
EMA	Economically Motivated Adulteration
ESI	Electrospray Ionization
ESI-QTOF	Electrospray Ionization-Quadrupole Time-Of-Flight
FCP	Free Choice Profile
FID	Free Induction Decay

GC	Gas Chromatography
GC-MS	Gas Chromatography Mass Spectrometry
GD	Golden Drop
GSH	Glutathione
HMDB	Human Metabolome Database
HPLC	High-Performance Liquid Chromatography
HPLC/FD	High-Performance Liquid Chromatography/ Fluorescence Detection
HPLC-DAD	High-Performance Liquid Chromatography- Diode-Array Detection
HPLC-MS	High-Performance Liquid Chromatography-Mass Spectrometry
ICP-MS	Inductively Coupled Plasma Mass Spectrometry
IR	Infrared
MA	Methyl Anthranilate
MS	Mass Spectrometry
MVA	Multivariate
NEM	N-ethylmaleimide
NIR	Near Infrared
NMR	Nuclear Magnetic Resonance
NOESY	Nuclear Overhauser Spectroscopy
OPLS-DA	Orthogonal Partial Least Square Discriminate Analysis
PCA	Principal Component Analysis
PLS	Partial Least Square
PLSDA	Partial Least Square Discriminant Analysis
ptPLS2	Post-transform Partial Least Square
RES	Resveratrol

RF	Radio Frequency
RT	Retention Time
SDBS	Spectral Database for Organic Compounds
SEM	Standard Error of the Mean
SIMCA	Soft Independent Modeling of Class Analogy
SSA	Sulphosalicylic Acid
Tdds	Turbo Depending Data Scanning
TLC	Thin Layer Chromatography
TOCSY	Total Correlation Spectroscopy
TSP	3-(trimethylsilyl)propionate-2,2,3,3,-d4
UPLC	Ultra-Performance Liquid Chromatography
UPLC-HRMS	Ultra-Performance Liquid Chromatography-High Resolution Mass Spectrometry
UV	Ultra-Violet
VIP	Variables Important in the Project
γ-Glu-Glu	γ-glutamyl-glutamic acid

Table of Contents

Chapter 1. General Introduction

Chapter 2. New Findings on the *in vivo* Antioxidant Activity of *Curcuma longa* Extract by an Integrated ¹H NMR and HPLC–MS Metabolomic Approach.

Chapter 3: Metabolic Effects of *Polygonum Cuspidatum* Supplementation on Urinary Composition of Healthy Rats: a Longitudinal Study.

Chapter 4: An NMR-based Metabolomics Approach to Seek Reliable Markers to Distinguish between Citrus and Clementine Italian Honey with Chemometric Analysis.

XX

Chapter 1: General Introduction

1.1 Metabolomics in Systems Biology

Cellular metabolites are the final outcome of the response of biological systems to genetic changes, stress or other environmental influence. The study of their concentration is therefore central to understand biological systems and to characterize different physiological and pathological conditions [1]. In the last twenty years, the technological progress in the field of biological sciences made it possible to analyze complete sets of genes (genomics), mRNA (transcriptomics), proteins (proteomics) and metabolites (metabolomics) rather than to study single genes, single mRNA, single proteins or single metabolites in a living system. While other "omics" sciences are very well established, metabolomics is still in its growing stages. The detailed investigations of the entire ''omics cascade'', from genomics to transcriptomics to proteomics to metabolomics will play a vital role in the field of systems biology [2, 8]. **Figure 1.1** shows how the process of information is successfully passed on from genes to metabolites in a living system. [1, 2]
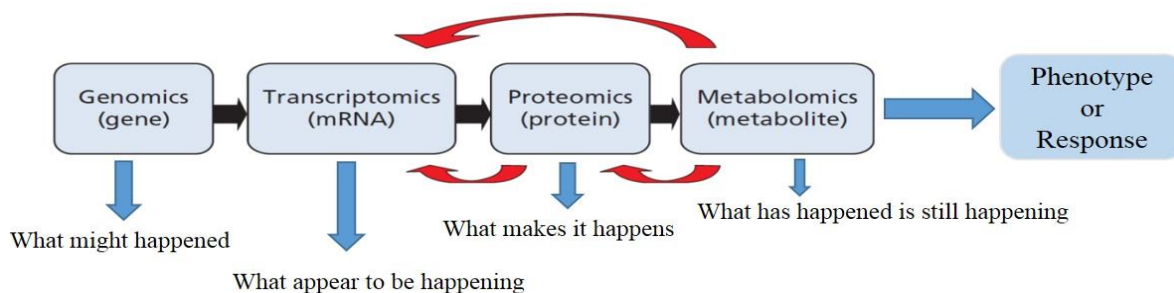
Figure 1.1: The "omics" cascade

1.2 Metabolomics

Metabolomics is the endpoint of the ''omics cascade'' as shown in the **Figure 1.1** and is the closest to the phenotype, which gives metabolomics the advantage to serve as a direct signature of biochemical activity and earned it the description of "the apogee of omics" [3]. The term metabolome was first used by Oliver [2] as the set of low-molecular-mass compounds synthesized by an organism. A few years later, the term metabolomics was introduced by Fiehn in 2001 as "the identification and quantification of every single metabolite presents in a

biological system" [4]. Before this in 1999s, the term metabonomics was used by J. K. Nicholson and colleagues as follows: "the quantitative measurement of the time related multi-parametric metabolic response of living systems to pathophysiological stimuli or genetic modification" [5]. There has been some conflict over the exact differences between the two terms, metabolomics and metabonomics. This difference is mainly philosophical, rather than technical; although there are some differences in concept, in practice, the analytical and modeling procedures are the same and the two terms are often used interchangeably [6]. The term metabolomics will be used in this thesis throughout to cover in vivo study of anti-oxidants compounds in rat urine and honey origin discrimination.

Metabolomics attempts to systematically identify and quantifies metabolites from biological samples. The small molecules characterize the end result of complexity of biological processes in a given cell, tissue or organ and thus form attractive candidates to understand disease phenotypes as shown in **Figure 1.2**. Metabolites represent a diverse group of low-molecular weight structures including lipids, amino acids, peptides, nucleic acids and organic acids, which makes comprehensive analysis a difficult analytical challenge [7].
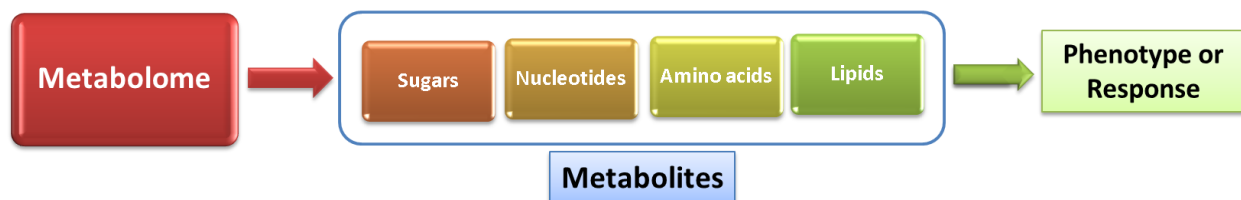
Figure 1.2: Complexity of Metabolome**.**

Metabolomics has been growing rapidly since its inception in the late 2000's (Figure 1.3) and the increasing number of papers published in the field of metabolomics makes it an emerging tool to study changes in phenotype under different biological conditions.
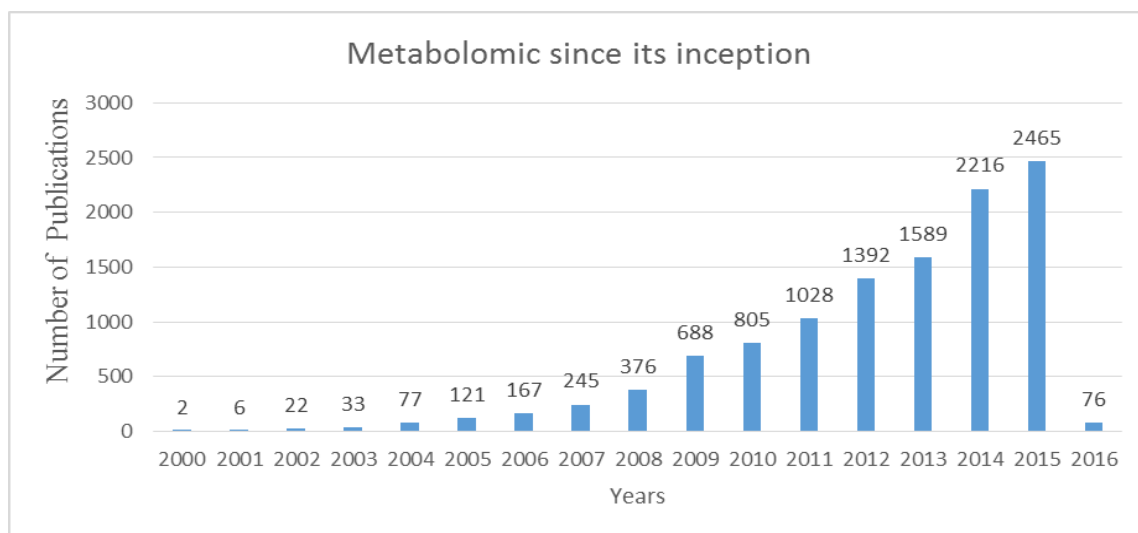
Figure 1.3: PubMed search result of the number of publications with "metabolomics" as a key word.

1.3 Approaches in Metabolites Analysis

Currently, four complementary approaches are used for metabolomic investigations: Metabolic profiling, Metabolic fingerprinting, Metabolite target analysis and Metabolomics [1, 8]. A summary of metabolomics-related definitions are depicted in **Figure 1.4**.

Fiehn [1] gave the first detailed definition of metabolomics and its subsections: (1) Metabolite target analysis, aims at quantitative analysis of substrate and/or product metabolites of a target protein; (2) metabolic profiling, aims to quantify a pre-defined set of metabolites belonging to a class of compounds or members of particular pathways or a linked group of metabolites; (3) metabolomics, attempts an unbiased overview of whole-cell metabolic patterns. For a more rapid analysis, (4) metabolic fingerprinting can be used, which reduces the analytical effort to the analysis of these intra-cellular metabolites with biochemical relevance [8, 9].

Metabolic Fingerprinting
Crude metabolite mixtures for classification. (FT-IR/Raman/DIMS)

Selection of technology is a compromise between speed, selectivity and sensitivity.

Metabolite target analysis
Analysis of specific metabolites.

Four Approaches

Metabolic profiling
Quantification of pre-defined targets.
(GC-MS, LC-MS, NMR, HPLC, LC/MS/MS)

Particular interest in **low molecular weight** compounds – the substrates and products in pathways.

Metabolomics
Unbiased identification of all metabolites in sample.

Figure 1.4: Different approaches used in metabolomics.

1.4 Analytical Platforms Used in Metabolomics

Metabolomics aims at the comprehensive and quantitative analysis of wide arrays of metabolites in biological samples. These numerous analytes have diverse physical-chemical properties and occur at different abundance. Consequently, comprehensive metabolomic investigations are primarily a challenge for analytical chemistry [9]. From the beginning, a variety of different analytical platforms have been used for metabolomics. Among them are thin-layer chromatography (TLC) [10], infrared spectroscopy (IR) [92], near infrared (NIR) [93], gas chromatography (GC) [12], nuclear magnetic resonance (NMR) [13, 14], high-performance liquid chromatography (HPLC) [14], mass spectrometry (MS) by direct-infusion [15] or coupled to chromatographic techniques, such as GC [16, 17], capillary electrophoresis (CE) [18] and liquid chromatography [19-20]. Among these different analytical techniques and tools used within metabolomics, Mass Spectrometry (MS) and NMR spectroscopy are the most popular and can profile the impact of time, stress, nutritional status, and environmental perturbation on hundreds of metabolites simultaneously resulting in massive, complex data sets [8, 21-22]. **Figure 1.5** shows a typical metabolomic workflow for an untargeted metabolomic study. This

strategy starts with the processing of the spectral data to generate the sample metabolic information. Different analytical techniques, such as NMR or MS, can be used to acquire the data. The main focus of this thesis is NMR-based metabolomics and its applications in food chemistry, as described in detail in the next section. Once the complete set of metabolomic features has been produced, multivariate data analysis methods are applied to investigate the general structure of the metabolomic data and how the different metabolic features are related with the phenotypic data associated with the samples. Finally, significant metabolites observed are identified with the help of online databases and their biological interpretation in biological system discussed [23].
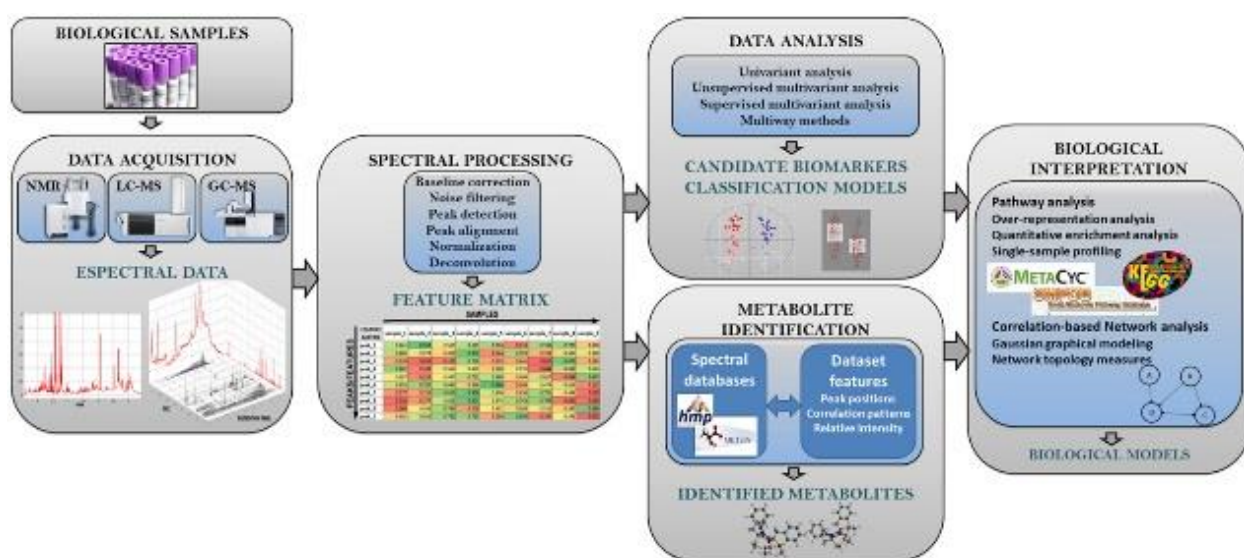
Figure 1.5: Analytical workflow in metabolomics and different steps involved. [23]

1.5 Nuclear Magnetic Spectroscopy

NMR, with the help of other complementary spectroscopic techniques, such as Infrared (IR), Ultraviolet (UV) and MS, is generally used to determine an organic compound's unique structure. In this section, ^{1}H-NMR spectroscopy will be briefly described.

Many atomic nuclei have a property called *spin*. Nuclei that possess either odd mass or odd atomic number or both have a quantized spin angular momentum and a magnetic moment. The magnetic moment (μ) is directly proportional to the angular momentum (P): $\mu = \gamma P$, where γ, the gyromagnetic ratio, is a constant for each nucleus. Each nucleus with spin has a nuclear spin quantum number, I, which determines the allowed spin states. For each nucleus, I is a physical

constant, and there are $2I+1$ allowed spin states with integral differences ranging from $+I$ to $-I$. Nuclei with $I = 0$ cannot be seen in NMR. In the absence of an external applied magnetic field, the nuclear spins are oriented in random directions as shown in **Figure 1.6(A).** As a result the samples do not have any macroscopic magnetization. When an external magnetic field (B_0) is applied, the nuclei align themselves either with or against the field of the external magnet, as in **Figure 1.6(B).** [24].
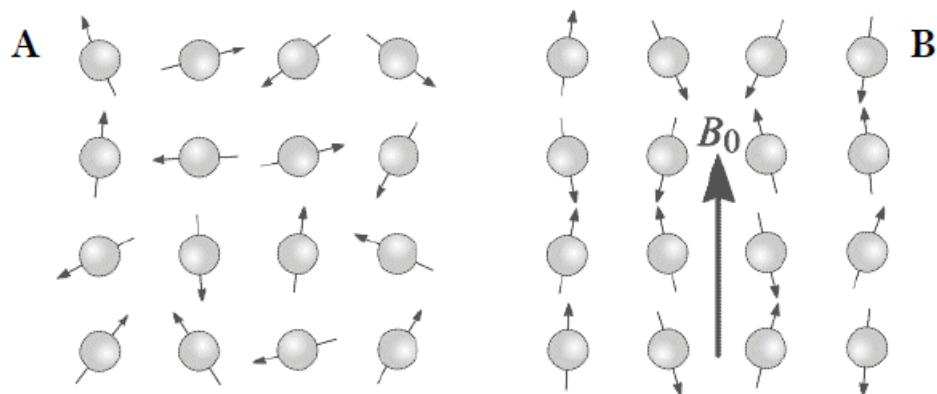
Figure 1.6 (A-B): (A) Nuclei in the absence of a magnetic field, the individual nuclear magnetic moments; (B) In the presence of an applied magnetic field, the nuclear magnetic moments are aligned with the applied field.

For $I = \frac{1}{2}$ nuclei, two spin states are possible: one at lower energy, called α, and one at higher energy, called β (Figure 1.7). The nuclear magnetic resonance phenomenon takes place when nuclei aligned with an applied magnetic field absorb energy and change state according to the direction of the applied magnetic field.
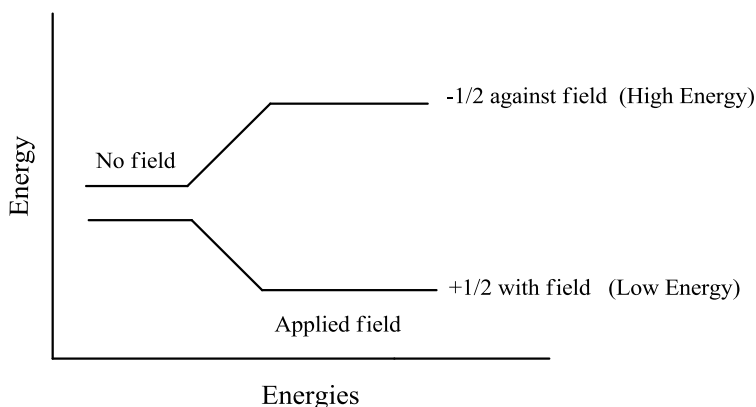
Figure 1.7: The spin state of the proton in the absence and in the presence of magnetic field.

6

The magnitude of the energy difference between the two spin states depends on the applied magnetic field. As shown by the **Figure 1.7**, the greater the strength of the applied magnetic field, the larger the energy difference. Therefore, the magnitude of the energy level separation is directly proportional to the applied magnetic field B_0 and to the particular nucleus involved, according to its γ. The energy difference between the two states is equal to $\Delta E = \gamma \ (h/2\pi) \ B_0$ [25].

The difference in spin population between the α and the β state determines the magnitude of a bulk magnetization vector created by the alignment of nuclear spins along the B_0 magnetic field (Figure 1.8).
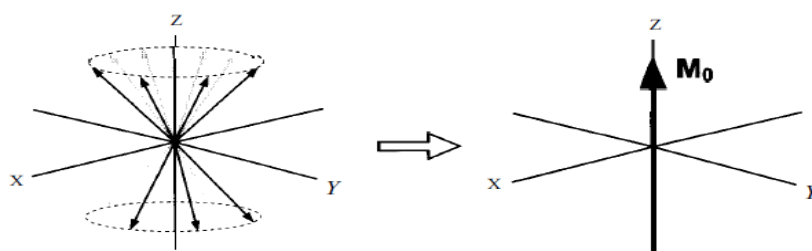
Figure 1.8: Macroscopic magnetization resulting from the sum of individual nuclear magnetic moments.

Application of an orthogonal energy pulse in the form of radio frequency (RF) imposes a torque on the bulk magnetisation vector in a direction that is perpendicular to the direction of the field (the 'motor rule') which rotates the vector from the z-axis toward the x-y plane (Figure 1.9).The bulk vector will precess (rotate) about the central Z axis, defined by the direction of the B_0 magnetic field, at a frequency that depends not only on the strength of the applied magnetic field and that nucleus' specific gyromagnetic ratio, γ, but also on the neighborhood and chemical environment of the nucleus under observation. The magnitude of these bulk magnetization vectors represents the number of nuclei under the influence of the magnetic field. The frequency of precession is called the Larmor frequency (ω_0) and is defined as $\omega_0 = \gamma B_0$.
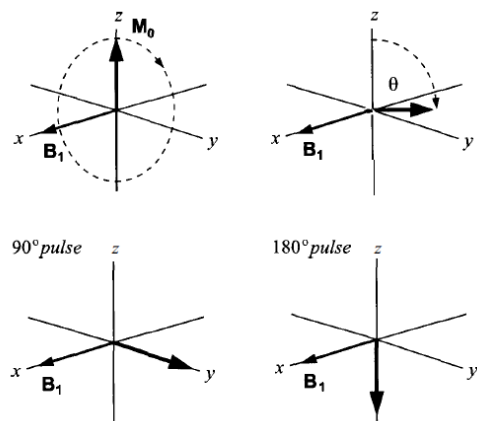
Figure 1.9. An RF pulse applies a torque to the bulk magnetization vector and drives it toward the x-y plane from equilibrium [26].

After the radiofrequency (RF) field is switched off, the system will return to equilibrium. Consequently, the readable magnetization in the xy plane decays to zero in a certain time during which a signal is recorded by the spectrometer in the form of a free induction decay (FID). During an NMR experiment, the signal is measured in the time domain, *i.e.*, as a function of time, and, then, Fourier transformed to obtain the spectrum in the frequency domain. Thereby, the FIDs from active NMR nuclei in a sample, which superimpose in the time domain, are sorted out according to their Larmor frequency by the Fourier transformation [26]. **Figure 1.10** A schematic diagram of the main components of an NMR spectrometer.
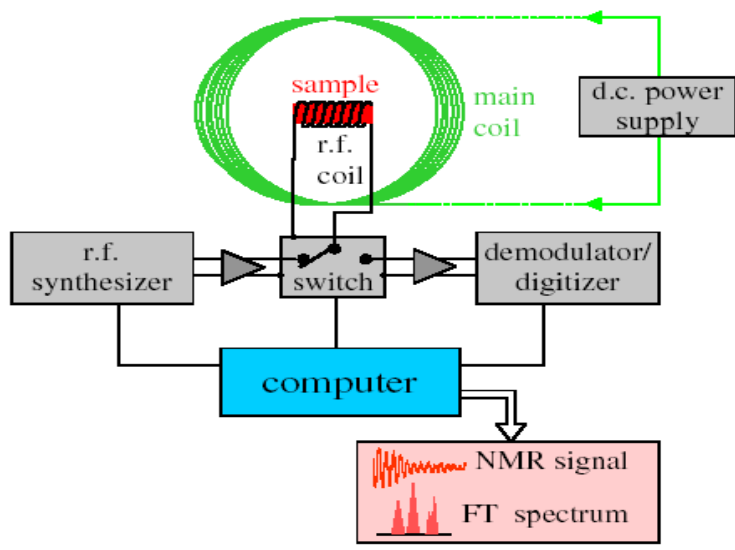
Figure 1.10: A schematic diagram of the main components of an NMR spectrometer.

1.6 NMR-based Metabolomics

Different techniques have their own strengths and weakness, resulting in different coverage of metabolomics. Therefore, suitable choice of instrumentation is important to address specific questions in metabolomic analysis. The selection of proper instrumentation depends on the availability of these analytical platforms and on the possibility to combine different techniques to obtain a comprehensive metabolic profile. NMR is a powerful spectroscopic method, traditionally used as a very important tool in chemistry for structure verification, elucidation and purity analysis [27]. Sensitivity of NMR has always been a concern compared to MS, which has also developed better protocols for separating the components in a complex mixtures. Nonetheless, NMR spectra can be recorded and quantified more easily. The spectral complexity of the MS fragments can complicate the quantification process of ions or metabolites, and MS can identify only a small fraction of the metabolome. NMR is a robust and reliable technique for metabolomic applications, in which high reproducibility is paramount [28-29]. It allows the detection of a wide range of structurally diverse metabolites simultaneously, providing a metabolic 'snapshot' at a particular time point [30]. NMR spectroscopy is therefore a powerful complementary technique for the identification and quantitative analysis of metabolites either *in vivo* or in tissue extracts and biological fluids [28]. The sensitivity issue of NMR has decreased recently with advancement in NMR spectrometer hardware such as used of cryoprobes, cold-probes or by using high field instruments, *i.e.*, 800 MHz. Using high field instruments enhances both the sensitivity and resolution, but cost of the instrumentation is still an issue [31].

In 2013, Wishart and colleagues introduced the Human Metabolome Database (HMDB) using 6 different analytical platforms: NMR; GC-MS; DFI/LC-MS/MS; HPLC/UV; HPLC/FD and ICP-MS. Today, the HMDB is the largest and most comprehensive, organism-specific metabolomics database [32]. It provides information on the composition of human biological fluids, along with custom-derived spectral libraries and targeted assays. In the case of urine, 445 distinct metabolites were identified and 378 quantified distinct metabolites. NMR spectroscopy was able to identify and quantify 209 compounds, as depicted in **Figure 1.11,** which shows that ^{1}H-NMR has the potential to detect and identify a large number of compounds; as such, it is emerging as a leading technique in the area of metabolomic studies [12-13, 21-22].
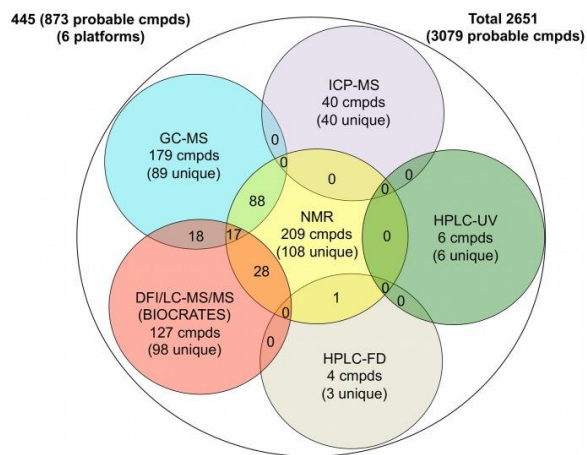
Figure 1.11: Venn diagram showing the broad range of NMR in comparison to other analytical techniques used in urine metabolomics [32].

Compared with MS, NMR spectroscopy yields relatively low-sensitivity measurements, with limits of detection on the order of 10 μM or a few nmol at high fields using new cryoprobes. Nevertheless, NMR has several prominent advantages over MS; for instance, biofluids can be analyzed without or with minimal sample preparation; also, NMR is highly quantitative and reproducible. Moreover, NMR is fast, convenient, and effective in discriminating between groups of related samples and it identifies the most important regions of the spectrum for further analysis [28]. These properties are especially important when metabolomic data are analyzed by multivariate statistical methods [33]. Compared with other techniques, NMR-based metabolomics is becoming a useful tool in the study of body fluids and has a strong potential to be particularly useful for the noninvasive diagnosis of different diseases [21], different types of cancer [34], diabetes [35], inborn error metabolism [36], kidney disorders [37] and also in early diagnosis of disease [22].

There are two different approaches used in NMR-based metabolomics, depending on the aim and the information required from the experiment. In cases when the interest is limited to certain defined set of metabolites, the approach of choice is usually known as "targeted metabolomics"; in other instances, it may be of interest to examine as many metabolites as possible by using an "untargeted" approach [3].

10

1.6.1 NMR-based Targeted Metabolomics

Targeted metabolomic approaches are usually initiated by a specific biochemical theory or hypothesis that encourages the investigation of a particular pathway [3]. The aim of NMR-based targeted metabolomics is to quantitate several selected metabolites, typically few to hundreds of known compounds. This involves the ability to differentiate the targeted analytes from other intrusive compounds, which may be achieved based on chemical shift in an NMR spectrum [38]. Conventionally, quantitative analysis by NMR has been restricted to relatively simple mixtures with minimal peak overlap. In this context, 1D ^{1}H-NMR is a natural choice, because its peaks scale linearly with concentration and its analytical precision is usually independent of the chemical properties of target molecules [20]. ^{1}H-NMR provides spectroscopic fingerprints in which the spectral intensity distribution is determined by the relative concentrations of solutes, and in some cases by their intermolecular interactions. This approach is generally useful, for example, for pharmacokinetic studies of drug metabolism as well as for the study of the activity of specific enzymes under the influence of therapeutics or subject of genetic modification [40]. The same experimental and statistical standards apply to a targeted metabolomics approach as to any quantitative biological analysis. In order to obtain meaningful results, multiple biological replicates must be analyzed and subjected to appropriate statistical tests to establish the significant changes in metabolite concentrations between groups [41].

1.6.2 NMR-based Untargeted Metabolomics

The untargeted approach in metabolomics is the comprehensive analysis of all the measurable metabolite in a biological sample, including unknown chemicals. Due to its comprehensive nature, untargeted metabolomics must be coupled to advanced chemometric techniques, such as multivariate analysis, to reduce the large datasets generated into a smaller set of manageable variables [41]. The nature of many compounds of interest in untargeted metabolomics is unknown, hence several solvents and extraction methods should be applied and compared between the groups of samples [42]. The aim of the untargeted metabolomics is to measure the many metabolites as possible from biological sample in unbiased fashion. This unbiased approach make it global in scope and sometime is also known as global metabolite profiling [3]. Chapters 2 and 3 of this thesis are examples of untargeted metabolomic approaches; thus, more detail will be discussed later in chapters 2 and 3.

1.6.3 NMR-based Metabolomics in Food Chemistry

Today, the need to ascertain and certify the quality of food is increasing because of increasing demand from industry and consumers. To improve and monitor the quality of our food, new tools and methods need to be developed. Metabolomics is a new technology with the potential to become one such tool of major value in this field. Phytochemicals cover a large proportion of the food metabolome. Some of the popular phytochemicals, such as lycopene in tomatoes, isoflavones in soy, proanthocyanidins in cranberries and polyphenols in fruits or in nutraceutical products, are responsible not only for the organoleptic properties, such as flavors and aromas of the plants, but also for their health properties [43-45]. Specifically, phytochemicals are rarely absorbed and excreted in their ingested forms; rather, they are extensively metabolized in the body. Therefore, only a limited portion of the theoretical 200,000 structures of phytochemicals have been characterized, suggesting that many remain to be discovered [44]. Metabolomic tools have become the method of choice and are being applied to the analysis of food components, the identification of their metabolites in body fluids and biological tissues, the evaluation of their bioavailability and metabolism, the role of gut microflora, and the physiological response to a particular diet regimen, food, or nutraceutical product [46-47]. The metabolic compositions of various biofluids reflect their diverse biological purposes and the functional integrity of the organs that are communicating with them, and eventually with the physiological status of the whole organism. NMR-based metabolomics has been employed extensively for multivariate metabolic profiling of cells, tissues and biological fluids [48-49]. Biofluids such as urine are very complex matrices containing a large number of potential biomarkers that can report on changes of endogenous metabolites in response to xenobiotic exposure [49-50]. Urine is a very popular biofluid for metabolomic investigations due to non-invasive collection, the complex metabolic nature of the fluid, and the ability to collect multiple samples over a period of time. The detection of metabolites and the information gained by tracing metabolic fluxes provides information regarding an organism's physiological response following an environmental insult or the pathophysiology of a disease [51-52].

Based on the metabolomics approach, ^{1}H-NMR screening has rapidly expanded in recent years in the area of food quality control *i.e.* food processing, quality, and safety. Advances in

compound extraction, separation, detection, identification, and data analysis enabled metabolomic reliable platform in the field of food sciences [45, 94]. NMR based metabolomic techniques assess the impact of food quality control in a reproducible manner for example adulteration of oranges juices with grape-fruit make it difficult for consumer to assess even via taste or aroma whether the juices are adulterated or not. Analysis of this kind of juice blending with HPLC-MS analysis is lengthy and labor process whereas, NMR-based metabolomic take less than half time as compare to chromatographic approach [53].

The quality or origin of food matrices is mostly determined by its biochemical composition, *i.e.*, its biochemical, metabolic profile [55]. Recently, NMR-based metabolomics coupled with chemometric analysis has been applied to obtain metabolic profiles of various kinds of food including honey [56-57], olive oil [58-59], meat [55, 60] salmon [61], milk [62], wine [63, 64], apple [65], mango [66], Italian sweet cherries [67], wheat [68], and other plants [69-70]. Despite the considerable amount of reviews and articles published in the field of food metabolomic, applications of metabolomics in food and nutrition research are still scarce. The main contribution of this thesis is to show the application of NMR-based metabolomics in food chemistry. The next section of the thesis will briefly introduce the role and application of chemometrics in metabolomics.

1.7 Chemometrics in Metabolomics

Chemometrics was first introduced by Svante Wold in 1972. Further progress was made by Isenhour and Jurs when they published the first pattern recognition article in 1972. However, first real article on chemometrics was published by Kowalski, Massart and Wold [71]. Chemometrics is rather new, but it has a huge impact on the spectroscopic field as we know it today, to such an extent that chemometric software is currently integrated with laboratory and process instrumentation as a standard. Chemometrics is still developing and the definition may have to be modified or improvised from time to time in order to include all new developments [72]. In the 1980s, John demonstrated the use of chemometrics (multivariate data analysis) in combination with NMR spectroscopy. Later on, further advancement and potential was put forward by Nicholson and his group in the field of NMR-based metabolomics [5, 74].

The overwhelming size and complexity of the 'omics' technologies has driven biology toward the adoption of chemometric methods. Extensive use of chemometrics in NMR-based

metabolomics first became computationally practical with the introduction of data-reduction methods such as the integration of spectral variables (preprocessing of data) into a limited number of regions, a technique that both reduces the computational time required to calculate models, and helps soften the effect of shifting resonances in NMR [74]. Therefore, it is necessary to perform some data reduction or preprocessing of any large data prior to transport the data for Multivariate data analysis (MDA). A typical workflow of chemometric analysis of NMR-based metabolomic data is shown in **Figure 1.12.**
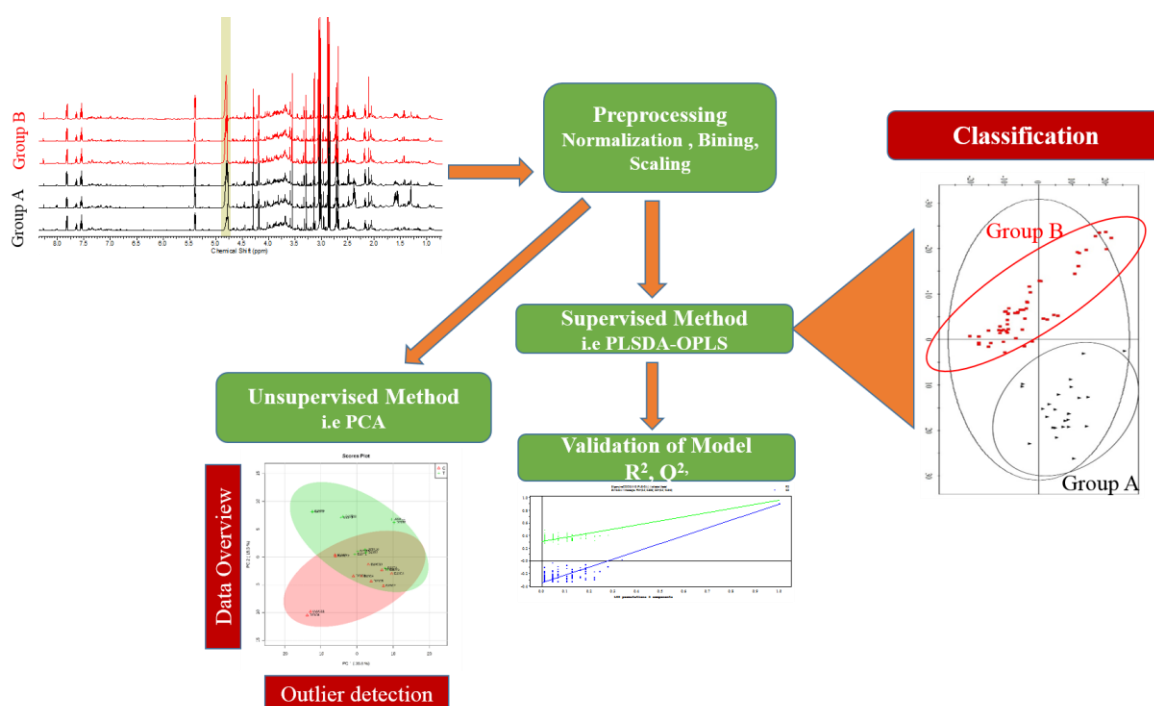
Figure 1.12: A typical chemometric workflow in an NMR-based metabolomic approach.

1.7.1 Data Preprocessing

Pre-treatment of raw spectral data is critical for generating reliable, interpretable models using multivariate analysis techniques. Data preprocessing is an intermediate step between raw spectra and data analysis. The main objective of data preprocessing is to transform the data in such a way that the samples in the dataset are more comparable in order to ease and improve data analysis. For NMR spectra, preprocessing usually involves: baseline correction, alignment, binning, normalization and scaling. Several reviews discuss the procedures involved for

preprocessing of metabolomics datasets, and efforts have been made to standardize the processes [75].

1.7.1.1 Baseline Correction

Generally, the first step of data preprocessing is the baseline removal. Baseline distortions in 1D NMR spectra are mostly due to the corruption of the first few data points in the FID (free induction decay). These corrupted data points add low frequency modulations in the Fourier-transformed spectrum, giving rise to a distorted baseline [76]. Correction of these distortions is a necessary step in NMR spectra data processing because they offset the intensity values and result in inaccuracy in peak assignment and quantification. In turn, this affect not only the statistical analysis but also the quantification of the metabolites. These distortions can be corrected in many different ways; usually, an automated baseline correction is applied. In this thesis, all the NMR spectra were processed with manual baseline correction, performed with the ACD/LAB NMR software. In urine metabolomics, the entire region between 0.2-10.0 ppm is frequently used to see the effects of endogenous metabolites belonging to different classes. Although solvent suppression techniques are used to reduce the signal of water, the residual water signal might interfere with data analysis. For this reason, the region of the spectrum between 4.7 ppm and 5.0 ppm is excluded from further data processing.

1.7.1.2 Alignment

Chemical and physical parameters such as ionic strength, concentration of some earth alkali, pH, and overall dilution of the urine influence the chemical shift of the ^{1}H-NMR spectra, although not all the peaks are affected to the same extent. The resulting chemical shift variations, as large as 0.1 ppm, generate misalignments of homogeneous peaks, artifacts and misinterpretations [77]. One of the most annoying problems with NMR profiles, from a multivariate data analysis point of view, is the presence of peak shifts among different spectra. Therefore, an essential step in preprocessing should be careful adjustment of the peaks shifts, *i.e.*, alignment. Usually, NMR spectra are first aligned by spectral referencing. This simple, global method for peak alignment sets the internal reference signal of each spectrum to 0.0 for TSP (as an internal standard) or 7.27 ppm (CD_3Cl) for the residual non-deuterated solvent. Nonetheless, this global method is not sufficient to overcome the problem of alignment [78]. Various methods have been reported so far [79] for the alignment of NMR spectra. In chapter 2 of this thesis, we used the Cluster-based

Peak Alignment (CluPA), a recent method for aligning NMR spectra, proposed by Vu et al [80]. The algorithm builds a hierarchical cluster tree from peaks of reference and target spectra and it aligns two spectra using this tree [80]. Although algorithms used for peak alignment might solve one problem, they can alter peak areas. Therefore, it is necessary to analyze unaligned spectra for the absolute biomarker quantification [81]. A representative example of an aligned ^{1}H-NMR spectrum is given in **Figure 1.13**.
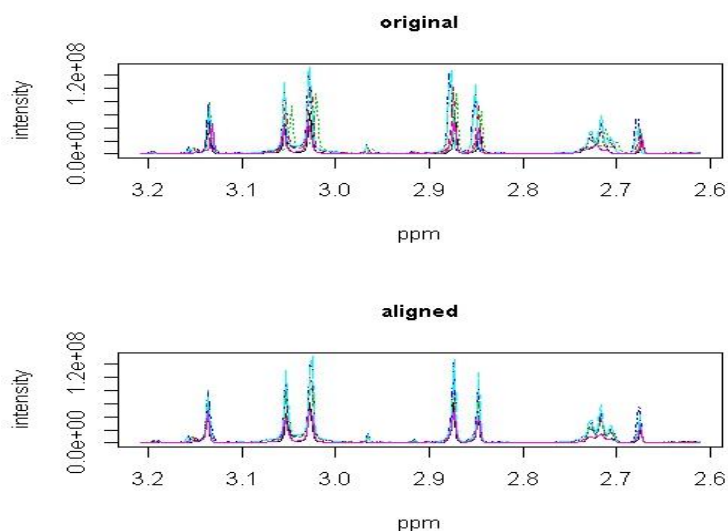
Figure 1.13: Example of peak position variation: unaligned ^{1}H-NMR citric acid peaks (Upper panel) and profiles processed by CluPA (lower panel).

1.7.1.3 Binning

In order to reduce the data dimensionality, binning (also called bucketing) is commonly used. The binning procedure helps in filtering noisy spectra, masks elusive chemical shift differences, and it also hides potentially significant changes of low-intensity peaks nearby strong signals. Among the many types of spectral binning available, the most commonly used one is an equidistant binning of 0.04 ppm: every spectrum is divided into evenly spaced integral regions with spectral width of 0.04 ppm [83]. This equal size binning suffers from lack of flexibility of the boundaries. For instance, if a peak is split between two bins, this may significantly influence the data analysis. In order to avoid peaks being split by the boundaries of bins, methods which

are based on non-equidistant spacing have been proposed, *e.g.*, adaptive-intelligent binning, Gaussian binning, adaptive binning using wavelet transform and dynamic adaptive binning [81].

1.7.1.4 Scaling and Normalization

Another important preprocessing step for NMR spectra of metabolomic studies is scaling and normalization. However, it should be noted that normalization and scaling operations offer different purposes and in fact, in usual practice it is also possible to use a combination of normalization and scaling methods [83]. The normalization steps are performed to account for variations of the overall concentrations of samples caused by different dilutions. Particularly for urine, unintended variations of the overall concentrations of samples are very distinctive. The dilution of human and animal urine can arise from food deprivation or drug effects and can exceed a factor of 10 [85]. Being the biological collection of waste material, urines reflect large variability in metabolite concentrations within control subjects. Therefore, data normalization and scaling between variables is typically applied to remove unwanted systematic variations meanwhile retaining the interesting biological information. Other inter-sample variations, such as different relaxation times or variations in RF pulse calibration, can be corrected by normalization. Normalization also involves measurement of urinary volumes and animal weights as a study proceeds. [82, 85]. A number of methods for normalizing metabolite profiles has been reported and every method is based on certain assumptions as to the nature of the data [83]. Some of the most common normalizations methods are: (1) Integral normalization; the simplest form of normalization, also called constant-sum normalization, whereby each spectrum is normalized such that its global integral is 1. While this accounts for variable dilutions each sample may possess, it can cover truly biologically relevant changes and obscure interpretation of loadings [75]; (2) probabilistic quotient/median fold-change [84]; (3) a reference feature present in all samples at a constant level. Normalization may be accomplished internally by computational means using internal standards (*e.g.*, TSP) [86] (4) The use of Creatinine peak for normalization is also a common practice in urine metabolomics under the assumption that it scales well with muscle mass [95].

The variation in metabolite levels is often linked to concentration in such a way that higher concentration metabolites have higher variation. This variation has a strong impact on the multivariate statistical analysis. Therefore, it is essential to scale metabolite intensities before

analysis to prevent selection of the most abundant metabolites as significant contributors. A number of scaling methods are commonly used, namely, mean centering, autoscaling, Pareto scaling, range scaling [81].

Mean centering is used to adjust the differences between high- and low-concentration metabolites by converting all values to vary around zero instead of around the mean of the metabolite level. Mean centering is commonly used in combination with other scaling methods. In case of autoscaling, all metabolites are subjected to unit variance and therefore the data is analyzed based on the correlations instead of covariances. Autoscaling is sensitive to noisy data, and noisy variables tend to be overvalued. Pareto scaling is widely used in metabolomics since it is an intermediate option. It uses the square root of the standard deviation as the scaling factor instead of the standard deviation. This scaling method is sensitive to large changes in the data but keeps closer to the real data. Range scaling uses the difference between the minimal and the maximal concentration of each metabolite which makes all metabolites in the data equally important. However, it is important to keep in mind that range scaling is sensitive to outliers because only two values are used to calculate the range. [75, 81-82]

1.7.2 Multivariate Statistical Analysis

The first step in Multivariate statistical analysis (MVA) is to perform an unsupervised analysis, which occurs without prior knowledge of neither the nature nor the group membership of the samples. This is often performed using principal component analysis (PCA) [87]. Data have to be preprocessed or "cleaned" before this exploratory treatment.

1.7.2.1 Principal Component Analysis:

PCA is the most common statistical analysis tool in metabolomics [73], invented by Pearson in 1901 [88] and later on rediscovered by Wold [89]. Hotelling further developed PCA to its present shape [87]. PCA is a multivariate analysis based on projection methods used for dimension reduction of data. It is an unsupervised technique and is used in chemometrics as workhorse to extract and display the systematic variations in the data. A PCA model provides a summary, or overview, of all observations or the samples in the data table. PCA converts the multidimensional data space into a low dimensional model plane that approximates all rows (e.g., objects or observations) in X, that is, the group of points. This technique expresses most of

18

the variance within a data set using a smaller number of factors, called Principal Components (PCs) (tp^T) as depicted in **Figure 1.14**. The first PCA model component ($t_1p_1^T$) describes the largest variation in the swarm of points. The second component models the second largest variation and so on. Each PC is a linear combination of the original variables; thus, each successive PC explains the maximum amount of variance, which was not accounted for the previous PCs. Each PC is orthogonal to the other PCs and therefore exhibits different information. The variation in spectral data is described by a few PCs, compared to the number of original variables. Moreover, it enables to find trends, groupings, and extend outliers in the data [73, 81, 89]. Mathematically, PCA can be written as follows:

$$X = TP^T + E = t_1p_1^T + t_2p_2^T + \ldots\ldots\ldots E$$

Where X is the data matrix representing samples and variables decomposed into a score matrix (T) and a transposed loading matrix (P^T). The E matrix represents the residuals, the part of the data not 'explained' by the PC model [89]. There are two important matrices obtained from the PCA model, known as score and loading plot (Figure 1.14).
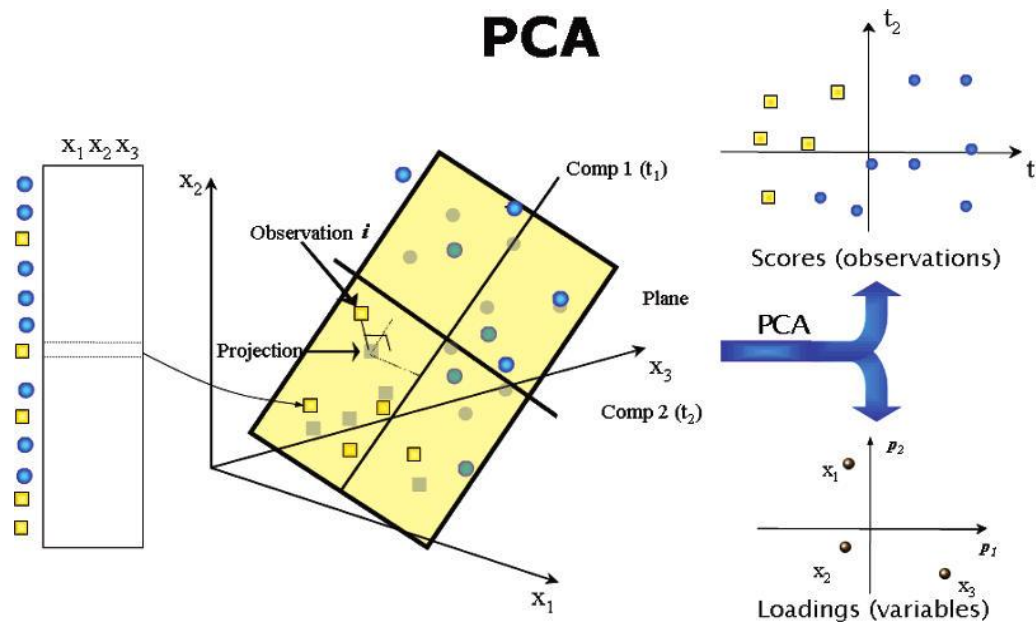
Figure 1.14: A representation of a Principal Component Analysis Model showing the reduction of data into score and loading plot [73].

In a score plot, each point represents a single spectrum. It provides a summary of all spectra and shows how they are related to each other. If the observations are close to each other, it means

they are similar whereas, observation far from each other are dissimilar. The loadings describe the way in which the old variables are linearly combined to generate new variables (PCs) and indicate which variables have the greatest contribution in transforming to the new variables. Thus, any spectral clustering observed in the score plot is interpreted by examination of the loadings. This is a powerful tool for understanding the underlying patterns in the data [73, 81].

1.7.2.2 Partial Least Square (PLS) Analysis

Unlike PCA, PLS is a supervised method, commonly used in metabolomics. Metabonomic studies typically include sets of controls and treated samples. In these situations, a more focused evaluation and analysis of the data is possible. The basic principle is similar to PCA, but in PLS, a second piece of information is used, namely, the labeled set of class identities. In addition to an X-matrix of observations (*e.g.*, spectra/samples) and variables (bins) that is used in PCA, a Y-matrix is created that consists of the same observations but the variables are classes, *e.g.*, treated or controls [90]. This enables the establishment of a linear model that can predict Y from the measured spectra in X. Observations belonging to the class have a value of one and observations not belonging to the class have a value of zero. Like PCA, PLS regression generates a linear model of the data, but where PCA models the principal variations in the data itself, PLS derives a model that describes the correlation between the X variables and a feature (Y variable) of interest [73].

PLS can also be used in discriminant analysis, that is, PLS-DA. The Y matrix then contains qualitative values, for example, class belonging, gender, and treatment of the samples. OPLS can, analogously to PLS-DA, be used for discrimination (OPLS-DA) [91]. In **Figure 1.15**, the advantages with OPLS-DA compared to PLS-DA are described. The between-class variation and the within-class variation are separated by OPLS-DA but not by PLS-DA, and this facilitates the OPLS-DA model interpretation [73].
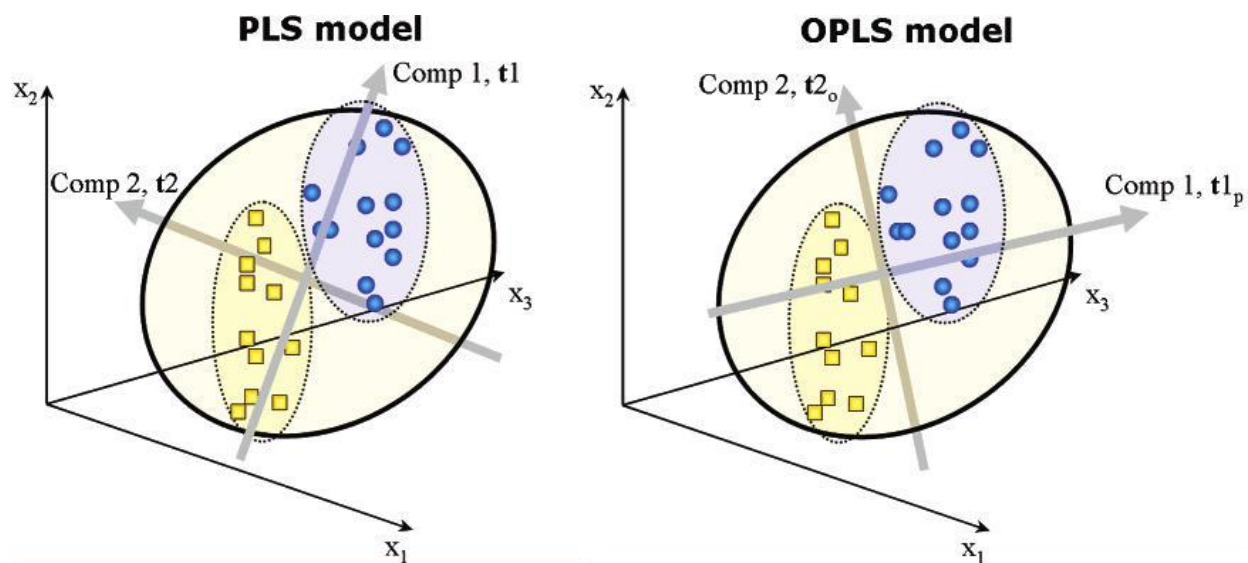
Figure 1.15: A geometrical illustration of the difference between the PLS-DA and OPLS-DA models. In the left panel, the PLS components cannot separate the between-class variation from the within-class variation [73].

1.8 References

[1]. Fiehn, O. Metabolomics – the link between genotypes and phenotypes. *Plant Mol. Bio.* **2002**, 48, 155–171.

[2]. Oliver, S. G., Winson, M. K., Kell, D. B., and Baganz, F. Systematic functional analysis of the yeast genome. *Trends Biotechnol.* **1998**, 16, 373-378.

[3]. Patti, G. J., Yanes, O. and Siuzdak, G. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat. Rev. Mol .Cell Biol.* **2012**, 13, 263-269.

[4]. Fiehn, O. Combining genomics, metabolome analysis, and biochemical modelling to understand metabolic networks. *Comp. Funct. Genomics,* **2001**, 2, 155–168.

[5]. Nicholson, J. K., Lindon, J. C., Holmes, E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR data. *Xenobiotica* **1999,** 29, 1181-1189.

[6]. Nicholson, J. K., Lindon J.C., System biology: Metabonomics. *Nature*, **2008**, 455, (7216), 1054-1056.

[7]. Wang, J. H., Byun, J. and Pennathur, S. Analytical approaches to metabolomics and applications to systems biology. *Semin. Nephrol.* **2010**, 30 (5), 500-511.

[8]. Dettmer, K, Aronov, P. A., and Hammock, B. D. Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.* **2007**, 26 (1), 51-78.

[9]. Oldiges, M., Lütz, S., Pflug, S., Schroer, K., Stein, N., and Wiendahl C. Metabolomics: current state and evolving methodologies and tools. *Appl. Microbiol. Biotechnol.* **2007**, 76 (3), 495 511.

[10]. Tweeddale, H., Notley-McRobb, L., Ferenci, T. Effect of slow growth on metabolism of Escherichia coli, as revealed by global metabolite pool ("metabolome") analysis. *J. Bacteriol.* **1998**, 180, 5109–5116.

[11]. Wang, Y., Hu, H., Su, Y., Zhang, F., and Guo Y. Potential of monitoring isotopologues by quantitative gas chromatography with time-of-flight mass spectrometry for metabolomic assay. *J Sep Sci*. **2016** (accepted online).

[12]. Solanky, K. S., Bailey, N. J., Holmes, E., Lindon, J. C., Davis, A. L., Mulder, T. P., Van Duynhoven, J. P., and Nicholson, J. K. NMR-Based Metabonomic Studies on the Biochemical Effects of Epicatechin in the Rat*J. Agric. Food Chem.* **2003,** 51, 4139-4145.

[13]. Daykin, C. A., Duynhoven, J. P., Groenewegen, A., Dachtler, M., Amelsvoort, J. M. M. V., and Mulder, T. P. J. Nuclear Magnetic Resonance Spectroscopic Based Studies of the Metabolism of Black Tea Polyphenols in Humans *J Agric Food Chem.* **2005,** 53, 1428-1434

[14]. Groussac, E., Ortiz, M., and Francois, J. Improved protocols for quantitative determination of metabolites from biological samples using high performance ionic-exchange chromatography with conductimetric and pulsed amperometric detection. *Enzyme Microb. Technol.* **2000**, 26:715–723.

[15]. Castrillo J. I., Hayes, A., Mohammed, S., Gaskell, S. J., and Oliver, S. G. An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry. *Phytochemistry* **2003**, 62, 929–937.

[16]. Kopka, J. Current challenges and developments in GC-MS based metabolite profiling technology. *J Biotechnol.* **2006**, 124, 312–322.

[17]. Roessner, U., Wagner, C., Kopka, J., Trethewey, R. N., and Willmitzer L. Simultaneous analysis of metabolites in potato tuber by gas chromatography-mass spectrometry. *Plant J.* **2000,** 23, 131–142.

[18]. Ramautar, R., Somsen, G. W., de Jong, and G. J. CE-MS for metabolomics: developments and applications in the period 2012-2014. *Electrophoresis*. **2015**, 36 (1), 212-24.

[19]. Ma, Y., Tanaka, N., Vaniya, A., Kind, T., and Fiehn, O. Ultrafast polyphenol mtabolomics of red wines using MicroLC-MS/MS. *J. Agric. Food Chem.* **2016**, (accepted online)

[20]. Gu, H., Zhang, P., Zhu, J., and Raftery, D. Globally optimized targeted mass spectrometry: reliable metabolomics analysis with broad coverage. *Anal. Chem.* **2015**, 15, 87(24), 12355-12362.

[21]. Abdul-Hamid M. Emwas, AH. M., Salek R. M., Griffin, L. J., and Merzaban J. NMR-based metabolomics in human disease diagnosis: applications, limitations, and recommendations. *Metabolomics* **2013**, 9, 1048–1072.

[22]. Duarte, F. I., Diaz, O. S., Gil, M. A. NMR metabolomics of human blood and urine in disease research. *J. Pharma. Biomed. Ana.* **2014**, 93, 17–26.

[23]. Alonso, A., Marsal, S., and Antonio Julià, A. Analytical methods in untargeted metabolomics: state of the art in 2015. *Front Bioeng. Biotechnol.* **2015**, 3, 23.

[24]. Abragam, A. The Principles of Nuclear Magnetism; Oxford University Press, London, **1961**

[25]. Pavia, D. L., Lampman, G. M., Kriz, G. S., and Vyvyan R. J. Introduction to Spectroscopy. Brooks/Cole Cengage Learning. 4th Edition, **2009**, 3, 107-109.

[26]. Claridge, D. W. T. High-Resolution NMR techniques in organic chemistry. Dyson Perrins Labortory, Oxford, Pergamon, Elsevier, 1999, 2, 16-21.

[27]. Reo, N. V. NMR-Based metabolomics. *Drug Chem. Toxicol.* **2002**, 25, 375-382.

[28]. Krishnan, P., Kruger, N. J., and Ratcliffe, R. G. Metabolite fingerprinting and profiling in plants using NMR. *J. Exp. Bot.* **2005,** 56, (410), 255-265.

[29]. Liu, M., Nicholson, J. K., and Lindon, J. C. High resolution diffusion and relaxation edited one- and two-dimensional ^{1}H NMR spectroscopy of biological fluids. *Anal. Chem.* **1996**, 68, 3370–3376.

[30]. Dumas, M. E. Assessment of analytical reproducibility of ^{1}H-NMR spectroscopy based metabonomics for large-scale epidemiological research: the INTERMAP study. *Anal. Chem.* **2006**, 78, 2199–2208.

[31]. Lenz, E. M., and Wilson, I. D. Analytical strategies in metabonomics. *J. Proteome Res.* **2007**, 6, 443-458.

[32]. Bouatra, S., Aziat, F., Mandal, R., Guo, A.C., Wilson, M.R., Knox, C., Bjorndahl, T.C., Krishnamurthy, R., Saleem, F., Liu, P. and Dame, Z.T. The human urine metabolome. *PloS one*, **2013**, 8(9), p.e73076.

[33]. Nicholson, J. K., and Wilson, I. D. High resolution proton magnetic resonance spectroscopy of biological fluids. *Prog. Nucl. Magn. Reson. Spectrosc.* **1989**, 21, 449–501.

[34]. Spratlin J. L., Serkova, N. J., Eckhardt, S. G. Clinical applications of metabolomics in oncology: a review. *Clin. Cancer Res.* **2009,** 15(2), 431-440.

[35]. Bain, J. R., Stevens, R. D., Wenner, B. R., Ilkayeva, O., Muoio, D. M., and Newgard, C. B. Metabolomics applied to diabetes research: moving from information to knowledge. *Diabetes.* **2009**, 58 (11), 2429-2443.

24

[36]. Constantinou, M. A. et al. ^{1}H-NMR-based metabonomics for the diagnosis of inborn errors of metabolism in urine. *Analytica. Chimica Acta.* **2005**, 542, 169–177.

[37]. Ganti, S., Weiss, R. H. Urine metabolomics for kidney cancer detection and biomarker discovery. *Urol Oncol.* **2011**, 29 (5), 551-557.

[38]. Lu, W., Bennett, B. D., and Rabinowitz, J. D. Analytical strategies for LC–MS-based targeted metabolomics. *J. Chromato. B* **2008**, 871, 236–242.

[39]. Gu, H., Zhang, P., Zhu, J., and Raftery, D., Globally optimized targeted mass spectrometry: reliable metabolomics analysis with broad coverage. *Anal. Chem.* **2015**, 87, 12355−12362.

[40]. Nicholson, J. K. Connelly, J., Lindon, J. C., and Holmes E. Metabonomics: a platform for studying drug toxicity and gene function. *Nat Rev Drug Discov.* **2002**, 1(2):153-161.

[41]. Weljie, A. M., Newton, J., Mercier, P., Carlson, E., and Slupsky, C. M. Targeted profiling: quantitative analysis of 1H NMR metabolomics data. *Ana. Chem.* **2006**, 78(13), 4430-4442.

[42]. Alonso, A., Marsal, S., and Julià, A. Analytical methods in untargeted metabolomics: state of the art in 2015. *Frontiers bioeng. Biotechn.* **2015**, 3.

[43]. Barnes, S., Prasain, J., and Kim, H. In Nutrition, Can We "See" What Is Good for Us? *Advances in Nutrition*: *An International Review Journal.* **2013**, 4(3), 327S-334S.

[44]. Manach, C., Hubert, J., Llorach, R., and Scalbert, A. The complex links between dietary phytochemicals and human health deciphered by metabolomics. *Mol. Nutr. Food Res*. **2009**, 53(10), 1303-1315.

[45]. Cevallos-Cevallos, J. M., Reyes-De-Corcuera, J. I., Etxeberria, E., Danyluk, M. D. and Rodrick, G. E. Metabolomic analysis in food science: a review. *Trends Food Sci. Techn*. **2009**, 20(11), 557-566.

[46]. Astarita, G., and Langridge J. An emerging role for metabolomics in nutrition science. J *Nutrigenet Nutrigenomics* **2013**, 6, 181–200.

[47]. Herrero, M., Simo, C., Garcia-Canas, V., Ibanez, E., and Cifuentes, A. Foodomics: MS-based strategies in modern food science and nutrition. *Mass Spectrom Rev* **2012**, 31, 49–69.

[48]. Beckonert, O., Keun, H. C., Ebbels, T. M., Bundy, J., Holmes, E., Lindon, J. C., and Nicholson, J. K. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature protocols.* **2007**, 2(11), 2692-2703.

[49]. Saude, J. E., and Sykes, D. B., Urine stability for metabolomic studies: effects of preparation and storage. *Metabolomics* **2007**, 3, (1) 19-27.

[50]. Zhang, A., Sun, H., Wu, X., and Wang, X. Urine metabolomics. *Clinica Chimica Acta*, **2012**, 414, 65-69.

[51]. Chadha, V., Garg, U., Alon, U. S. Measurement of urinary concentration: a critical appraisal of methodologies. *Pediatr. Nephrol*. **2001** 16, 374–382.

[52]. Lindon, J. C., Holmes, E., Bollard, M. E., Stanley, E. G., and Nicholson, J. K. Metabonomics technologies and their applications in physiological monitoring, drug safety assessment and disease diagnosis. *Biomarkers* **2004,** 9, 1–31.

[53]. Wishart, D.S. Metabolomics: applications to food science and nutrition research. *Trends Food Sci. Tech.* **2008**, 19(9), 482-493.

[54]. Kim, S., Kim, J., Yun, E. J., and Kim, K. H. Food metabolomics: From farm to human. *Curr. Opin. Chem. Biol.* **2016**, 37, 16-23.

[55]. Jung, Y., Lee, J., Kwon, J., Lee, K. S., Ryu, D. H., & Hwang, G. S. Discrimination of the geographical origin of beef by ^{1}H NMR-based metabolomics. *J. Agric. Food Chem*. **2010**, 58, 10458–10466.

[56]. Schievano, E., Peggion, E., and Mammi, S. [1]H nuclear magnetic resonance spectra of chloroform extracts of honey for chemometric determination of its botanical origin. *J. Agric. Food Chem.* **2010**, 58, 57–65.

[57]. Schievano, E., Stocchero, M., Morelato, E., Facchin, C. and Mammi, S. An NMR-based metabolomic approach to identify the botanical origin of honey. *Metabolomics*, **2012**, 8(4), 679-690.

[58]. Fauhl, C., Reniero, F. and Guillou, C., ^{1}H-NMR as a tool for the analysis of mixtures of virgin olive oil with oils of different botanical origin. Magnetic Resonance in Chemistry, *Magn. Reson. Chem.* **2000**, 38, 436–443.

[59]. Longobardi, F., Ventrella, A., Napoli, C., Humpfer, E., Schütz, B., Schäfer, H., Kontominas, M.G. and Sacco, A. Classification of olive oils according to geographical origin by using 1 H NMR fingerprinting combined with multivariate analysis. *Food Chem.* **2012**, 130(1), 177-183.

[60]. Sacco, D., Brescia, M.A., Buccolieri, A. and Jambrenghi, A.C. Geographical origin and breed discrimination of Apulian lamb meat samples by means of analytical and spectroscopic determinations. *Meat Sci.* **2005**, 71(3), 542-548.

[61]. Aursand, M., Standal, I.B., Prael, A., McEvoy, L., Irvine, J. and Axelson, D.E. ^{13}C-NMR pattern recognition techniques for the classification of Atlantic salmon (Salmo salar L.) according to their wild, farmed, and geographical origin. *J. Agric. Food Chem.* **2009**, 57, 3444–3451.

[62]. Sacco, D., Brescia, M.A., Sgaramella, A., Casiello, G., Buccolieri, A., Ogrinc, N. and Sacco, A. Discrimination between Southern Italy and foreign milk samples using spectroscopic and analytical data. *Food Chem.* **2009**, 1559–1563.

[63]. Godelmann, R., Fang, F., Humpfer, E., Schütz, B., Bansbach, M., Schäfer, H. and Spraul, M. Targeted and nontargeted wine analysis by ^{1}H NMR spectroscopy combined with multivariate statistical analysis. Differentiation of important parameters: grape variety, geographical origin, year of vintage. *J. Agric. Food Chem*. **2013,** 61, 5610−5619.

[64]. Son, H.S., Hwang, G.S., Kim, K.M., Ahn, H.J., Park, W.M., Van Den Berg, F., Hong, Y.S. and Lee, C.H. Metabolomic studies on geographical grapes and their wines using ^{1}H NMR analysis coupled with multivariate statistics. *J. Agric. Food Chem.* **2009**, 57, 1481–1490.

[65]. Belton S. P. et al. Application of chemometrics to the ^{1}H NMR spectra of apple juices: discrimination between apple varieties. *Food Chem.* **1998**, 61, 207-213.

[66]. Koda, M., Furihata, K., Wei, F., Miyakawa, T. and Tanokura, M. Metabolic discrimination of mango juice from various cultivars by band-selective NMR spectroscopy. *J. Agric. Food Chem.* **2012**, 60, 1158−1166.

[67]. Longobardi, F., Ventrella, A., Bianco, A., Catucci, L., Cafagna, I., Gallo, V., Mastrorilli, P. and Agostiano, A. Non-targeted ^{1}H-NMR fingerprinting and multivariate statistical analyses for the characterisation of the geographical origin of Italian sweet cherries. *Food Chem.* **2013**, 141(3), 3028-3033.

[68]. Lamanna, R., Cattivelli, L., Miglietta, M.L. and Troccoli, A. Geographical origin of durum wheat studied by ^{1}H-NMR profiling. *Magn. Reson. Chem.* **2011**, 49, 1-5.

[69]. Choi, H.K., Choi, Y.H., Verberne, M., Lefeber, A.W., Erkelens, C. and Verpoorte, R. Metabolic fingerprinting of wild type and transgenic tobacco plants by ^{1}H-NMR and multivariate analysis technique. *Phytochemistry*, **2004**, 65(7), 857-864.

[70]. Kang, J., Lee, S., Kang, S., Kwon, H.N., Park, J.H., Kwon, S.W. and Park, S., NMR-based metabolomics approach for the differentiation of ginseng (Panax ginseng) roots from different origins. *Arch. Pharm. Res.* **2008**, 31(3), pp.330-336.

[71]. Geladi, P. and Esbensen, K. The start and early history of chemometrics: Selected interviews. Part 1. *J. Chemometr.* **1990**, 4(5), 337-354.

[72]. Geladi, P. Chemometrics in spectroscopy. Part 1. Classical chemometrics. *Spectrochimica Acta Part B* **2003**, 58, 767−782.

[73]. Trygg, J., Holmes, E. and Lundstedt, T. Chemometrics in metabonomics. *J. Proteo. Res.* **2007**, 6, 469-479.

[74]. Spraul, M., Neidig, P., Klauck, U., Kessler, P., Holmes, E., Nicholson, J.K., Sweatman, B.C., Salman, S.R., Farrant, R.D., Rahr, E. and Beddell, C.R. Automatic reduction of NMR

spectroscopic data for statistical and pattern recognition classification of samples. *J. pharma. Biomed. Ana.* **1994**, 12(10), 1215-1225.

[75]. Worley, B. and Robert Powers, R. Multivariate Analysis in Metabolomics. *Curr. Metabolomics*, **2013,** 1, 92-107.

[76]. Xi, Y. and Rocke, D.M., Baseline correction for NMR spectroscopic metabolomics data analysis. *BMC bioinformatics*, **2008**, 9 (1), 324.

[77]. Beneduci, A., Chidichimo, G., Dardo, G. and Pontoni, G. Highly routinely reproducible alignment of ^{1}H-NMR spectral peaks of metabolites in huge sets of urines. *Anal. Chim. Acta.* **2011**, 685(2), 186-195.

[78]. Giskeødegård, G.F., Bloemberg, T.G., Postma, G., Sitter, B., Tessem, M.B., Gribbestad, I.S., Bathen, T.F. and Buydens, L.M. Alignment of high resolution magic angle spinning magnetic resonance spectra using warping methods. *Anal. Chim. Acta.* **2010**, 683(1), 1-11.

[79]. Vu, T.N. and Laukens, K. Getting your peaks in line: a review of alignment methods for NMR spectral data. *Metabolites*, **2013**, 3(2), 259-276.

[80]. Vu, T.N., Valkenborg, D., Smets, K., Verwaest, K.A., Dommisse, R., Lemière, F., Verschoren, A., Goethals, B. and Laukens, K., An integrated workflow for robust alignment and simplified quantitative analysis of NMR spectrometry data. *BMC bioinformatics*, **2011**, 12(1), 405.

[81]. Smolinska, A., Blanchet, L., Buydens, L.M. and Wijmenga, S.S. NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery: a review. *Anal. Chim. Acta*. **2012**, 750, 82-97.

[82]. Craig, A., Cloarec, O., Holmes, E., Nicholson, J.K. and Lindon, J.C. Scaling and normalization effects in NMR spectroscopic metabonomic data sets. *Ana. Chem.* **2006**, 78(7), 2262-2267.

[83]. Kohl, S.M., Klein, M.S., Hochrein, J., Oefner, P.J., Spang, R. and Gronwald, W. State-of-the art data normalization methods improve NMR-based metabolomic analysis. *Metabolomics*, **2012**, 8(1), 146-160.

[84]. Dieterle, F., Ross, A., Schlotterbeck, G. and Senn, H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in ^{1}H NMR metabonomics. *Ana. Chem.* **2006**, 78(13), 4281-4290.

[85]. Savorani, F., Rasmussen, M.A., Mikkelsen, M.S. and Engelsen, S.B. A primer to nutritional metabolomics by NMR spectroscopy and chemometrics. *Food Res. Int.* **2013**, 54(1), 1131-1145.

[86]. Sysi-Aho, M., Katajamaa, M., Yetukuri, L. and Orešič, M. Normalization method for metabolomics data using optimal selection of multiple internal standards. *BMC bioinformatics*, **2007**, 8(1), 93.

[87]. Roussel, S., Preys, S., Chauchard, F. and Lallemand, E.J. Multivariate Data Analysis (Chemometrics). *In Process Analytical Technology for the Food Industry*. Springer New York. **2014**, 7-59.

[88]. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *Edinb.Dubl.Phil.Mag J. Sci.* **1901**, 2(11), 559-572.

[89]. Wold, S., Esbensen, K. and Geladi, P. Principal component analysis. *Chemometr. Intell. Lab.* **1987**, 2(1), 37-52.

[90]. Wold, S., Sjöström, M., and Eriksson, L. PLS-regression: a basic tool of chemometrics. *Chemometr. Intell. Lab.* **2001**, 58, 109-130.

[91]. Bylesjö, M., Rantalainen, M., Cloarec, O., Nicholson, J.K., Holmes, E. and Trygg, J. OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. *J. Chemometrics*. **2006**, 20 (8-10), 341-351.

[92]. Ellis, D.I. and Goodacre, R. Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and Raman spectroscopy. *Analyst.* **2006**, 131(8), 875-885.

[93]. Seli, E., Sakkas, D., Scott, R., Kwok, S.C., Rosendahl, S.M. and Burns, D.H. Noninvasive metabolomic profiling of embryo culture media using Raman and near-infrared spectroscopy correlates with reproductive potential of embryos in women undergoing in vitro fertilization. *Fertil. Steril.* **2007**, 88(5), 1350-1357.

[94]. Cevallos-Cevallos, J.M. and Reyes-De-Corcuera, J.I. Metabolomics in food science. Adv. *Food Nutr. Res.* **2012**, 67, 1-24.

[95]. Holmes, E., Foxall, P.J.D., Nicholson, J.K., Neild, G.H., Brown, S.M., Beddell, C.R., Sweatman, B.C., Rahr, E., Lindon, J.C., Spraul, M. and Neidig, P. Automatic data reduction and pattern recognition methods for analysis of ^{1}H nuclear magnetic resonance spectra of human urine from normal and pathological states. *Anal. Biochem.* **1994**, 220, 284-296.

Chapter 2. New findings on the *in vivo* antioxidant activity of *Curcuma longa* extract by an integrated ^{1}H NMR and HPLC–MS metabolomic approach.

This chapter is taken from the following article: Stefano Dall'Acqua, Matteo Stocchero, Irene Boschiero, Mariano Schiavon, Samuel Golob, Jalal Uddin, Dario Voinovich, Stefano Mammi, Elisabetta Schievano 'New findings on the in vivo antioxidant activity of *Curcuma longa* extract by an integrated ^{1}H NMR and HPLC-MS metabolomic approach', published in *Fitoterapia* **2016, 109, 125-131**.

2.1 Abstract

Curcuminoids possess powerful antioxidant activity as demonstrated in many chemical in vitro tests and in several in vivo trials. Nevertheless, the mechanism of this activity is not completely elucidated and studies on the *in vivo* antioxidant effects are still needed. Metabolomics may be used as an attractive approach for such studies and in this paper, we describe the effects of oral administration of a *Curcuma longa* L. extract (150 mg/kg of total curcuminoids) to 12 healthy rats with particular attention to urinary markers of oxidative stress. The experiment was carried out over 33 days and changes in the 24-h urine samples metabolome were evaluated by ^{1}H NMR and HPLC–MS. Both techniques produced similar representations for the collected samples confirming our previous study. Modifications of the urinary metabolome lead to the observation of different variables proving the complementarity of ^{1}H NMR and HPLC–MS for metabolomic purposes. The urinary levels of allantoin, *m*-tyrosine, 8-hydroxy-2′-deoxyguanosine, and nitrotyrosine were decreased in the treated group thus supporting an *in vivo* antioxidant effect of the oral administration of *Curcuma* extract to healthy rats. On the other hand, urinary TMAO levels were higher in the treated compared to the control group suggesting a role of curcumin supplementation on microbiota or on TMAO urinary excretion. Furthermore, the urinary levels of the sulphur containing compounds taurine and cystine were also changed suggesting a role for such constituents in the biochemical pathways involved in *Curcuma* extract bioactivity and indicating the need for further investigation on the complex role of antioxidant curcumin effects.

2.2 Introduction

Food supplements and nutraceuticals are largely used for health promoting purposes mainly ascribed to the antioxidant properties of the phytochemicals contained in these products. Nevertheless, it is well accepted that the antioxidant activity is poorly related to the radical scavenging properties that can be demonstrated with in vitro chemical assays. Furthermore, studies related to the real in vivo antioxidant activities of these chemicals as well as to their effects on healthy subjects are still missing. In general, the study of in vivo antioxidant activity is difficult due to the complex multiple targets of purified natural products or extracts possessing this effect [1, 2]. Current studies of antioxidant phytochemicals are generally focused on specific compounds and their effects are evaluated on a limited number of markers [1].

Curcuma longa L. is extensively used in Ayurveda, Unani, Siddha, and Chinese medicine for the management of various diseases. This spice is highly regarded for its numerous biological activities especially related to antioxidant, anti-inflammatory and cancer preventive properties [3–11]. The effects of *C. longa* are ascribed to the presence of diarylheptanoid compounds known as curcuminoids (namely curcumin, demethoxycurcumin, and bisdemethoxycurcumin), which are considered the main active principles of the plant, although their bioavailability is poor because of scarce absorption, rapid metabolism and systemic elimination [12, 13]. Extremely low serum levels of curcumin after oral administration were observed [12], making it difficult to explain its antioxidant properties on the basis of simple radical scavenging action. Nevertheless, extensive scientific research over the past decade [6, 14–17] has shown that this compound is able to modulate multiple cellular targets and hence that it possesses preventive and therapeutic value against a wide variety of diseases thus showing the need for new approaches in the study of this natural product. Previous studies in rats have shown the ability of curcumin to upregulate the transcription factor nuclear factor erythroid 2-related factor 2 (Nrf2), which is responsible for phase 2 antioxidant and detoxification genes expression, indicating that this compound increases the total superoxide dismutase and glutathione peroxidase activities [18,19].

Metabolomics can offer new opportunities in this research area since it allows the observation of changes in particular bio-fluids caused by the overall effect of a natural product on different biochemical pathways. Urinary biomarkers of oxidative status present a great opportunity to study redox balance because specimen collection is non-invasive [20] and long-term observation

experiments are possible. Therefore, studies using urinary metabolome analysis are attractive especially for the evaluation of antioxidants in healthy subjects or in healthy *in vivo* models.

In a previous study, we used a metabolomic approach to study the changes of the urinary metabolic profile after the administration of *C. longa* extract in rats. Compared to the control group, the treated animals were characterized by decreased levels of allantoin, a urinary biomarker of oxidative stress [21].

As a continuation of our previous study, we evaluated the effect of oral daily administration of standardized *Curcuma longa* L. extract (corresponding to 150 mg/kg of total curcuminoids) to 12 healthy rats by untargeted metabolomics. Treatment was carried out over 33 days and changes in the urinary metabolome were evaluated by monitoring the 24-h urine composition by ^{1}H NMR and HPLC–MS. Urinary collections at 42 days (after stopping the treatment at day 33) were also analysed. We attempted to use the combined potential of NMR and MS in a unified metabolomic approach as a powerful tool to assess the modification of urine composition caused by curcumin supplementation in a healthy animal model. Both techniques produced similar representations for the collected samples confirming our previous study using similar methodology. The two different approaches were able to detect variations in the urinary metabolome, leading to the observation of different variables thus proving the complementarity of these two analytical techniques for metabolomic purposes.

2.3 Experimental

2.3.1 Materials

Curcumin standard, methanol, acetonitrile, formic acid, hydrochloric acid, deuterated water, methanol, glutathione (GSH), sulphosalicylic acid (SSA), γ-glutamyl-glutamic acid (γ-Glu-Glu), ethylenediaminetetracetic acid (EDTA), and N-ethylmaleimide (NEM) were obtained from Sigma-Aldrich(Milan, Italy). Curcumin glucuronide was synthesized in our laboratory using a previously published protocol [22]. *C. longa* L. dried extract was purchased from a local market; the total curcuminoid content was measured as previously described [21,23] as 94%; specifically, 71.0%, curcumin, 20.5% demethoxycurcumin, and 2.5% bisdemethoxycurcumin were determined using HPLC–MS and HPLC-DAD measurements [21].

2.3.2 Animals and Urine Collection

All experimental protocols involving animals were reviewed and approved by the Ethical Committee for animal experiments of the University of Padua (CEASA, protocol number 49571). The study involved 12 Sprague–Dawley rats: 6 males and 6 females, 8 ± 1 weeks of age, at the beginning of the experiments; male animals weighted 78.0 ± 2.3 g and female animals 79.5 ± 4.0 g. They were caged in a temperature- and photoperiod-controlled (12-h light/dark cycle) room with rodent maintenance diet and water *ad libitum*. Rats were randomly divided into a control (three males and three females) and a curcumin-treated group (three males and three females). No differences were observed between the two groups at the beginning of the experiment, based on HPLC–MS and NMR preliminary data. Six hundred milligrammes of *C. longa* extract were suspended in 12 mL of water. The treated group received a daily dose of 160 mg/kg of *C. longa* extract (corresponding to 150 mg/kg total curcuminoids or 112 mg/kg of curcumin) orally by gavage for 33 days. An equal dose of water was given to the control group. At day 0, 6, 15, 22, 28, 33 and 42 (10 days after the end of the treatment), the animals were housed individually in metabolic cages for the collection of the 24-h urine outputs. The collected samples were stored at -80 °C until ^{1}H NMR and HPLC–MS analysis.

2.3.3 HPLC–MS Urine Analysis

The UPLC-HRMS analysis was conducted by the group of Prof. Stefano Dall'Acqua at the Department of Pharmaceutical Sciences, University of Padova. To obtain a metabolic profiling of urine, an HPLC–MS full scan method was used. A Varian MS 500 equipped with a prostar 430 autosampler and binary chromatograph 212 series (Varian, Palo Alto, USA), was used as HPLC–MS system. An Agilent (Milan, Italy) Eclipse XDB C-8 column (2.1×150 mm 3.5 µm) was used as stationary phase. The mobile phase was composed of solvent A (acetonitrile with 0.5% acetic acid) and solvent B (water with 2% formic acid). Linear gradients of A and B were used, as follows: 0 min, 10% A; 20 min, 85% A; 21 min, 100% A, 21.30 min, 10% A; 27 min, 10% A. The flow rate was 200 µL/min and the injection volume was 10 µL. The mass range explored was 50–1000 m/z. MS were recorded both in positive standard mode and in turbo depending data scanning (tdds) mode that allows the elucidation of the fragmentation patterns of the detected ions. Collected urine samples were centrifuged (13,000 g for 10 min) and directly injected in the HPLC. Each HPLC–MS data set was processed with MZmine 2.9 software [24];

from the raw data files, we obtained a data set composed of 102 variables. Median Fold Change normalization was applied to take into account the effects of sample dilution. Data were log-transformed and mean centred.

2.3.4 ^{1}H–NMR Urine Analysis

Aliquots of 700 μL of urine at pH 2.50 ± 0.05 were centrifuged at 13,000 rpm for 10 min and mixed with 70 μL of 2 mM 3-(trimethylsilyl)propionate-2,2,3,3,-d4 (TSP) in D2O solution. ^{1}H NMR spectra were recorded at room temperature using a Bruker (Rheinstetten, Germany) Avance DMX 600 spectrometer. One-dimensional spectra were acquired using the NOESYGPPR1D pulse sequence. Parameters used were: 64 scans, 32 k data points, spectral width of 8389.26 Hz, 2 s relaxation delay, 50 ms mixing time, 1.95 s acquisition time. Prior to Fourier transformation, the FIDs were zero-filled to 64 k points and an exponential line broadening factor of 0.3 Hz was applied. All spectra were manually corrected for phase and baseline distortions using ACD/NMR Workbook software (Advanced Chemistry Development, Inc. Toronto, Ontario, Canada) and were referenced to the CH_3 resonance of creatinine at 3.13 ppm. Spectra were aligned using the CluPA algorithm (VU T.N., Laukens K., Valkenborg D. (2012) speaq: an R-package for NMR spectrum alignment and quantitation. R package version 1.1.). The spectral region between 4.7 and 5.0 ppm was removed prior to statistical data analysis to avoid variability due to the residual water signal. Data were reduced to 470 bins by intelligent bucketing; the obtained data set was normalized by Total Sum Normalization and mean centering and Pareto scaling were applied.

2.3.5 Blood Sample, Glutathione and Curcumin Quantification

Whole blood samples were collected at day 34 and stored in heparinised tubes at −20 °C until analysis. GSH was measured using a previously described method [25]. Briefly, a 20 mM GSH stock solution in water was used to prepare calibration curves. The precipitating solution was prepared by mixing 150 μL of a solution containing NEM, EDTA and γ-Glu-Glu (in water/methanol, 85/15 (v/v)) with 50 μL of SSA; the final concentrations in the precipitating solution were 20 mM, 2 mM, 250 μM and 2% (w/v) for NEM, EDTA, γ-Glu-Glu and SSA, respectively. Curcumin was measured using a previously published method using SPE extraction [23].

36

2.3.6 Statistical Data Analysis

Multivariate data analysis based on projection methods was applied for statistical data analysis. Specifically, exploratory data analysis was performed by principal component analysis (PCA) while a new projection to latent structures (PLS)-based method was applied to study the changes in the urinary metabolome during the experiment. While PCA is a well-known technique used in multivariate data analysis [26], the PLS-based approach applied to model the data collected during our longitudinal study was recently published by our group [21]. Projection to latent structures by partial least squares regression (PLS) [27] is an effective and robust regression technique used to investigate the relationships existing between two blocks of data, usually called X- and Y-block. In metabolomics applications, PLS often produces a large number of latent components with the result to compromise a clear interpretation of the model. For this reason, we elaborated a post-transformation method, called post-transformation of PLS2, able to decompose the structured variation of the X-block discovered by PLS into two main blocks corresponding to the variations correlated (the so called parallel or predictive block) and orthogonal to the Y-block by a suitable rotation of the weights of the PLS model. Post-transformation of PLS2 is a three step approach. In the first step, a PLS regression model is built on the data; in the second step, the weight matrix of the model is rotated while in the third step a regression model is rebuilt by using the same framework of the PLS algorithm but the new weight matrix to project the data. The relationships between the X-block and the Y-block can be investigated by exploring only the parallel part of the model by using suitable correlation loading plots. As a result, the model obtained by post-transformation of PLS2 maintains the same power in prediction and regression coefficients of the unrotated PLS model but can be easily interpreted because the number of components useful to interpret the model is usually reduced. Post-transformation of PLS2 can be applied to model longitudinal studies by considering the experimental data and the design matrix as X- and Y-block, respectively. In our study, we supposed an interaction model to define the design matrix. The significance of the terms time, treatment and time× treatment included in the model was evaluated by permutation tests.

To avoid over-fitting and prove the robustness of the obtained models, we performed N-fold full cross-validation with different values of N (N=6, 7, 8) and a permutation test on the response (500 random permutations) according to good practice for model validation. Data set comparison was performed by Bidirectional Orthogonal Projections to Latent Structures (O2PLS) [28]. PCA

and PLS models were built using SIMCA 13 (Umetrics, Umea, Sweden) while the platform R 3.0.2 (R Foundation for Statistical Computing) was used to perform t-test and Mann–Whitney test, to post-transform the PLS model (user-written R function) and to build the O2PLS model (user-written R function).

2.4 Results

2.4.1 Animal Weight and Urinary Output

No differences in the treated vs control group were observed in animal weight and 24-h urinary output during the experiment. The differences between the two groups were not statistically significant according to t-test and Mann-Whitney test (both p-values were >0.10). Data are summarized in **Table 2.1**.

	Body weight (g)		Urine volume (mL)	
Day	Control	Treated	Control	Treated
0	78.8 ± 3.7	78.8 ± 3.1	6.3 ± 4.0	7.0 ± 2.4
6	110.0 ± 5.0	107.6 ± 7.3	10.7 ± 3.0	15.3 ± 6.5
15	190.3 ± 20.4	180.8 ± 20.8	12.7 ± 2.2	15.7 ± 5.3
22	230.8 ± 38.1	217.0 ± 39.6	13.8 ± 2.0	16.5 ± 2.7
28	254.8 ± 44.8	236.8 ± 45.1	11.2 ± 1.8	11.8 ± 2.1
33	278.5 ± 60.8	262.8 ± 55.0	11.5 ± 1.5	10.6 ± 1.2
42	300.3 ± 68.7	290.3 ± 64.0	10.0 ± 2.9	10.3 ± 3.1

Table 2.1. Variations of body weight and 24-hour urine output of the rats during the course of the experiment.

2.4.2 Data Analysis of the ^{1}H NMR and HPLC–MS Data Sets

Exploratory data analysis on the two data sets did not show the presence of outliers in the data. PCA models of the urines collected on day 0 did not show differences between rats belonging to the control or the treated group.

As a first step of our data modelling strategy, the two data sets were compared to investigate the common information shared by NMR and HPLC–MS. To this end, we scaled both data sets to unit variance and applied O2PLS (**Fig. 2.1**). The obtained model showed a joint overlapping variation described by 3 latent components ($R^2 = 0.59$ for NMR and $R^2=0.45$ for HPLC–MS), a unique systematic variation for NMR having 4 latent components ($R^2=0.19$) and a unique systematic variation of 4 latent components for HPLC–MS ($R^2=0.29$). As a consequence, we can conclude that a large part of the systematic variation of the two data sets contains the same information while only a small part is unique and non-overlapping in the two data sets. **Figure 2.1** shows the score scatter plot for the first and the second latent components describing the joint co-variation of the two data sets: the common information can be qualitatively interpreted in terms of the effects of time evolution and curcumin supplementation. This first model considers only the correlation structure existing between the NMR and the HPLC–MS data sets, and ignores the chemical identity of the variables used. For this reason, we cannot consider the two data sets as equivalent as will be proven with the following models.
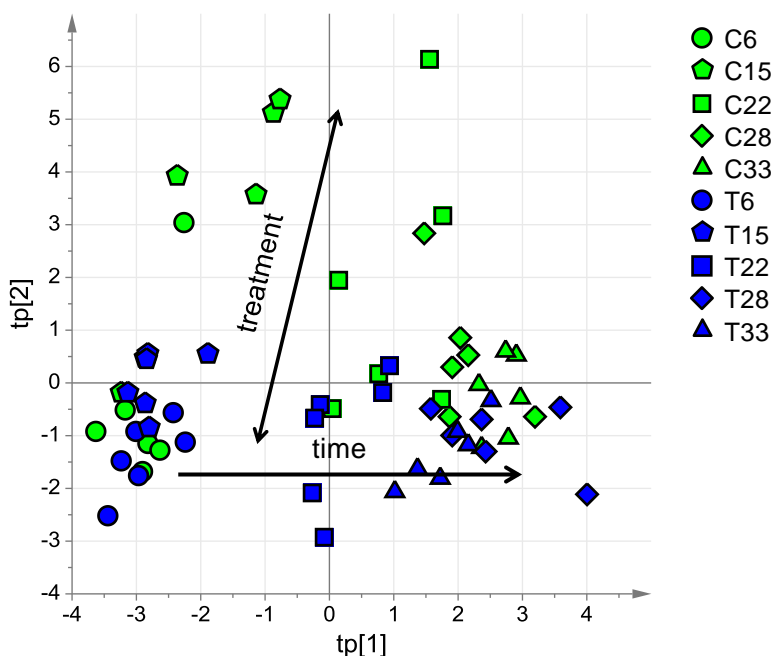
Figure 2.1. Score scatter plot describing the joint systematic variation explained by the first and the second latent component of the O2PLS model; different symbols and colours were used to allow the interpretation of the observed patterns in terms of time evolution and treatment effect. C = control; T = treated; the numbers refer to the day of urine collection.

In the second part of our data modelling strategy, post-transformed PLS2 models of the data related to urine samples collected from day 6 to day 33 proved that the interaction term time × treatment was not significant at the level of 95% and that a simple linear model can be used for the design matrix. Both the model obtained for the ^{1}H NMR data set and that obtained for the HPLC–MS data set clearly showed the effects of time and treatment on the metabolic profile of the urine as it can be observed in the score scatter plots of **Figure 2.2 and 2.3**. The model for the ^{1}H NMR data set had A = 2 + 4 components, R^2 = 0.74 and Q^2 = 0.52 for treatment and R^2 = 0.77 and Q2 = 0.70 for time while the model for the HPLC–MS data set had A = 2 + 2 components, R^2 = 0.73 and Q^2 = 0.44 for treatment and R^2 = 0.85 and Q^2 = 0.64 for time.
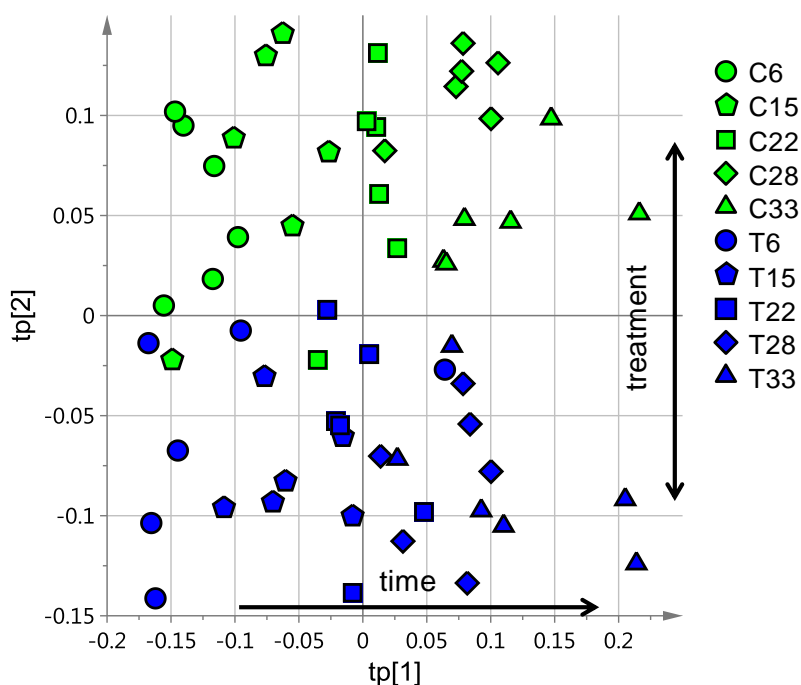
Figure 2.2. ^{1}H-NMR data set: score scatter plot for the post-transformed PLS2 model; the time evolution of the composition of the collected urines is described by the horizontal axis (tp[1]) while the vertical axis (tp[2]) represents the effects of the different treatment. C = control; T = treated; the numbers refer to the day of urine collection.

By analyzing the correlation loading plots of the obtained models, it was possible to find variables characterizing the time evolution of the samples and the effects of curcumin administration on the rat urine metabolome.

Days 15, 28 and 33 were selected as key points of the experiment to check the presence of significant modifications of control vs. treated groups in the metabolites highlighted by our analysis.
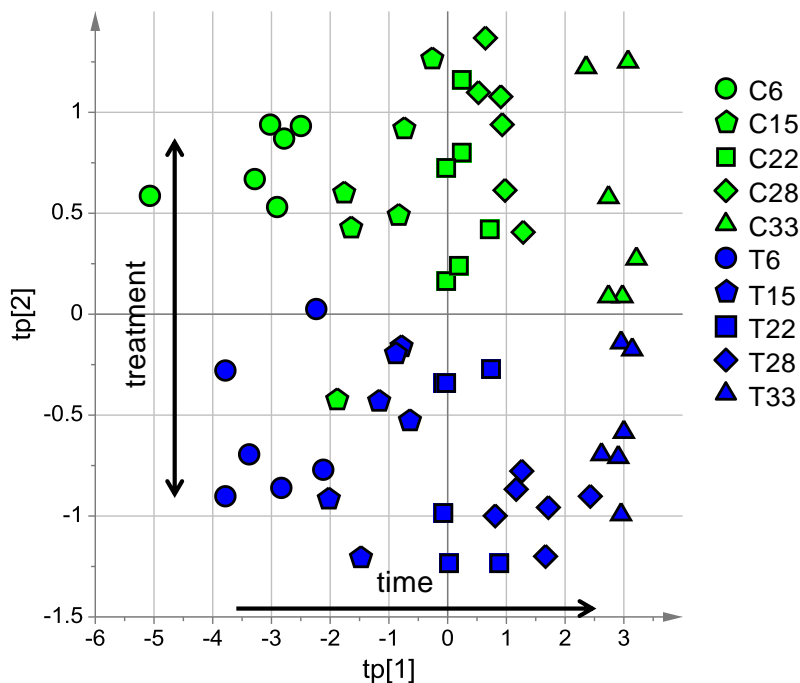
Figure 2.3. HPLC-MS data set: score scatter plot for the post-transformed PLS2 model; the time evolution of the urinary metabolome can be observed along the horizontal axis (tp[1]) while the effects of the treatment are included in tp[2] (vertical axis). C = control; T = treated; the numbers refer to the day of urine collection.

Considering the ^{1}H NMR data, resonances associated with the discrimination between control and treated groups were observed in the range of deshielded protons (8.17–8.50 ppm), but it was not possible to assign any known metabolite to those signals (**Figure 2.4**). Tentative assignments were deduced on the basis of spectral data for bins as reported in **Table 2.2**. Hippuric acid, 2-oxoglutarate and trimethylamine N-oxide (TMAO) levels were higher in the treated group compared to the control. On the other hand, the urinary marker of oxidative stress, 8-hydroxy-2′-deoxyguanosine (8-*OHdG*), was higher in the control group. The correlation loading plot obtained for the model of the HPLC–MS data set highlighted significant differences in a set of variables that were chemically identified and are related to oxidative stress. Specifically, our untargeted approach confirmed the metabolites found in our previous targeted approach [21].
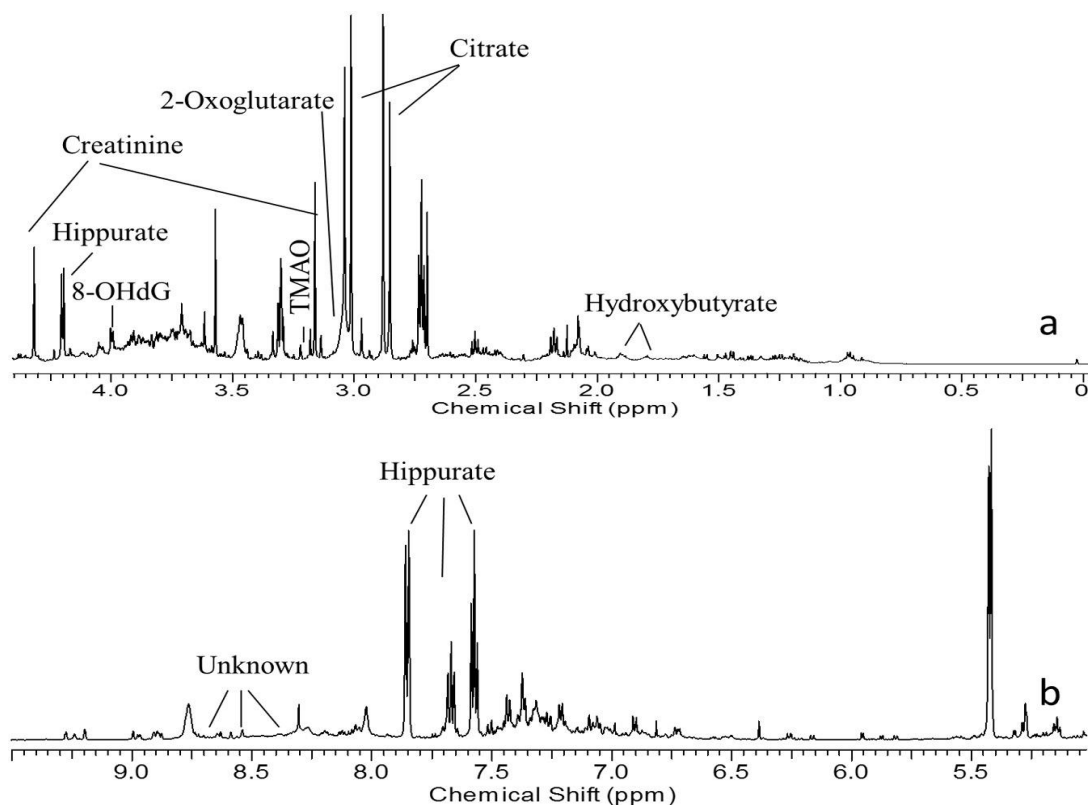
Figure 2.4. Representative 600 MHz 1H NMR spectra of urine after oral administration of curcumin in rats. (a) Aliphatic region (0.0–4.5 ppm). (b) Aromatic region (5.0–9.5 ppm). The vertical scale in the aromatic region was magnified four times compared with that in the aliphatic region.

Indeed, four urinary markers of the oxidative status of the animals, i.e., allantoin, *m*-tyrosine, 3-nitrotyrosine, and 8-*OHdG*, were identified in the urine on the basis of their m/z value and fragmentation patterns compared to those registered in the Human Metabolome Database and Mass Bank Database. Furthermore, urinary levels of two sulphur containing compounds, namely taurine and cystine were modified during the experiment. In **Table 2.3**, the calculated reduction of these metabolites as average measured data of treated vs. control animals on days 15, 28 and 33 are reported. Allantoin, 3-nitrotyrosine, *m*-tyrosine, and 8-OHdG levels were significantly decreased in the treated compared to the control group starting from day 6 of urine collection, and also at the end of the treatment (**Table 2.3**). Surprisingly, on day 28 the average *m*-tyrosine levels were higher (+13%) in the treated group.

ppm	Assignment	day 15	day 28	day 33
8.50-8.17	not assigned	>C*	>C#	>C*
7.57-7.55	hippuric acid	>T#	>T*	>T#
4.30	Creatinine	>C#	>C*	>C*
3.95	8-*OHdG*	>C#	>C*	>C*
3.32	not assigned	>T*	>T#	>T#
3.23-3.22	TMAO	>T*	>T*	>T#
3.19	2-oxoglutarate	>T*	>T#	>T*
2.81	Citrate	>C#	>T*	>T*
1.81-1.79	hydroxybutyrate	>C*	>C*	>C#

Table 2.2. Comparison of selected bins observed in the NMR data set; differences between treated and control group are reported in terms of increase or decrease of the NMR integral. *p-value <0.10 for both t-test and Mann-Whitney test; #difference in the mean values, but not significant at the level of 90% for both t-test and Mann-Whitney test.

On day 33, allantoin, *m*-tyrosine, and 8-*OHdG* reductions in the treated group compared to controls were significant (p-value b 0.10 for both t-test and Mann–Whitney test). Significant (p-value b 0.10 for both t-test and Mann–Whitney test) changes were also observed for taurine (−24%) and cystine (−40%) in the treated group compared to controls (**Table 2.3**). The levels of these metabolites were measured also 10 days after the administration of Curcuma extract stopped (day 42). At this point (day 42) no significant differences in urinary composition were observed between treated and control groups showing that the observed changes were reversible with the interruption of curcumin extract administration.

Assignment	day 15	day 28	day 33
allantoin	-31.0*	-16.0#	-34.0*
3-nitrotyrosine	-23.0#	-18.0#	-27.0#
m-tyrosine	-0.2#	+13.0#	-20.0*
8-*OHdG*	-0.7#	-19.0#	-28.0*
taurine	*-4.4#*	*-26.0**	*-24.0**
cystine	*-24.0#*	*-29.0**	*-40.0**

Table 2.3. Oxidative stress urinary markers reduction in the treated group as a percentage of the value of the control group; *p-value <0.10 for both t-test and Mann-Whitney test, #difference in the mean values, but not significant at the level of 90% for both t-test and Mann-Whitney test.

2.4.3 Whole Blood Glutathione (GSH) and Curcumin Levels

Blood samples were collected at the end of the treatment (day 34) and GSH and curcumin levels were measured. We did not find any difference between the average concentration of whole blood GSH in the control and the treated group on day 34. Curcumin, curcumin glucuronide and curcumin sulphate were not detectable in either the treated or control groups.

2.5 Discussion

Food supplements and nutraceuticals with claimed antioxidant properties are enjoying a growing diffusion because of their health-promoting effects. Polyphenols are a large group of phytochemicals that present strong chemical antioxidant properties. Their health benefits are often claimed based on their antioxidant properties in vitro, but evidence for *in vivo* antioxidant effects is still limited since no validated in vivo biomarkers have been identified and no long-term studies are available [1, 29–31]. Metabolomics offers new opportunities for the evaluation of in vivo antioxidant properties of complex mixtures such as natural products [32–34]. The study of urinary metabolome and urinary biomarkers of oxidative stress is attractive because sample collection is simple and non-invasive [20] and may lead to the observation of modified levels of metabolites that can be considered as a starting points for depicting new mechanisms of in vivo antioxidant activity. In this paper, we report significant changes in the urinary metabolome of healthy rats, orally treated with curcumin, compared to controls in data sets obtained both by NMR and HPLC–MS. These results confirm our previous findings obtained using a different experiment design: a lower dose of curcumin extract (80 mg/kg) was administered, and a targeted HPLC–MS approach was used, by selecting 25 metabolites in the urine chromatogram [21]. In the present work, we identified a larger number of metabolites that are related to curcumin supplementation in healthy rats.

The metabolic changes revealed by NMR data are related to phenolic compound metabolism. Previously published papers reported increase in urinary hippuric acid levels after administration of fruit and polyphenol rich foods [35]. Other authors reported that the oral administration of the

44

flavonoid quercetin to rats, in an NMR-based metabolomic study, resulted in an increase in choline, creatinine, dimethylglycine, hippurate, taurine, and TMAO and in the reduction in acetate, alanine, and lactate [36]. These authors considered the general changes of such compounds as modification in osmolyte levels thus suggesting that these data may indicate improved glomerular or general renal function and correlated their observation in kidney osmolyte activity to the potential beneficial effects of quercetin on kidney function and hypertension [36]. Our data showed a higher level of creatinine and taurine in the control group. It is difficult to correlate the decreased urinary creatinine levels in the treated group with any biological meaning. On the contrary, the reduced urinary taurine levels, confirmed our previous observations [21] and other published papers that reported lowered taurine brain concentrations in curcumin-treated rats [37]. Taurine has been previously reported to decline in a number of tissues with advancing age and also in rats, the urinary levels were significantly reduced with ageing [38]. On the other hand, increased urinary taurine levels have been indicated as a specific marker of liver toxicity [37, 38]. Cystine levels were also significantly reduced on day 33, but this metabolite can also derive from cysteine modification during 24-h urine collection so that it is difficult to assess its meaning in this bio-fluid.

TMAO is an oxidation product of trimethylamine (TMA), and both these compounds are products of choline metabolism. The methylamine pathway is a typical example of microbial–mammalian co-metabolism and is well known that intestinal microbiota plays a role in the catabolism of choline in humans and rodents [39]. Dietary choline is converted in TMA by gut microbiota and TMA is mainly oxidized to TMAO [39-41]; thus, an increase of such metabolite may be related to the influence of the oral curcumin treatment on intestinal microbiota. There is strong interest in the evaluation of TMAO plasmatic or urinary levels due to various physio-pathological functions that have been proposed for this compound. In fact, recent animal studies have shown a link between intestinal microbial metabolism of the choline moiety in dietary phosphatidylcholine and coronary diseases through the production of TMAO, considered as a proatherogenic compound. In humans, the production of TMAO from dietary phosphatidylcholine is dependent on metabolism by the intestinal microbiota. Furthermore, ingestion of different types of foods, such as eggs or fish, may influence TMAO plasma levels [41–43]. Other authors reported the ability of oral broad-spectrum antibiotics to temporarily suppress the production of TMAO suggesting that intestinal microorganisms participate in

phosphatidylcholine metabolism to form circulating and urinary TMAO [41, 44]. Increased plasma TMAO levels are associated with an increased risk of incidence of major adverse cardiovascular events [41, 44]. The role of TMAO appears to be complex but our data indicate that curcumin oral administration modifies the microbiota or influences the urinary excretion of this compound thus leading to changed urinary levels in the treated group. This result suggests a role of the intestinal microbial population in the mechanism of action of curcuminoids. Recently, the ability of curcumin supplementation to modulate colonic microbiota during colitis and colon cancer prevention was studied showing an increase in microbial diversity and restoration of colonic microbial composition to that observed in healthy WT animals compared to mouse model of IBD-associated colon cancer [45].

Urinary levels of some markers of oxidative stress were significantly reduced because of the treatment, as demonstrated by HPLC–MS results. Allantoin is considered a urinary marker of oxidative stress, because it is the predominant product of non-enzymatic oxidation of uric acid by many types of free radicals, and it is considered a valid biomarker of oxidative state especially in humans [20].

m-Tyrosine is considered a promising biomarker for oxidative damage to proteins [46]. The highly reactive hydroxyl radical oxidizes phenylalanine residues to o-tyrosine and m-tyrosine and increased levels of these metabolites are correlated to an increased ROS production from normal metabolic processes or from exposure to exogenous factors. Also reactive nitrogen species react readily with tyrosine and protein-associated tyrosine to form free 3-nitrotyrosine and protein associated 3-nitrotyrosine, respectively [47]. Urinary 3-nitrotyrosine is a potential biomarker that may reflect the enhanced generation of reactive nitrogen species and it has been proposed as a biomarker to detect changes in oxidative stress and to evaluate the efficacy of therapeutic interventions aimed at reducing oxidative stress [48].

Urinary 8-*OHdG* is considered a biomarker of generalized cellular oxidative stress because it is one of the predominant products of oxidized DNA repair [49]. Because it is fairly water-soluble, it will be excreted into the urine without being further metabolized and it is considered a stable end product of non-enzymatic DNA oxidation [20]. Therefore, increased urinary levels of 8-*OHdG* could be correlated to an increase of oxidative DNA damage [20]. Our experiment showed significant reduction of some of these urinary markers of oxidative stress after 33 days

46

of treatment (see **Table 2.3**), suggesting an *in vivo* antioxidant effect of curcumin supplementation in the healthy rat.

To have a further parameter related to the oxidative state of the animals at the end of the treatment, blood samples were collected on day 34 and GSH levels were measured. The average blood GSH levels showed no difference (t-test p-value b 0.01) between the control and the treated group. Other authors reported similar results in a study that evaluated the effect of curcumin and one analogue [bis-1,7- (2-hydroxy-phenyl)-hepta-1,6-diene-3,5-dione] (BDMC-A) on carbon tetrachloride-induced hepatotoxicity in rats. Control and curcumin treated groups presented the same GSH values, while significant increase in plasma GSH levels was observed in the animals group treated with CCl_4 and curcumin compared with the CCl4 treated ones [50].

However, a previous study reported a role for curcumin against GSH depletion-mediated mitochondrial dysfunction *in vitro* and *in vivo* [51]. Another study reported a significant effect of curcumin on GSH biosynthesis in alveolar epithelial cells [52]. Other researchers have considered that some of the most important effects of curcumin, such as the anticarcinogenic, antimutagenic, antioxidant and cytoprotective activities can be explained by its inhibitory effect on glutathione S-transferase (GST) [15]. The capacity of curcumin to protect rats from adriamicin (ADR) nephrotoxicity was demonstrated [53]. Curcumin protected against ADR-induced renal injury by suppressing oxidative stress and increasing kidney glutathione content and glutathione peroxidase activity; nevertheless, in the same paper, kidney GSH levels of animals treated either with saline or curcumin were similar showing that curcumin treatment does not increase kidney GSH levels but can restore ADR induced GSH depletion. Our results indicate that curcumin supplementation does not increase blood GSH levels in healthy subject. Considering this result the low plasma concentration of curcumin due to its poor absorption, must be underlined. Previous studies have reported that, due to poor bio-availability, only traces of curcumin are detected in plasma after oral administration [12]. Also our data showed no detectable curcumin and curcumin conjugated metabolites in blood 24 h after the last administration (day 34) indicating rapid elimination of the compound from the bloodstream, in agreement with previously published results that reported the curcumin plasma peak 40 min after oral administration of 500 mg/kg in rats [54]. A large amount of orally administered curcumin, due to poor absorption, is present at the intestinal level so that the intestinal mucosa is exposed to higher concentrations of curcumin and for this reason, interactions of curcumin with GSH and

GST are likely to be more significant in intestinal epithelial cells rather than in plasma [55]. In our healthy animal model, curcumin supplementation is not likely to play a prominent role to change GSH levels in plasma.

2.6 Conclusions

In this work we used both ^{1}H NMR and HPLC–MS techniques to study the modification of urinary composition in rats treated with *C. longa* extract correlated with *in vivo* antioxidant activity. Multivariate analysis on ^{1}H NMR and HPLC–MS data produced similar representations for the collected samples. The two different approaches were able to detect variations in the urinary metabolome, leading to the observation of different components, showing the complementarity of these two analytical techniques for metabolomic purposes. The results of the present study are in agreement with our previously published data obtained with a lower curcumin dose and using a targeted ^{1}H NMR and HPLC–MS approach [21]. The evaluation of the effects of Curcuma extract on urinary composition in healthy rats by a metabolomic approach led us to observe evidence for an *in vivo* antioxidant effect caused by a significant reduction in the amount of urinary biomarkers of oxidative stress such as allantoin, m-tyrosine, 8-*OHdG*. A tendency to the reduction of 3-nitrotyrosine was also observed. Our metabolomics-based study supports an *in vivo* antioxidant effect of the oral administration of *C. longa* extract to healthy rats. The observation that urinary TMAO levels are increased in the treated compared to the control group may be related to the influence of curcumin supplementation on microbiota, as recently indicated by other research groups, or on the urinary excretion of this metabolite. Urinary levels of taurine and cystine, sulphur containing compounds, were also changed suggesting a role for such constituents in the biochemical pathways involved in *C. longa* extract bioactivity and indicating the need for further investigation on curcumin effects. The undetectable plasmatic levels of curcumin and its conjugates confirmed its rapid elimination from the bloodstream after oral administration indicating that the contribution to the whole antioxidant activity by a direct radical scavenging action is negligible. The unchanged plasmatic GSH amount in treated and control group indicates that curcumin supplementation in the health subject does not increase this endogenous antioxidant's levels.

48

2.7 References

[1]. Manach, C., Hubert, J., Llorach, R. and Scalbert, A. The complex links between dietary phytochemicals and human health deciphered by metabolomics. *Mol. Nutr. Food Res.* **2009**, 53, 1303-1315.

[2]. Ndhlala, A.R., Moyo, M. and Van Staden, J. Natural antioxidants: Fascinating or mythical biomolecules?. *Molecules.* **2010**, 15, 6905-6930.

[3]. Teiten, M.H., Eifes, S., Dicato, M. and Diederich, M. Curcumin-the paradigm of a multi-target natural compound with applications in cancer prevention and treatment. *Toxins.* **2010**, 2, 128-162.

[4]. Goel, A., Kunnumakkara, A.B. and Aggarwal, B.B. Curcumin as "Curecumin": From kitchen to clinic. *Biochem. Pharmacol.* **2008**, 75, 787-809.

[5]. Hatcher, H., Planalp, R., Cho, J., Torti, F.M. and Torti, S.V. Curcumin: From ancient medicine to current clinical trials. *Cell. Mol. Life Sci.* **2008**, 65, 1631-1652.

[6]. Aggarwal, B.B., Sundaram, C., Malani, N. and Ichikawa, H. Curcumin: The Indian solid gold, *Adv. Exp. Med. Biol.* **2007**, 59, 1-75.

[7]. Maheshwari, R.K., Singh, A.K., Gaddipati, J. and Srimal, R.C. Multiple biological activities of curcumin: A short review. *Life Sci.* **2006**, 78, 2081-2087.

[8]. Matchanickal, R.A. and Rafi, M.M. Curcumin: Potential health benefits, molecular mechanism of action, and its anticancer properties in vitro and in vivo, *ACS Symp. Ser.* **2006**, 925, 92-107.

[9]. Duvoix, A., Blasius, R., Delhalle, S., Schnekenburger, M., Morceau, F., Henry, E., Dicato, M. and Diederich. Chemopreventive and therapeutic effects of curcumin. *Cancer Lett.* **2005,** 223, 181-190.

[10]. Sharma, R.A., Gescher, A.J. and Steward, W.P. Curcumin: The story so far. *Eur. J. Cancer.* **2005**, 41, 1955-1968.

[11]. Chattopadhyay, I., Biswas, K., Bandyopadhyay, U. and Banerjee, R.K. Turmeric and curcumin: Biological actions and medicinal applications, *Curr. Sci.* **2004**, 87, 44-53.

[12]. Anand, P., Kunnumakkara, A.B., Newman, R.A. and Aggarwal, B.B. Bioavailability of curcumin: Problems and promises, *Mol. Pharm.* **2007**, 4, 807-818.

[13]. Ireson, C., Orr, S., Jones, D.J., Verschoyle, R., Lim, C.K., Luo, J.L., Howells, L., Plummer, S., Jukes, R., Williams, M. and Steward, W.P. Characterization of metabolites of the chemopreventive agent curcumin in human and rat hepatocytes and in the rat in vivo, and evaluation of their ability to inhibit phorbol ester-induced prostaglandin E2 production, *Cancer Res.* **2001**, 61, 1058-1064.

[14]. Aggarwal, B.B. and Harikumar, K.B. Potential therapeutic effects of curcumin, the anti-inflammatory agent, against neurodegenerative, cardiovascular, pulmonary, metabolic, autoimmune and neoplastic diseases. *Int. J. Biochem. Cell Biol.* **2009**, 41, 40-59.

[15]. Aggarwal, B.B., Kumar, A. and Bharti, A.C. Anticancer potential of curcumin: Preclinical and clinical studies. *Anticancer Res*. **2003**, 23, 363-398.

[16]. Anand, P., Sundaram, C., Jhurani, S., Kunnumakkara, A.B. and Aggarwal, B.B. Curcumin and cancer: An "old-age" disease with an "age-old" solution. *Cancer Lett.* **2008**, 267, 133-164.

[17]. Bharti, A.C., Donato, N., Singh, S. and Aggarwal, B.B. Curcumin (diferuloylmethane) down-regulates the constitutive activation of nuclear factor-κB and IκBα kinase in human multiple myeloma cells, leading to suppression of proliferation and induction of apoptosis. *Blood*. **2003**, 101, 1053-1062.

[18]. Yang, C., Zhang, X., Fan, H. and Liu, Y. Curcumin upregulates transcription factor Nrf2, HO-1 expression and protects rat brains against focal ischemia. *Brain Res.* **2009**, 1282, 133-141.

[19]. Carmona-Ramírez, I., Santamaría, A., Tobón-Velasco, J.C., Orozco-Ibarra, M., González-Herrera, I.G., Pedraza-Chaverrí, J. and Maldonado, P.D. Curcumin restores Nrf2 levels and prevents quinolinic acid-induced neurotoxicity. *J Nutr Biochem*, **2013**, 24(1), 14-24.

[20]. Il'yasova, D., Scarbrough, P. and Spasojevic, I. Urinary biomarkers of oxidative status. *Clinica Chimica Acta*. **2012**, 413, 1446-1453.

[21]. Dall'Acqua, S., Stocchero, M., Clauser, M., Boschiero, I., Ndoum, E., Schiavon, M., Mammi, S. and Schievano, E. Changes in urinary metabolic profile after oral administration of curcuma extract in rats. *J. Pharm. Biomed. Anal.* **2014**, 100, 348-356.

[22]. Pfeiffer, E., Hoehle, S.I., Walch, S.G., Riess, A., Sólyom, A.M. and Metzler, M. Curcuminoids form reactive glucuronides in vitro. *J. Agric. Food Chem.* **2007**, 55, 538-544.

[23]. Marczylo, T.H., Steward, W.P. and Gescher, A.J. Rapid analysis of curcumin and curcumin metabolites in rat biomatrices using a novel ultraperformance liquid chromatography (UPLC) method. *J. Agric. Food Chem.* **2009**, 57, 797-803.

[24]. Katajamaa, M., Miettinen, J. and Orešič, M. MZmine: Toolbox for processing and visualization of mass spectrometry based molecular profile data. *Bioinformatics.* **2006**, 22, 634-636.

[25]. Steghens, J.P., Flourié, F., Arab, K. and Collombel, C. Fast liquid chromatography-mass spectrometry glutathione measurement in whole blood: Micromolar GSSG is a sample preparation artefact. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2003**, 798, 343-349.

[26]. Jackson, J.E. Users Guide to Principal Components, John Wiley, New York, **1991**.

[27]. Wold, S., Sjöström, M. and Eriksson, L. PLS-regression: A basic tool of chemometrics, *Chemometr. Intell. Lab.* **2001**, 58, 109-130.

[28]. Trygg, J. and Wold, S. O2-PLS, a two-block (X-Y) latent variable regression (LVR) method with an integral OSC filter. *J. Chemometrics.* **2003**, 17, 53-64.

[29]. Manach, C., Williamson, G., Morand, C., Scalbert, A. and Rémésy, C. Bioavailability and bioefficacy of polyphenols in humans. I. Review of 97 bioavailability studies. *Am. J. Clin. Nutr.* **2005**, 81, 230S-242S.

[30]. Halliwell, B. Are polyphenols antioxidants or pro-oxidants? What do we learn from cell culture and in vivo studies?, *Arch. Biochem. Biophys.* **2008**, 476, 107-112.

[31]. Halliwell, B., Rafter, J. and Jenner, A. Health promotion by flavonoids, tocopherols, tocotrienols, and other phenols: direct or indirect effects? Antioxidant or not?, *Am. J. Clin. Nutr.* **2005**, 81, 268S-276S.

[32]. Godzien, J., Ciborowski, M., Angulo, S., Ruperez, F.J., Paz Martínez, M., Señorans, F.J., Cifuentes, A., Ibañez, E. and Barbas, C. Metabolomic approach with LC-QTOF to study the effect of a nutraceutical treatment on urine of diabetic rats. *J. Proteome Res.* **2011**, 10, 837-844.

[33]. Krastanov, A. Metabolomics - The state of art. *Biotechnol. Biotechnol. Equip.* **2010**, 24, 1537-1543.

[34]. Llorach, R., Garcia-Aloy, M., Tulipani, S., Vazquez-Fresno, R. and Andres-Lacueva, C. Nutrimetabolomic strategies to develop new biomarkers of intake and health effects. *J. Agric. Food Chem.* **2012**, 60, 8797-8808.

[35]. Mulder, T.P., Rietveld, A.G. and van Amelsvoort, J.M. Consumption of both black tea and green tea results in an increase in the excretion of hippuric acid into urine. *Am. J. Clin. Nutr.* **2005**, 81, 2565-2605.

[36]. An, D., Zhang, Q., Wu, S., Wei, J., Yang, J., Dong, F., Yan, X. and Guo, C. Changes of metabolic profiles in urine after oral administration of quercetin in rats. *Food Chem. Toxicol.* **2010**, 48, 1521-1527.

[37]. Pyrzanowska, J., Piechal, A., Blecharz-Klin, K., Lehner, M., Skórzewska, A., Turzyńska, D., Sobolewska, A., Plaznik, A. and Widy-Tyszkiewicz, E. The influence of the long-term administration of Curcuma longa extract on learning and spatial memory as well as the concentration of brain neurotransmitters and level of plasma corticosterone in aged rats. *Pharmacol. Biochem. Behav.* **2010**, 95, 351-358.

[38]. Dawson, R., Liu, S., Eppler, B. and Patterson, T. Effects of dietary taurine supplementation or deprivation in aged male Fischer 344 rats. *Mech. Ageing Dev.* **1999**, 107, 73-91.

[39]. Russell, W.R., Hoyles, L., Flint, H.J. and Dumas, M.E. Colonic bacterial metabolites and human health. *Curr. Opin. Microbiol.* **2013**, 16, 246-254.

[40]. Zeisel, S.H., daCosta, K.A., Youssef, M.E.R.V.A.T. and Hensey, S. Conversion of dietary choline to trimethylamine and dimethylamine in rats: Dose-response relationship. *J. Nutr.* **1989**, 119, 800-804.

[41]. Tang, W.H. and Hazen, S.L. The contributory role of gut microbiota in cardiovascular disease. *J. Clin. Invest*. **2014**, 124, 4204-4211.

[42]. Miller, C.A., Corbin, K.D., da Costa, K.A., Zhang, S., Zhao, X., Galanko, J.A., Blevins, T., Bennett, B.J., O'Connor, A. and Zeisel, S.H. Effect of egg ingestion on trimethylamine-N-oxide production in humans: A randomized, controlled, dose-response study. *Am. J. Clin. Nutr.* **2014**, 100, 778-786.

[43]. O'Gorman, A., Gibbons, H. and Brennan, L. Metabolomics in the identification of biomarkers of dietary intake. *Comput. Struct. Biotechnol. J.* **2013**, 4 (5), 1-7.

[44]. Tang, W.W., Wang, Z., Levison, B.S., Koeth, R.A., Britt, E.B., Fu, X., Wu, Y. and Hazen, S.L. Intestinal microbial metabolism of phosphatidylcholine and cardiovascular risk. *N. Engl. J. Med.* **2013**, 368, 1575-1584.

[45]. McFadden, R.M.T., Larmonier, C.B., Midura-Kiela, M.T., Ramalingam, R., Harrison, C.A., Besselsen, D.G., Chase, J., Caporaso, G., Ghishan, F.K. and Kiela, P.R. 275 The Role of Curcumin in Modulating Colonic Microbiota During Colitis and Colon Cancer Prevention. *Gastroenterology*. **2014**, 146, S-66.

[46]. Orhan, H., Vermeulen, N.P., Tump, C., Zappey, H. and Meerman, J.H. Simultaneous determination of tyrosine, phenylalanine and deoxyguanosine oxidation products by liquid chromatography-tandem mass spectrometry as non-invasive biomarkers for oxidative damage, *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2004**, 799, 245-254.

[47]. Tsikas, D., Mitschke, A., Suchy, M.T., Gutzki, F.M. and Stichtenoth, D.O. Determination of 3-nitrotyrosine in human urine at the basal state by gas chromatography-tandem mass spectrometry and evaluation of the excretion after oral intake, *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **2005**, 827, 146-156.

[48]. Schwemmer, M., Fink, B., Köckerbauer, R. and Bassenge, E. How urine analysis reflects oxidative stress — nitrotyrosine as a potential marker. *Clinica Chimica Acta.* **2000**, 297, 207-216.

[49]. Wu, L.L., Chiou, C.C., Chang, P.Y. and Wu, J.T. Urinary 8-*OHdG*: A marker of oxidative stress to DNA and a risk factor for cancer, atherosclerosis and diabetics. *Clinica Chimica Acta.* **2004**, 339, 1-9.

[50]. Kamalakkannan, N., Rukkumani, R., Varma, P.S., Viswanathan, P., Rajasekharan, K.N. and Menon, V.P. Comparative effects of curcumin and an analogue of curcumin in carbon tetrachloride-induced hepatotoxicity in rats. *Basic Clin. Pharmacol. Toxicol.* **2005**, 97, 15-21.

[51]. Jagatha, B., Mythri, R.B., Vali, S. and Bharath, M.S. Curcumin treatment alleviates the effects of glutathione depletion in vitro and in vivo: Therapeutic implications for Parkinson's disease explained via in silico studies. *Free Radic. Biol. Med.* **2008**, 44, 907-917.

[52]. Biswas, S.K., McClure, D., Jimenez, L.A., Megson, I.L. and Rahman, I. Curcumin induces glutathione biosynthesis and inhibits NF-κB activation and interleukin-8 release in alveolar epithelial cells: mechanism of free radical scavenging activity. *Antioxid. Redox Signal*. **2005**, 7, 32-41.

[53]. Venkatesan, N., Punithavathi, D. and Arumugam, V. Curcumin prevents adriamycin nephrotoxicity in rats. *Br. J. Pharmacol.*, **2000**, 129(2), 231-234.

[54]. Yang, K.Y., Lin, L.C., Tseng, T.Y., Wang, S.C. and Tsai, T.H. Oral bioavailability of curcumin in rat and the herbal analysis from Curcuma longa by LC–MS/MS. *J. Chromatogr. B*, **2007**, 853(1), 183-189.

[55]. Awasthi, S., Pandya, U., Singhal, S.S., Lin, J.T., Thiviyanathan, V., Seifert, W.E., Awasthi, Y.C. and Ansari, G.A.S. Curcumin–glutathione interactions and the role of human glutathione S-transferase P1-1. *Chem. Biol. Interact.* **2000**, 128(1), 19-38.

Chapter 3: Metabolic effects of *Polygonum cuspidatum* supplementation on urinary composition of healthy rats: a longitudinal study.

3.1 Introduction

Resveratrol (3,5,4'-trihydroxystilbene), is a simple polyphenolic compound that in the past twenty years has received widespread attention due to its supposed beneficial health effects. It is found in grapes, berries and peanuts [1] and in various plants, such as in *Vitis vinifera* and in *Polygonum cuspidatum* (Japanese knotweed). In traditional Chinese and Japanese medicine, *P. cuspidatum* has been used to treat several diseases such as hyperlipidemia, inflammation, infections and cancer [2]. *P. cuspidatum* is considered one of the best sources of resveratrol because it is a world-diffused invasive plant and it contains higher amounts of this compound than other plants or fruits [3]. This plant is also used as dried extract in the formulation of food supplements and herbal medicines, because of its content in resveratrol.

Numerous resveratrol derivatives, such as piceid (resveratrol-3-O-β-glucoside) and other glucopyranoside-conjugates [2] have been detected in many botanical sources, but until now, only few studies have focused on their possible biological effects.

Scientific interest in resveratrol has grown starting from the late 1990s, when it was first demonstrated to prevent carcinogenesis in mice [4]. Since then, its potential health promoting effects have been studied through numerous *in vitro* and *in vivo* experiments. A large number of pharmacological targets were discovered and numerous health benefits and disease-preventing activities, such as chemopreventive [5] and anti-inflammatory [6] ones, were related to resveratrol administration in animal models. Furthermore, many papers have considered resveratrol as a possible treatment or a preventing-agent with usefulness in cancer, cardiovascular disease, ischaemic injuries, as well as a possible enhancer of stress resistance. Finally, some papers considered this compound for its effects on the lifespan extension of various organisms from yeast to vertebrates [7-8]. Because of its antioxidant properties and its caloric restriction mimetic role [8], a prevention role in many age-related and metabolic diseases has been attributed to resveratrol, most importantly in obesity-related disorders such as diabetes and cardiovascular diseases [9-10].

The mechanism by which resveratrol exerts such a range of beneficial effects is not yet clear. There are numerous papers describing *in vitro* and potential *in vivo* resveratrol effects, but doubts on its effectiveness are present because of its low bioavailability. Resveratrol is absorbed by the intestine but it is rapidly metabolized both in the intestinal lumen and in the liver in more hydrophilic derivatives that are then excreted in the urine as glucuronides and sulfates. Moreover, gut microbiota transform resveratrol in dihydroresveratrol by reducing the molecule's double bond and also this metabolite is absorbed by the intestine and further metabolized and conjugated [11]. Despite extremely low bioavailability and rapid clearance from the circulation, the evidence of health promoting effects of resveratrol in the prevention or delay of cardiovascular, metabolic and inflammation related diseases is growing [10].

One explanation can be that resveratrol may exert such different biological activities through multiple targets and multiple molecular mechanisms. Since several biochemical pathways are potentially affected, metabolomic approaches can provide unexplored information in the study of the bioactivity of natural products. The combined use of NMR and mass spectrometry coupled to liquid chromatography is a fast, effective and convenient tool to elucidate the alterations in metabolic pathways in biological fluids under different physiological conditions. Although mass spectrometry offers better sensitivity and better developed protocols, the use of NMR for the metabolic profiling of cells, tissues and biological fluids has grown extensively in the recent years [12].

The analysis of urinary metabolic changes can provide information on the effects of food supplements or health promoting products on healthy subjects or animal models. Specimen collection is non-invasive and long-term experiments can be easily conducted; some urinary biomarkers of oxidative stress can be measured offering the opportunity to evaluate the redox status of the considered organism [13]. Sample preparation is more straightforward in the case of urine than with other biological fluids (plasma for example) because of lower sample complexity and lower protein/peptide content [14]. Finally, urine samples the metabolic end-products from the organism destined for excretion; therefore, compared to blood, it is prone to higher biological variations [14]. Thus, the entire urine metabolome can be described by an untargeted approach and the subtle differences between control and treated groups can be highlighted by multivariate data analysis methods.

56

Food supplements and so called "nutraceutical" products are used by people with the aim to maintain or improve their state of health; nevertheless, scarce information is available in the published literature about the effects of these products on healthy animal models or subjects. In the case of resveratrol, many studies are available on its activity on pathological animal models, such as cancerous or diabetic rats or animals fed with a high-fat-diet [15-18], but no published papers deal with urine metabolomics of healthy animal models supplemented with *P. cuspidatum* extract.

In the proceeding of metabolomics-based works on natural products performed by Dall'Acqua and collaborators [19-20], with the aim to evaluate the possible effects of *P. cuspidatum* on a healthy animal model, we supplemented rats with a *P. cuspidatum* extract (100 mg/kg containing 20% of resveratrol) for 49 days. The variations of urinary composition were studied using ^{1}H-NMR, UPLC-HRMS and multivariate statistical analysis. Specific attention was devoted to two oxidative stress urinary biomarkers, namely allantoin and 8-hydroxydeoxyguanosine (*8-OHdG*).

3.2 Experimental

3.2.1 Materials

P. cuspidatum extract was purchased from a local market. The resveratrol content was measured by HPLC-DAD and HPLC-MS analysis and the amount found was $20.0 \pm 0.1\%$ w/w. Allantoin, 8-OHdG and resveratrol standards were purchased from Sigma Aldrich. HPLC-grade acetonitrile and formic acid were purchased from Sigma Aldrich. Deionized water used in HPLC and UPLC analysis was filtered through a Milli-Q system equipped with a 0.22 μm cut-off filter (Millipore).

3.2.2 Experimental Design

A longitudinal design was implemented to evaluate the effects of *P. cuspidatum* supplementation on the metabolite content of the urine collected during 49 days of treatment of 2 groups of rats. The treated group received, orally by gavage, a daily dose of 100 mg/kg of *P. cuspidatum* extract suspended in water while an equal dose of water was given to the control group. Rat urine samples were collected at day 0, 7, 14, 21, 28, 35 and 49 during the experiment and they were subjected to UPLC-HRMS and ^{1}H-NMR analysis. The *P. cuspidatum* dose (100 mg/kg) was selected on the basis of the resveratrol content ($20.0 \pm 0.1\%$ w/w corresponding to 20 mg/kg resveratrol), and considering previous published studies on rats showing both efficacy [21-23]

and safety of such oral dose [21].

All animal procedures were approved and conformed to the directives of the Ethical Committee for animal experiments of the University of Padova (CEASA). Six male and six female Sprague-Dawley rats, 6 ±2 weeks of age and weighing 108.0 ± 24.4 g, where used for the study. Animals were fed standard laboratory diets and water ad libitum and were maintained in a temperature- and photoperiod- controlled room (12-h light/dark cycle). Rats were divided randomly in a *P. cuspidatum* treated group (three males and three females) and a control group (three males and three females). No differences were observed between control and treated groups at the beginning of the experiment on the basis of a preliminary metabolomics investigation.

Rats were housed individually in metabolic cages for the collection of the 24-h urine outputs. Urine samples were then stored at -80 °C to avoid chemical degradation.

3.2.3 UPLC-HRMS Urine Analysis

The UPLC-HRMS analysis was conducted by the group of Prof. Stefano Dall'Acqua at the Department of Pharmaceutical Sciences, University of Padova. To obtain a metabolic profile of urine, a UPLC-HRMS full-scan method was used. An Agilent 1290 Infinity UPLC system equipped with a Waters Xevo G2 Q-TOF mass spectrometer was employed. The detector was equipped with an electrospray (ESI) ionisation source and was operating in positive resolution mode. The sampling cone voltage was adjusted at 40 V, the source offset at 80 V. The capillary voltage was adjusted to 1500 V. The nebulizer gas used was N_2 at a flow rate of 800 L/h. The desolvation temperature was 450 °C. The mass accuracy and reproducibility were maintained by infusing lockmass (leucine–enkephalin, *m/z* 556.2771) thorough Lockspray at a flow rate of 20 µL/min. Centroided data were collected for each sample in the mass range 50 − 1200 Da, and the *m/z* value of all acquired spectra was automatically corrected during acquisition based on lockmass. An Agilent XDB C-8 column (2.1 mm x 150 mm, 3.5 µm) was used as stationary phase. The mobile phase was composed of solvent A (acetonitrile with 0.1% formic acid) and solvent B (water with 0.1% formic acid). Linear gradients of solvents A and B were used, as follows: 0 min, 8% A; 14 min, 48% A; 16 min, 100% A; 17 min, 100% A; 17.5 min, 8% A; 24 min, 8% A. The flow rate was 200 µL/min and the injection volume was 1 µL.

Collected urine samples were centrifuged at 13.000 rpm for 3 minutes prior to analysis and

directly injected in the UPLC.

Centroided and integrated chromatographic mass data were processed by MarkerLynx Applications Manager Version 4.1 (Waters) to generate a multivariate data matrix. An appropriate method for data deconvolution, alignment and peak detection was created and the data were subsequently elaborated by the software. The parameters used were retention-time range 2.00–17.00 min, mass range 100–650 Da, mass tolerance 0.01 Da, noise elimination level was set to 10.00, minimum intensity was set to 15% of base peak intensity, maximum masses per RT was set to 6 and, finally, RT tolerance was set at 0.01 min. Isotopic peaks were excluded from analysis. A list of the ion intensities of each peak detected was generated, using retention time and the *m/z* data pairs as the identifier for each ion. The resulting three-dimensional matrix contains arbitrarily assigned peak index (retention time-*m/z* pairs), sample names (observations), and ion intensity information (variables). After the exclusion of the variables having more than 30% of missing data in both groups under investigation and log-transformation, the obtained data set composed of 483 time x mass variables was normalized by Median Fold Change normalization and mean centred.

3.2.4 ^{1}H-NMR Urine Analysis

Sample preparation was performed as described elsewhere [24], with minor modifications. Frozen urine samples were thawed in a fuming hood for 30 min and then 1 mL urine was transferred into a 1.5 mL eppendorf vial. The pH value was adjusted to $1.2 \leq pH \leq 2$ by adding 3 M HCl (120 ± 80 μL) and the sample was centrifuged at 13,000 rpm for 10 min. To 700 μL of supernatant, 70 μL of 2 mM sodium 3-trimethylsilylpropionate-2,2,3,3,-d4 (TSP) in D_2O were added. Finally, 700 μL were transferred into an NMR tube and the pH was monitored by inspection of the chemical shift of the citrate signal and sometimes adjusted by adding 5 μL of 1 M NaOH or HCl. 2D-experiments were performed on a urine sample three times more concentrated, obtained by pooling several samples and partially evaporating the solvent.

All ^{1}H-NMR spectra were acquired on a Bruker Avance DMX600 MHz spectrometer equipped with a 5 mm TXI xyz gradient inverse probe at room temperature. The probe was tuned, matched and shimmed manually for each sample. The NOESYGPPR1D pulse sequence was used to acquire one-dimensional ^{1}H-NMR spectra. Parameters used were: 64 scans; 32k data points; spectral width, 8389.26 Hz; relaxation delay, 2 s; mixing time, 50 ms; acquisition time, 1.95 s.

^{1}H-^{1}H NOESY spectra were recorded with the NOESYGPPRTP pulse sequence, with a spectral window of 13 ppm in both dimensions, 2048×512 data points, 1.2 s relaxation delay, 0.6 s mixing time, and 256 scans.

Prior to Fourier transformation, the FIDs were zero-filled to 64K points and an exponential line broadening factor of 0.3 Hz was applied. All spectra were processed using ACD/NMR Manager 12.01 (Advanced Chemistry Development Inc.). Spectra were referenced to 0.0 ppm using the resonance of TSP and manually corrected for phase and baseline distortions. Intelligent bucketing was applied to the region between 0.6 ppm and 9.5 ppm excluding the regions where resveratrol derivatives (6.24-7.50 ppm, 5.10-5.22 ppm, 4.10-4.26 ppm, 3.56-3.96 ppm, 2.74-2.82 ppm) and water (4.70-4.95 ppm) resonate. Total sum normalization was applied to compensate for differences in overall concentration between individual urine samples. The obtained data set composed of 229 variables was mean centred and Pareto scaled.

3.2.5 Statistical Data Analysis

Data modelling was performed by applying multivariate techniques based on projection. Specifically, Principal Component Analysis (PCA) was used for exploratory data analysis and for highlighting the presence of outliers while Projection to Latent Structures by partial least squares regression (PLS) was applied to investigate the relationships between the metabolic content of the urines collected during the experiment and the time of treatment and type of diet. To perform an efficacy data modelling able to extract the whole information contained into the collected data, a design matrix supporting an interaction model including time, treatment and time x treatment effects, was explicitly considered. The measured variables and the design matrix were used as X-block and Y-block, respectively, in the PLS regression. Following good practice for model validation, N-fold full cross-validation with different values of N (N = 6, 7, 8) and permutation test on the responses (1000 random permutations) were performed, in order to avoid over-fitting and prove the robustness of the obtained models. The significance of the effects included in the design matrix was estimated by permutation test. The number of components of the PLS models was estimated on the basis of the first maximum of Q^2 calculated by 7-fold full cross-validation ($Q^2_{7\text{-fold CV}}$) under the constraint to pass the permutation test on the responses. To better interpret the obtained model, a new method for rotating the PLS model, called post-

60

transformation of PLS2 (ptPLS2), was applied [25]. The main advantage to use the post-transformation of the PLS model is the possibility to obtain a new model where the structured variation of the X-block discovered by PLS is decomposed into two blocks, corresponding to the variations correlated (predictive part of the model) and orthogonal (non-predictive part) to the Y-block. The post-transformed model maintains the same power in prediction of the PLS model but simplifies model interpretation because the dimension of the predictive latent space is usually lower than that of the whole latent space discovered by PLS.

Recently, stability selection was introduced in metabolomics to avoid false discoveries due to overtraining [26]. The central idea of stability selection is that real differences or effects should be present consistently, and therefore should be found even under perturbation of the data by subsampling or bootstrapping. In our study, we performed stability selection by Monte-Carlo subsampling using PLS VIP-based as regression technique [27] for identifying putative markers. Specifically, 200 random subsamples were extracted by Monte Carlo sampling of the collected urine samples (with prior probability of 0.70), and then PLS VIP-based applied to each subsample, obtaining a set of 200 regression models. Within this set of PLS VIP-based models, variables related to the effects of time and/or treatment were identified as the most frequently selected variables. The advantage to use PLS VIP-based is that the VIP selection is able to extract a reduced subset of variables for each subsample and the most frequently used variables can be selected by calculating the median of the VIP for each variable in all sub models excluding the variables having median equal to zero. The threshold of the VIP score to use in variable selection within PLS VIP-based was estimated by maximizing the $Q^2_{\text{7-fold CV}}$. As a result, a small number of metabolites changing during the experiment was extracted and the behaviour of each single metabolite was studied by linear mixed-effects model for longitudinal studies [28]. In this step of our data analysis, the covariance structure was modelled by considering a first-order autoregressive structure with homogenous variances.

PCA was performed using SIMCA 13 (Umetrics, Umea, Sweden) while the platform R 3.0.2 (R Foundation for Statistical Computing) was used for statistical data analysis on single variable, to build the ptPLS2 model and to perform Monte-Carlo stability selection.

The results of the levels of oxidative stress urinary markers are expressed as means ± Standard Error of the Mean (SEM). Statistical evaluation and p calculation were obtained using GraphPad

Prism 5 software (San Diego California).

3.2.6 Urinary Allantoin and 8-Hydroxydeoxyguanosine Quantification

Allantoin and 8-OHdG were quantified in urine samples from both treated and control rats using HPLC-MS/MS methods. A Varian 212 HPLC system equipped with a Varian MS 500 IT detector was used for HPLC-ion trap mass spectrometry. A Phenomenex Kinetex EVO C-18 column (3 mm x 100 mm, 5 μm) was used as stationary phase. The detector was equipped with an ESI source and was operating in positive resolution mode. For allantoin quantification, the operating parameters used were as follows: nebulizing gas, nitrogen; nebulizing pressure, 35.0 psi; needle voltage, 4500 V; capillary voltage, 600.0 V; drying gas temperature, 350 °C; drying gas pressure, 10.0 psi. The mobile phase was composed of solvent A (acetonitrile with 0.1% formic acid) and solvent B (water with 0.1% formic acid). Linear gradients of A and B were used, as follows: 0 min, 15% A; 7 min, 100% A; 7.06 min, 15% A; 10 min, 15% A. The flow rate was 200 μL/min and the injection volume was 10 μL. An MS^n experiment was performed, monitoring the fragmentation of allantoin precursor ion ($[M+H]^+$, m/z 159) in the product ion mass range 65-169 Da. Allantoin was finally quantified on the basis of its major fragment (m/z 116) using a standard titration curve obtained by eluting 0.3 - 1 μg/mL allantoin solutions in water (y = 97216x + 51629; R^2 = 0.9833). LOD and LOQ were 0.1 and 0.5 ng/mL, respectively.

For 8-OHdG quantification, the operating parameters used were as follows: nebulizing gas, nitrogen; nebulizing pressure, 20.0 psi; needle voltage, 4500 V; capillary voltage, 600.0 V; drying gas temperature, 280 °C; drying gas pressure, 15.0 psi. The mobile phase was composed of solvent A (acetonitrile with 0.5% formic acid) and solvent B (water with 0.1% formic acid). Linear gradients of A and B were used, as follows: 0 min, 10% A; 1 min, 10% A; 10 min, 60% A; 12 min, 10% A; 16 min, 10% A. The flow rate was 200 μL/min and the injection volume was 10 μL. An MS^n experiment was performed, monitoring the fragmentation of 8-OHdG precursor ion ($[M+H]^+$, m/z 284) in the product ion mass range 155-175 Da. 8-OHdG was finally quantified on the basis of its major fragment (m/z 168) using a standard titration curve obtained by eluting 3 - 100 ng/mL 8-OHdG solutions in water (y = 48.853x + 1002.5; R^2 = 0.9981). LOD and LOQ were 1 and 5 ng/mL, respectively.

3.3 Results

3.3.1 Animal Weight and Urinary Output

No significant differences were observed either in body weight or in 24-h urine volume in the treated *vs* control group during the experiment. Data reported in **Table 3.1** show body weights and urinary volumes during the experiment.

	Body weight (g)		Urine volume (mL)	
Day	Control	Treated	Control	Treated
0	108.5±	107.5±	5.5±0.7	6.4±1.4
7	162.8±	164.8±	9.3±2.8	11.4±4.6
14	205.8±	208.6±	12.3±3.7	10.1±1.8
21	250.3±	248.0±	12.2±2.3	13.5±3.1
28	283.8±	287.7±	10.8±1.4	13.4±1.8
35	309.0±	303.5±	13.8±3.8	12.7±1.9
49	332.0±	340.3±	15.8±8.0	12.9±1.9

Table 3.1. Variations of body weight and 24-h urine volume of control and treated rats during the course of the experiment.

3.3.2 UPLC-HRMS Measurements on 24-h Collected Urines

Twenty four hour urine samples collected during the experiment were analysed by UPLC coupled with ESI-QTOF-MS to obtain urinary metabolite profiles of both treated and control animals. Exemplificative chromatograms are reported in **Figure 3.1.** Exploratory data analysis by PCA revealed no presence of outliers while PLS-DA did not detect differences in the metabolite content of the urines at day 0 between the two groups of rats.
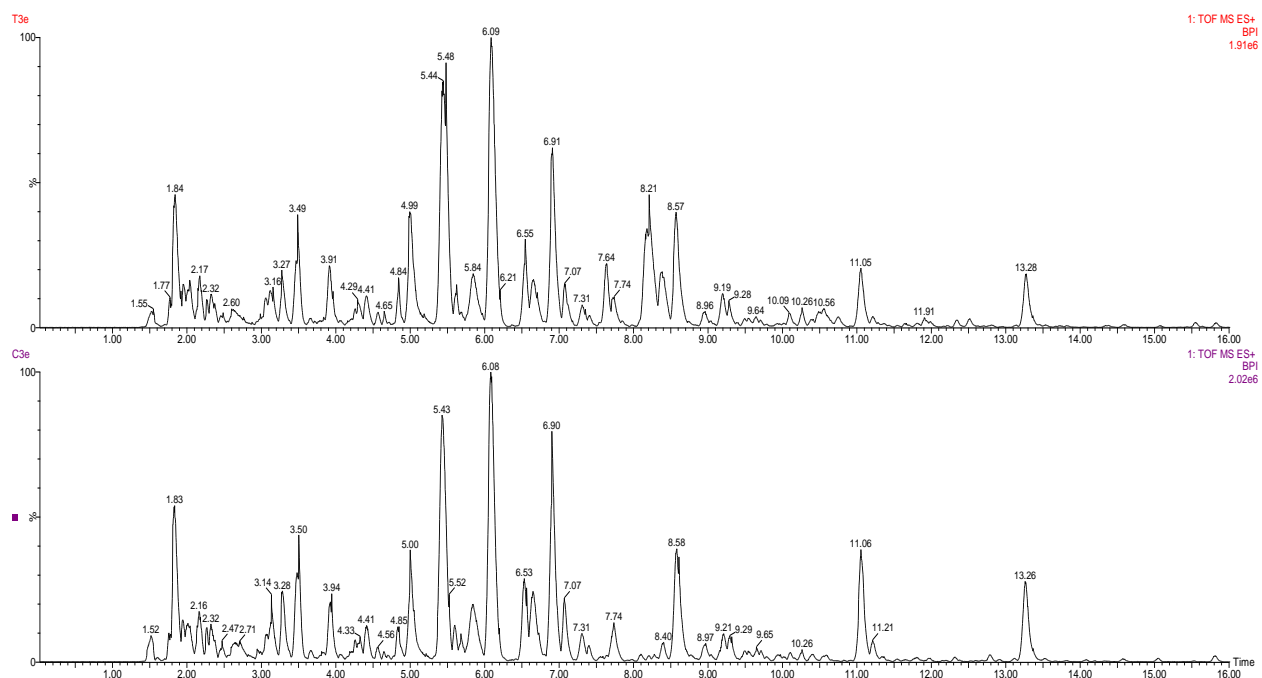
Figure 3.1. Typical base peak intensity (BPI) chromatograms obtained from UPLC-HRMS analysis of rat urines from both treated (upper panel) and control rats (lower panel).

As a first step of data analysis, we searched for expected metabolites directly related to the resveratrol metabolism in the treated group. We found resveratrol phase II metabolites, namely resveratrol-glucuronide (*m/z* 405.1180 and the fragment at *m/z* 229.0866), resveratrol-sulfate (*m/z* 309.1314 and the fragment at *m/z* 229.0866) and dihydroresveratrol glucuronide (*m/z* 407.1344 and the fragment at *m/z* 231.1022), a metabolite derived from the action of gut microbiota and a phase II metabolic processing in the liver [29] (Table 3.2). No significant traces of un-metabolized resveratrol was detected in rat urines.

Molecular identity	Retention time (min)	m/z value	Adduct ion type
Resveratrol-glucuronide	8.19	405.1180	[M+H]$^+$
Resveratrol-sulfate	8.83	309.1314	[M+H]$^+$
Dihydroresveratrol-glucuronide	8.38	407.1344	[M+H]$^+$

Table 3.2. Identified resveratrol metabolites obtained by the analysis of UPLC-HRMS data of treated animals. The reported m/z values are from experimental data.

After the exclusion of the time x mass variables directly related to the metabolism of resveratrol, data modelling was performed by ptPLS2. The obtained model highlighted a clear effect of both time and treatment on the urine composition. The interaction term time x treatment was insignificant ($Q^2_{7\text{-fold}}$ CV = 0.003, p-value = 0.29). The model showed 2 parallel and 3 orthogonal components. Specifically, the model showed R^2 = 0.89 and $Q^2_{7\text{-fold}}$ CV = 0.76 (p-value<0.001) for time while for the treatment resulted R^2 = 0.89 and $Q^2_{7\text{-fold}}$ CV = 0.68 (p-value<0.001). **Figure 3.2** reports the score scatter plot of the model, where it is possible to observe that the evolution of the urine composition during the experiment can be explained by considering two main effects, one related to the time passing and one related to the dietary intervention.
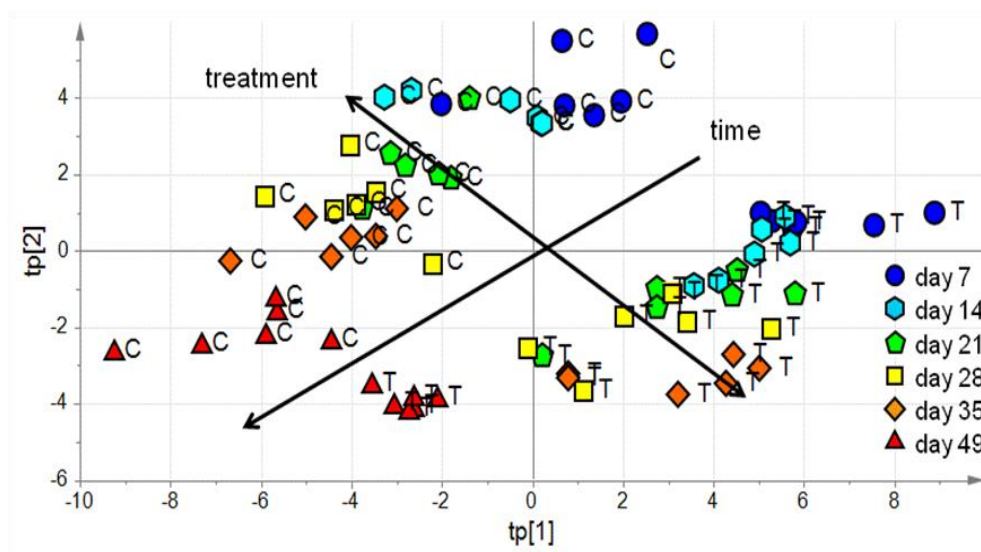
Figure 3.2. Score scatter plot of the ptPLS2 model for UPLC-HRMS data. The plot shows modifications of the urine composition both due to treatment and to aging. Symbols of different shapes represent metabolic changes in urine along time.

By Monte-Carlo stability selection, a reduced set of 34 measured variables was detected as relevant in the explanation of the effects of time and treatment in the metabolite content of the collected urines. Chemical characterization of the selected time x mass variables was tentatively performed by comparing the exact *m/z* value and fragmentation patterns with web-available databases (Human Metabolome Database, Metlin Database, Mass Bank Database) and interpreting their MS^e spectra.

A linear mixed-effects model considering first-order autoregressive structure with homogenous variances was applied to quantify the effects of time and treatment on the selected variables. As a

result, hypoxanthine, indole-3-carboxylic acid, 6-hydroxymelatonin, tetrahydrocortisol, 17-beta-methylestra-1,3,5(10)-trien-3-ol, sebacic acid, 3-oxo-5beta-chol-6-en-24-oic acid, 11-dehydrocorticosterone and 19-hydroxytestosterone predominantly contributed to the observed urinary metabolite differences due to *P. cuspidatum* extract supplementation as summarized in **Table 3.3** while the most significant variables related to aging are reported in **Table 3.4**.

Retention Time (min)	HR *m/z* value	Identification	Treated vs. Control
2.26	137.0460	Hypoxanthine	↓
6.92	162.0554	Indole-3-carboxylic acid	↓
8.58	249.1242	6-Hydroxymelatonin	↑
11.17	271.2060	17-beta-methylestra-1,3,5(10)-trien-3-ol	↓
11.17	733.4891	Tetrahydrocortisol ($[2M+H]^+$)	↓
11.34	225.1102	Sebacic acid ($[M+Na]^+$)	↓
12.77	355.2634	3-oxo-5beta-chol-6-en-24-oic acid ($[M+H-H_2O]^+$)	↑
12.80	345.2064	11-Dehydrocorticosterone	↓
13.17	305.2116	19-Hydroxytestosterone	↓

Table 3.3. Significant variables (p-value<0.01) related to treatment for UPLC-HRMS data set. Variations of the metabolites are reported as increase (↑) or decrease (↓) in treated compared to control rats.

Retention Time (min)	HR m/z value	Identification
7.06	242.1757	(±)-Hexanoylcarnitine ($[M+H-H_2O]^+$)
5.80	260.1501	Unknown
13.34	335.2376	Unknown
13.39	410.2693	Unknown
8.85	409.1832	Unknown
4.45	318.1555	Isoleucyl-Tryptophan ($[M+H]^+$)
13.50	448.3060	Unknown
10.53	435.1285	Unknown
13.17	359.1929	Unknown
4.65	372.2385	Unknown
3.08	413.1239	Unknown
13.50	488.2987	Glycocholic acid ($[M+Na]^+$)

Table 3.4. Significant variables related to rat aging, obtained from the ptPLS-DA VIP-based model for UPLC-HRMS data.

3.3.3 ^{1}H-NMR Measurements on 24-h Collected Urines

In the 1D spectra (**Figure 3.3**) of treated rats, the resonances of resveratrol derivatives are evident. The NOESY spectrum confirms the presence of resveratrol-glucuronide and dihydroresveratrol glucuronide as shown in **Figure 3.4**.
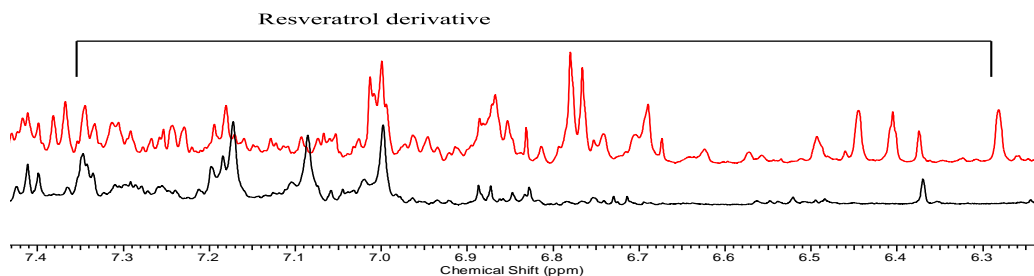
Figure 3.3: Representative urine ^{1}H-NMR spectra of a control rat and of a rat treated with *P. cuspidatum.* In the aromatic region (6.20 – 7.40 ppm), the presence of resveratrol derivatives in the treated (red) spectrum is evident.
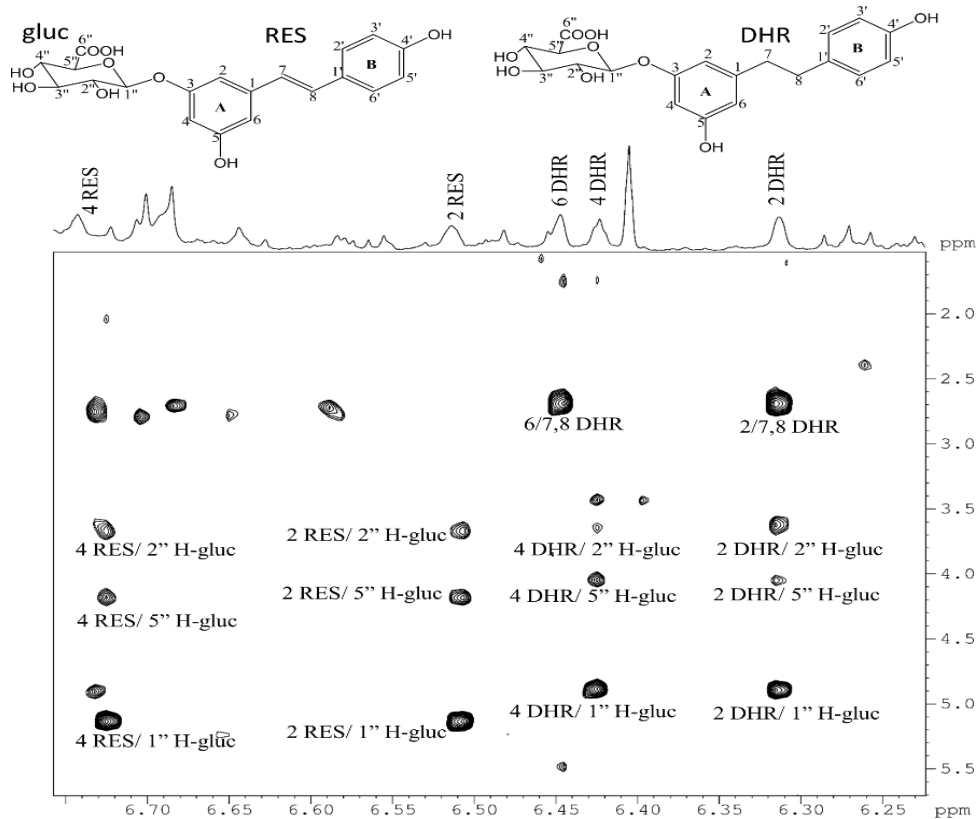
Figure 3.4. Portion of a NOESY spectrum in which NOE correlations between resveratrol (RES)

67

or dihydroreveratrol (DHR) aromatic protons with protons from glucuronic acid are shown.

Pattern recognition was performed by PCA in order to highlight outliers or differences in the metabolite content of the urines collected at the beginning of the experiment. No outliers were detected or differences at day 0 related to the groups under investigation. Data modelling performed by ptPLS2 highlighted a clear effect of both time and treatment on the urine composition. The interaction term time x treatment was insignificant ($Q^2_{7\text{-fold CV}}$ = -0.02, p-value = 0.14). The model showed 2 parallel and 1 orthogonal components, R^2 = 0.87 and $Q^2_{7\text{-fold CV}}$ = 0.81 (p-value<0.001) for time and R^2 = 0.60 and $Q^2_{7\text{-fold CV}}$ = 0.28 (p-value<0.001) for treatment. **Figure 3.5** reports the score scatter plot of the model. It is possible to observe that two main effects, one related to the time passing and one related to the dietary intervention, to determine the evolution of the urine composition during the experiment.
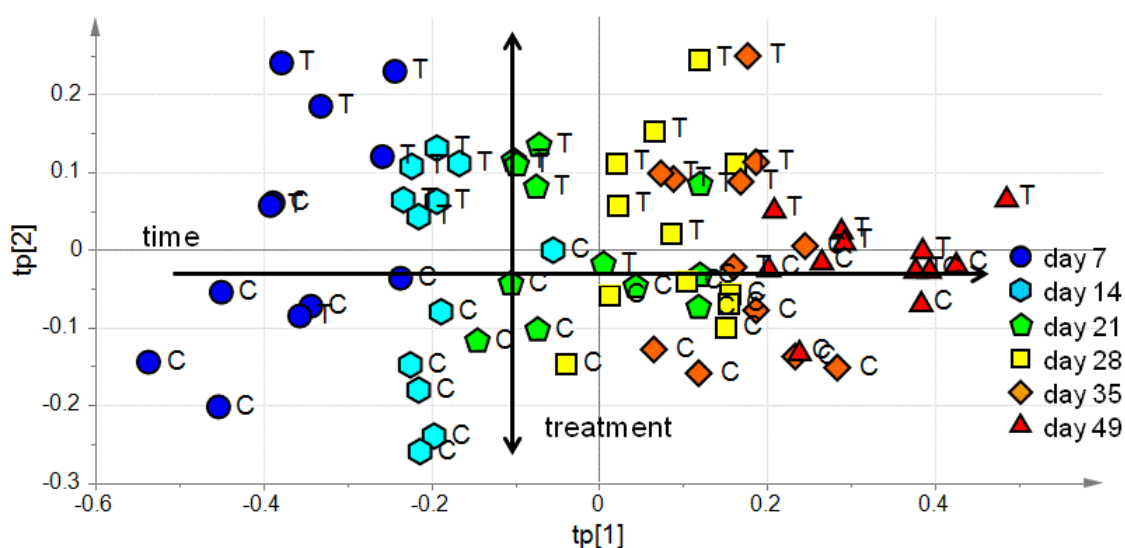
Figure 3.5. Score scatter plot of the ptPLS2 model for ^{1}H-NMR data set. The plot shows modifications of the urine composition both due to treatment and to aging. Symbols of different shapes represent metabolic changing in urines along time.

Monte-Carlo stability selection allowed us to select a reduced set of 21 bins. **Table 3.5** reports the assignment of the selected bins related to the treatment effect while **Table 3.6** that of the bins explaining the aging effect.

Some metabolites were identified with the help of an online database (HMDB), literature [30]

and 2D-NMR spectroscopy. The metabolites affected by the treatment are 3-hydroxybutryate (3-HB), glutamate, glycine, glucose, formate and two unassigned metabolites (Table 3.4). Dicarboxylic acid, lactate, acetate, succinate, 2-oxogluturate (2-OG), citrate, dimethyl glycine (DMG), creatinine, taurine and allantoin significantly change over the time period (Table 3.5). Interestingly, hippurate and betaine changed significantly both with time and treatment. **Figure 3.6** shows a typical ^{1}HNMR spectrum in which metabolites responsible from time and treatment are highlighted.
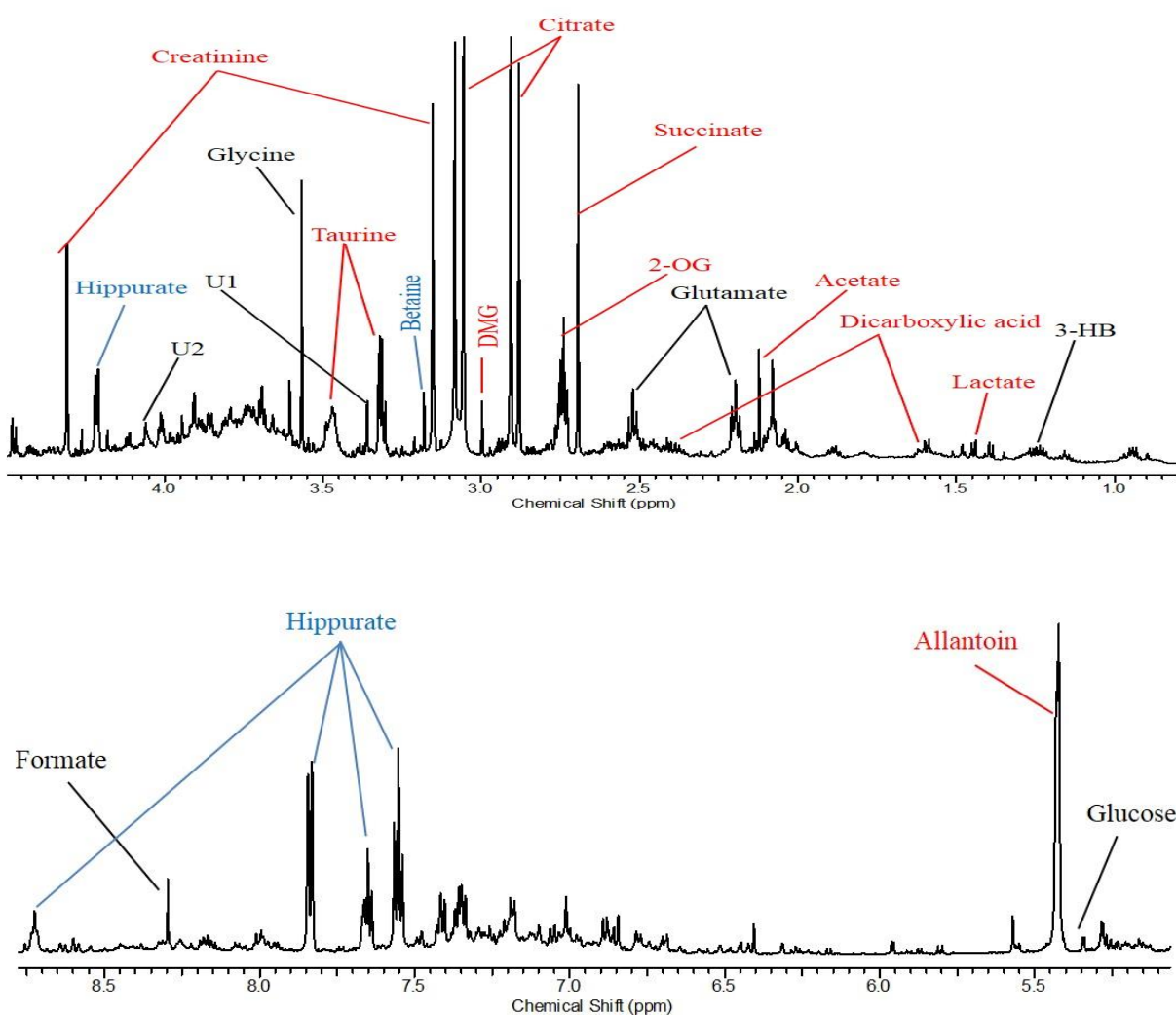
Figure 3.6. Typical 1D ^{1}H-NMR spectrum of rat urine with enlargements of the aliphatic (0.6 – 4.5 ppm, above panel) and aromatic (5.5 – 9.5 ppm, lower panel) regions. The resonances according to the time (red), treatment (black) and both (blue) effects are indicated.

Bin	Chemical shift (ppm)	Metabolite	Treated vs. Control
[1.22- 1.28]	1.23 (d)	3-HB	↑
[2.46- 2.51]	2.48 (m)	Glutamate	↓
[3.20- 3.24]	3.22 (s)	Betaine	↓
[3.30- 3.34]	3.32 (s)	U1	↑
[3.52- 3.56]	3.56 (s)	Glycine	↑
[4.00-4.04]	4.02 (m)	U2	↑
[5.22- 5.26]	5.25 (d)	Glucose	↑
[8.25- 8.29]	8.26 (s)	Formate	↑
[8.75- 8.79]	8.76 (t)	Hippurate	↑

Table 3.5 Significant variables (p-value<0.05) related to treatment effect for the ^{1}H-NMR data set.

Bin	Chemical shift (ppm)	Metabolite	Time effect
[1.40-1.44]	1.42 (d)	Lactate	↑
[2.07-2.11]	2.10 (s)	Acetate	↑
[2.27-2.31]	2.28 (brs)	Dicarboxylic acid	↓
[2.65-2.69]	2.68 (s)	Succinate	↓
[2.69-2.73]	2.71 (t)	2-Oxogluturate	↓
[2.82-2.86]	2.86 (d)	Citrate	↓
[2.94-2.97]	2.95 (s)	Dimethyl glycine	↑
[3.11-3.14]	3.13 (s)	Creatinine	↑
[3.20-3.24]	3.22 (s)	Betaine	↑
[3.25-3.30]	3.28 (t)/ 3.29 (s)	Taurine/TMAO	↑
[3.41-3.46]	3.44 (t)	Taurine	↑
[8.75- 8.79]	8.76 (t)	Hippurate	↓

Table 3.6. Significant variables (p-value<0.01) related to rat aging for the ^{1}H-NMR data set.

3.3.4 Allantoin and 8-OHdG Quantification in Urine Samples

Because of the reported antioxidant activity of resveratrol, we decided to measure the urinary levels of two well-known oxidative stress biomarkers by HPLC-MS/MS, namely allantoin and 8-OHdG as shown in **Figure 3.7**. The total amounts of each biomarker in the collected urines were assessed monitoring the major fragments of allantoin and 8-OHdG at the beginning and at the end of the experiment (day 49) by MS/MS method ([M+H]$^+$ at m/z 159/116 and [M+H]$^+$ at m/z 284/168 respectively). Data are summarised in **Table 3.7**. The content of 8-OHdG in the urines weakly decreases for the treated rats (t-test p-value = 0.05) while it does not change for the

70

control rats (t-test p-value = 0.47). At the beginning of the experiment, urinary 8-OHdG concentration was 10.83 ± 0.54 ng/mL for the treated rats and it weakly decreased to 8.64 ± 0.92 ng/mL at the 49th day of treatment. Allantoin concentration significantly decreased for both control and treated rats (t-test p-value = 0.02 and test p-value<0.001, respectively). Specifically, for control rats the allantoin concentration decreased from 1.75 ± 0.08 µg/mL to 1.01 ± 0.18 µg/mL while for treated rats from 1.54 ± 0.06 µg/mL to 0.41 ± 0.05 µg/mL. We can conclude that the reduction of the concentration of these two markers was more relevant for treated rats than control rats suggesting a reduction of oxidative stress in the treated group.
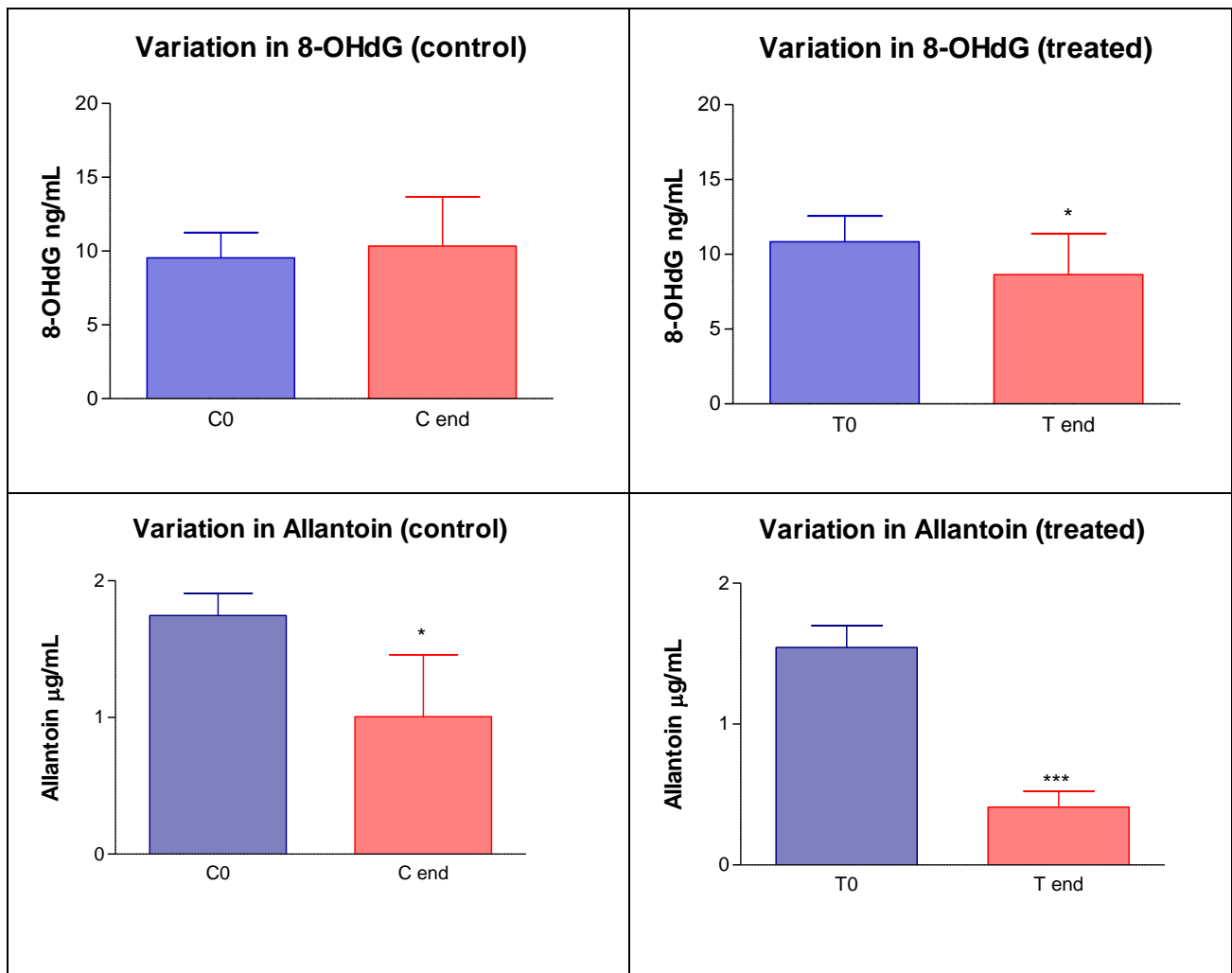
Figure 3.7. Urinary variations of 8-*OHdG* and allantoin from the beginning of the experiment to the end. Results obtained from rats supplemented with *P. cuspidatum* extract are compared to results obtained from the control group. Values are represented as means ± SEM. p-values calculated in unpaired t test. p< 0.05 (*), p < 0.001 (***).

8-OHdG	Day 0	Day 49	t-test p-value
Control	9.54 ± 0.49 ng/mL	10.35 ± 1.05 ng/mL	0.47
Treated	10.83 ± 0.54 ng/mL	8.64 ± 0.92 ng/mL	0.05*
Allantoin	**Day 0**	**Day 49**	**t-test p-value**
Control	1.75 ± 0.08 µg/mL	1.01 ± 0.18 µg/mL	0.02*
Treated	1.54 ± 0.06 µg/mL	0.41 ± 0.05 µg/mL	<0.0001***

Table 3.7. Urinary variations of 8-*OHdG* and allantoin from the beginning of the experiment to the end. Results obtained from rats supplemented with *P. cuspidatum* extract (treated) are compared to results obtained from the control group (control). Values are represented as means ± SEM. p-values calculated in unpaired t test. $p < 0.05$ (*), $p < 0.001$ (***).

3.4 Discussion

P. cuspidatum extract (100 mg/kg) containing 20% of resveratrol was administered to rats to study the effect of this natural product in a healthy animal model through the observation of metabolic urinary changes. The underlying physiological changes responsible for these results are currently under examination. Below, a partial analysis is presented.

Urines from treated animals showed decreased levels of hypoxantine and indole-3-carboxylic acid (ICA). Hypoxanthine is normally detected in urines and blood, but increased levels could be related to oxidative stress [31-32]. Also increased urinary levels of indole-3-carboxylic acid (ICA), a tryptophan metabolite normally detected in urines [33], can be related to oxidative stress and were reported in diet-induced hyperlipidemic rats [34].

Thus, our results suggest that *P. cuspidatum* extract supplementation in healthy rats reduces oxidative stress. This observation was confirmed by specific measurement of two urinary markers of oxidative stress. Allantoin is formed from the stepwise degradation of uric acid due to the reaction with oxidative species [35]. Allantoin is present also in urine of healthy subjects and its level is increased with increasing oxidative stress in the body [36]. 8-OHdG is a product of mitochondrial and nuclear DNA oxidative damage formed by the 8-hydroxylation of the guanine base by reactive oxygen species [37]. The oxidized nucleoside is enzymatically cleaved from the

DNA chains, is released in the bloodstream, and is eventually eliminated in the urines [37]. This biomarker is present in urines from healthy subjects, and variations of its urinary levels suggest alterations of the oxidative status. Thus, the observed reduction of both these metabolites in treated groups support an *in vivo* antioxidant effect of *P. cuspidatum* supplementation.

The treated group showed increased levels of urinary 6-hydroxymelatonin, a hydroxylated hepatic metabolite of melatonin, mainly excreted in urine. Previous studies reported that melatonin and its metabolites exert both immunomodulatory and antioxidant activities [38-39]. Its meaning in the *P. cuspidatum* mode of action need to be deeply investigated.

Different metabolites were identified as steroid derivatives. Among all, amounts of tetrahydrocortisol, 17-beta-methylestra-1,3,5(10)-trien-3-ol, 11-dehydrocorticosterone and 19-hydroxytestosterone, a metabolite derived from androgen and estrogen metabolism [40], were detected in lower amount in treated rat urines. A biliary acid derivative was also identified, (3-oxo-5beta-chol-6-en-24-oic acid), and its urinary amount was increased in the treated group.

Tetrahydrocortisol is a urinary metabolite of cortisol [41] and its urinary levels can be increased by ACTH administration [42]. This preliminary observation can suggest a role for *P. cuspidatum* extract as stimulant of corticosurrenal hormones thus being related to anti-stress or "adaptogenic" activity as Ginseng or similar herbal products [43]. The role of resveratrol on cortisol biosynthesis was previously investigated *in vitro* showing that the compound is able to stimulate cortisol biosynthesis by activating SIRT-dependent deacetylation of P450scc [44]. The change in sterol and corticosterol levels may be related to some of the health promoting effects of resveratrol, but these are the data first indicating a role of this natural compound on the levels of these physiological metabolites, and this aspect needs deeper investigations.

Sebacic acid is a medium chain dicarboxylic acid normally detected in urines. It is involved in lipid and carbohydrates metabolism. MS data show that *P. cuspidatum* supplementation decreased its urinary amounts, suggesting a role on energy metabolism in rats. The ^{1}H-NMR data show a significant decrease of a dicarboxylic acid (possibly sebacic or suberic acid) only with time. Dicarboxylic acids such as suberic acid are mainly metabolized in the liver by the ω-oxidation pathway; thus, their increase in urine can indicate an influence of aging and of the treatment on liver function. Specifically, the decrease of the urinary levels of such compounds may indicate a decreased fatty acid ω-oxidation. Previously published studies reported that variations in urinary excretion of dicarboxylic acids and urinary dicarboxylic aciduria are related

to defective fatty acid metabolism [45]. Oddly, our previously published study on curcumin indicated lower levels of dicarboxylic acid in the control group [19]. These results are globally not completely consistent and require further analysis.

Hippuric acid was found to be higher in *P. cuspidatum* treated rats as compared to the control rats and it also decreased with time. Hippuric acid is a normal component of urine and is typically increased with increased consumption of phenolic compounds (tea, wine, fruit juices) [46-47]. The increased urinary excretion of hippurate suggests alterations in gut microflora induced by *P. cuspidatum* because urinary hippuric acid excretion was not detected in volunteers without colon [53].

3.5 Conclusions

Nutraceuticals and supplements containing herbal remedies are mainly used by healthy subjects as disease preventive or health maintaining purposes. Nevertheless, scientific investigations on the effects of such natural products on healthy subjects are largely missing. *P. cuspidatum* is a natural source of resveratrol, a stilbenoid claimed to be a powerful antioxidant used as food supplement or nutraceutical in hyperlipidaemia, inflammation, and age related diseases. In this study, using both ^{1}H-NMR and UPLC-MS measurements, we performed a longitudinal study to assess the urinary metabolomic changes induced by 49 days supplementation of *P. cuspidatum* to healthy rats. This approach may offer a new view of the possible mode of action of herbal remedy supplementation containing resveratrol. Our data showed that resveratrol is subject to extensive metabolism and urinary excretion that can be detected both with MS and ^{1}H-NMR techniques. No oxidized resveratrol or oxidised resveratrol metabolites were detected suggesting a negligible role of its "direct" antioxidant activity. Resveratrol induced changes in the endogenous urinary profile, and the levels of urinary markers of oxidative stress confirmed the *in vivo* antioxidant effect of this supplement. We detected changes that could be ascribed to treatment or to aging. The many changes observed await a complete interpretation in terms of modification of biochemical pathways.

74

3.6 References

[1]. Jasinski, M., Jasinska, L. and Ogrodowczyk, M. Resveratrol in prostate diseases - A short review. *Cent. European J. Urol.* **2013**, 66, 144-149.

[2]. Peng, W., Qin, R., Li, X. and Zhou, H. Botany, phytochemistry, pharmacology, and potential application of Polygonum cuspidatum Sieb.et Zucc.: A review. *J. Ethnopharmacol.* **2013**, 148, 729-745.

[3]. Wang, D. G., Liu, W. Y., and Chen, G. T. A simple method for the isolation and purification of resveratrol from *Polygonum cuspidatum*. *J. Pharm. Ana*. **2013**, 3, 241-247.

[4]. Jang, M., Cai, L., Udeani, G.O. et al. Cancer chemopreventive activity of resveratrol, a natural product derived from grapes. *Science.* **1997**, 275, 218-220.

[5]. Chen, Y., Tseng, S.H., Lai, H.S., and Chen, W.J. Resveratrol-induced cellular apoptosis and cell cycle arrest in neuroblastoma cells and antitumor effects on neuroblastoma in mice. *Surgery*. **2004**, 136, 57-66.

[6]. Tomé-Carneiro, J., Gonzálvez, M., Larrosa, M. et al. One-year consumption of a grape nutraceutical containing resveratrol improves the inflammatory and fibrinolytic status of patients in primary prevention of cardiovascular disease. *Am. J. Cardiol.* **2012**, 110, 356-363.

[7]. Baur, J.A., Pearson, K.J., Price, N.L. et al. Resveratrol improves health and survival of mice on a high-calorie diet. *Nature.* **2006**, 444, 337-342.

[8]. Pearson, K.J., Baur, J.A., Lewis, K.N. et al. Resveratrol delays age-related deterioration and mimics transcriptional aspects of dietary restriction without extending life span. *Cell Metab*. **2008**, 8, 157-168.

[9]. Howitz, K.T., Bitterman, K.J., Cohen, H.Y. et al. Small molecule activators of sirtuins extend Saccharomyces cerevisiae lifespan. *Nature.* **2003**,425, 191-196.

[10]. Baur, J.A. and Sinclair, D.A. Therapeutic potential of resveratrol: The in vivo evidence. *Nat. Rev. Drug Discov.* **2006**, 5, 493-506.

[11]. Juan, M.E., Alfaras, I. and Planas, J.M. Determination of dihydroresveratrol in rat plasma by HPLC. *J Agric Food Chem.* **2010**, 58, 7472-7475.

[12]. Brennan, L. NMR-based metabolomics: From sample preparation to applications in nutrition research. *Prog. Nucl. Magn. Reson. Spectrosc.* **2014**, 83, 42-49.

[13]. Il'yasova, D., Scarbrough, P. and Spasojevic, I. Urinary biomarkers of oxidative status. *Clinica Chimica Acta.* **2012**, 413, 1446-1453.

[14]. Kemperman, R.F.J., Horvatovich, P.L., Hoekman, B., Reijmers, T.H., Muskiet, F.A.J. and Bischoff, R. Comparative urine analysis by liquid chromatography-mass spectrometry and multivariate statistics: Method development, evaluation, and application to proteinuria. *J. Proteome Res.* **2007**, 6, 194-206.

[15]. Seeni, A., Takahashi, S., Takeshita, K. et al. Suppression of prostate cancer growth by resveratol in the transgenic for adenocarcinoma of prostate (TRAP) model. *Asian Pac. J. Cancer Prev.* **2008**, 9, 7-14.

[16]. Kalra, N., Roy, P., Prasad, S. and Shukla, Y. Resveratrol induces apoptosis involving mitochondrial pathways in mouse skin tumorigenesis. *Life Sci.* **2008**, 82, 348-358.

[17]. Su, H.C., Hung, L.M., and Chen, J.K. Resveratrol, a red wine antioxidant, possesses an insulin-like effect in streptozotocin-induced diabetic rats. *Am. J. Physiol. Endocrinol. Metab.* **2006**, 290, E1339-E1346.

[18]. Palsamy, P. and Subramanian, S. Modulatory effects of resveratrol on attenuating the key enzymes activities of carbohydrate metabolism in streptozotocin-nicotinamide-induced diabetic rats. *Chem Biol Interact.* **2009**, 179, 356-362.

[19]. Dall'Acqua, S., Stocchero, M., Clauser, M. et al. Changes in urinary metabolic profile after oral administration of curcuma extract in rats. *J. Pharm. Biomed. Anal.* **2014**, 100, 348-356.

[20]. Dall'Acqua, S., Stocchero, M., Boschiero, I. et al. New findings on the in vivo antioxidant activity of Curcuma longa extract by an integrated ^{1}H NMR and HPLC–MS metabolomic approach. *Fitoterapia.* **2016**, 109, 125-131.

[21]. Emĺia Juan, M., Pilar Vinardell, M. and Planas, J.M. The daily oral administration of high doses of trans-resveratrol to rats for 28 days is not harmful. *J Nutr.* **2002**, 132, 257-260.

[22]. Lee, E.S., Shin, M.O., Yoon, S. and Moon, J.O. Resveratrol inhibits dimethylnitrosamine-induced hepatic fibrosis in rats. *Arch. Pharmacal. Res.* **2010**, 33, 925-932.
76

[23]. Tiwari, V. and Chopra, K. Resveratrol prevents alcohol-induced cognitive deficits and brain damage by blocking inflammatory signaling and cell death cascade in neonatal rat brain. *J Neurochem.* **2011**, 117, 678-690.

[24]. Beneduci, A., Chidichimo, G., Dardo, G. and Pontoni, G. Highly routinely reproducible alignment of 1H NMR spectral peaks of metabolites in huge sets of urines. *Anal Chim Acta.* **2011**, 685, 186-195.

[25]. Stocchero, M., and Paris, D. Post-transformation of PLS2 (ptPLS2) by orthogonal matrix: a new approach for generating predictive and orthogonal latent variables. *J. Chemometrics,* [accepted for publication].

[26]. Wehrens, R., Franceschi, P., Vrhovsek, U. and Mattivi, F. Stability-based biomarker selection. *Anal Chim Acta.* **2011**, 705, 15-23.

[27]. Chong, I.G., and Jun, C.H. Performance of some variable selection methods when multicollinearity is present. *Chemometr. Intelligent. Lab. Syst.* **2005**, 78, 103-112.

[28]. Laird, N.M. and Ware, J.H. Random-effects models for longitudinal data. *Biometrics.* **1982**, 38, 963-974.

[29]. Andres-Lacueva, C., MacArulla, M.T., Rotches-Ribalta, M. et al. Distribution of resveratrol metabolites in liver, adipose tissue, and skeletal muscle in rats fed different doses of this polyphenol. *J. Agric. Food Chem.* **2012**, 60, 4833-4840.

[30]. Van Dorsten, F.A., Daykin, C.A., Mulder, T.P.J. and Van Duynhoven, J.P.M. Metabonomics approach to determine metabolic differences between green tea and black tea consumption. *J Agric Food Chem.* **2006**, 54, 6929-6938.

[31]. Saiki, S., Sato, T., Kohzuki, M., Kamimoto, M. and Yosida, T. Changes in serum hypoxanthine levels by exercise in obese subjects. *Metab. Clin. Exp.* **2001**, 50, 627-630.

[32]. Gerritsen, W.B., Aarts, L.P., Morshuis, W.J. and Haas, F.J. Indices of oxidative stress in urine of patients undergoing coronary artery bypass grafting. *Eur. J. Clin. Chem. Clin. Biochem.* **1997**, 35, 737-742.

[33]. Byrd, D.J., Kochen, W., Idzko, D. and Knorr, E. The analysis of indolic tryptophan metabolites in human urine. Thin-layer chromatography and situ quantitation. *J. Chromatogr. A* **1974**, 94, 85-106.

[34]. Miao, H., Chen, H., Zhang, X. et al. Urinary metabolomics on the biochemical profiles in diet-induced hyperlipidemia rat using ultraperformance liquid chromatography coupled with quadrupole time-of-flight SYNAPT high-definition mass spectrometry. *J. Anal. Methods Chem*. **2014**.

[35]. Kim, K.M., Henderson, G.N., Frye, R.F. et al. Simultaneous determination of uric acid metabolites allantoin, 6-aminouracil, and triuret in human urine using liquid chromatography-mass spectrometry. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci*. **2009**, 877, 65-70.

[36]. Tolun, A.A., Zhang, H., Il'yasova, D., Sztáray, J., Young, S.P. and Millington, D.S. Allantoin in human urine quantified by ultra-performance liquid chromatography-tandem mass spectrometry. *Anal. Biochem.* **2010**, 402, 191-193.

[37]. Zhang, S., Sun, X., Wang, W. and Cai, L. Determination of urinary 8-hydroxy-2'-deoxyguanosine by a combination of on-line molecularly imprinted monolithic solid phase microextraction with high performance liquid chromatography-ultraviolet detection. *J. Sep. Sci.* **2013**, 36, 752-757.

[39]. Ozkan, E., Yaman, H., Cakir, E. et al. Plasma melatonin and urinary 6-hydroxymelatonin levels in patients with pulmonary tuberculosis. *Inflammation.* **2012**, 35, 1429-1434.

[40]. Watanabe, S., Tohma, Y., Chiba, H., Shimizu, Y., Saito, H. and Yanaihara, T. Biochemical significance of 19-hydroxytesterone in the process of aromatization in human corpus luteum. *Endocr J.* **1994**, 41, 421-427.

[41]. Pavlovic, R., Cannizzo, F.T., Panseri, S. et al. Tetrahydro-metabolites of cortisol and cortisone in bovine urine evaluated by HPLC-ESI-mass spectrometry. *J. Steroid Biochem. Mol. Biol.* **2013**, 135, 30-35.

[42]. Gomez-Sanchez, C.E., Clore, J.N., Estep, H.L. and Watlington, C.O. Effect of chronic adrenocorticotropin stimulation on the excretion of 18-hydroxycortisol and 18-oxocortisol. *J Clin Endocrinol Metab.* **1998**, 67, 322-326.

[43]. Nocerino, E., Amato, M. and Izzo, A.A. The aphrodisiac and adaptogenic properties of ginseng. *Fitoterapia.* **2005**, 71, S1-S5.

[44]. Li, D., Dammer, E.B. and Sewer, M.B. Resveratrol stimulates cortisol biosynthesis by activating SIRT-dependent deacetylation of P450scc. *Endocrinology.* **2012**, 153, 3258-3268.

[45]. Yang, J., Sun, X., Feng, Z. et al. Metabolomic analysis of the toxic effects of chronic exposure to low-level dichlorvos on rats using ultra-performance liquid chromatography-mass spectrometry. *Toxicol. Lett.* **2011**, 206, 306-313.

[46]. Daykin, C.A., Van Duynhoven, J.P.M., Groenewegen, A., Dachtler, M., Van Amelsvoort, J.M.M. and Mulder, T.P.J. (2005). Nuclear magnetic resonance spectroscopic based studies of the metabolism of black tea polyphenols in humans. *J. Agric. Food Chem.* **2005**, 53, 1428-1434.

[47]. Krupp, D., Doberstein, N., Shi, L. and Remer, T. Hippuric acid in 24-hour urine collections is a potential biomarker for fruit and vegetable consumption in healthy children and adolescents. *J. Nutr.* **2012**, 142, 1314-1320.

Chapter 4: An NMR based metabolomics approach to seek reliable markers to distinguish between Citrus and Clementine Italian honey with chemometric analysis.

4.1 Aims of the work

NMR-based metabolomics is a fast, convenient, and effective tool for origin discrimination and biomarker discovery in food analysis. Traditionally, the determination of the floral origin of honey is made from palynological analysis. The method is based on the identification of pollen by microscopic inspection. However, melissopalynological analysis needs expertise and also it is not very reliable for several botanical origins. The aim of this project was to develop an NMR-based metabolomic approach that used multivariate statistical analysis to discriminate the botanical origin of different types of honey. Multivariate statistical analysis helped us to identify the most relevant signals to differentiate honey botanically. The obtained data sets were useful in the search of markers responsible for the discrimination of two types of citrus honey.

4.2 Introduction

Honey has a long history of human consumption; it is taken as medicine, eaten as food, or incorporated as an additive in a variety of foods and beverages. Honey is mainly composed of a complex mixture of carbohydrates and other minor substances, such as organic acids, amino acids, proteins, minerals, vitamins, and lipids. In almost all honey types, fructose predominates, glucose being the second main sugar. These two account for nearly 85–95% of the honey carbohydrates. Di- or oligo-saccharides constitute the remaining carbohydrates, except for traces of polysaccharides. Honey also contains volatile substances which are responsible for the characteristic flavor [1]. A number of tools and methods can be used to ensure that a given honey belongs to a botanical variety that complies with the label. The most often used method is pollen analysis (melissopalynology) [2]. By studying the pollen in a sample of honey, it is possible to gain evidence of the genus of the plants that the bees visited. Generally, melissopalynology is used to combat fraud and inaccurate labelling of honey. However, melissopalynology is not an easy technique because it requires very experienced analysts and it is very time consuming. According to some authors, melissopalynology is valid only for the determination of geographical origin of honey while it is less valid to determine its botanical origin [3]. Because

80

of the high price of certain honey types, adulteration with materials of low cost and nutritional value or mislabeling regarding the botanical origin sometimes occurs. Nonetheless, the discrimination between different types of honey is important for honeys that possess discrete aroma and taste and special nutritional properties which make them preferable for consumers.

Honey made primarily from the nectar of one type of flower is called monofloral honey, whereas honey made from many types of flowers is called polyfloral honey. Theoretically, a monofloral honey can be produced from every plant that produces nectar. However, in practice, monofloral honeys are not so easy to produce. Thus, their price is, in most cases, higher than polyfloral ones, especially for certain types of monofloral honeys. Also, monofloral honey typically has a higher value in the marketplace due to its distinctive flavor. The need to find reliable marker compounds to discriminate between monofloral honeys is necessary [4]. Marker compounds can not only characterize a certain type of honey, but they can also show adulteration in honey [5]. Several analytical platforms such as gas chromatography/mass spectrometry (GC/MS) [6], nuclear magnetic resonance spectroscopy (NMR) [7], infrared spectroscopy (IR) [8] have been used for the chemical analyses of honeys and their classification.

4.2.1 Citrus Honey

Citrus honeys are considered to be among the best monofloral honeys. They are highly appreciated by consumers because of their distinctive and pleasant flavor and aroma and they usually sell at a premium [9]. They are obtained from cultivated *Citrus* species (botanical family of *Rutaceae*), mostly orange but also, to a lesser extent, lemon, tangerine, grapefruit, lime, clementine, etc. Citrus honey is produced in several Mediterranean countries [10]. In Italy, it is very important due to its high economical value; citrus honey is one of the most important and valuable honey production in Calabria and it is one of the few Italian honeys which is exported. This production, however, suffers the competition from similar products coming from other areas of the world, especially Spain, but also Mexico, Israel, the United States, and Brazil. Citrus honey is commonly commercialized without distinction between the different citrus species (*Citrus sinensis L., Citrus deliciosa Ten., Citrus limon L.,* etc.) or at best as orange blossom honey. However, some honeys from botanical varieties of the citrus genus, such as Clementine, have excellent individual properties and could be commercialized independently and not just as citrus honey. Like other monofloral honeys, the botanical classification of *Citrus reticulata* must

be carried out by identification and quantification of the percentage of pollen by microscopic examination. The problem is that pollen analysis is considered of little value for the citrus genus as it is one of the several types of honey whose pollen is considered to be "under-represented" [11]. In other words, the amount of pollen present in citrus honey on occasions is lower than the quantity expected considering organoleptic characteristics such as flavor. This is because anther maturity does not always coincide with maximum secretion of nectar. Therefore, due to the difficulties in proving the botanical source of citrus honey using pollen analysis, the classification is based on the identification of specific chemical components present in different types of citrus honey, such as antioxidants or aromatic fractions [12]. Few reports indicate the presence of methyl anthranilate (MA) as a marker of citrus honeys in analyses of Spanish and Mediterranean citrus honeys [13]. However, this compound may also be present in honeys of other origins [14]. Other compounds, including hotrienol and the isomers of lilac aldehyde and α-4-dimethyl-3-cyclohexen-1-acetaldehyde, have been proposed for differentiating citrus honeys [15]. Moreover, a report concludes that MA content cannot be used as a discriminating parameter for *Citrus* honey, and should be used only as a further descriptive element. Among the monofloral samples, the MA content was lower in those produced in Italy than in the countries, and mostly below the 2 mg/kg limit that some European laboratories require to accept *Citrus* honey [16]. It is therefore of great importance to improve the extraction techniques used so far, as well as to develop new ones to enable the analysis of honey non-volatiles to become a routine procedure. In this regard, the purpose of this study was also to establish an extraction method apart from the already reported ones [17, 18] to characterize and enhance the MA content measured in Italian citrus honey. We were not able to identify the MA in the chloroform extraction and therefore we directly extracted citrus honey in $CDCl_3$ rather than first dissolve it in water.

4.2.2 European Commission Legislation

Honey regulation in the EU (Codex Alimentarius Commission 2001; European Commission 2002) states that the botanical and geographical origins of the product must be printed on the label. The control of commercially sold honey therefore requires the determination of parameters that unambiguously establish the origin as part of the overall quality of honey. Honey fraud involves adding either industrial sugar syrups or mixing several floral origins, and selling the

82

product under a false name. The Commission of the EU is encouraging the development of harmonized analytical methods to permit the verification of compliance with the quality specifications for the different honeys.

4.3 Methods and Materials

4.3.1 Chemicals

$CDCl_3$ (99.96%-d, H_2O 0.005% Euriso-Top, Gif sur Yvette, France), D_2O (99.9%-d,) were purchased from Sigma Aldrich (Milan, Italy).

4.3.2 Honey Samples

Eighty two samples of citrus honey were obtained from an exhibition called "GD" (golden drop) an annual exhibition in Italy where honey producers from all over Italy exhibit honeys from different botanical and geographical origin. We particularly acknowledge Lucia Piana for providing these samples after verification of their botanical origins in most of the cases by melissopalynlogic analysis. Details of the honey samples is given in **Table 4.1.**

S. No.	Honey type	Botanical origin	No. Of samples
1	Citrus	*C. sinensis*	31
2	Clementine	*C. reticulate*	51

Table 4.1: Details of citrus honey samples used in this study.

4.3.3 Sample Preparation:

4.3.3.1 Direct CDCl₃ Extraction of Honey

Six g of honey are dissolved in a Teflon tube with 1.5 mL of D_2O and 1.5 mL of $CDCl_3$ and mechanically stirred for 15 min. The resultant solution is centrifuged for 15 min at 10,000 rpm and 4 °C. After centrifugation, the lower $CDCl_3$ phase (600 μL) is collected and transferred into an NMR tube for analysis.

4.3.3.2 Direct D₂O Extraction of Honey

One hundred ± 0.03 mg of citrus honey are dissolved in 600 μL of D_2O and mechanically stirred until complete dissolution to obtain a homogenous mixture; the pH of the resulting solution is

then adjusted in the range of 2-3 by adding 8 µL of 1 M HCl. Finally, 600 µL of honey solution is transferred into an NMR tube for analysis.

4.3.4 ^{1}H-NMR Spectroscopy

For $CDCl_3$ extraction spectra were recorded on a Bruker Avance 600DMX instrument, operating at 599.99 MHz for ^{1}H and equipped with a 5 mm TXI xyz gradient inverse probe. The 1D spectra were acquired using a modified double pulsed field gradient spin echoes (DPFGSE) sequence [17]. The typical acquisition parameters of this experiment were as follows: temperature, 298 K; recycle time, 2 s; spectral window, 6000 Hz; number of scans, 256; data points, 32K; receiver gain, 8K. The parameters for the 1D spectra obtained with the standard single-pulse sequence were as follows: temperature, 298 K; recycle time, 2 s; spectral window, 6000 Hz; number of scans, 1024; data points, 32K; receiver gain, 256. In the case of D_2O extraction, all the parameter were same except that 8K data points were acquired and the number of scans was used 512.

Data for $CDCl_3$ extraction samples were processed using the ACD software (ACD/Specmanager 7.00 software, Advanced Chemistry Development Inc., 90 Adelaide Street West, Toronto, Ontario, Canada M5H 3V9). Fourier transformation was performed after zero-filling the FID data to 128K points and after apodization using a decreasing exponential with line broadening of 0.5 Hz. The spectra were phased and baseline-corrected using the ACD manual routine, and the ^{1}H NMR chemical shifts were referenced to the residual $CHCl_3$ signal at 7.27 ppm. Each ^{1}H spectrum was segmented into identical intervals ("buckets") of 0.04 ppm, and the signal intensity in each interval was integrated. The spectra were normalized to the total sum of integral covering the δ interval 13-2.16 and excluding the δ region 7.26-7.28, which contains the residual solvent peak. The resulting normalized integrals composed the data matrix that was submitted to multivariate analysis.

For D_2O extraction, spectra were processed using the ACD software (ACD/Specmanager 7.00 software, Advanced Chemistry Development Inc., 90 Adelaide Street West, Toronto, Ontario, Canada M5H 3V9). Fourier transformation was performed using interactive FT after zero-filling the FID data from 8K to 32K points and after apodization using a Sq. cosine function. The spectra were phased and baseline-corrected using the ACD manual routine; we used the signal of Tyrosine at 6.80 ppm as a reference. Each ^{1}H spectrum was segmented by using intelligent

84

bucketing and the signal intensity in each interval was integrated. The spectra were normalized to the total sum of integral covering the δ interval 6-9 ppm and excluding the δ region 3-6 and 9-13 ppm which contains the sugars and empty region (without any important peak), respectively. The resulting normalized integrals composed the data matrix that was submitted to multivariate analysis.

4.3.5 Statistical Analysis:

Principal component analysis (PCA) and (Ortho partial least square discriminate analysis) OPLSDA using "mean centering" as data pretreatment was performed using the software SIMCA-P13 (Umetrics, Umea, Sweden). Data were visualized by plotting either the PC scores, where each point in the score plot represents an individual sample, or the loading plot, which permits us to identify the spectral regions with the greatest influence on the separation and clustering of the samples and, therefore, to deduce which compounds are responsible for such clustering (markers).

4.4 Results and Discussion

4.4.1 D_2O extraction of honey:

Figure 4.1A shows a typical ^{1}H-NMR spectrum of citrus honey dissolved in deuterated water, showing the dominant resonances of fructose and glucose, honey's main components. These sugars signals are not relevant for our study. Therefore, we exclude the aliphatic and sugar region from our consideration and emphasize the aromatic region (6-9 ppm). **Figure 4.1B** shows the expansion of the aromatic region of only one citrus honey magnified 6 times compared to the original spectrum. **Figure 4.2** shows the aromatic region of all the 82 spectra of clementine and citrus honey. Several signals in the NMR spectra differ in intensity. To specify differences and similarities, multivariate data analysis was used.
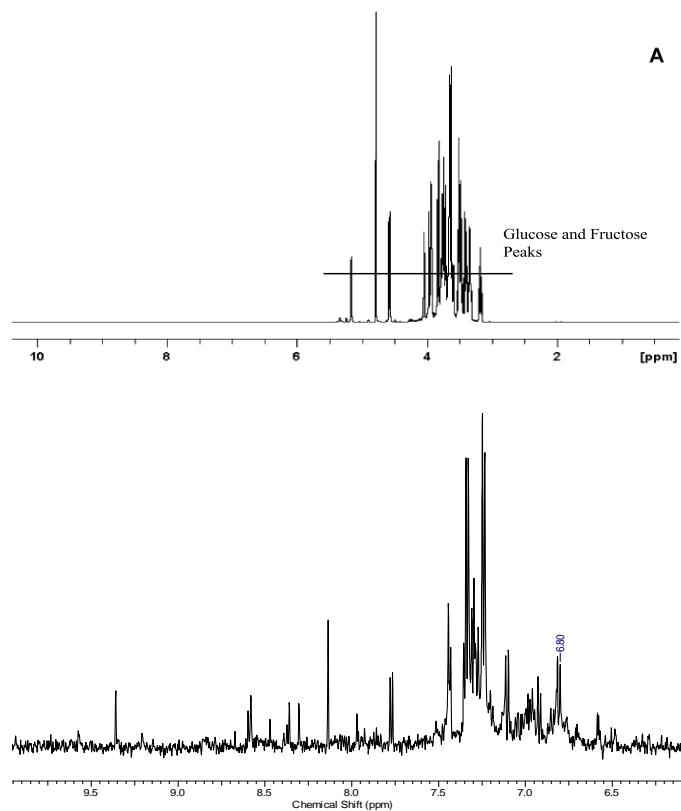
Figure 4.1: **(A)** Complete ^{1}H NMR spectrum of a citrus honey sample dissolved in D$_2$O. Expansion of the aromatic region **(B)**.
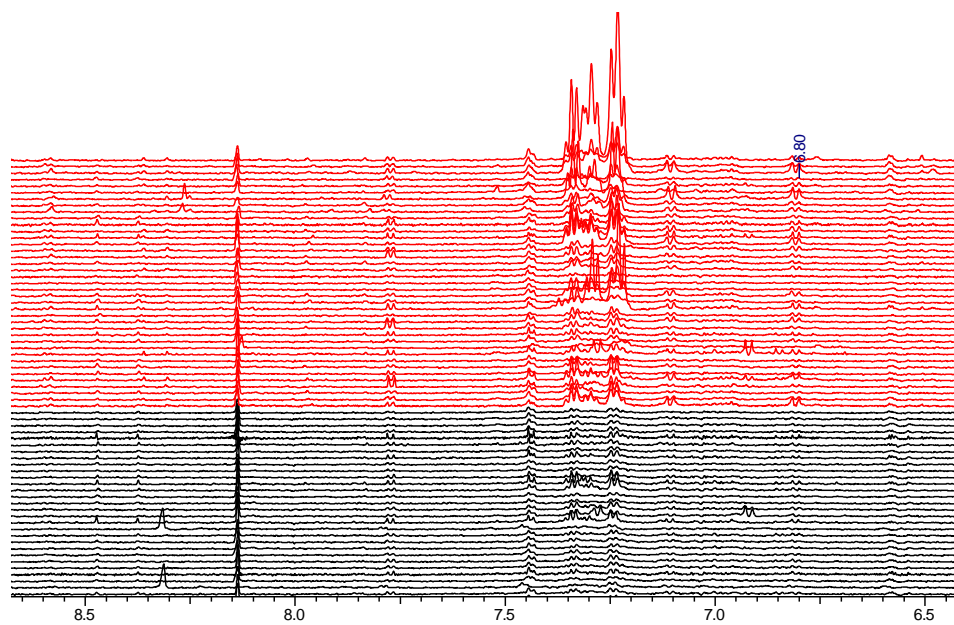

Figure 4.2: Tile plot of all spectra of citrus honeys: region between 6 and 9 ppm of clementine (black) and citrus (red) honeys.

86

As a first step, an unsupervised approach by means of PCA was applied to overview the data, to find out outlier and pattern of data set (Figure 4.3). The plots of the first two PCs of all the PCAs showed low sensitivity and selectivity and also show poor separation of the two groups.
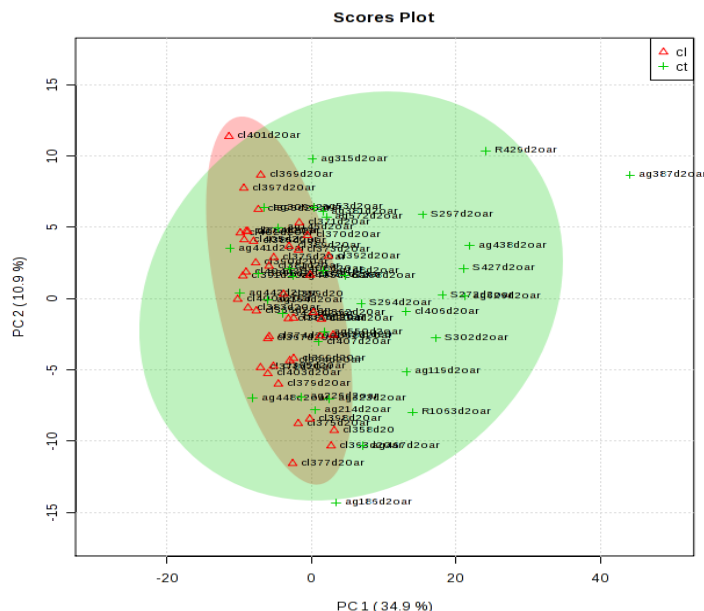

Figure 4.3: PCA score plot derived from the aromatic region of ^{1}H NMR spectra of honey dissolved in deuterated water. The plot shows the poor discrimination between **clementine (red)** and **citrus honeys (green).**

This result prompted us to perform a supervised technique, *i.e.*, OPLS-DA. The score plot of the OPLS-DA model obtained, reported in **Figure 4.4A**, shows good separation of honey samples into clementine (red) and citrus (blue) honey, with $R^2 X = 79.5\%$, $R^2 Y=39.2\%$ and $Q^2 = 54.5\%$. On the basis of the Hotelling's T2 test at the confidence level of 95%, we identified five strong outliers, which were excluded from the dataset. In **Figure 4.4B,** the S-plot allowed us to specify the variables (chemical shifts), which are responsible for the observed clustering and hence led us to the identification of markers responsible for the two different origins. The marker compounds were identified, using literature [19] and online free NMR metabolomics databases such as HMDB (http://www.hmdb.ca/) and SDBS (http://sdbs.db.aist.go.jp/sdbs/cgi-bin/cre_index.cgi). According to the S-plot, phenylacetic acid and tyrosine are strong discriminants for citrus honey while formic acid is less discriminating for clementine honey, although it is consistently more intense in clementine than in citrus honey. Chemical structures

87

and chemical shifts are given in **Table 4.2**. Since we used Pareto scaling, the discrimination is based on consistently different intensity of the same signal in the two types of honey. **Figure 4.5**, shows two NMR spectra of orange and clementine honeys in comparison. To the best of our knowledge, it is the first time that Tyrosine, Phenylacetic acid and Formic acid were reported in citrus honey. Tyrosine and Formic acid have been reported as markers for Polish buckwheat honey and Polish heather honey shows phenylacetic acid as a marker [19].
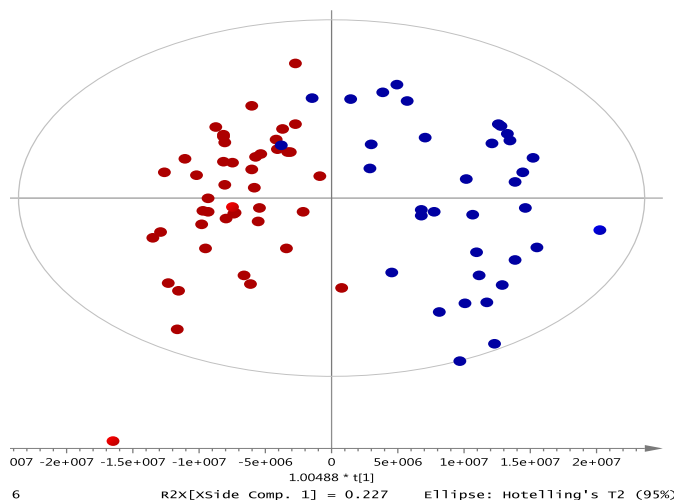

Figure 4.4A: OPLS-DA score plot derived from the aromatic region of ^{1}H NMR spectra of honey dissolved in deuterated water. The plot shows the clear discrimination between clementine (red) and citrus honeys (blue).
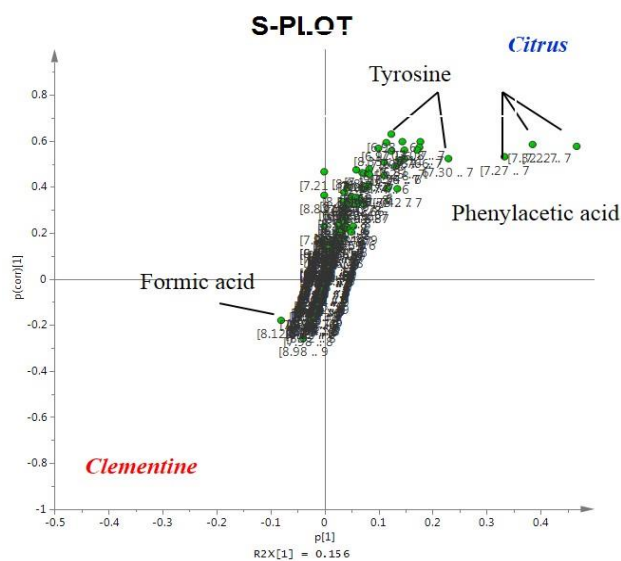

Figure 4.4B: S-Plot showing the important resonances for the distinction of the two honey types.
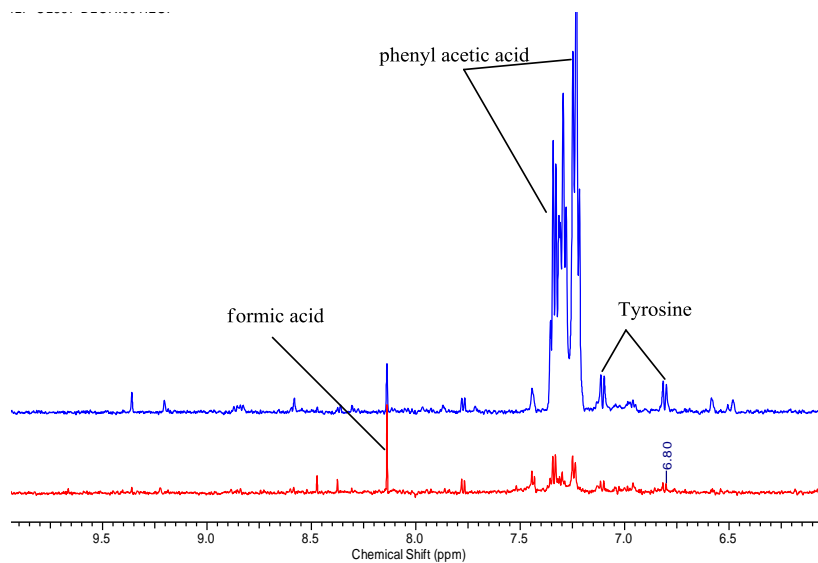
88

Figure 4.5: Aromatic region (6-9 ppm) of the ^{1}H-NMR spectra of clementine (red) and citrus (blue) honeys in D$_2$O.

Name	Structure	Position	^{1}H Chemical shift
Phenylacetic acid		4	7.37 ppm
		3,5	7.27 ppm
		2,6	7.30 ppm
		7	3.37 ppm (overlapped by sugar region)
Formic acid		2a	8.12 ppm
Tyrosine		2,6	6.80 ppm
		3,5	7.10 ppm
		8,9	Overlapped with carbohydrate region

Table 4.2: Assignment of marker compounds with their structure.

Model validation was performed following good practice, N-fold full cross-validation with value of N (N = 10) and permutation test on the responses (100 random permutations), in order to avoid over-fitting and prove the robustness of the obtained models. **Figure 4.6** represents the obtained model by permutation tests based on prediction accuracy. The p-value based on permutation is p < 0.01 (0/100). The number of components of the PLSDA models was estimated on the basis of the first maximum of Q^2 calculated by 10-fold full cross-validation under the constraint to pass the permutation test on the responses.
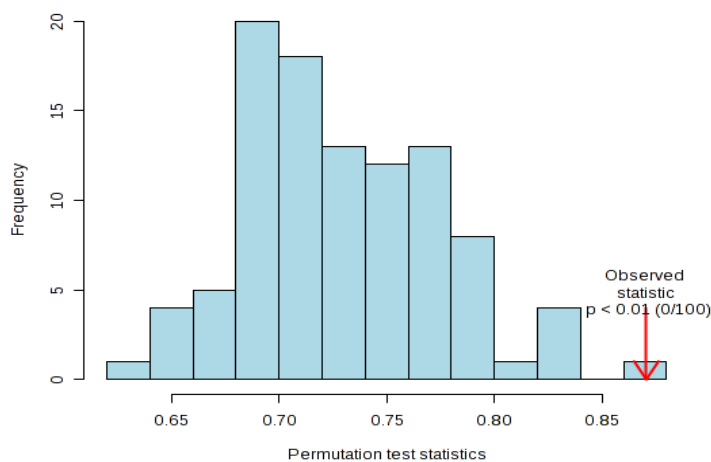

Figure 4.6: PLS-DA model validation by permutation tests based on prediction accuracy. The p value based on permutation is p < 0.01 (0/100).

4.4.2 CDCl₃ Extraction of Honey

Methyl anthranilate (MA) is widely accepted as a marker compound of citrus honey [17, 20-21]. In work conducted previously in our lab, we did not find MA using the chloroform extraction method previously developed [17], so we decided to perform a direct extraction of honey in deuterated chloroform to analyze the chloroform layer directly without evaporation. However, we were not able to see MA in these samples either. Sesta et al. [20] concluded that the MA content cannot be used as a discriminating parameter for *Citrus* honey, and should be used only as a further descriptive element. Among the monofloral samples, the MA content was lower in those produced in Italy than in other countries, and mostly below the 2 mg/kg limit that some European laboratories require to accept *Citrus* honey [20]. The NMR spectra acquired with direct

CDCl₃ extraction showed better resolution and signal to noise ratio in comparison with the spectra obtained using previously established methods [17, 18]. **Figure 4.7** shows all the NMR spectra of citrus honey acquired by direct CDCl₃ extraction.
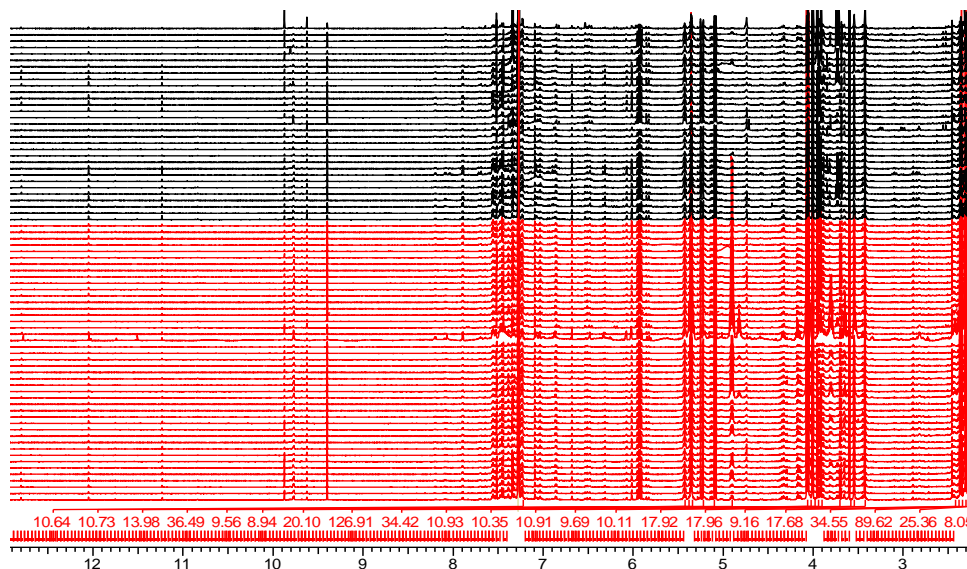

Figure 4.7: Tile plot of all spectra of citrus honeys extracted directly in CDCl₃: Full length spectra between 2.6 and 13 ppm of clementine (red) and citrus (black) honey honeys.

Figure 4.8 represents the loading plot of the OPLS-DA model of all the honey samples using UV scaling. The first two PC's on the basis of Hotelling's T2 95% revealed clear separation of two types of honey, even better than the one obtained in D₂O. The model diagnostics were summarized by the fit goodness, R^2 (84.9%) and the prediction goodness parameter Q^2 (77.7%).
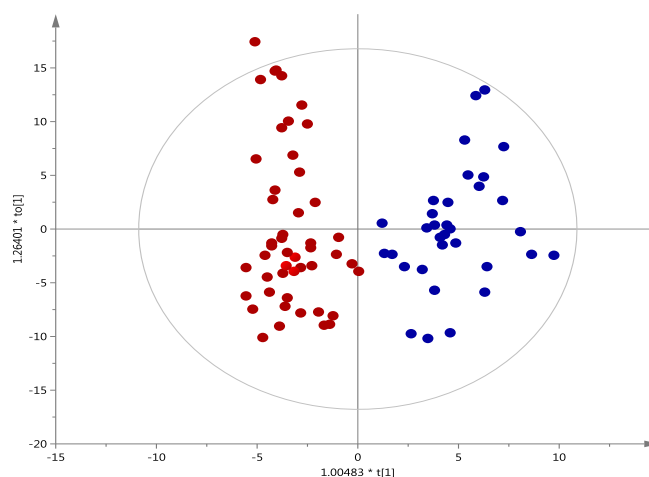
Figure 4.8: OPLS-DA loading plot derived from the ¹H NMR spectra of CDCl₃ extracts of honeys. The plot shows the clear discrimination between clementine (red) and citrus honeys (blue).
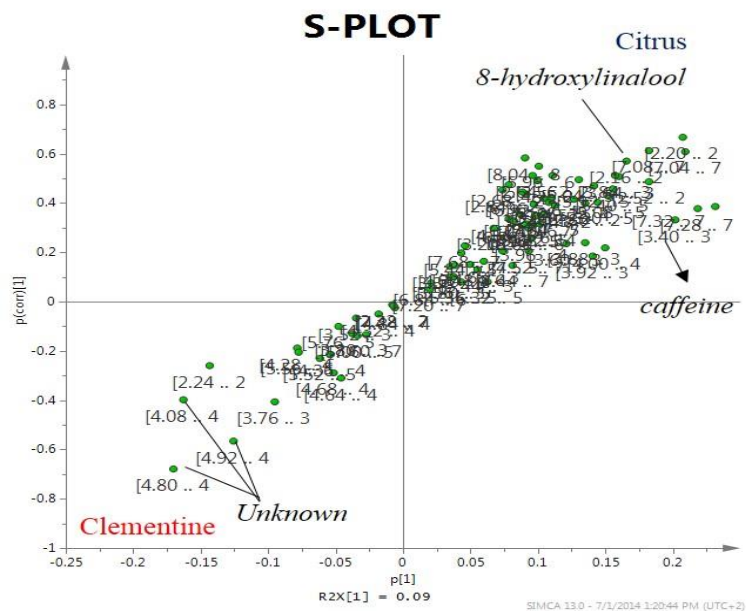
Figure 4.9: S-plot showing the important resonances for the distinction of the two honey types.
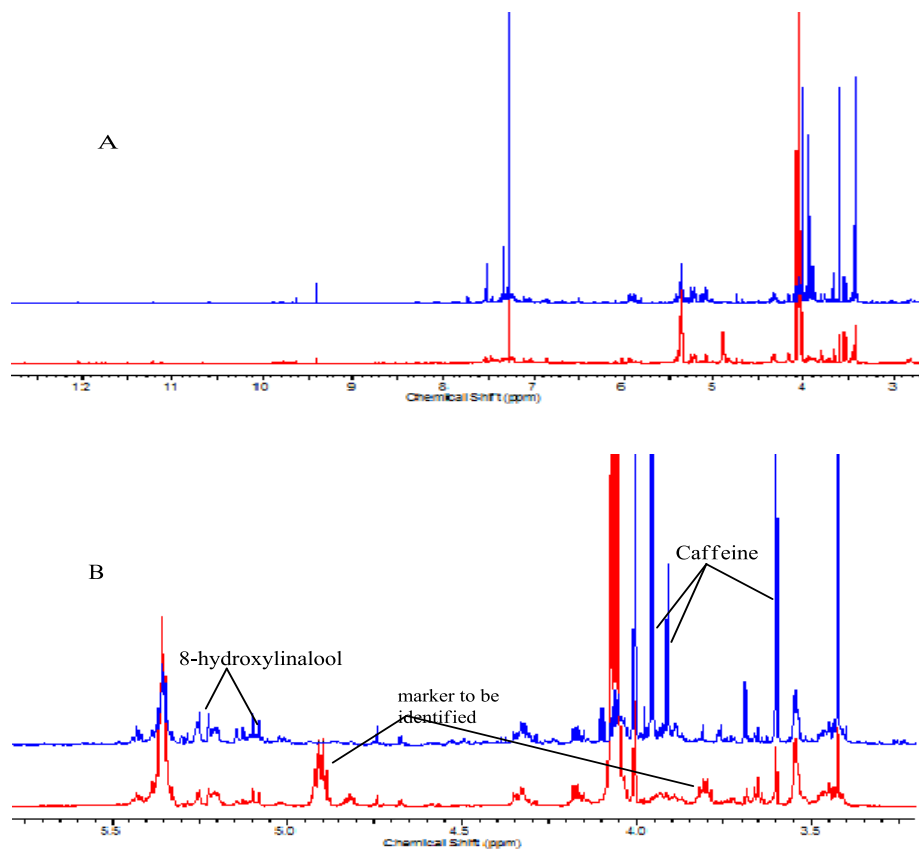

Figure 4.10: Comparison of **citrus (blue)** and **clementine (red)** ^{1}H NMR spectra.
a) Whole NMR spectra. **b)** Expansion of the spectral region 5.8-3.0 ppm.

These resonance can be recognized as caffeine and 8-dihydroxylinalool from already reported work [17]; two resonances, assigned to an unknown compound which are present in the clementine samples. A single representative spectrum from each type of citrus honey is shown in **Figure 4.10**. All the marker resonances for citrus honey extracted in deuterated chloroform are presented in **Table 4.3**.
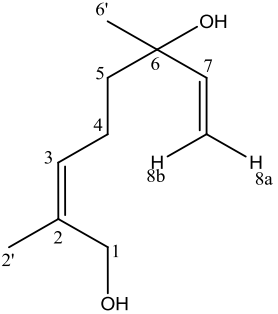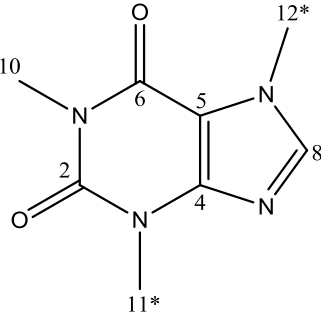
Name	Structure	Position	^{1}H Chemical Shift
8-hydroxylinalool		1	4.00
		3	5.43
		4	2.08
		5	1.61
		7	5.93
		8a	5.09
		8b	5.24
		2`	1.68
		6`	1.30
Caffeine		8	7.55
		10	3.42
		11	3.61
		12	4.01
Unknown	-	-	4.88 & 3.76

Table 4.3: Assignment of the marker compounds extracted in CDCl$_3$.

4.5 Conclusions

In total, we analyzed 82 samples for each extraction method adopted for citrus honey. Herein, we only discussed the results obtained from direct dissolution in D$_2$O and from direct CDCl$_3$ extraction. The results from chloroform (CHCl$_3$) extraction have already been reported from our lab [17-18] and were not described here. Among possible multivariate statistical tools, the

hierarchical OPLS-DA demonstrated high efficiency in NMR data analysis in terms of classification capability, as already demonstrated [17]. Using our new method of direct $CDCl_3$ extraction, we achieved better signal to noise ratio and also discovered a new marker (Unknown). According to our knowledge, this is the first attempt to discriminate clementine and citrus honeys. Our D_2O extraction strategy revealed phenylacetic acid and tyrosine as marker compounds in citrus honey whereas higher amounts of formic acid characterized clementine honey. We did not find MA into either honey types studied; however, high amounts of phenylacetic acid and tyrosine distinguish citrus honey from clementine honey. NMR-based metabolomics gave a descriptive view of two types of honey belonging to the same botanical species.

4.6 References

[1]. Finola, M.S., Lasagno, M.C. and Marioli, J.M. Microbiological and chemical characterization of honeys from central Argentina. *Food Chem.* **2007**, 4(100), 1649-1653.

[2]. Cotte, J.F., Casabianca, H., Chardon, S., Lheritier, J. and Grenier-Loustalot, M.F. Chromatographic analysis of sugars applied to the characterisation of monofloral honey. *Anal. Bioanal. Chem.* **2004**, 380(4), 698-705.

[3]. Stephens, J.M., Schlothauer, R.C., Morris, B.D., Yang, D., Fearnley, L., Greenwood, D.R. and Loomes, K.M. Phenolic compounds and methylglyoxal in some New Zealand manuka and kanuka honeys. *Food Chem*. **2010**, 120(1), 78-86.

[4]. Anklam, E. A review of the analytical methods to determine the geographical and botanical origin of honey. *Food Chem.* **1998**, 63(4), 549-562.

[5]. Alissandrakis, E., Tarantilis, P.A., Harizanis, P.C. and Polissiou, M. Aroma investigation of unifloral Greek citrus honey using solid-phase microextraction coupled to gas chromatographic–mass spectrometric analysis. *Food Chem.* **2007**, 100(1), pp.396-404.

[6]. Tananaki, C., Thrasyvoulou, A., Giraudel, J.L. and Montury, M. Determination of volatile characteristics of Greek and Turkish pine honey samples and their classification by using Kohonen self organising maps. *Food Chem.* **2007**, 101(4), 1687-1693.

[7]. Lolli, M., Bertelli, D., Plessi, M., Sabatini, A.G. and Restani, C., 2008. Classification of Italian honeys by 2D HR-NMR. *J. Agric. Food Chem.* **2008**, 56(4), 1298-1304.

[8]. Ruoff, K., Luginbühl, W., Künzli, R., Iglesias, M.T., Bogdanov, S., Bosset, J.O., von der Ohe, K., von der Ohe, W. and Amadò, R. Authentication of the botanical and geographical origin of honey by mid-infrared spectroscopy. *J. Agric. Food Chem.* **2006**, 54(18), 6873-6880.

[9]. White, J.W. and Bryant, V.M. Assessing citrus honey quality: pollen and methyl anthranilate content. *J. Agric. Food Chem.*, **1996**, 44(11), 3423-3425.

[10]. Terrab, A., Díez, M.J. and Heredia, F.J., 2003. Palynological, physico-chemical and colour characterization of Moroccan honeys. II. Orange (Citrus sp.) honey. *Int. J. Food Sci. Tech.* **2003**, 38(4), 387-394.

[11]. Rodriguez, I., Salud, S., Hortensia, G., Luis, U.J. and Jodral, M. Characterisation of Sierra Morena citrus blossom honey (Citrus sp). *Int. J. Food Sci. Tech*, **2010**, 45(10), 2008-2015.

[12]. Yao, L. H., Datta, N., Tomás-Barberán, F. A., Ferreres, F., Martos, I., and Singanusong, R. Flavonoids, phenolic acids and abscisic acid in Australian and New Zealand Leptospermum honeys. *Food Chem.* **2003**, 81(2), 159–168.

[13]. Ferreres, F., Giner, J. M., and Tomas-Barberan, F. A. A comparative study of hesperetin and methyl antranilate as markers of the floral origin of citrus honey. *J. Agric. Food Chem.* **1994**, 65, 371–372.

[14]. Verzera, A., Campisi, S., Zappalà, M. and Bonaccorsi, I. SPME-GC-MS analysis of honey volatile components for the characterization of different floral origin. *Am. Lab.* **2001**, 33(15), pp.18-21.

[15]. Castro-Vazquez, L., Diaz-Maroto, M. C., and Perez-Coello, M. S. Aroma composition and new chemical markers of Spanish citrus honeys. *Food Chem.* **2007**, 103, 601–606.

[16]. Sesta, G., Piana, L. M., Oddo, P. L., Lusco, L., and Belligoli P. Methyl anthranilate in Citrus honey. Analytical method and suitability as a chemical marker. *Apidologie* **2008**, 39, 334–342.

[17]. Schievano, E., Peggion, E. and Mammi, S. [1]H nuclear magnetic resonance spectra of chloroform extracts of honey for chemometric determination of its botanical origin. *J. Agric. Food Chem.* **2009**, 58(1), 57-65.

[18]. Schievano, E., Stocchero, M., Morelato, E., Facchin, C. and Mammi, S. An NMR-based metabolomic approach to identify the botanical origin of honey. *Metabolomics*, **2012**, 8(4), 679-690.

[19]. Zieliński, Ł., Deja, S., Jasicka-Misiak, I. and Kafarski, P. Chemometrics as a tool of origin determination of polish monofloral and multifloral honeys. *J. Agric. Food Chem.* **2014**, 62(13), pp.2973-2981.

[20]. Sesta, G., Piana, M.L., Oddo, L.P., Lusco, L. and Belligoli, P. Methyl anthranilate in Citrus honey. Analytical method and suitability as a chemical marker. *Apidologie*, **2008**, 39(3), 334-342.

Chapter 5. Determination of auuthenticity and entomological origin of Ecuadorian honey by NMR

5.1 Aim of the work

Increasing production of stingless-bee honey and the prospect of broader marker for natural and organic products indicate the need to establish parameters to determine the entomological origin and authenticity of honey. In this context, we received from Dr. Patricia Vit 83 Ecuadorian commercial honeys from four different entomological origins, i.e., *Apis mellifera*, *Geotrigona, Melipona* and *Scaptotrigona*, plus eight false or adulterated honey samples. A simple extraction in chloroform followed by acquisition of a ^{1}H-NMR spectrum of each honey sample yields a database amenable to chemometric analysis. In a first approach, NMR results were used to distinguish between the genuine *Apis mellifera* honey and false honey. In a second step, the NMR spectra were analyzed with a chemometric approach to distinguish the entomological origin of honey produced by *Apis mellifera* and pot-honey produced by the genera *Geotrigona*, *Melipona,* and *Scaptotrigona*. The important compounds for the authentication of the honey determined with the chemometric approach were identified with the aid of online databases and 2D NMR spectroscopy.

5.2 Introduction

Honey has a great importance in both ancient and modern civilizations with many functional applications. Apart from being simply used as a sweetener, honey is widely known both as food with significant nutritional properties and as a natural product with valuable therapeutic applications, topically or orally administered [1]. Stingless bees, sometimes called stingless honey bees or simply meliponines, are a large group of bees (about 500 species), comprising the tribe Meliponini [2]. However, stingless bee honeys are better known and mainly used in South America, Africa and Australia, while honeys derived from *A. mellifera* are mainly produced and distributed in Europe and Asia [1]. Beekeeping with honey bees belonging to the genus *Apis* is more widespread than meliponiculture – beekeeping with stingless bees (e.g., *Melipona* spp.) – which is a well-known tradition in tropical countries. Both groups of bees belong to the family *Apidae*, order Hymenoptera. They are differentiated at the subfamily level, *Apinae* for honey bees and *Meliponinae* for stingless bees. The main structural differences between honey bees and

stingless bees is nest construction: stingless bees construct horizontal combs made of cerumen (a mixture of propolis and beeswax) for their nests, and honey pots instead of honeycombs for storing honey [3-4]. Stingless bee honey is completely different from that produced by the bees of the genus *Apis* and it reaches higher prices in the market compared with traditional honey. Nevertheless, standards have not been created for pot-honey [5]. The Codex Alimentarius Commission international honey standards for honey are designed only for *Apis mellifera* [6]. To authenticate the entomological origin, chemical makers such as fructose, glucose and maltose have been used in a multivariate analysis [7], physical-chemical parameters [8], sensory Free Choice Profile (FCP) [9], NMR and chemometrics [10], flavonoid C-glycosides [11] and O-glycosyl flavones [12], and more recently electrophoretic patterns because different species of stingless bees produce honey with distinctive protein bands [2].

Pot-honey price can reach up to 1,100% that of *Apis mellifera* honey. This great difference is a motivation to declare a false entomological origin, causing an Economically Motivated Adulteration (EMA), to be added to the four types of EMAs causing vulnerabilities in the US honey market: 1. Dilution with syrups, 2. Feeding sugar to bees during nectar offer, 3. Masking countries of origin, and 4. Antibiotic residues after excessive veterinary doses [13]. Therefore, it is of great importance to develop a method to certify the entomological origin of honey and to create a rapid method to avoid adulteration of the pot honey.

5.3 Method and Materials

5.3.1 Honey Samples

Eighty three commercial honeys from Ecuador were purchased in markets or received from local stingless bee keepers. The entomological origin were declared by the beekeeper.

5.3.2 Sample Preparation:

Precisely weighed amounts (6.00 ± 0.01 g) of honey were placed in Teflon tubes and dissolved in 15 mL of deionized water; the pH of each sample was maintained between 2 and 3 using small volumes of HCl. Fifteen mL of $CHCl_3$ were added and the mixture was mechanically stirred for 15 min and then centrifuged at 10,000 rpm for 15 min at 4 °C. The lower chloroform phase was collected in a glass vial and the solvent was evaporated under a gentle stream of nitrogen. The solid residue was reconstituted in 600 µL of $CDCl_3$ and placed in an NMR tube for the analysis.

5.3.3 ^{1}H-NMR Spectroscopy

Spectra were recorded on a Bruker Avance 600DMX instrument, operating at 599.99 MHz for ^{1}H and equipped with a 5 mm TXI xyz gradient inverse probe. The 1D spectra were acquired using a modified double pulsed field gradient spin echoes (DPFGSE) sequence. The typical acquisition parameters were as follows: temperature, 298 K; recycle time, 2 s; spectral window, 6000 Hz; number of scans, 256; data points, 32K; receiver gain, 8K.

5.3.4 ^{1}H-NMR Data Reduction and Processing

ACD software (ACD/Specmanager 7.00 software, Advanced Chemistry Development Inc., 90 Adelaide Street West, Toronto, Ontario, Canada M5H 3V9) was used for data reduction and processing. Fourier transformation was performed after zero-filling the FID data to 128K points and after apodization using a decreasing exponential with line broadening of 0.5 Hz. The spectra were phased and baseline-corrected using the ACD manual routine, and the ^{1}H-NMR chemical shifts were referenced to the residual $CHCl_3$ signal at 7.27 ppm. Each ^{1}H spectrum was segmented into intelligent bucketing, and the signal intensity in each interval was integrated. The spectra were normalized to the total sum of integrals covering the δ interval 9.45-2.16 and excluding the δ region 7.26-7.28, which contains the residual solvent peak. The resulting normalized integrals composed the data matrix that was submitted to multivariate analysis.

5.3.5 Statistical Analysis

Principal component analysis (PCA) and partial least square-discriminate analysis (PLS-DA) were conducted using the software SIMCA-P10 (Umetrics, Umea, Sweden). Data were mean centered and Pareto scaled. Data are visualized by plotting either the PC scores, where each point in the scores plot represents an individual sample, or the loadings plot, which permits us to identify the spectral regions with the greatest influence on the separation and clustering of the samples.

5.4 Results and Discussion

5.4.1 Authentication of *Apis mellifera* Honey Hamples:

For the authentication of genuine honey, the study was carried out on 24 genuine *Apis mellifera* honey and 8 false honey samples. **Figure 5.1 (A-B)** shows the spectra of a false honey in comparison with genuine *Apis mellifera* honey. There is appreciable difference between the two

spectra. The genuine *Apis mellifera* honey complex spectrum derives from the biodiversity and complexity of plants and bees. Compounds distinguishing genuine *Apis mellifera* and false honey (**Fig. 5.1**) were identified with the help of our previously published report [14] and standard 2D NMR experiments. These compounds are listed in **Table 5.1**
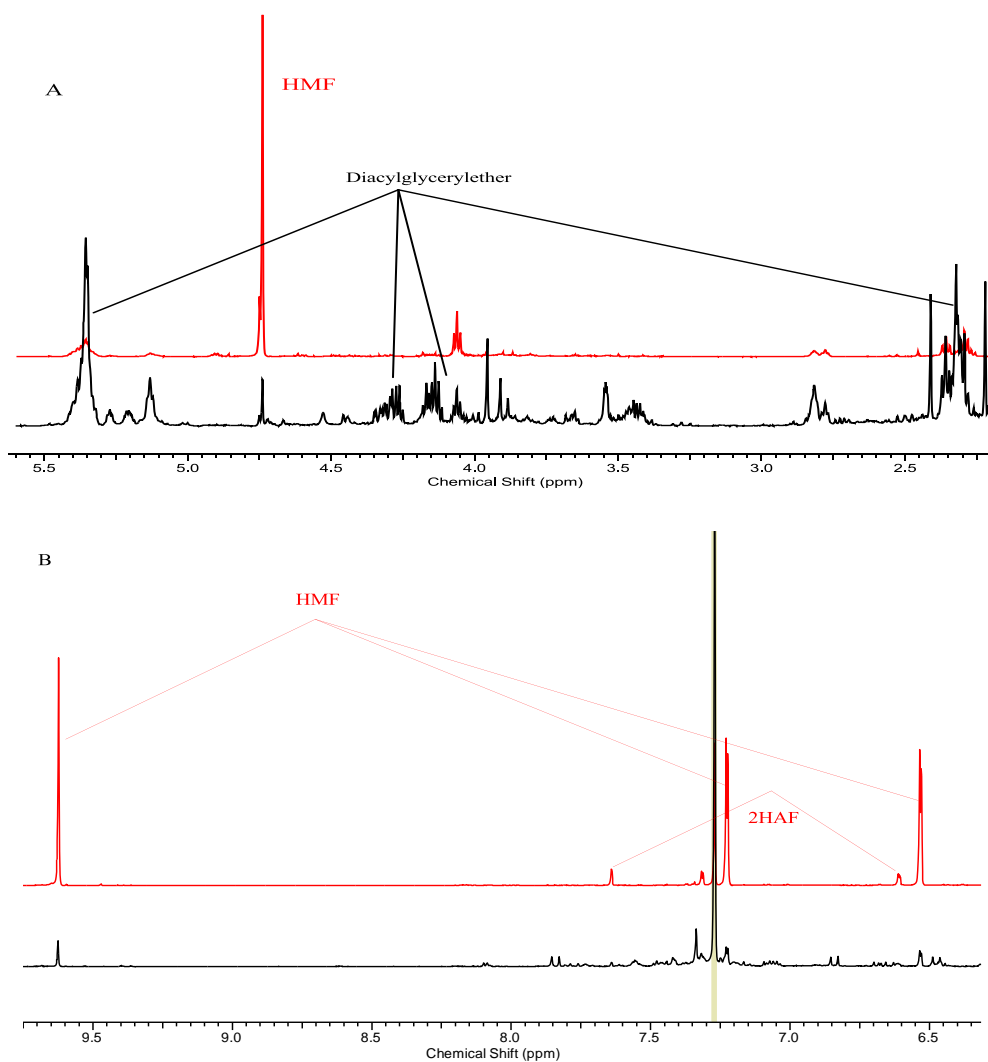

Figure 5.1: (A) Complete ^{1}H-NMR spectrum of a false honey sample in chloroform (red) and of an authentic *Apis Mellifera* sample (black). **(B)** Expansion of the aromatic region (4.5-9.2 ppm).

Figure 5.2 A shows ^{1}H-NMR spectra of 8 false (AMF) honeys in CDCl$_3$, whereas **Figure 5.2 B** shows the main compounds that characterize them: 5-Hydroxymethyl furfural (5-HMF), 2-Hydroxyacetyl-furan (2-HAF), Benzoic acid, and Sorbic acid. 5-HMF is also present in genuine

honey, but its amount is much higher in false honey. Benzoic and Sorbic acid were also present in 2 samples as preservatives.

The Presence of 5-HMF in honey is unavoidable since low amounts of this compound naturally occur in fresh sugar containing food, such as honey and milk [15]. The content of 5-HMF depends on the temperature of the beehives, and it can increase at higher temperatures, becoming dangerous also for the health of the honey bee [16]. Moreover, its amount also depends on storage time and pH; therefore, it is normal practice to use the 5-HMF content as an indicator of adulteration and quality [15]. Both the Codex Alimentarius Commission (Alinorm 01/25, 2000) and the European Union (Directive 110/2001) established that its concentration usually should not exceed 40 mg/kg (80 mg/kg for tropical honeys) [17]. The presence of 2-HAF along with 5-HMF is an indication of adulteration since both compounds are products of caramelization reactions.

Preservatives are added into food deliberately with the purpose to save nutritional value and to increase the shelf life. Benzoic and Sorbic acid are two commonly used preservatives in the food industry and possesses the anti-microbial property to inhibit yeast and mold growth as well as to prevent the attack of bacteria [18]. Although widespread use of these preservative is common, still they have an adverse effect on human and animal health, such as metabolic acidosis, convulsions and hypernoea at high doses [19]. All the *Apis mellifera* genuine samples were characterize by the presence of an insoluble ether compound; Diacylglyceryl ether. **Figure 5.3** shows the structure of the marker of *Apis mellifera* genuine honey. All the resonances were assigned and identified with the help of previous work published by our group [14]. In that work, a metabolomic approach was used to discriminate honey belonging to different botanical origins but produced by *Apis mellifera*. Interestingly, this type of molecule has been found only in the muscles and liver of a giant squid and certain species of sharks. Diacylglyceryl ethers are fundamental molecules for these species as a short-term energy resource, and they may also be important in imparting buoyancy regulation [14].
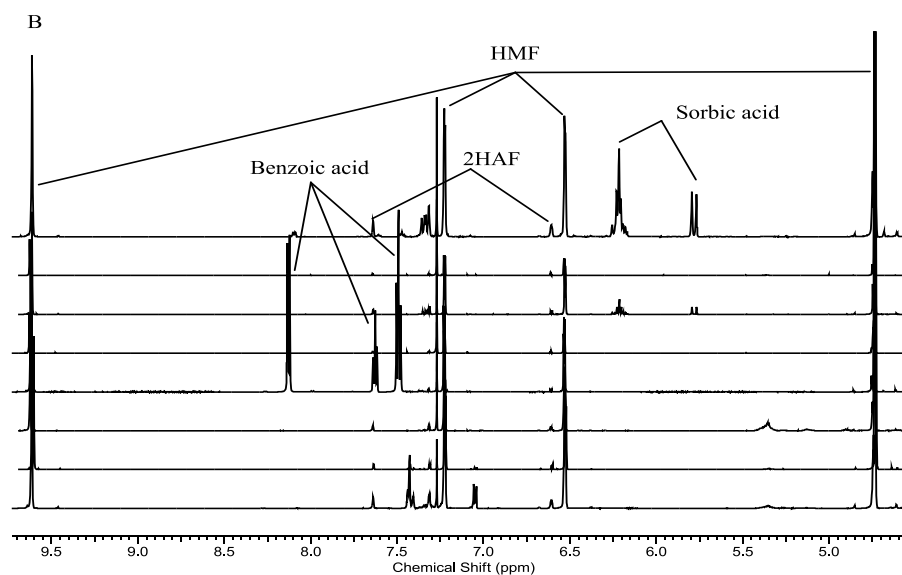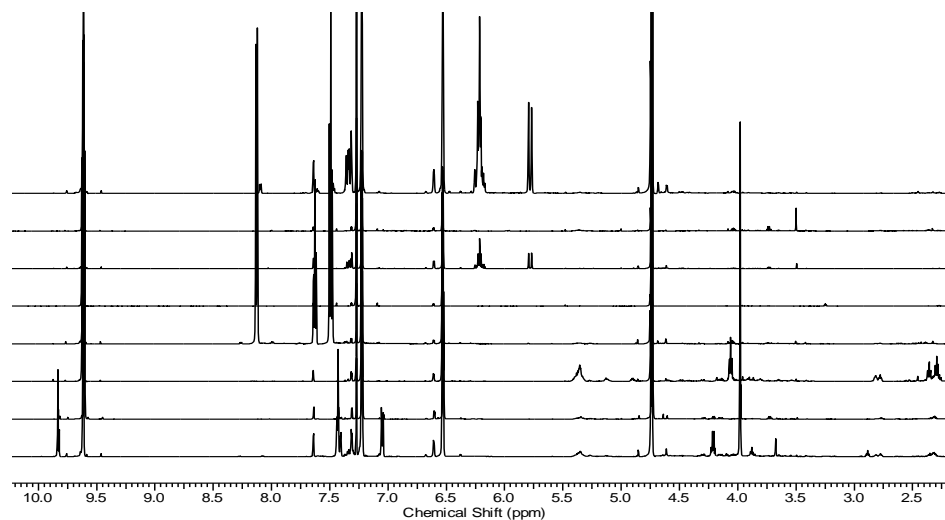
Figure 5.2: (**A**) Complete ^{1}H NMR spectrum of 8 false honey samples in chloroform. (**B**) Expansion of the aromatic region (4.5-9.2 ppm).

Honey samples	Compound	Chemical shifts (δ)
False Honey	5-Hydroxymethyl furfural	4.76 (s); 6.56 (d); 7.25 (d); 9.64 (s)
	2-Hydroxyacetyl-furan	4.77 (s); 6.63 (dd); 7.34 (d); 7.66 (d);
	Sorbic acid	1.90 (d); 5.81 (d); 6.24 (m); 7.36 (d)

	Benzoic acid	7.52 (t); 7.65 (t); 8.15 (d)
Genuine *Apis Mellifera* honey	Diacylglycerylether	0.91 (t); 1.3 (m); 1.56 (m); 1.64 (m); 2.04 (br); 3.45 (m); 3.56 (dd); 4.18 (dd); 4.35 (dd); 5.21 (m); 5.36 (m)

Table 5.1: ^{1}H-NMR data of identified compounds in false and genuine honey in CDCl$_3$.
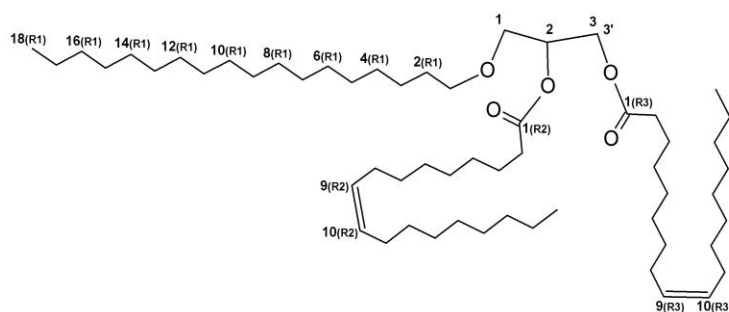

Figure 5.3: Diacylglyceryl ether, a marker of *Apis mellifera* genuine honey.

5.4.2 Authentication of Honey with Chemometric Analysis.

Hierarchical Cluster Analysis (**Figure 5.4**) was applied to classify the samples based on spectral differences. In data mining and statistics, hierarchical clustering is a method which seeks to build a hierarchy of clusters. In general, the merges and splits are determined in a greedy manner. The results of hierarchical clustering are usually presented in a dendrogram or a heatmap. For the determination of spectral differentiation among studied groups, cluster analysis was performed via MetaboAnalyst (http://mirror.metaboanalyst.ca/). Total sum normalization was used for each spectrum in the range of 2.16- 9.50 ppm. Spectral distances were calculated between pairs of spectra as Pearson's correlation coefficients and Euclidean distance was used to calculate the sample similarities and to indicate the complete linkage clustering by Ward's algorithm.

Use of chemometrics together with classical methods for the classification of different honey samples has been proposed in previous researches. PCA-like techniques can be preferred primarily for the determination of general relationship among data [20]. However, to group similar data gathered from different samples, cluster analysis must be performed [21]. Similar samples incline to be classified in the same cluster and the level of difference between the

clusters is indicated with heterogeneity values. Ward's algorithm was previously reported to give one of the best predictions, among the different methods used in cluster analysis [20, 22].
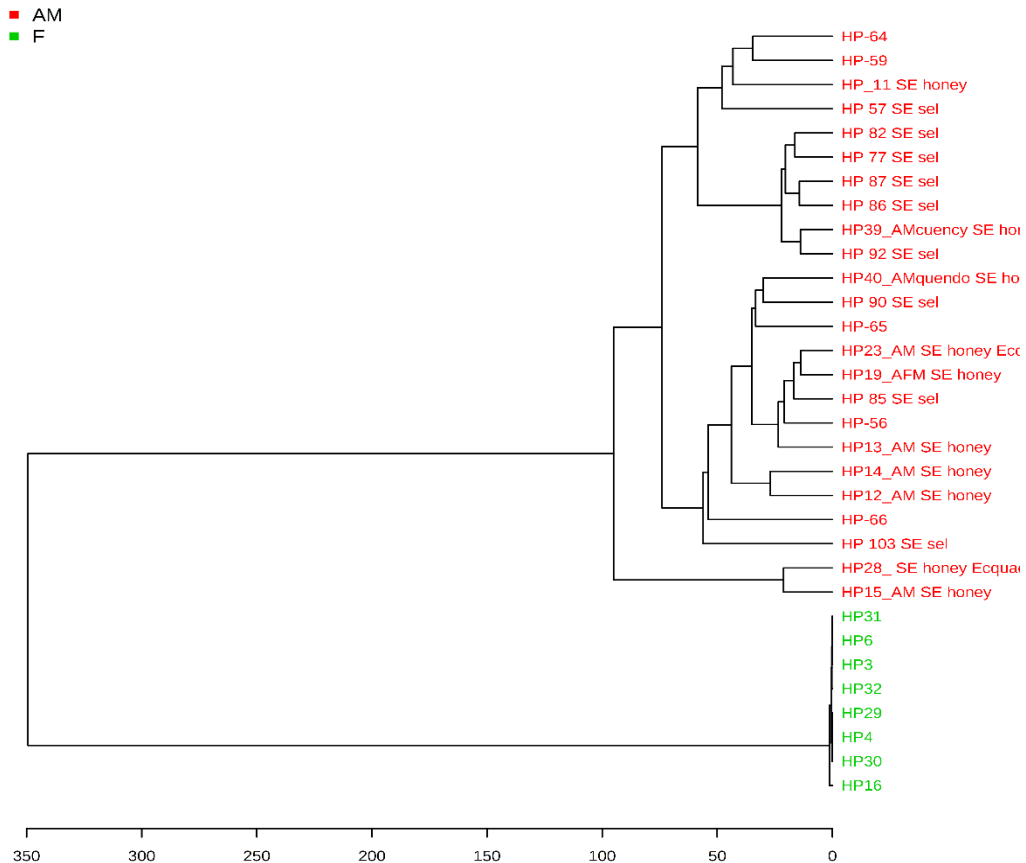

Figure 5.4: Clustering dendrogram showing differences between *Apis mellifera* in red and false honey in green (distance measure using Euclidean, and clustering algorithm using Ward).

The dendrogram shows clearly two clusters which represent the inter-sample similarity of genuine honey and dissimilarity with false honey. Hierarchical Cluster Analysis also provides the heatmap shown in **Figure 5.5** where the top 10 important variables that discriminate false vs. genuine honey are indicated. A heat map is a graphical representation of data where the individual values contained in a matrix are represented as colors.
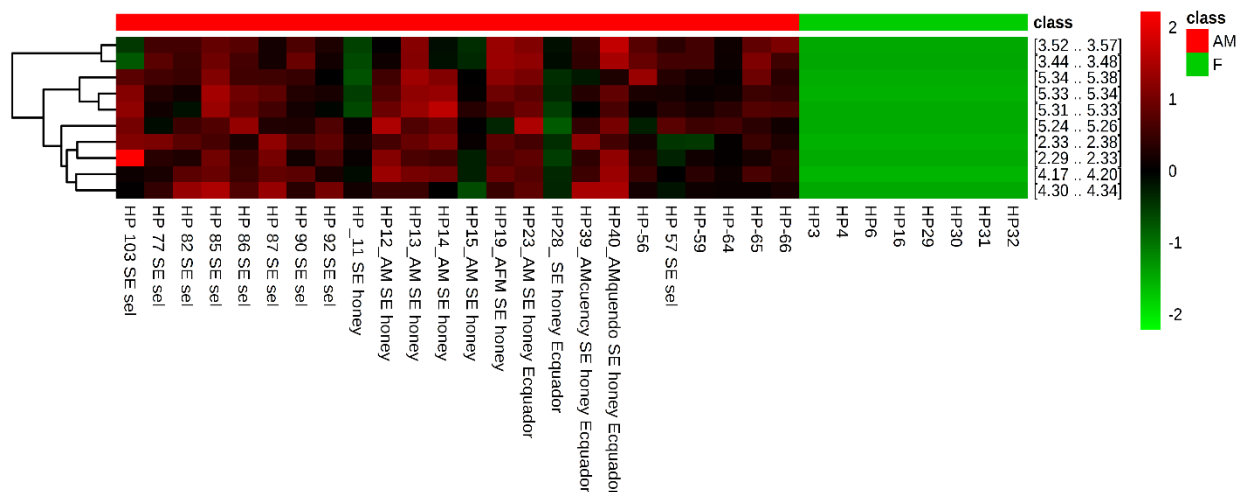
Figure 5.5: Clustering result shown as heatmap (distance measure using Euclidean, and clustering algorithm using Ward).

The right hand side of the heatmap show the scale of the intensity of a particular variable in a particular sample. These features were selected on the basis of the VIP score plot (top 10). These variables represent the resonances of the Diacylglyceryl ether (from bees wax), a marker for the genuine when compared to the false honey. The heatmap shows absence (green color) of this compound in fake honey as also evident from the NMR spectrum of false honey (Figure 5.1). **Table 2** summarizes the classification rate of the two honey types used for the authentication of genuine and false honey: the model classifies the honey samples with 100% accuracy.

Classes	*Apis mellifera* (AM)	False honey (F)	Class. Error
Apis mellifera (AM)	24	0	0.00
False honey (F)	0	8	0.00

Table 5.2: Classification of hierarchical clustering analysis model.

5.4.3 Chemometric Approach for Entomological Origin Discrimination

For the analysis of entomological origin of honey, we used 83 commercial honey samples belonging to four different entomological origin, namely *Apis mellifera* (25), *Geotrigona* (16), *Mellipona* (18), and *Scaptotrigona* (24). By using principal component analysis (PCA) and supervised techniques (projection to latent structures by partial least squares-discriminant

analysis, PLS-DA), a classification model according to entomological origin was obtained, with high predictability power.

In the first step, we applied the PCA analysis to the four different types of honey samples; the scores plot is shown in **Figure 5.6**. Three outliers were detected using PCA analysis, which were eliminated from further analysis.
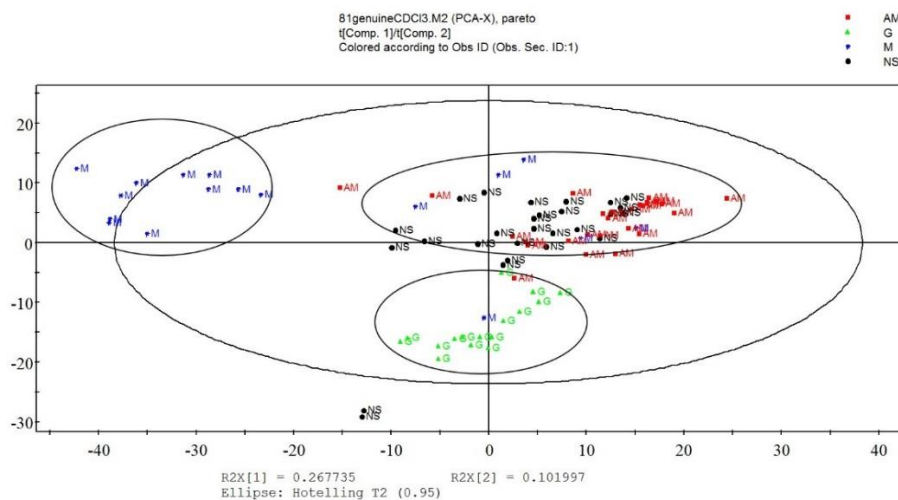

Figure 5.6: PCA scores plot of all samples: *Apis mellifera* (red), *Scaptotrigona* (black), *Melipona* (blue) and *Geotrigona* (green).

The PCA scores plot shows good separation between *Geotrigona* and *Melipona* groups, but there is not much difference between *Scaptotrigona* and *Apis mellifera*. Next, PLSDA was applied to visualize the subtle differences between the four groups; the scores plot is shown in **Figure 5.7**. A clear differentiation of samples according to the entomological origin was achieved. The model diagnostics were summarized for three components by the fit goodness, R^2X (0.664), R^2Y (0.929) and the prediction goodness parameter, Q^2 (0.854). These values are considered good for a valid model.
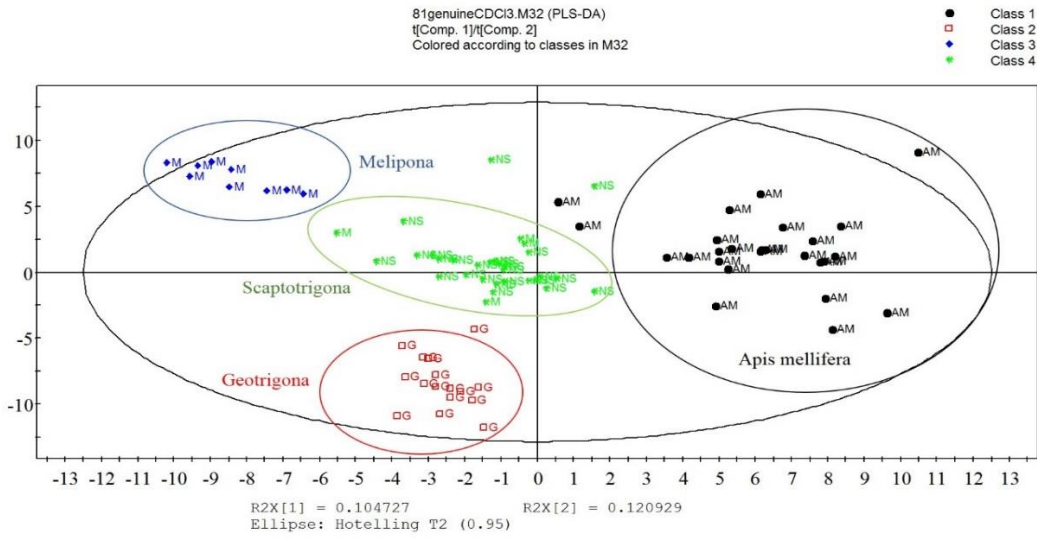
Figure 5.7: PLSDA score plot of all samples: *Apis mellifera* (black), *Scaptotrigona* (green), *Melipona* (blue) and *Geotrigona* (G)**.**

The next step is to determine which resonances differentiate the four groups; therefore, we used a *one-vs-all* strategy through a PLSDA model where each class of honey was compared with the other honeys considered together as one class. **Figure 5.8 (a-d)** shows the scores plot of each entomological origin honey versus all as one class with their corresponding loadings plot to visualize the resonances important to differentiate between them.

Meliponi (M) vs All

Meliponi (M) vs All

Scaptotrigona (NS) vs All

Scaptotrigona (NS) vs All

108
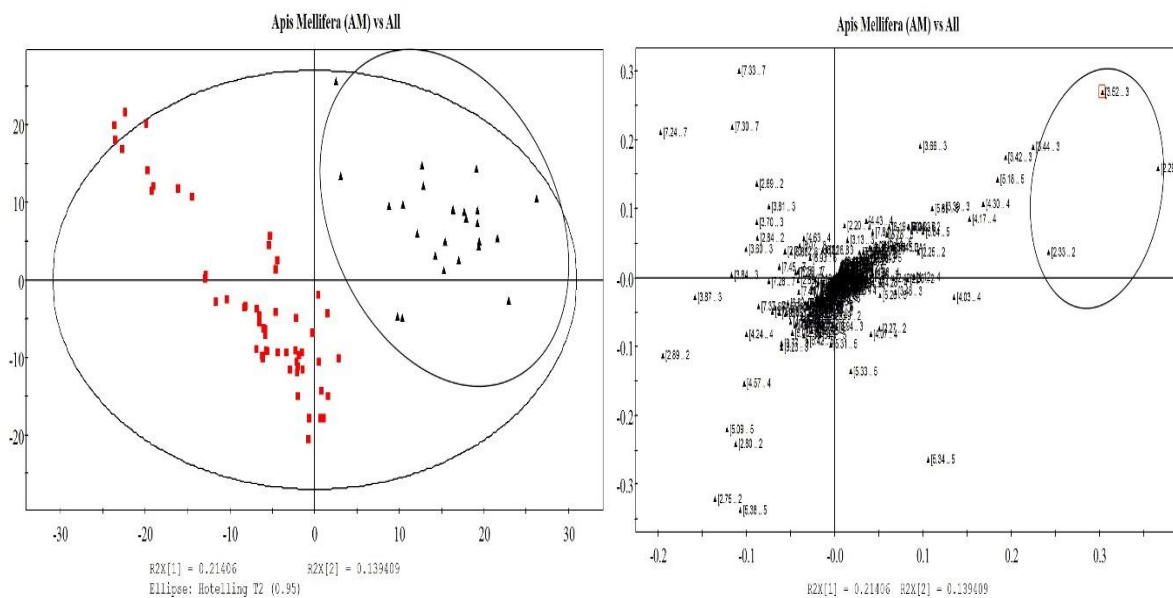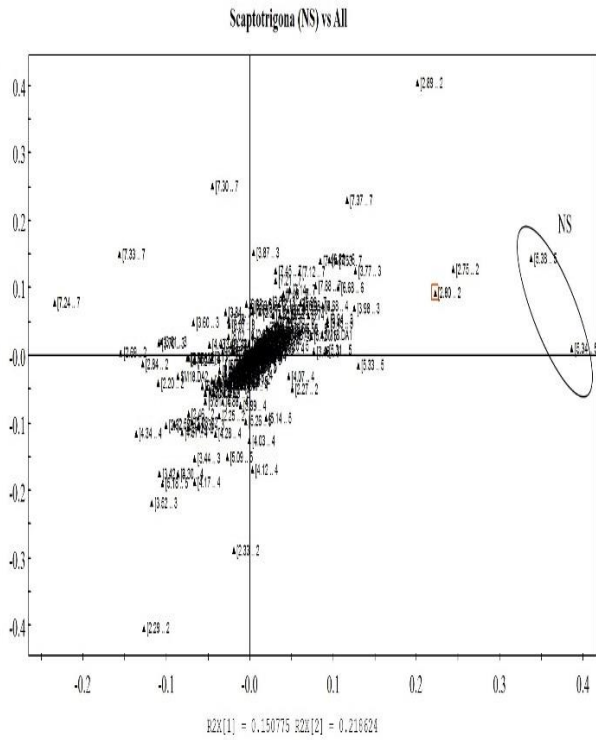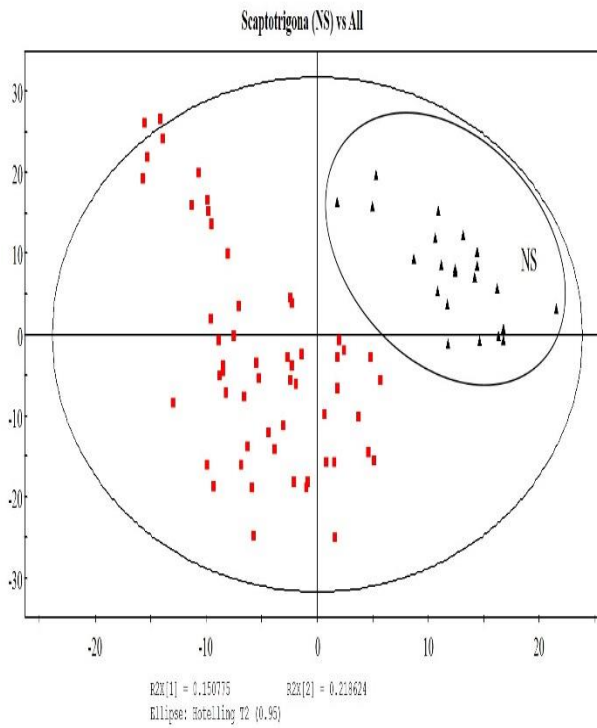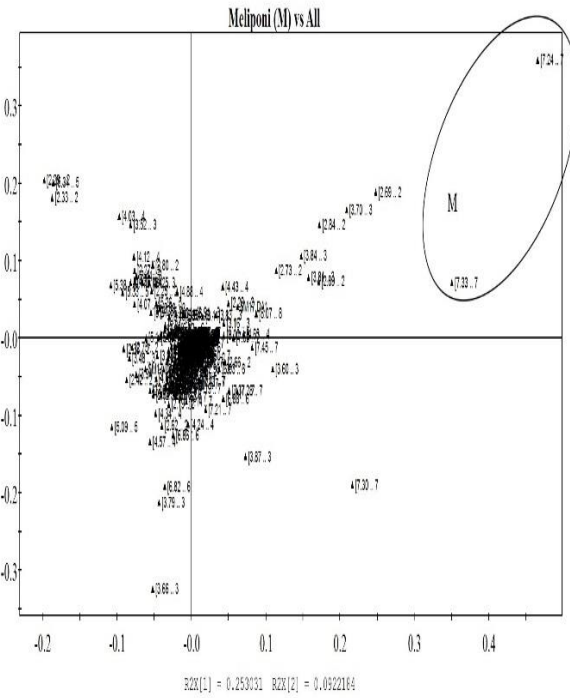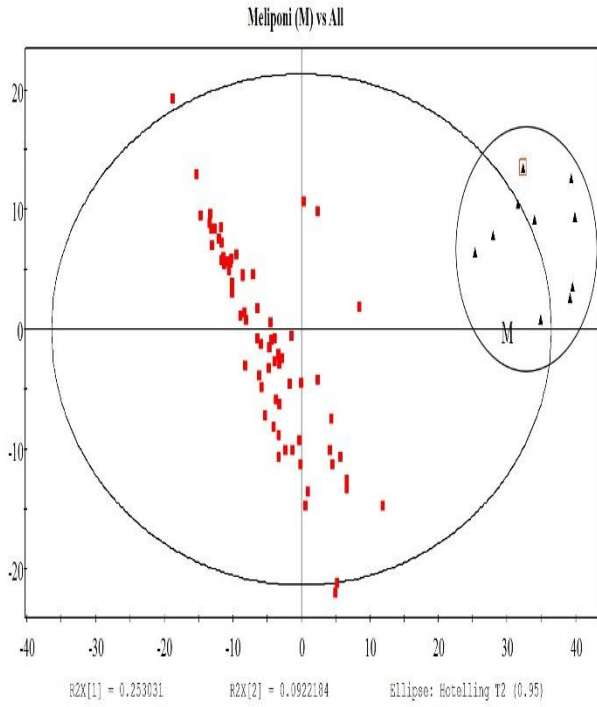
Figure 5.8: PLSDA scores plot of 4 honey types (left) and corresponding loadings plot (right). (a) *Apis mellifera* vs all; (b) *Melipona* vs all; (c) *Scaptotrigona* vs all; (d) *Geotrigona* vs all.

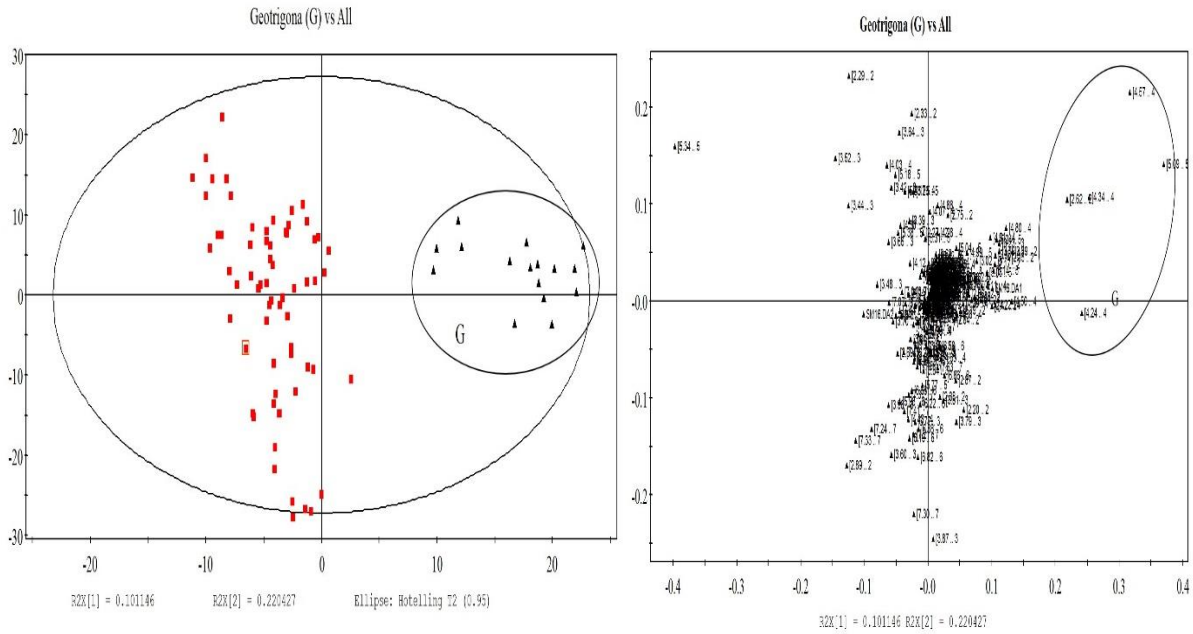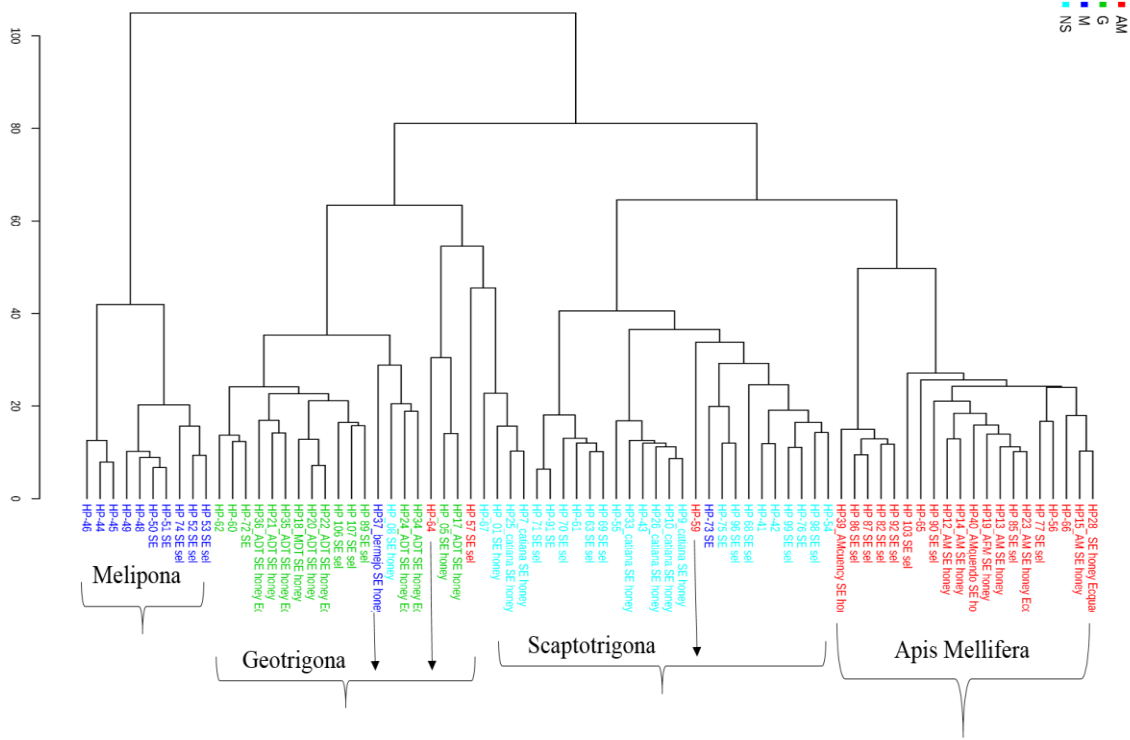**Figure 5.9:** Clustering result showing differences between four observations: *Apis mellifera*, *Scaptotrigona*, *Geotrigona* and *Melipona* as a dendrogram (distance measured using Euclidean, and clustering algorithm using Ward).

The discrimination among the four different honey types was supported with cluster analysis (dendogram) after PCA and PLSDA analysis. The dendrogram in **Figure 5.9** shows two main clusters (top to bottom). Three entomological classes of honey, namely *Apis mellifera*, *Scaptotrigona* and *Geotrigona* are merged into one big cluster whereas all *Melipona* honey samples are merged into one cluster. The two sub-cluster are merged into one big cluster through two nodes (bottom to top). These two sub-clusters are divided into 4 individual clusters as four unique observations represent each entomological origin. Two samples of *Melipona* honey, coded as HP73 and HP37 and indicated with arrows, are clustered along with *Geotrigona* and *Scaptotrigona*, respectively. Three samples of *Apis mellifera*, coded as HP64, HP57 and HP59 are also merged with *Scaptotrigona* and *Geotrigona*, respectively. Out of these five samples, the ones coded as HP64, HP37 and HP57 were declared as adulterated samples using a classical approach physical-chemical and melissopalynological analysis (performed by Dr. Patricia Vit, sample provider). The other two honey samples coded as HP73 and HP59 are misclassified.

5.5 Conclusions

As a continuation of our work focused on the origin discrimination of Italian honey with a chemometric approach by NMR spectrometry, we received 91 samples from Ecuador, 83 genuine honey and 8 false honey. In Ecuadorian markets, is it possible to find false honeys, composed entirely of inexpensive syrups or with a very little amount of genuine honey. Therefore, a method is needed which can be used as an alternative to classical ones to authenticate the honey and also to be used for entomological origin determination.

A chemometric analysis of ^{1}H-NMR data can discriminate between genuine and false honey. We found 5-Hydroxymethyl furfural, 2-Hydroxyacetyl-furan as main indicators of false honey. Moreover, Sorbic acid and Benzoic acid were not present in all false honey, but could be used as descriptors of false honey. We found a diacylglyceryl ether as the marker of genuine honey. The present study also demonstrates that a metabolomic approach based on ^{1}H-NMR and associated with PLS-DA and cluster analysis is a very suitable tool to distinguish the entomological origin of honeys. The PLS-DA analysis further identified resonances belonging to different compounds responsible for the discrimination of entomological origin. The proposed method of chloroform

110

extraction requires very little sample preparation, is fast, reproducible, allows one to obtain a fingerprint of different classes of compounds at the same time, and is more objective than melissopalynological analysis. To the best of our knowledge, there is no report published on the use of NMR-based metabolomics to discriminate the entomological origin. We are aware of one recently published method based on electrophoresis to discriminate the stingless bee species from Mexico [2]. The initial findings in this work would be very helpful in the future to identify the unknown resonances to establish the entomological origin of Ecuador honey.

5.6 References

[1]. Guerrini, A., Bruni, R., Maietti, S., Poli, F., Rossi, D., Paganetto, G., Muzzoli, M., Scalvenzi, L. and Sacchetti, G. Ecuadorian stingless bee (Meliponinae) honey: A chemical and functional profile of an ancient health product. *Food Chem.* **2009**, 114(4), 1413-1420.

[2]. Roubik, D.W. Ecology and natural history of tropical bees. **1992**. Cambridge University Press.

[3]. Ramón-Sierra, J.M., Ruiz-Ruiz, J.C. and de la Luz Ortiz-Vázquez, E. Electrophoresis characterisation of protein as a method to establish the entomological origin of stingless bee honeys. *Food Chem.* **2015**, 183, 43-48.

[4]. Jaffé, R., Pope, N., Carvalho, A.T., Maia, U.M., Blochtein, B., de Carvalho, C.A.L., Carvalho-Zilse, G.A., Freitas, B.M., Menezes, C., de Fátima Ribeiro, M. and Venturieri, G.C. Bees for development: Brazilian survey reveals how to optimize stingless beekeeping. *PloS one*, **2015**, 10(3).

[5]. Vit, P. La miel precolombina de abejas sin aguijón (Meliponini) aún notiene normas de calidad. *Rev. Bol. Centro Investigaciones Biológicas*, **2008**, 42, 415-423.

[6]. Instituto Colombiano de ormas Técnicas y Certificación. ICONTEC. NTC 1273. Norma Técnica Colombiana de Miel de Abejas. *Bogotá*: ICONTEC, **2007**, 1-21.

[7]. Vit, P., Fernández-Maeso, M.C., Ortiz-Valbuena, A. Potential use of the three frequently occurring sugars in honey to predict stingless bee entomological origin. *J. Applied Entomol.* **1997**, 122, 5-8.

[8]. Vit, P., Persano Oddo, L., Marano, M.L., Salas de Mejías, E. Venezuelan stingless bee honeys characterised by multivariate analysis of compositional factors. *Apidologie*, **1998**, 29, 377-389.

[9]. Vit, P., Deliza, R., Pérez, A. How a Huottuja (Piaroa) community perceives genuine and false honey from the Venezuelan Amazon, by free-choice profile sensory method. *Rev. Bras. De Farmacogn*, **2011**, 21, 786-792.

[10]. Schievano, E., Mammi, S., Menegazzo, I. Nuclear Magnetic Resonance as a method for prediction of geographical and entomological origin of pot honey". Pot-honey: A legacy of stingless bees. Vit. P.; Pedro, S.R.M.; Roubik, D., Eds. New York: Springer, **2013**, 429-445.

[11]. Truchado, P., Vit, P., Ferreres, F., Tomás-Barberán, F. Liquid chromatography-tandem mass spectrometry analysis allows the simultaneous characterization of C-glycosyl and O-glycosyl flavonoids in stingless bee honeys. *J. Chromatogr. A*. **2011**, 1218, 7601-7607.

[12]. Truchado, P., Vit, P., Heard, T., Tomás-Barberán, F.A., Ferreres, F. Determination of interglycosidic linkages in O-glycosil flavones by high-performance liquid chromatography/photodiode-array detection coupled to electrospray ionization ion trap mass. Its application to Tetragonula carbonaria honey from Australia. *Rapid Commun. Mass Spectrom.* **2015**, 29, 948-954.

[13]. Alves, R.M.O. "Production and marketing of pot-honey". Pot-honey: A legacy of stingless bees. Vit. P.; Pedro, S.R.M.; Roubik, D., Eds. New York: Springer, **2013**, 541-556.

[14]. Schievano, E., Morelato, E., Facchin, C. and Mammi, S. Characterization of markers of botanical origin and other compounds extracted from unifloral honeys. *J. Agri. Food Chem.* **2013**, 61(8), 1747-1755.

[15]. Basumallick, L. and Rohrer, J. Determination of Hydroxymethylfurfural in Honey and Biomass. *Thermo Scientific Dionex Appl. Note Update*, **2011**, 270, 1-6.

[16]. LeBlanc, B.W., Eggleston, G., Sammataro, D., Cornett, C., Dufault, R., Deeby, T. and St. Cyr, E. Formation of hydroxymethylfurfural in domestic high-fructose corn syrup and its toxicity to the honey bee (Apis mellifera). *J. Agri. Food Chem.* **2009**, 57(16), 7369-7376.

[17]. Zappala, M., Fallico, B., Arena, E. and Verzera, A. Methods for the determination of HMF in honey: a comparison. *Food Control.* **2005**, 6(3), 273-277.

[18]. Tfouni, S.A.V. and Toledo, M.C.F. etermination of benzoic and sorbic acids in Brazilian food. *Food Control.* **2002**, 13(2), pp.117-123.

[19]. Saad, B., Bari, M.F., Saleh, M.I., Ahmad, K. and Talib, M.K.M. Simultaneous determination of preservatives (benzoic acid, sorbic acid, methylparaben and propylparaben) in foodstuffs using high-performance liquid chromatography. *J. Chromatogr. A.* **2005**, 1073(1), 393-397.

[20]. Gok, S., Severcan, M., Goormaghtigh, E., Kandemir, I. and Severcan, F. Differentiation of Anatolian honey samples from different botanical origins by ATR-FTIR spectroscopy using multivariate analysis. *Food Chem.* **2015**, 170, 234-240.

[21]. Wang, L. and Mizaikoff, B., 2008. Application of multivariate data-analysis techniques to biomedical diagnostics based on mid-infrared spectroscopy. *Anal. Bioanal. Chem.* **2008**, 391(5), 1641-1654.

[22]. Ward Jr, J. H. Hierarchical grouping to optimize an objective function. *JASA*, **1963**, 58(301), 236-244.

Chapter 6. General Conclusions

The research that I have conducted at University of Padova over the course of three years was divided into two main topics under the common theme of "NMR-based metabolomics" in food chemistry: 1) The *in vivo* effects of nutraceutical products containing antioxidants on rat urine were evaluated using both NMR and MS; 2) An NMR-based metabolomic approach was applied to the discrimination of honey of different origins.

The role of NMR-based metabolomics in food sciences has been expanding in the last decade after it was successfully applied to diseases, toxicology and in the medicinal area. The information available in the field of food chemistry is still scarce, therefore its potential and capability need to be explored. Recently, NMR-based metabolomics started to gain appreciable popularity in food stuff where it can be applied to food authentication. There is no argument for the low sensitivity of NMR as compared to MS; still, NMR is a method of choice for metabolomic studies because of its reproducibility and quantitative and non-destructive nature. Beside food authentication studies, *i.e.*, origin discrimination of food matrices, NMR-based metabolomics recently also established itself in the assessment of phytochemical products and their metabolism in human and animal urine. Urine is a very convenient biological fluid to study because of the ease and non-invasive nature of multiple collections. Also, it is a biological waste which possesses important information of the living organism under study. The combination of NMR spectroscopy with data reduction techniques such as chemometrics to convert large datasets into valuable information have become a powerful tool in metabolomic studies, as we also prove in this thesis.

In the first two chapters, we used both ^{1}H-NMR and HPLC–MS techniques to study the modification of urinary composition in rats treated with *C. longa* L. and *P. cuspidatum* extracts and tried to correlate these changes with *in vivo* antioxidant activity. Multivariate analysis on ^{1}H NMR and HPLC–MS data produced similar but not identical representations of the collected samples. Both approaches were able to detect variations in the urinary metabolome, leading to the observation of different components, showing the complementarity of these two analytical techniques for metabolomic purposes.

The evaluation of the effects of *Curcuma* extract (Chapter 2) on urinary composition in healthy rats by a metabolomic approach led us to observe evidence for an *in vivo* antioxidant effect caused by a significant reduction in the amount of urinary biomarkers of oxidative stress, such as allantoin, m-tyrosine, 8-OHdG. A tendency to the reduction of 3-nitrotyrosine was also observed. Our study supports an *in vivo* antioxidant effect of the oral administration of *C. longa* extract to healthy rats. The observation that urinary TMAO levels are increased in the treated compared to the control group may be related to the influence of curcumin supplementation on microbiota, as recently indicated by other research groups, or on the urinary excretion of this metabolite. Urinary levels of taurine and cystine, sulphur containing compounds, were also changed suggesting a role for such constituents in the biochemical pathways involved in *C. longa* extract bioactivity and indicating the need for further investigation on curcumin effects.

In the third chapter, again both ^{1}H-NMR and UPLC-MS measurements were performed in a longitudinal study to assess the urinary metabolome changes induced by 49 days supplementation of *P. cuspidatum* to healthy rats. There were significant metabolic effects that could be ascribed to treatment or to aging. Our data showed that resveratrol from *P. cuspidatum* induced changes in the endogenous urinary profile, confirming the bioactivity of this antioxidant. The measurement of oxidative stress urinary markers, *i.e.,* allantoin and 8-OHdG confirmed the *in vivo* antioxidant effect of this *P. cuspidatum*. Cortisol derivatives were increased in the treated group suggesting a stimulating activity of this extract on steroid metabolism that need further investigations in future. Nevertheless, the analysis highlights the potential of NMR and MS based metabolomic analysis as a tool for *in vivo* investigations and the possibility that integrated metabolomic studies of other biofluids (*i.e.*, plasma) and tissues could provide a clearer metabolic picture in the future with the same antioxidants.

In chapters 4 and 5, we successfully used an NMR-based metabolomic approach in the origin discrimination of Italian and Ecuador honey with different botanical and entomological origins, respectively.

Specifically, two methods were adopted for 82 citrus honey samples (in chapter 2) to distinguish clementine from citrus honeys. A simple, direct CDCl$_3$ extraction method was developed that results in a better signal-to-ratio as compared to our old reported method with CHCl$_3$ liquid-liquid extraction. This new method requires less time in sample preparation and maintains all the

necessary information in the ^{1}H-NMR spectra. To the best of our knowledge, this is the first attempt to discriminate clementine from citrus honeys. Our D_2O extraction strategy revealed phenylacetic acid and tyrosine as marker compounds in citrus honey whereas higher amounts of formic acid characterized clementine honey. In the case of the $CDCl_3$ extraction, 8-hydroxylinalool and caffeine were found to be higher in the citrus honey whereas, clementine honey was characterized by the resonances of a yet unknown compound.

In chapter 5, a chemometric approach was used in combination with NMR spectrometry to discriminate the entomological origin of different honey samples from Ecuador. The study included also the distinction of false honey from genuine honey. In total, we received 91 samples, out of which 8 were false and 83 were genuine honeys produced by different bee species. Not surprisingly, we found 5-hydroxymethyl furfural and 2-hydroxyacetyl-furan as main indicators of false honey. Sorbic acid and benzoic acid were not present in all false honey, but could be used as descriptors of false honey. We found a diacylglyceryl ether as the marker of genuine honey. This study also demonstrated that a metabolomic approach based on ^{1}H-NMR and associated with PLS-DA and cluster analysis is a very suitable tool to distinguish the entomological origin of honeys. The multivariate statistical analysis also revealed yet unknown markers of different entomological origins. These preliminary results will be further pursued in the future to identify the resonances responsible for discrimination of different honeys.

In general, in chapter 4 and 5 we proposed that a ^{1}H-NMR-based metabolomic approach for origin discrimination of honey could be used as an alternative to the classical method of origin assessment through melissopalynological analysis and physical-chemical methods. The proposed method of chloroform extraction requires very little sample preparation, is fast, reproducible, allows one to obtain a fingerprint of different classes of compounds at the same time, and is more objective than melissopalynological analysis.