**UNIVERSITY OF NAPLES FEDERICO II**

DEPARTMENT OF PHARMACY

*PhD in Pharmaceutical Science – XXIX cycle*

# NMR-based metabolomics applications: *from food to human biofluids*

PhD Thesis

**Nunzia Iaccarino**

Tutor:
Prof. ANTONIO RANDAZZO

PhD Coordinator:
Prof. MARIA VALERIA D'AURIA

March 2017

# PREFACE

This Ph.D. thesis is submitted as a requirement for obtaining the Ph.D. Double Degree at the **University of Naples Federico II** (Italy) and **University of Copenhagen** (Denmark) with the further mention of *Doctor Europaeus*. It is written based on three-years research conducted by the author, Nunzia Iaccarino, both at the Department of Pharmacy in Naples, under the supervision of Prof. Antonio Randazzo, and at the Department of Food Science (Chemometrics and Analytical Technology Section) in Copenhagen, under the supervision of Prof. Søren Balling Engelsen.

The PhD research activity has been financed by Campania Regional Government (Regional Operational Programme of Campania, European Social Fund 2014-2020).

*Nunzia Iaccarino*

*Naples, March 2017*

# ABSTRACT

Metabolomics is the scientific discipline that identifies and quantifies endogenous and exogenous metabolites in different biological samples. Metabolites are crucial components of a biological system and they are highly informative about its functional state, due to their closeness to the organism's phenotype. This approach finds an increasing number of applications in many areas including medical, pharmaceutical, food and environmental sciences. The combined use of NMR spectroscopy and chemometrics techniques, is able to provide the metabolic "*fingerprint"* of the various samples.

This PhD project focused on the analysis of various samples covering a wide range of fields, namely, food and nutraceutical sciences, cell metabolomics and medicine using a metabolomics approach. Indeed, the first part of the thesis describes two exploratory studies performed on Algerian extra virgin olive oil and apple juice from ancient Danish apple cultivars. Both studies revealed variety-related peculiarities that would have been difficult to detect by means of traditional analysis. The second part of the project includes four metabolomics studies performed on samples of biological origin. In particular, the first study is related to a recent emerging field: cell metabolomics. Indeed, tumour cells (HTC116) were treated with novel anticancer drugs in order to understand their *in vitro* action. The aim of this study was also the development of a reliable experimental protocol for an efficient harvesting, quenching and extraction of cellular metabolites of human adherent cancer cell lines.

The second and the third studies concern the evaluation of the effects of functional food ingredients, namely β-glucans and phytosterols, on *in vivo* animal models. In particular, the hypocholesterolemic action of β-glucans was investigated by analysing rat plasma and faecal samples. This study confirmed the role of barley β-glucans in increasing faecal bile acids excretion in hypercholesterolemic rats and showed, for the first time, a modulation of the primary and secondary bile acid excretion, depending on the molecular weight of the β-glucan employed. In the other study, the effects of phytosterols on a murine colitis model, was investigated. NMR measurements on the liver metabolome revealed the role of these

plant sterols in restoring the homeostatic equilibrium of the living system. Thus, in both cases, the results suggest the appropriate use of these nutraceutical products. The last study explores the differences in the follicular fluid metabolome of hyper- and normoinsulinemic women affected with Polycystic Ovary Syndrome (PCOS). The study provides preliminary but interesting relationships between serum hormones and metabolites in follicular fluids.

# LIST OF PUBLICATIONS

**Paper I**

"Development of an optimized protocol for NMR metabolomics studies of human colon cancer cell lines and first insight from testing of the protocol using DNA G-Quadruplex ligands as Novel Anti- Cancer Drugs". I. Lauri, F. Savorani, N. Iaccarino, P. Zizza, L.M. Pavone, E. Novellino, S.B. Engelsen and A. Randazzo. *Metabolites*, 2016, 6 (1), pii: E4. doi: 10.3390/metabo6010004.

**Paper II**

"Characterization of monovarietal extra virgin olive oils from the province of Béjaïa (Algeria)". F. Laincer, N. Iaccarino, J. Amato, B. Pagano, A. Pagano, G. Tenore, A. Tamendjari, P. Rovellini, S. Venturini, G. Bellan, A. Ritieni, L. Mannina, E. Novellino, and A. Randazzo. *Food Research International*, 2016, 89, 1123-33.

**Paper III**

"Application of recursive partial least square regression for prediction of apple juice sensory attributes from NMR spectra". N. Iaccarino, C. Varming, M.A. Petersen, F. Savorani, A. Randazzo, B. Schütz, T.B Toldam-Andersen and S.B. Engelsen. *Proceedings of the XIII International Conference on the Applications of Magnetic Resonance in Food Science,* 2016, 7-11.

**Paper IV**

"$^{1}$H NMR-based metabolomics study on follicular fluid from patients with PolyCystic Ovarian Syndrome (PCOS)". N. Iaccarino, J. Amato, B. Pagano, A. Pagano, L. D'Oriano, S. Pelliccia, M. Giustiniano, D. Brancaccio, F. Merlino, E. Novellino, C. Alviggi, A. Randazzo. *Biochimica Clinica.* Submitted.

**Paper V**

"Barley β-Glucan supplement can modulate the primary and secondary bile acid excretion in rats depending on the β-glucan fine structure". N. Iaccarino, B. Khakimov, M. Skau Mikkelsen, T. Skau Nielsen, M.G. Jensen, A. Randazzo, and S.B. Engelsen. Manuscript draft ready for submission.

**Other publications**

1. "Identification of an acetal derivative of the piperonyl methyl ketone in tablets seized for suspected drug trafficking". J. Amato, N. Iaccarino, B. Pagano, M. Maglieri, G. Persico, R. Russo, C. De Caro, A. Calignano, E. Novellino, and A. Randazzo. *Forensic Toxicology*, 2014, 32, 311-316.
2. "Bis-indole derivatives with antitumor activity turn out to be specific ligands of human telomeric G-quadruplex". J. Amato, N. Iaccarino, B. Pagano, R. Morigi, A. Locatelli, A. Leoni, M. Rambaldi, P. Zizza, Biroccio A, E. Novellino, and A. Randazzo. *Frontiers in Chemistry*, 2014, 2:54. doi: 10.3389/fchem.2014.00054.
3. "Noncanonical DNA secondary structures as drug targets: the prospect of the i-motif". J. Amato, N. Iaccarino, A. Randazzo, E. Novellino, and B. Pagano. *ChemMedChem,* 2014, 9, 9, 2026-2030.
4. "Identification of novel interactors of human telomeric G-quadruplex DNA". B. Pagano, L. Margarucci, P. Zizza, J. Amato, N. Iaccarino, C. Cassiano, E. Salvati, E. Novellino, A. Biroccio, A. Casapullo, and A. Randazzo. *Chemical Communications*, 2015, 51, 14, 2964-2967.
5. "Looking for efficient G-quadruplex ligands: evidence for selective stabilizing properties and telomere damage by drug-like molecules". B. Pagano, J. Amato, N. Iaccarino, C. Cingolani, P. Zizza, A. Biroccio, E. Novellino, and A. Randazzo. *ChemMedChem,* 2015, 10, 4, 640-649.
6. "Discovery of the first dual G-triplex/G-quadruplex stabilizing compound: a new opportunity in the targeting of G-rich DNA structures?". J. Amato, A. Pagano, S. Cosconati, G. Amendola, I. Fotticchia, N. Iaccarino, J. Marinello, A. De Magis, G. Capranico, E. Novellino, B. Pagano, and A. Randazzo. *Biochimica et Biophysica Acta*, 2017, in press.

# LIST OF ABBREVIATIONS

| | |
|---|---|
| ADP | Adenosine diphosphate |
| AMP | Adenosine monophosphate |
| ATP | Adenosine triphosphate |
| BG | Mixed-linkage (1→3),(1→4)-β-D-glucan |
| CPMG | Carr-Purcell-Meiboom-Gill (NMR pulse sequence) |
| DNA | Deoxyribonucleic acid |
| DSS | Sodium 2,2-dimethyl-2-silapentane-5-sulfonate |
| EVOO | Extra-virgin olive oil |
| FID | Free induction decay |
| GC-MS | Gas chromatography-mass spectrometry |
| G4-DNA | G-quadruplex DNA |
| HMDB | Human Metabolome Database |
| NMR | Nuclear Magnetic Resonance |
| NOESY | Nuclear Overhauser Spectroscopy |
| PARAFAC | Parallel Factor Analysis |
| PC | Principal Component |
| PCA | Principal Component Analysis |
| PCOS | Polycystic Ovary Syndrome |
| PLS | Partial Least Squares regression |
| rPLS | recursive weighted Partial Least Squares regression |
| STOCSY | Statistical Total Correlation Spectroscopy |

# TABLE OF CONTENTS

# Chapter 1
# INTRODUCTION

## 1.1 **Objectives of the project**

The overall goal of this PhD project was to explore the different fields of application of the NMR-based metabolomics, using both traditional and more recent developed chemometrics tools for the data analysis.

Different studies, ranging from Food to the Pharmaceutical Sciences, have been carried out aiming to retrieve useful information employing the "*omics*" approach. Indeed, both food systems (olive oil and apple juice) as well as biological samples (cells extracts, tissue extracts, plasma, follicular fluids and faeces) have been analysed.

## 1.2 **Brief description of the outline of the thesis**

The thesis is subdivided as follow:

*Chapter 1* describes the project and its main objectives.

*Chapter 2* presents a general description of metabolomics science, focusing in particular on the different scientific approaches available, the metabolomic workflow and the main analytical platforms employed in metabolomics studies.

*Chapter 3* presents an overview of the principles of the nuclear magnetic resonance spectroscopy, the main analytical platform used during this PhD project, and the most common processing and pre-processing tools needed to perform the data analysis.

*Chapter 4* presents an overview of the main chemometrics techniques used in the studies presented in this thesis. The objective is to guide the reader into the understanding of the basic principles of the chemometrics and multivariate data analysis.

*Chapter 5* describes the importance of the NMR-based metabolomics in food science. Two studies are presented. The first concerns the characterization of Algerian extra virgin olive oil, while the second is about the NMR fingerprinting of Danish apple juices.

*Chapter 6* presents four metabolomics studies performed on various biological samples, showing the wide range of applications that NMR and chemometrics allow to explore.

*Chapter 7* presents the main conclusions of the thesis work.

Finally the *Appendix* reports the published publications (Papers I, II and III) and the submitted one (Paper IV) as cited throughout the thesis, while Paper V is not included since it has not been submitted at present time.

# Chapter 2
# METABOLOMICS

## 2.1 **General remarks**

The term *metabolomics* was introduced for the first time by Oliver Fiehn (Fiehn 2002). It refers to the identification and quantification of the collective set of low molecular weight metabolites (<1.5 kDa), such as, fatty acids, nucleic acids, carbohydrates, amino acids organic acids, vitamins, polyphenols and small lipids (metabolome). However, in 1999, Jeremy Nicholson and coworkers had defined the term *metabonomics* as "the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification" (Nicholson et al. 1999). The distinction between terms *metabolomics* and *metabonomics* has been widely discussed (Ramsden 2009). Even if the concepts are slightly different, the two terms are often used interchangeably by scientists and organizations (Nicholson & Lindon 2008).

'-*Omic*' sciences adopt a holistic view of the molecules contained in cells, tissue or organism. The suffix "-*omic*" has been added to denote studies performed on a very large-scale data analysis (i.e. measuring/profiling a large number of variables simultaneously). Their primary aim is the detection of genes (genomics), mRNA (transcriptomics), proteins (proteomics) and metabolites (metabolomics) in a specific biological sample in a non-targeted and non-biased manner (Berry et al. 2011).

The metabolome can be considered as a snapshot of the physiological state of an organism, being the downstream product of the "-*omic"* cascade, as well as the ultimate response to disease or environmental influences (Figure 2.1). Indeed, whereas genes and proteins are subjected to regulatory epigenetic processes and post-translational modifications, respectively, metabolites are the closest ones to the phenotype. Thus, it is easier to correlate metabolomic profiles with phenotype compared to genomic, transcriptomic and proteomic profiles. Applications of metabolomics include disease diagnosis, monitoring the effects of medical interventions including drugs, detection of adulteration of food, and analysis of biochemical pathways and their perturbations resulting from mutations, aging, diet, exercise, or life style (Markley et al. 2017). The size of the metabolome is large and it goes from 600

metabolites of *S. cerevisiae* (Förster et al. 2003) to up to more than 200 000 metabolites of the plant kingdom (Fiehn 2001). The number of important metabolites in humans is greater than these values and it surely bigger than the currently represented on metabolic charts. The complexity of the metabolome is raised by the swapping of metabolites between pathways increasing the network of relevant reactions by an astonishing amount.

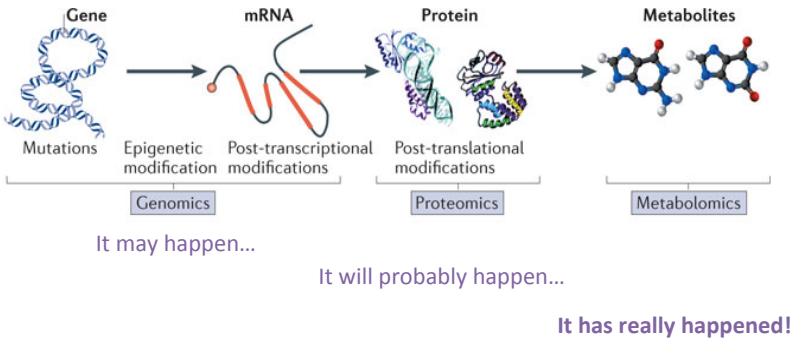


**Figure 2.1. The central dogma of biology and the omic cascade.** *[Adapted from* (Patti et al. 2012)*]*

The Human Metabolome Database (HMDB) (Wishart et al. 2013) lists 42 000 metabolites and the number of lipid variants is on the order of 100 000; thus, a lower limit of expected endogenous and exogenous human metabolites is around 150 000, but the actual number of metabolites could be much higher. Among these metabolites, only 1500 may be identified from global profiling, 200 – 500 from targeted profiling, and far fewer are routinely subjected to quantitative analysis (Markley et al. 2017). These numbers indicate, on one hand, the extreme complexity that is behind the metabolomics approach, and on the other hand, its extraordinary potential.

### 2.1.1 *Targeted and Untargeted approach*

Two different approaches can be used in metabolomics. The most suitable one for the study purpose depends on the kind of information that is sought from the metabolomic analysis.

A *targeted approach* is the proper choice when a defined set of metabolites has to be investigated. This method focuses on the absolute quantification of metabolites, that have been identified in advance, and which are highly related to a specific pathway or intersecting

pathways, after an appropriate sample preparation. This kind of analysis, also known as *targeted profiling* or *quantitative metabolomics,* is characterized as being a hypothesis-driven approach rather than a hypothesis-generating approach. The targeted approach has been successfully used in the β-glucan study (*Paper V)*, where a specific class of metabolites (bile acids and neutral sterols), assumed to be involved in the mechanism of action of the tested compound, has been first extracted and then analysed.

On the other hand, an *untargeted or global profiling approach* provides an unbiased overview of the metabolome that characterizes the sample and can reveal novel and unanticipated perturbations. It represents an unbiased tool to examine the relationship between interconnected metabolites from multiple pathways (Johnson et al. 2016). Most of the studies presented in this thesis were conducted using the untargeted approach.

### 2.1.2 *Metabolomics workflow*

All kinds of metabolomic studies need to follow a specific workflow in order to achieve appropriate and reliable results. A summary of the crucial steps to be performed is presented in Figure 2.2.

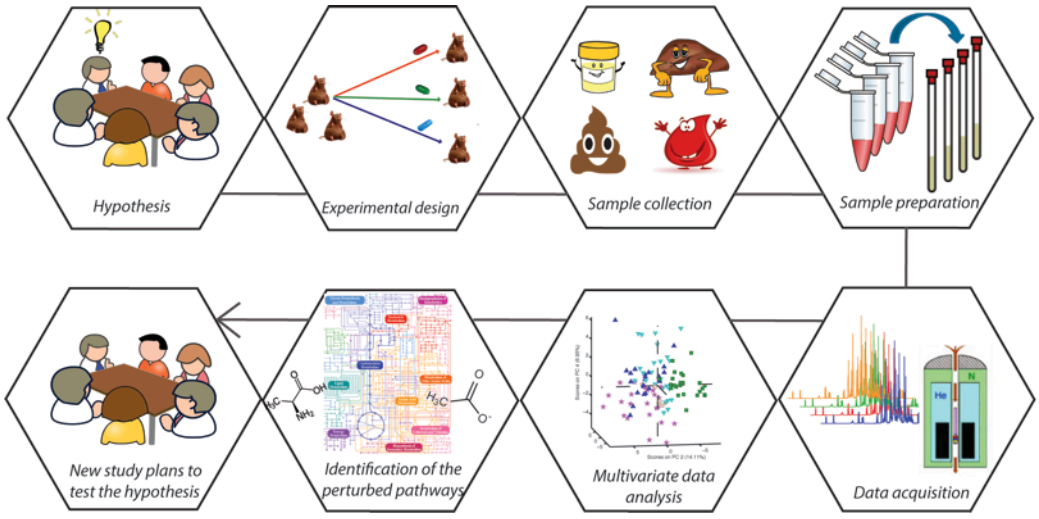


**Figure 2.2. Workflow of a metabolomic study: hypothesis, experimental design, sample collection, sample preparation, data acquisition, multivariate data analysis, interpretation of the results (i.e. identification of the perturbed metabolic pathways) and eventually test the new hypothesis.**

The first step is the formulation of a hypothesis/idea followed by the set up of the experimental design. Usually, the main aim is to discover a set of metabolites (known or unknown) that allow to distinguish samples from two different groups or populations and

thus, identifying the so-called biomarkers. As in any kind of experiment, a proper experimental design, and, in particular, the sample size is critical if statistically reliable information has to be obtained. Sample collection, preparation and storage are crucial steps since only a correct sampling provides a real snapshot of the metabolome at a certain point in time, thus strict protocols both for the collection and the preparation of the samples have to be followed by the all the operators involved in the project. For the most common analysed biofluids, such as plasma and urine, several protocols are available in literature (Beckonert et al. 2007). However, there is an increasing need of a unique standardized protocol for sample selection, collection, storage and preparation in NMR-based metabolomic studies in order to avoid the identification of spurious biomarkers due to a general lack of reproducibility between laboratories (Emwas et al. 2015). In the case of cells or tissues, the adopted procedures for metabolite extraction highly influence the nature and levels of the extracted metabolites. Generally, the procedure includes the disruption of cell walls and subsequent distribution of metabolites into polar (methanol, water) and non-polar (chloroform, hexane, ethyl acetate) solvents followed by the removal of the cellular residue.

### 2.1.3 *Analytical platforms*

Nuclear Magnetic Resonance (NMR) spectroscopy and Mass Spectrometry (MS), are the most commonly used platforms in metabolomics (Dunn & Ellis 2005). Both present advantages and drawbacks, thus there is not a single analytical technique completely suitable for metabolomic studies. Indeed, NMR and MS have been demonstrated to be complementary and powerful analytical approaches for the complete characterization of the metabolome (Pan & Raftery 2007).

Mass spectrometry measures the masses of molecules and their fragments to determine their identity. This information is gained by measuring the mass-to-charge ratio (*m/z*) of ions that are formed by inducing the loss or gain of a charge from a neutral species. The sample, comprising a complex mixture of metabolites, can be introduced into the mass spectrometer either directly or preceded by a separation approach, namely, liquid chromatography (LC) or gas chromatography (GC). In addition to *m/z* and retention time information, the identification of an ion is facilitated by fragmentation pattern information that is acquired by tandem mass spectrometry (Want et al. 2005). Usually LC is employed for the analysis of a wide range of non-polar compounds, while GC is applied for the analysis of low molecular weight volatile

compounds. Nevertheless, in GC-MS, the detection and quantification of many compounds requires a pre-analysis derivatization step and a pre-selection of the "expected" metabolites. This represents the main limitation of the methodology, since non-derivatized chemical classes are lost in the analysis. The main advantage of MS is the high sensitivity, and it allows to detect traces of a metabolite (*pg* level) in the sample (Johnson et al. 2016).

NMR spectroscopy offers many advantages over MS (Fan & Lane 2016; Nagana Gowda & Raftery 2015). It allows identification and quantification of the more abundant bulk metabolites present in biofluids, cell extracts, and tissues without any sort of sample pre-treatment or fractionation. Due to its non-selectivity, no prior knowledge of the samples is required. Moreover, it allows the identification of compounds having identical masses, including those with different isotopomer distributions. In spite of the high number of advantages, NMR spectroscopy has its main drawback in sensitivity, with limits of detection on the order of 10 μM or a few nmol at high fields using new cryoprobes (Pan & Raftery 2007).

# Chapter 3
# NMR BASED METABOLOMICS

Nuclear Magnetic Resonance (NMR) spectroscopy is a valuable analytical technique in metabolomics since it allows the qualitative and quantitative analysis of chemical compounds from complex mixtures as well as the structural elucidation of unknown compounds. It provides a complete picture of the chemical composition of the analysed sample since is not selective for special groups of molecules. Furthermore, no sample pre treatment and no sample destruction is necessary during the NMR experiment, allowing the re-use of the same sample for subsequent analysis.

The most important nuclei in biomolecular NMR studies are $^{1}H$, $^{13}C$, $^{15}N$, and $^{31}P$. Among these nuclei, $^{1}H$ is the most sensitive followed by $^{31}P$; for both of them the natural abundance is near 100%. However, the majority of the observed compounds in metabolomics studies are not phosphorylated, thus the proton is the most used nucleus in this field. This chapter will provide an overview both of the principles of Nuclear Magnetic Resonance and of the processing steps needed for its application to metabolomics studies.

## 3.1 **Principles of Nuclear Magnetic Resonance Spectroscopy**

NMR was first described by Isidor Rabi in 1938 that was then awarded the Nobel Prize in physics for his work. Rabi and his team employed a modification of Otto Stern's apparatus to measure the magnetic properties of various isolated nuclei such as hydrogen, deuterium, and lithium. However, it was only in 1945 that Felix Bloch at Stanford and Edward Mills Purcell at the Massachusetts Institute of Technology simultaneously demonstrated NMR in condensed matter (water and paraffin, respectively) (Bloch et al. 1946; Purcell et al. 1946). Seven years later, Bloch and Purcell jointly received the Nobel Prize for physics.

Being a spectroscopic method, NMR is based on the interaction between energy and matter (Weber and Thiele 2008), but differently from other spectroscopic techniques it needs a strong static magnetic field.

The usefulness of NMR in chemistry was not appreciated until 1950 when the importance of the so-called *chemical shift* was discovered (Proctor & Yu 1950). Indeed, it is in chemistry that NMR has attracted the greatest interest in the recent years.
In the following sections, the main theoretical aspects of NMR are discussed.

### 3.1.1 *Nuclear spin and resonances*

Nuclear Magnetic Resonance is a property of the nucleus of an atom, which consists of protons and neutrons, related to what is known as *nuclear spin* (*I*). This is equivalent to the nucleus acting like a miniature bar magnet. The spin of a nucleus depends on the mass of the isotope and nuclei with even mass and even charge numbers have no spin angular momentum ($I$=0). These kinds of nuclei are called "NMR inactive" or "NMR silent" since the nuclear spin property is fundamental to enable NMR. Nuclei as hydrogen ($^{1}$H), carbon ($^{13}$C), fluorine ($^{19}$F) and phosphorus ($^{31}$P), have $I=1/2$, thus they can be analysed using NMR.

As showed in Figure 3.1, when the nuclei are not affected by external magnetic field ($B_0$), the spin are randomly oriented in all direction, while if nuclei with $I\neq0$ are placed in a magnetic field, they will assume a possible number of different orientations that will correspond to specific energy levels; this number depends on the value of *I*, in particular it is equal to $2I+1$.

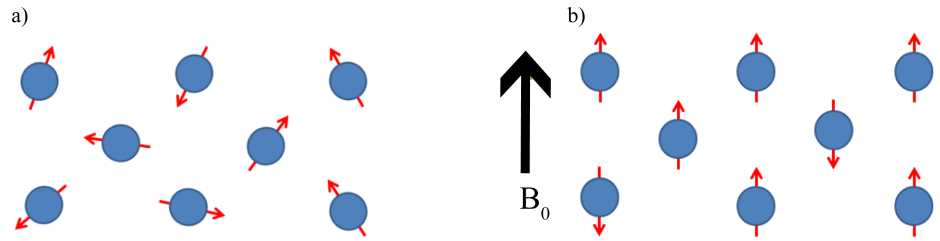


**Figura 3.1. a) Randomly oriented spins in the absence of a magnetic field. b) Aligned spins in the presence of an applied magnetic field ($B_0$)**

The proton ($^{1}$H) is the most abundant NMR nucleus and it has I= ½ , therefore when $B_0$ is applied, these nuclei can assume two possible orientations, α parallel (I = ½ ) or β antiparallel (I = - ½ ), each corresponding to an energy level.
The difference in the energy levels is equal to:

$$\Delta E = \gamma \cdot B_0 \tag{3.1}$$

where $\gamma$ is the *gyromagnetic ratio* that is constant for a given nucleus ($\gamma = 2.6752 \cdot 10^8$/T/s for protons).

This means that, given a nucleus, to stronger magnetic field will correspond higher gap in the energy levels, thus resulting in higher sensitivity in the NMR experiment.

The protons are distributed between the two energy states according to the Boltzmann distribution:

$$\frac{N_\alpha}{N_\beta} = \exp\left(\frac{\Delta E}{k \cdot T}\right) \tag{3.2}$$

Here $N_\alpha$ and $N_\beta$ represent the protons populations in the lower and upper energy levels, respectively, *k* is the Boltzmann constant and T is the temperature.

The number of nuclei is not equal in the two states: a small excess of protons will occupy the lower energy state (α) since it is, indeed, the more favourable state from an energetic point of view. This gives rise to a net magnetization $M_0$, aligned with the applied magnetic field $B_0$.

As showed in Figure 3.2, when a radio frequency pulse is applied, the nuclei will absorb energy and nuclear spins transitions from lower to higher energy levels will be induced, thus giving:

$$\Delta E = h \cdot v \tag{3.3}$$

where *h* is the Planck constant and *v* is the frequency of the excitation pulse that induces the transitions between the levels. This frequence is referred to as *Larmor frequency* and depends both on the nucleus and the magnetic field as showed in the following equation:

$$v = \frac{\gamma \cdot B_0}{2 \cdot \pi} \tag{3.3}$$

No NMR signal is observable when the sample is at equilibrium (in the static magnetic field), since the net magnetization vector has no component on the xy plane where the signal is detected by the detector coil. The duration of the pulse is usually measured in microseconds

(μs), and it is called 90° pulse when it brings the magnetization on the xy-plane, whereas a 180° pulse flips the net magnetization on the negative z-axis.

As illustrated in Figure 3.2, when the radiofrequency is switched off, the system will return to equilibrium. This return to equilibrium is referred to as relaxation and it causes the NMR signal to decay with time, producing the observed free induction decay (FID). The NMR signal is then Fourier transformed to be converted in the frequency domain.

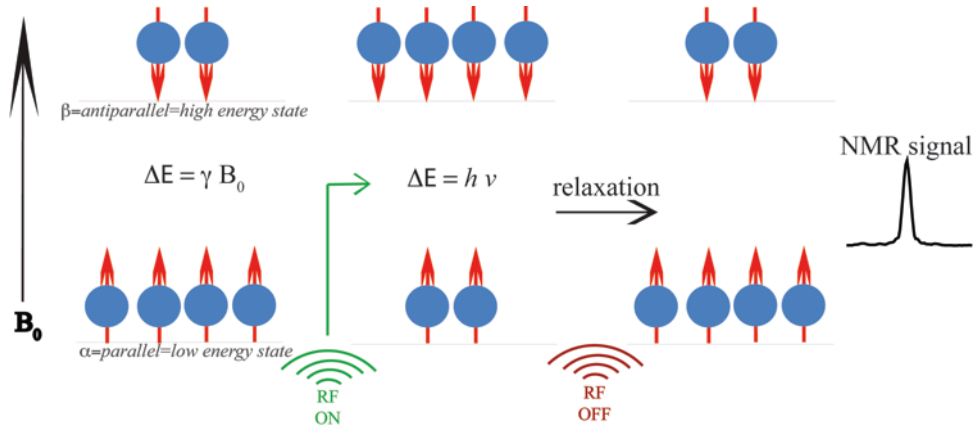


**Figure 3.2. Schematization of the NMR experiment.**

### 3.1.2 *Chemical shift and couplings*

Protons are situated in different positions in a molecule, thus they do not experience the same magnetic field for a phenomenon known as *shielding*. Indeed, the nucleus is surrounded by electrons that start a rotational motion when an external magnetic field ($B_0$) is applied. This gives rise to a small local magnetic field $B_{loc}$ , that may oppose the external field, and, as a consequence, the nucleus experiences a slightly reduced field, that can be called $B_{eff.}$ A specific parameter (σ, shielding constant) is used to indicate the density and the distribution of the electronic cloud that surrounds the nucleus. This constant ranges from $10^{-6}$, for the lighter nuclei, to $10^{-3}$ for the heavier ones. The changes in the σ value are given, for example, by the presence of functional groups in the vicinity of a nucleus. In particular, when an electronegative atom is present, it will withdraw electrons from the observed nucleus, reducing the density of the electronic cloud, thus causing a de-shielding effect. Thus, the nucleus will resonate at higher frequencies. Considering that $B_{loc}$ is equal to $B_0$ σ (Lenz rule), $B_{eff}$ is given by:

$$B_{eff} = B_0 - B_{loc} = B_0 - B_0\sigma = B_0(1-\sigma) \quad (3.4)$$

As a consequence the relationship between the degree of shielding and the resulting resonance frequency is:

$$\nu = \frac{\gamma \cdot B_0}{2 \cdot \pi} * (1-\sigma) \quad (3.5)$$

Therefore, protons situated in different chemical environments will experience different magnetic fields, thus meaning that they will resonate at different frequencies giving rise to the so-called chemical shift (δ). All the information retrievable from a NMR experiment is hidden in the chemical shift, thus it is crucial to only look at the frequencies of the different nuclei in a way that is independent from the magnetic field employed. For this reason a conventional way to calculate δ has been established:

$$\delta = \frac{\nu \cdot \nu_{ref}}{\nu_{ref}} * 10^6 \quad (3.6)$$

where $\nu$ is the frequency of the observed nucleus and $\nu_{ref}$ is the frequency of a reference compound. The most widely used reference compounds are tetramethylsilane, $Si(CH_3)_4$ (simply called TMS) for organic solvents and the sodium salt of trimethylsilyl proprionic acid (TSP) for aqueous solutions. Both compounds have maximum shielding, showing higher σ than the nuclei usually analysed via NMR. Therefore, the equation 3.6 converts the chemical shift frequencies into parts per million (ppm), allowing to display the NMR signals on a new axis on which a given spin always shows the same value independently of the magnetic field employed. In this new system the reference compound has δ=0 while the sample resonances have positive δ values.

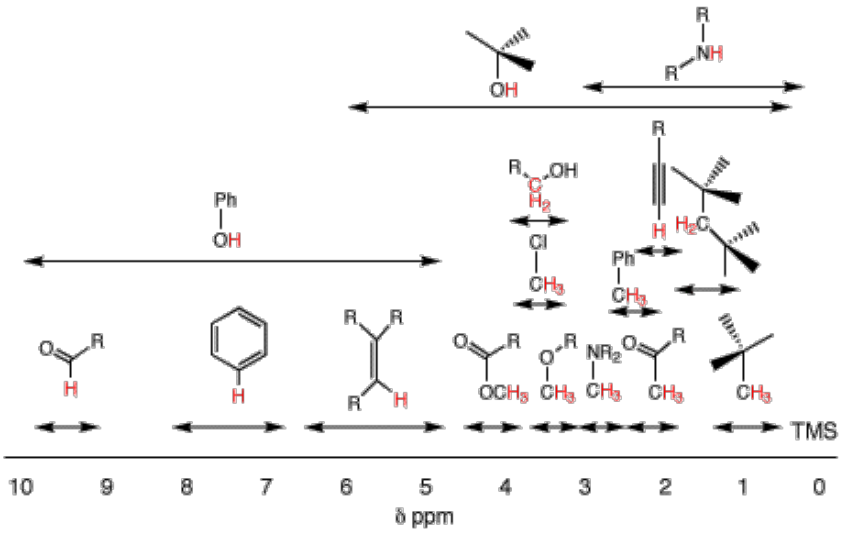


**Figure 3.3. Approximate proton chemical shift.** [*Adapted from Orgchemboulder.com*]

Chemical shift is not the only value to inspect to retrieve chemical information from the NMR experiment. Another valuable feature is the *J coupling* or *scalar coupling*. This phenomenon is due to the fact that the magnetic moments of the nuclei can influence each other in two possible ways: through space (dipolar coupling) or trough chemical bonds (scalar coupling). The effect of the dipolar coupling can be neglected since the interactions average to zero because of the rapid molecular tumbling while the scalar coupling is visible from the NMR spectrum.

Indeed, when an atom is influenced by another, this results in the split of its resonance signal. This is valid in both directions, thus when the perturbing nucleus becomes the observed one, it also exhibits signal splitting. In order to observe the splitting of the signal, the interacting nuclei must be bonded in proximity (e.g. vicinal and geminal positions), or be oriented in certain optimal configurations. Usually the signal splitting in proton spectra ranges from fractions of Hz to around 18 Hz.

### 3.1.3 *Detection and Fourier transform*

After the nuclei excitation by means of a radiofrequency pulse, the net magnetization will flip on the *xy* plane (90° pulse). Here the spins start to *precess* (rotate) both clockwise and counter-clockwise, thus in order to distinguish the spin with a frequency $-\nu$ and one with a frequency $+\nu$, a detection system with two simultaneous channels (reciprocally 90° out of phase), on each axis, is often employed. This is called *quadrature detection.*

Therefore, after the pulsed excitation, the nuclei return to the equilibrium distribution between the two energy levels trough a process called *relaxation* with relaxation times that are specific for different nuclei. This process generates the free induction decay (FID). The FID is a time-domain representation of the superimposition of the frequencies of all the nuclei present in the analysed sample, thus it is not easy to interpret. *Fourier transform* is a mathematical operation that can be carried out on the final FID data to produce the familiar frequency spectrum (Figure 3.4).

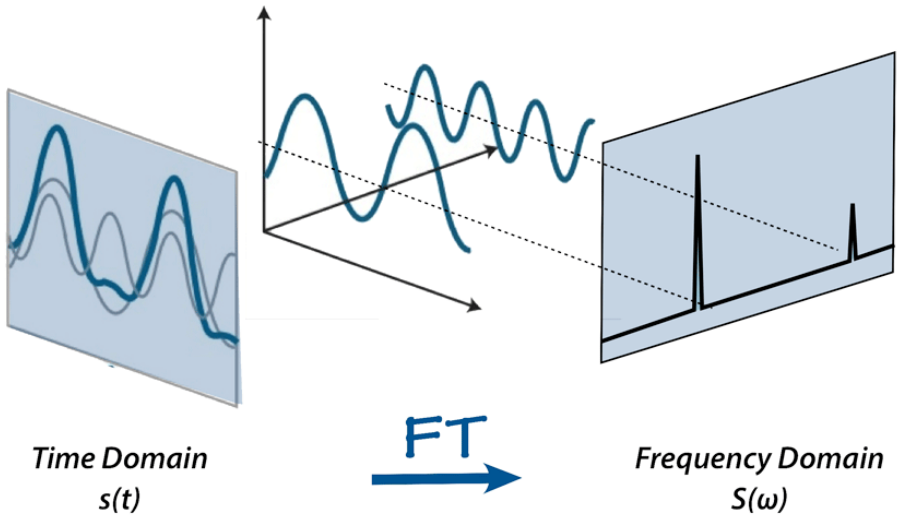


**Figure 3.4. Illustration of the Fourier Transformation from the the Time Domain s(t) to the Frequency Domain S(ω).** [*Adapted from mriquestions.com*]

### 3.1.4 *Processing tools*

In order to enhance the sensitivity and the resolution of the acquired data, a series of processing steps have to be performed before and after the Fourier transform of the FID.

In this section, zero filling, apodization, phase correction, and baseline correction will be briefly discussed.

*Zero-filling* allows to increase the digital resolution of the NMR spectra by adding zeros the end of the FID data points, just before performing the Fourier Transform. This will add data points to the FID without adding additional noise. It is important to note, however, that zero-filling does not improve true resolution; it only improves the apparent resolution. This can be very useful because fine coupling may not be visible due to low digital resolution. Furthermore, it has to be stressed that setting the acquisition time to a very short value, and

then use zero-filling to increase the digital resolution is not a good choice. Indeed, the FID has to naturally fall to zero and only after that, the zeros point can be added, otherwise baseline artifacts, as sinc wiggles, will appear after the FT.

*Apodization* consists in multiplying the FID with different window (weighting) functions (Lorentzian, Exponential, Gaussian, or Sine-bell function) which can be chosen either to enhance the sensitivity or resolution (or both if possible) in the final spectrum. For instance, exponential multiplication leads to line broadening and a reduction in noise, while other trigonometric functions produce narrowing of the spectral line while increasing the noise.

*Phase correction* allows the correction of phase errors in the spectra, usually generated by two main phenomena (i) delays (in terms of microseconds) between the RF pulse and the opening of the receiver for the FID acquisition (ii) off-resonance effects due to pulse inability to equally excite all the nuclei.

A spectrum that has not been phase-corrected has signal with a dispersive line shape as well as inverted signals. Therefore, a zero-order and first order phase corrections are usually employed to face this issue. Zero-order correction is chemical shift independent and thus, it is the same for all lines across the spectrum. Whereas the first-order correction is frequency dependent, therefore it applies a phase change that increases linearly with the distance to the reference signal.

*Baseline correction* is a crucial feature in metabolomics. Indeed, a flat baseline is fundamental for an accurate integration and therefore quantification of the chemical compounds. Furthermore small peaks could be hidden under a distorted baseline. This issue is usually caused by the corruption of the first few data points in FID that add low frequency modulations in the Fourier-transformed spectrum, thus forming the distorted baseline. The main reasons that cause this phenomenon are: too high signal amplification or a not complete recovery of the electronics from the RF.

The most common correction method is, first, the fitting of the baseline with a polynomial function and then the subtraction of the baseline from the spectrum.

### 3.1.5 *One dimensional and two dimensional NMR experiments*

One dimensional (1D) $^{1}$H NMR has formed the bedrock of metabolomics studies to date, for (i) the rapidity in the spectral acquisition (ii) the possibility to directly measure the metabolite concentration by integrating the peak area using an internal standard.

1D experiments performed in this PhD project include two of the most used pulse sequences in the analysis of biofluids and aqueous extracts: Carr-Purcell-Meiboom-Gill (CPMG) and 1D Nuclear Overhauser Effect Spectroscopy (NOESY) pulse sequences.

In particular, 1D NOESY-presat was employed for the analysis of the aqueous extracts of cancer cell (*Paper I*), apple juices samples (*Paper III*), and rat plasma (*Paper V*).

1D NOESY-presat employs the first increment of a NOESY pulse sequence, with water irradiation during the relaxation delay and also during the mixing time (on a Bruker instrument, this is called noesypr1d). This has the form −RD-90°-*t*-90°-$t_m$-90°-ACQ, where RD is the relaxation delay, *t* is a short delay typically of ~3 μs, 90° represents the RF pulse, $t_m$ is the mixing time and ACQ is the data acquisition period. Usually, gradients are also part of the pulse sequence to improve the solvent suppression quality (e.g., noesygppr1d)(Beckonert et al. 2007).

CPMG experiment is based on the spin-echo pulse sequence and consists of $-90°_x$-($t_E$-180°- $t_E$ )$_n$, where $t_E$ represents the so-called *echo time* and *n* is the number of repetition of the block in parenthesis (Carr & Purcell 1954). By carefully choosing these two parameters, the signals in the spectrum can be separated according to their spin-spin relaxation time ($T_2$). This experiment helps to attenuate, or even eliminate, the broad signals from macromolecules or bound small molecules, allowing the visualization of the sharp peaks given by the mobile small molecules (Tang et al. 2004). Figure 3.5 shows the comparison between 1D-NOESY and CPMG NMR experiments performed on rat plasma samples during the study described in the section 6.2 of this thesis. The low-molecular-weight metabolites (aminoacids, sugars and acids) are more clearly visible in the CPMG spectrum than in the regular $^1$H NMR spectrum, since the latter is dominated by the broad signals of the plasma lipids. Furthermore, CPMG has the great advantage to improve the baseline, allowing a much better relative or absolute quantification of the metabolites.

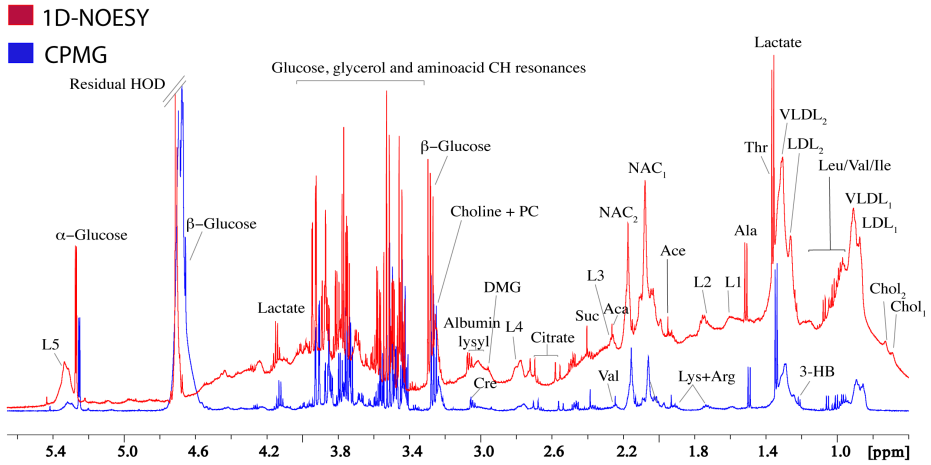


**Figure 3.5. Superimposition of the 1D-NOESY (red) and CPMG (blue) spectra of rat plasma (*Paper V*). Plasma lipids (L1-L5, LDL, VLDL) are the main responsible for the broad resonances.**

However, the main limitation of 1D NMR experiments is the overlap of the spectral resonances that seriously limits the clear identification of metabolites and their subsequent quantification. For this reason, two-dimensional (2D) NMR experiments are often performed after 1D $^{1}$H NMR measurement. 2D NMR methods allow the unambiguous identification of the metabolites in the mixture with the main drawback of being time-consuming. The most used experiments are $^{1}$H-$^{1}$H COSY (COrrelation SpectroscopY), $^{1}$H–$^{1}$H TOCSY (TOtal Correlation SpectroscopY), and $^{1}$H–$^{13}$C HSQC (Heteronuclear Single-Quantum Correlation spectroscopy), and HMBC (Heteronuclear Multiple-Bond Correlation spectroscopy). These approaches have been employed in *Paper I* and *Paper IV*. Another 2D homonuclear experiment called *J Resolved* (JRES) also represents a valuable tool for metabolite identification (Ludwig & Viant 2010), thus it has been used also in this PhD project (*Paper I*, *Paper III* and *Paper V*). JRES keeps the advantages of a simple 1D NMR experiment, further providing, along the second dimension, the proton-proton coupling. In this way the overlapping resonances are dispersed, simplifying the assignment, especially in case of the presence of minor compounds that could be completely hidden by the overlapping of main signals. A real example of JRES spectrum is given in Figure 3.6 where an enlargement of the apple juice 1D NOESY and JRES spectra (Section 5.2) is showed.

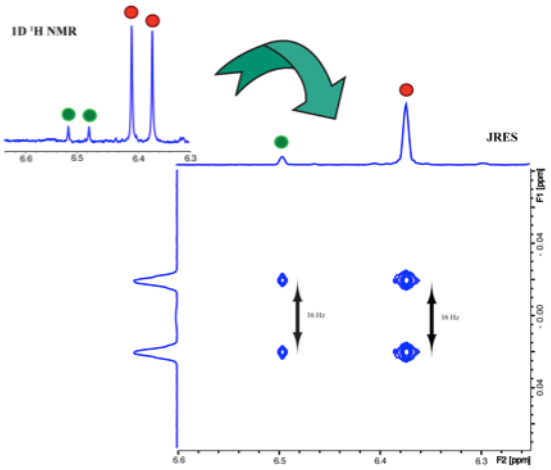


**Figure 3.6. Expanded areas of 1D NOESY (top-left) and JRES experiment on apple juices. Red labelled signals belong to chlorogenic acid while the green ones to the p-coumaric acid. The intensity of the signals is kept in the JRES.**

From the Figure 3.6, it is clear how a doublet in 1D spectrum becomes a singlet in JRES. Indeed, the projection along F2 can be considered as a decoupled proton spectrum that highly reduces the spectrum complexity. All the information about the multiplicity and the coupling constant is instead provided along the F1 dimension: in this case both signals are doublet (16 Hz).

## 3.2 **NMR data pre-processing for chemometric analysis**

NMR data are not readily suitable for the analysis by chemometric methods. Two different approaches can be adopted to analyse NMR spectra: (i) considering each intensity point of the spectrum and/or (ii) performing the integration of isolated peaks. This latter approach takes into account only the NMR signals of specific metabolites, making this analysis more targeted than that computed considering the entire spectrum.

However, it can happen that a peak belonging to the same analyte, and thus expected at the same chemical shift in all the samples, changes its position across the spectra. The most common reasons for this phenomenon are changes in experimental condition such as pH, temperature, ionic strength or background matrix as well as physicochemical interactions. In most cases, no useful information can be retrieved from these peak shifts thus it is preferable to perform an alignment that will allow a good chemometric modeling. Indeed, there is an important assumption that has to be fulfilled in techniques such as Principal Component

Analysis and Partial Least Square: the bilinearity. It requires that each column of the matrix to be analysed (if samples are stored in the rows) contains information about a signal originating from the same compound across all the samples. Thus, successful models are computed only when data are reproducible among the samples. For this reason peak shifts among different samples have to be avoided in different ways: keeping as homogeneous as possible the preparation protocol for all the samples, avoiding changing instrumental conditions and parameters or using a buffer solution in case of pH-based shifts. If variations in peak positions are still observed in the spectra, binning or alignment procedures should be computed.

Furthermore normalization and scaling steps play a fundamental role in the chemometric analysis, thus they also will be discussed in this section.

### 3.2.1 *Binning*

Binning consists of dividing the spectra into small buckets (typically 0.04 ppm), which are ideally large enough to include peak shift variations. The area under the curve is then calculated for each bucket, thus providing the bin intensity. Traditional binning masks chemical shift misalignments and filter noise in the spectra but it can also hide potentially significant changes of low-intensity peaks nearby huge signals. After binning, the statistical analysis is carried out on the extracted bin intensities, and peaks are assigned to metabolites. The main drawback of binning is the drastic loss in resolution that can lead to poor metabolite quantitation. This approach was employed only in the olive oil study (*Paper II)*, where a binning of 0.002 ppm was performed.

### 3.2.2 *Alignment*

When high resolution is required, other approaches are employed to solve the peak shift issue. Methods such as dynamic time warping (DTW) or correlation optimized warping (COW) (Tomasi et al. 2004; Bylund et al. 2002; Nielsen et al. 1998), have been demonstrated to be effective on chromatographic data and have also been employed for solving simple NMR alignments with satisfactory results (Flemming H. Larsen 2006). However, these methods are complex from a computational point of view and they align by stretching or compressing the signals and the baseline. This approach is optimal when a positive correlation between peak width and shift occur, thus it is not suitable for NMR signals. For NMR signals

various alignment algorithms have been proposed, among these, the *interval-based icoshift algorithm* (Savorani et al. 2010) has been employed in all the NMR studies performed during this PhD project.

The icoshift algorithm is based on COrrelation SHIFTing of spectral Intervals and employs a Fast Fourier Transformation (FFT) correlation engine to boost the algorithmic speed. Three main steps are computed by icoshift: interval definition, maximization of the cross-correlation of each interval by the FFT engine and signal reconstruction. The definition of the intervals is optional, therefore, the user can both decide to align the whole spectra as it is (global alignment), or the whole spectrum divided in equal size intervals or just to align only few customised intervals leaving untouched the rest of the spectrum. Then, the target for the alignment has to be defined, it can be an actual signal or a synthetic one like the average, or the median calculated from all the spectra included in the dataset.

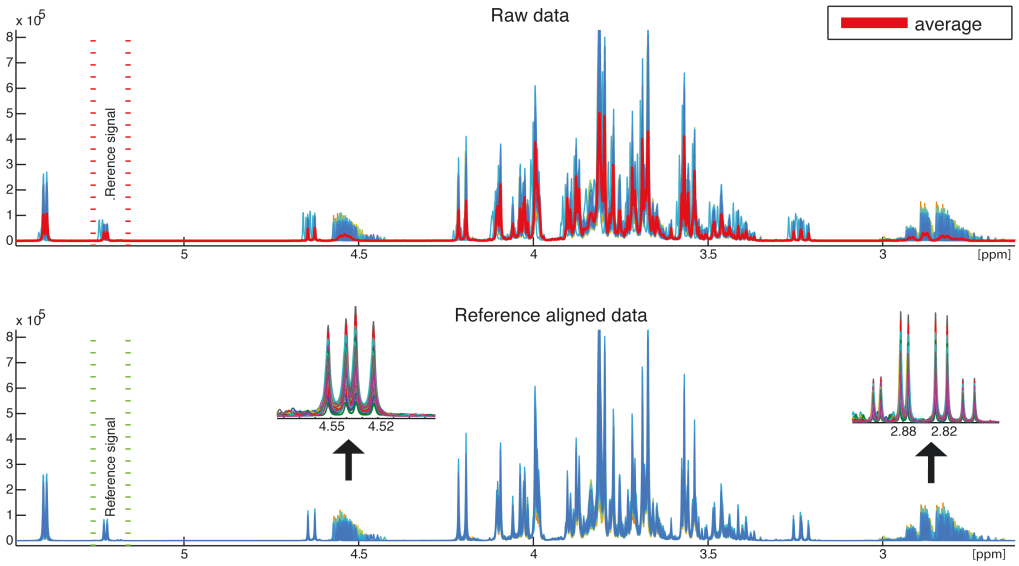


**Figure 3.7. icoshift alignment based on a reference signal of the apple juice NMR dataset. The chosen reference region (5.24–5.20 ppm) contains the anomeric doublet of α-D-glucopyranose which was chosen to drive the alignment according to a rigid shift of whole spectra. The average spectrum, used as the target, is highlighted in the top part of the figure which shows the raw data as well as the bottom shows the icoshift aligned data. Malic acid regions around 2.88 and 4.55 ppm have been aligned only after a further icoshift step that only included the interested regions.**

After target and interval definition, the algorithm works maximizing the cross-correlation between user-defined intervals and then it reconstructs the signal by adding missing values, in order to avoid spectral artifacts at the segment boundaries, or the first/last point in the

segment. An example of icoshift application is shown in Figure 3.7. Whole $^{1}$H NMR spectra of apple juice (*Paper III*) have been aligned according to a user defined reference signal (α-glucose anomeric proton at 5.24 ppm). The glucose signals are usually good candidates for this purpose since they don't show pH dependent peak shifting. The alignment was successful for almost all he signals, except for the pH sensitive malic acid peaks. To solve this issue, icoshift was performed a second time, only including the two malic acid regions, resulting in a perfect final alignment.

The Matlab code for icoshift is free and can be downloaded from *www.models.life.ku.dk*.

### 3.2.3 *Normalization*

Metabolomic responses are reflected in differences in concentration of specific metabolites. Therefore, variations in signal intensity attributable to the amount of material analysed or dilution effects are not desirable since they can be misleading for the data the interpretation.

Normalization methods compute a table row operation that aims to remove this effect to make spectra comparable with each other (Euceda et al. 2015).

Among the most used approaches used to compute this operation, there are:

1) Normalization to an *internal standard*
2) Normalization to a particular *reference peak*
3) Normalization to *total intensity* or *total area*
4) Probabilistic Quotient Normalization (*PQN*) (Dieterle et al. 2006)
5) Normalization to an artificial signal (Barantin et al. 1997)

In the studies presented in this thesis methods 2) and 3) have been employed. In particular, in *Paper III* and *V*, the area under an artificial peak around 12 ppm, was used as reference for the normalization. This peak is generated by a software-based protocol called QUANTAS (QUANTification by Artificial Signal) that aims to provide absolute quantification by NMR (Farrant et al. 2010).

The normalization to the total area was instead performed in the studies described in *Paper I*, *II* and *IV*. Generally, the normalization algorithm divides each data point (NMR variable) by a constant number that can be represented either by the integral of the reference peak or the

integral of the whole spectrum. In order to understand the benefits provided by the use of the normalization, a hypothetical case in which three different sets of spectral data points were acquired from different amounts of the same tissue sample is shown in Figure 3.8.

| Sample weight (mg) | Signal intensity | | | Sum |
|---|---|---|---|---|
| | Variable 1 | Variable 2 | Variable 3 | |
| 10 | 6.8 | 17.8 | 4.2 | 28.8 |
| 20 | 13.6 | 35.6 | 8.4 | 57.6 |
| 30 | 20.4 | 53.4 | 12.6 | 86.4 |

Divide by sum →

| Normalized Signal intensity | | |
|---|---|---|
| Variable 1 | Variable 2 | Variable 3 |
| 0.2 | 0.6 | 0.1 |
| 0.2 | 0.6 | 0.1 |
| 0.2 | 0.6 | 0.1 |

**Figure 3.8. Hypothetical case of area normalization to eliminate variance related to the amount of sample analysed. In order to make each spectrum comparable, the intensity of each signal is divided by the sum of the intensity in each set.**

The example showed in Figure 3.8 clearly explains how the unwanted variability related to the different weight of the samples analysed is eliminated by the normalization procedure. It is important to underline that before the normalization step, the removal of spectral regions containing undesired variability has to be performed. These regions, usually, include signals from residual water, solvents, or contaminants that only provide information about the goodness of the experimental procedure, while hiding or interfering with the biological variability.

### 3.2.4 *Scaling*

Scaling is a mathematical operation performed on the dataset to balance signal intensity variances that originate from difference in average abundance of metabolites.

Indeed, metabolites that are naturally more abundant in the sample analysed (i.e. fructose, sucrose and malic acid in apple juice) will dominate during the data analysis, masking the variance related to minor, but maybe more interesting, metabolites. Differently from the normalization methods, which are computed independently on each sample (row-wise operation), scaling is a column-wise operation that depends on all the samples included in the dataset (Hendriks et al. 2011).

Prior to scaling, an operation called *mean centering,* transforms all values letting them vary around zero instead of the mean value (Smolinska et al. 2012). Basically it subtracts the column mean intensity from each individual intensity value, allowing the correction for the gap between metabolites that have high differences in concentration values.

*Autoscaling*, also referred as *standardization or Unit Variance Scaling (UV)*, allows to all the variables to have unit variance, giving each variable equal chance to influence the model. Basically each spectral data point is divided by its standard deviation after mean centering:

$$UV = \frac{x - \bar{x}}{SD} \qquad (3.7)$$

where $x$ is the variable, $\bar{x}$ is the mean value of the variable $x$ in all the samples of the dataset, and *SD* is the standard deviation. This approach blows up baseline noise, thus it is not the optimal choice for processing the NMR spectra that contain many baseline regions with random noise. On the other hand, it is very helpful in case of the analysis of variables with different units, as described in *Paper IV*. Indeed, in this study integrated NMR peak areas, corresponding to specific metabolites, have been analysed together with clinical parameters (i.e. hormones values). Moreover, autoscaling has been successfully used on the GC-MS data in *Paper V*. The whole autoscaling procedure is illustrated in Figure 3.9. The importance of such a mathematical approach is even clearer if we look at the data representation on a plane that has equal axis, as reported in the bottom part of the Figure 3.9. When the Principal Component analysis is computed (see Chapter 4), the main objective is to look how data vary, and this can only be achieved by looking at the variations around the axis origins (0,0).

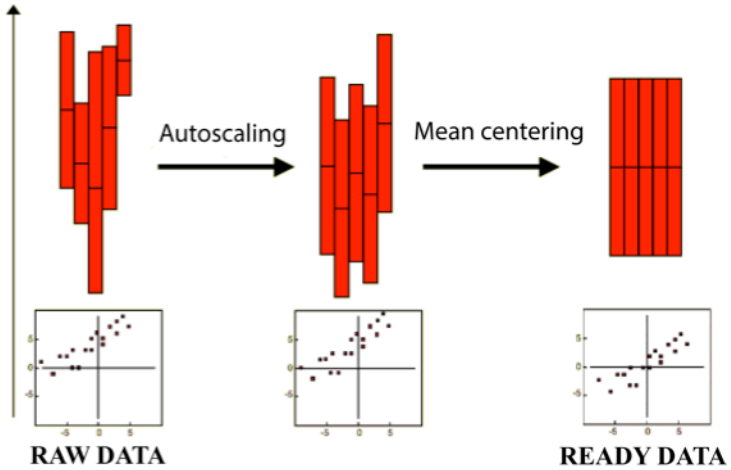


**Figure 3.9. Schematization of the Autoscaling procedure (top) and its effect on the data representation (bottom). The vertical axis represents the strength of the signal detected for each variable or metabolite. Each variable corresponds to a bar, with the mean value presented as a short, horizontal, black line.**

*Pareto scaling* is similar to autoscaling except that it employs the square root of the standard deviation as scaling factor, instead of the standard deviation alone (Smolinska et al. 2012).

$$Par = \frac{x - \bar{x}}{\sqrt{SD}} \tag{3.8}$$

This approach is the most used for the preprocessing of NMR data. It can be considered as an intermediate between the extremes of no scaling (only mean cenetering), where medium features (low abundance metabolites) are overwhelmed by big ones (high abundance metabolites), and autoscaling that enlarges baseline noise. Basically, Pareto scaling up weights medium features without inflating baseline noise as schematized in Figure 4.

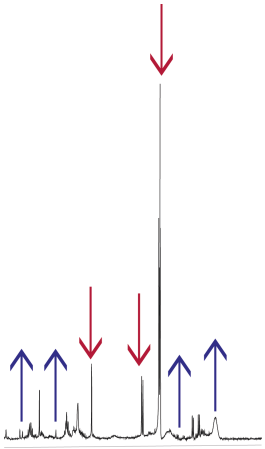

**Figure 4.0. Pareto scaling effect on NMR variables having different intensity. The influence of large biologically relevant variations is decreased (red arrows) while minor compounds variability is up weighted (blue arrows).**

# Chapter 4
# CHEMOMETRICS IN METABOLOMICS

Chemometrics is the application and development of mathematical and statistical methods to extract information from chemical data. The analytical techniques commonly used in Metabolomics are able to record thousands of variables for every sample analysed, thus providing very complex datasets. Chemometrics and, in particular, multivariate data analysis, provide valuable tools for the analysis and the recover of the maximum information from a complex data matrix.

The basic categories of analysis techniques are:

1) *Exploratory analysis*: unbiased overview of the data where patterns and outliers can be easily detected.
2) *Classification and discrimination*: allow to discriminate between groups and to find biomarker candidates.
3) *Regression*: compares blocks of data from a quantitative point of view.

In the works presented in this thesis the first and last approaches have been mainly employed. A brief description of the chemometrics techniques included in the Papers is provided in the following sections.

## 4.1 **Principal Component Analysis (PCA)**

PCA is one of the most commonly used approaches in metabolomics. It was invented in 1901 by Pearson (Pearson 1901) as an analogue of the principal axis theorem in mechanics. However, it was later developed and named by Harold Hotelling (Hotelling 1933).

PCA is an unsupervised pattern recognition method that allows the reduction of the dimensionality of a dataset consisting of a large number of interrelated variables, providing a visual representation of the major variance in the data. Thus, the starting point for a PCA is a matrix of data with $N$ rows (*observations*) and $K$ columns (*variables*).

The observations can be analytical samples, biological samples from different individuals, process time points of a continuous process, batches from a batch process and so on. The columns can be spectral or chromatographic variables as well as measurements of various origin (e.g. patients clinical parameters or food sample characteristics).

Basically, the original variables are transformed into a smaller set of new orthogonal (uncorrelated) variables, called principal components (PCs), which are ordered according to the explained variance that they are able to retain. Most of the relevant systematic information is usually calculated by the first few PCs, while the following ones are often computed considering chance variation and noise. The outcome of this analysis consists of two plots: a scores plot, where each point represents a single spectrum (sample), and a loadings plot that shows the variables (variables of the spectrum). These plots always need to be inspected together since the directions in the score plot correspond to direction in the loading plot. Thus, in order to understand the reason of a particular grouping observed in the score plot, it is just needed to look at the same direction in loadings.

Furthermore, score and loading plots also provide information about how the samples and the variables are related to each other. Hence, the scores that are close to each other have similar profiles while close variables are correlated. On the contrary, objects that lie in the opposite part of the plot have very different characteristics.

A more appropriate way to explain the PCA is possible using the mathematical interpretation of this method, as illustrated in Figure 4.1.

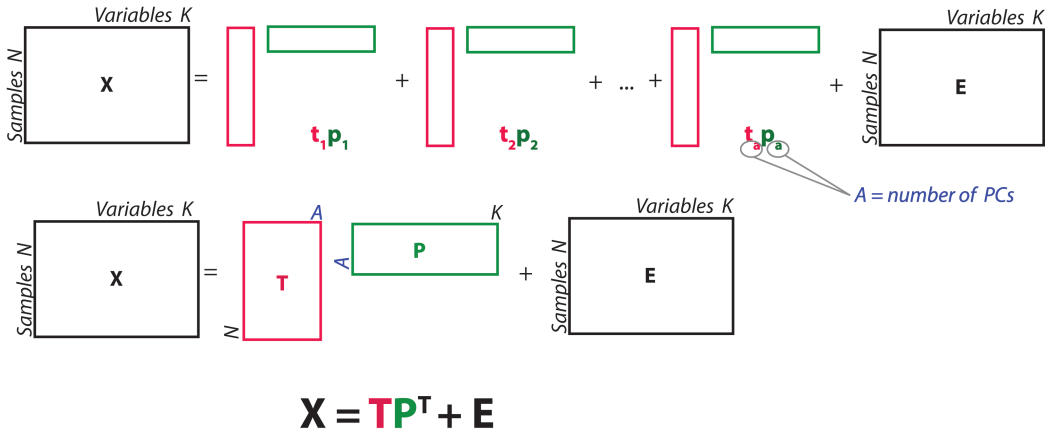


**Figure 4.1. Schematic illustration of a PCA model. T and P represent the scores and loading matrices, while E is the residual matrix. *N* and *K* are the number of samples and variables respectively. *A* is the number of Principal Components chosen to build the model.**

The data matrix is decomposed into a structure part and a noise part:

$$X = TP^T + E = Structure + Noise \tag{4.1}$$

where $T$ is the scores matrix, $P$ the transposed loadings matrix and $E$ the residual matrix. Another way to clarify the scores definition is considering them as element of the T matrix, where each row is an observation while each column represents the value that the observation has along each Principal Component.

The Loadings, also called weights, allow to understand the influence of the original variables on the scores T. The loading plot shows the loadings of a certain PC; also in this case it is possible to plot the loadings of a PC against the loading of another PC.

The residuals are not part of the model, thus this part should be "small" in order to not remove too much information from the original dataset. More PCs are included in the model, higher will be the variance that it is able to explain, thus smaller will be the residuals, following this calculation:

$$Explained\ Variance + Residual\ Variance = 100\% \tag{4.2}$$

However, retaining too many components has the drawback to include not relevant information in the model, thus affecting its goodness.

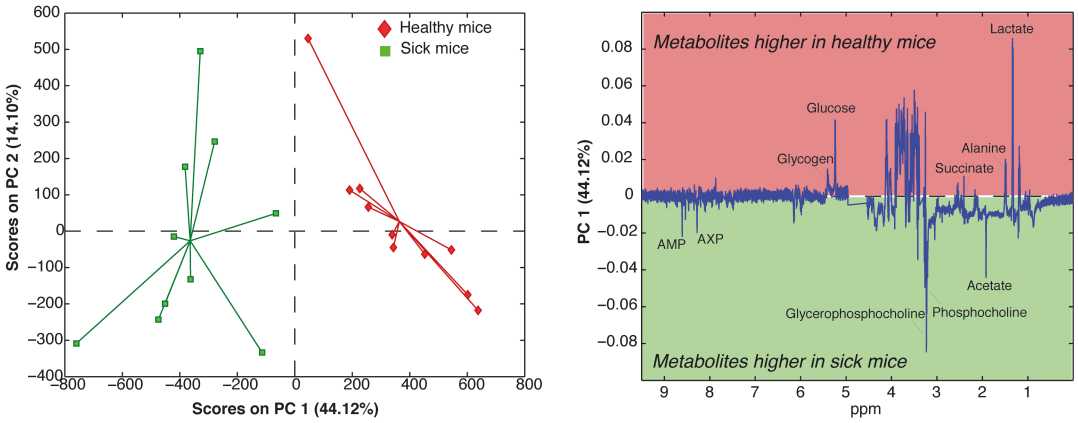


**Figure 4.2. PCA score plot (left) and loading plot (right) of $^1$H NMR spectra of mice liver extracts. The separation between the healthy and sick mice is visible along the PC1. The corresponding loading plot shows that the metabolites in the red area are higher in the healthy group while those in the green area are higher in the sick mice liver extracts.**

## 4.2 **Partial Lest Square regression (PLS)**

*Partial Least Squares (PLS) regression* is a supervised method that can be used when prior information about samples is available, such as class membership and quantitative information (Wold et al. 2001). It is a multivariate technique that assesses the relationship between two blocks of data: a descriptor matrix *X* (i.e. spectral data) and a response matrix *Y* (known sample information). The *Y* matrix can either contain quantitative (i.e. metabolite concentration) or qualitative information (i.e. class membership).

The model is then built only using the *Y*-related variance in *X*. PLS can be either used for calibration purposes (when *Y* contains quantitative information) or for class discrimination (when *Y* contains qualitative information). In the latter case, the PLS method is called *PLS Discriminant Analysis* (PLS-DA) (Ståhle & Wold 1987).

### 4.2.1 *Recursive PLS (r-PLS)*

Recursive PLS, or just rPLS (Rinnan et al. 2014) , is a recently developed variable selection method where the regression coefficients are recursively used as weights on the original data matrix. This concept is schematized in Figure 4.3.

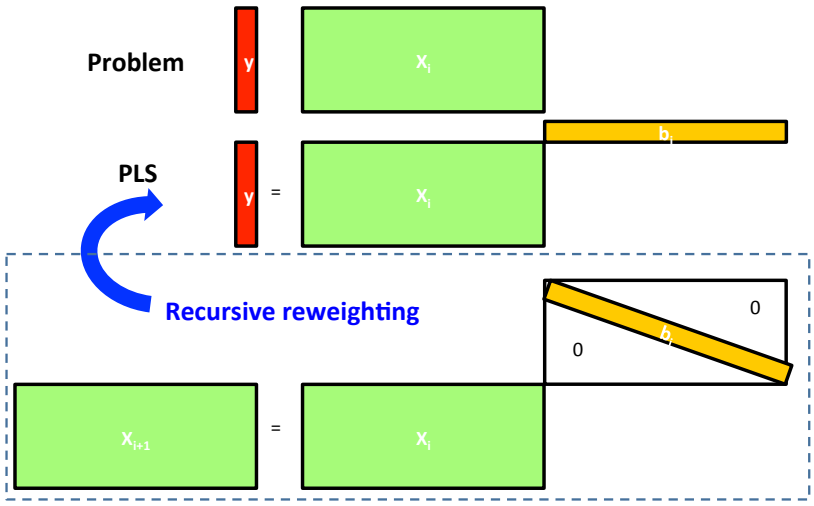


**Figure 4.3. Schematization of the recursive PLS procedure. $X_i$ represents the X data matrix; $X_{i+1}$ is the X data matrix weighed using the *b* vector; *b* is the regression vector.**

rPLS is similar to the method proposed by Forina (Forina et al. 1999) and jack-knifing (Martens et al. 2001), but instead of eliminating variables after each iteration, the rPLS

method iteratively uses the regression coefficients to magnify important variables and thus relatively down-weight less important ones. As showed in Figure 4.3 a normal PLS approach finds the Y-related variance in the X matrix by calculating the regression coefficients (b vector). Basically, rPLS is based on a process of repeated PLS models; for each repetition a different X matrix is employed and it is given by:

$$\mathrm{X_{I+1}} = X_I \times \mathrm{diag}(\mathbf{b})$$

where $X_I$ is the previous updated weighted X, b is the regression coefficient from the last model and $\mathrm{X_{I+1}}$ is the "new X" to be used in the subsequent PLS model. Reweighting each time the X matrix, will allow to converge to a very limited number of variables (good for interpretation), that will normally including covarying neighbour variables. This facilitates the interpretation and reducing the time consuming step of a thorough signal assignment

The regression vector, which is represented as a diagonal matrix only for mathematical reasons, is fundamental in this approach. It is important to remember that the regression vector reflects the importance of the variables: weights around 0 indicate variables not correlated with y, and weights with large absolute values indicate important variables. An application of the rPLS approach is described in section 5.2.2 where it has been applied for the prediction of apple juice sensory attributes from NMR spectra.

## 4.3 **Analysis of Variance-Simultaneous Component Analysis (ASCA)**

Analysis of variance (ANOVA) provides information about the variance between and among groups of samples in an experimental design and its statistical significance. It works by partitioning the variance of a variable into components that originate from different sources, depending on the experimental design. This statistical tool is generally used for univariate data, when only one variable or metabolite is monitored in samples from an experimental design. A multivariate version of ANOVA (MANOVA) is available, however, it doesn't work properly when a large number of variables is considered because of problems of singularity of covariance matrices and assumptions that are not fulfilled.

A valuable multivariate generalization of ANOVA is represented by ASCA (Smilde et al. 2005). This algorithm perfectly works on different types of data characterized by a balanced experimental design, or a temporal structure. Indeed, it separates the variance of the original dataset into different matrices from the design factors, allowing to separately investigate the

importance of each factor and to better interpret the data. This approach has been employed in *Paper V,* described in section 6.2, to separate the diet-related effects from other undesired variability coming from experimental design.

## 4.4 **Parallel Factor Analysis (PARAFAC)**

PARAFAC (Harshman 1970; Bro 1997; Bro et al. 2010) is a generalization of PCA to higher order arrays as illustrated in Figure 4.4.

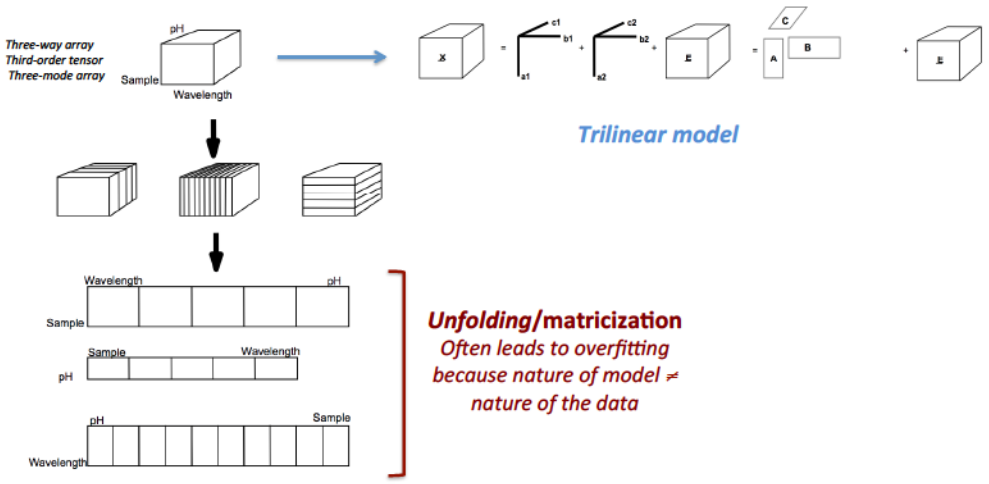


**Figure 4.4. Schematization of the possible ways to analyse a three-way array. PARAFAC builds a trilinear model, avoiding the two-dimensional unfolding of the "cube".**

Three-way arrays are the outcome of several types of measurements:

- Sensory analysis (Food sample, Judge, Attribute)
- Process Analysis (Batch, variable, Time)
- Image analysis (Sample, Image Pixel, Variable)
- Spectroscopy (Wavelength, Sample, Time)
- Chromatography (Sample, Retention Time, Variable)

PARAFAC decomposes the three-way array into three matrices, called loading matrices, one for each mode. For the sample mode, the loading matrix is usually referred to as a score matrix and holds the relative concentration of each chemical compound in the sample if the model successfully separates the individual chemical compounds into individual PARAFAC

factors. The elution-mode loading matrix correspondingly holds the estimated elution profiles of each analyte and the last mode, the estimated mass spectra of each factor.

Like PARAFAC, PARAFAC2 (Harshman 1972; Kiers et al. 1999; Bro et al. 1999) also decomposes three-way data arrays into loading matrices, but the main difference is that PARAFAC2 does not impose strong restrictions on the data structure. It does not assume that the shape (or even length) of the elution profile of an analyte is the same in each sample. Even though PARAFAC2 allows elution profiles to differ in shape in different samples, it still possesses uniqueness properties that are very similar to those of PARAFAC. This means that a successful PARAFAC2 model can separate mixture data into the contributions (concentrations, elution profiles and mass spectra) of the underlying analytes directly. PARAFAC2 represents an ideal technique for modeling GC-MS data, as the model allows one elution-time profile to be obtained for each factor of the sample, taking into consideration that each analyte has a mass spectrum that is consistent across all samples (Amigo et al. 2008). Figure 4.5 illustrates the steps carried out during the PARAFAC2 modeling in the β-glucan study described in Chapter 6.

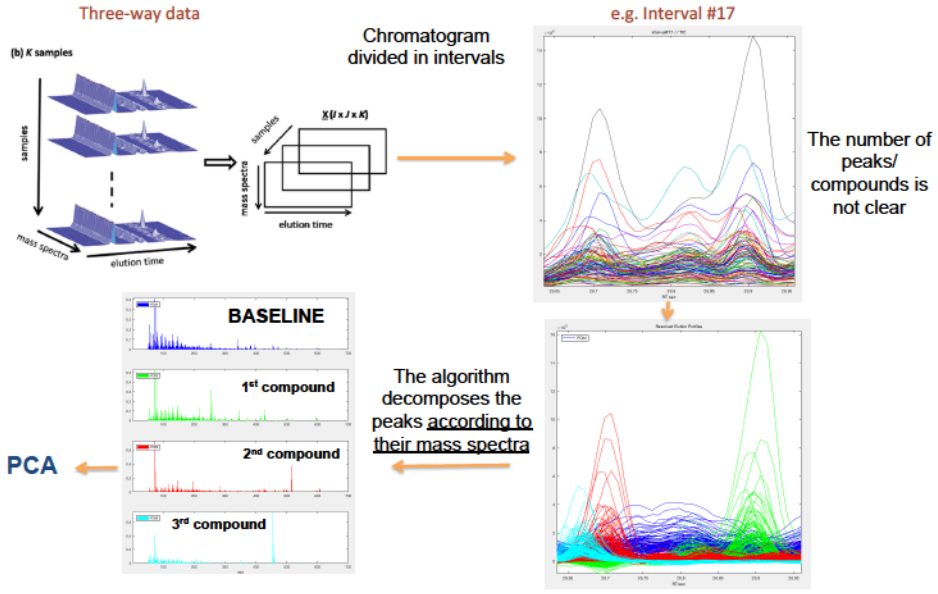


**Figure 4.5 Schematic representation of the PARAFAC2 modelling steps performed in *Paper V*.**

In order to reduce the complexity of the GC-MS dataset, the chromatograms were then manually divided into intervals in the elution time dimension, in correspondence of the peaks, leaving some baseline between them. Each interval was then individually submitted to PARAFAC2 modelling, thus ensuring a faster and more reliable model validation. For each

interval, a model with one to ten components, was built and validated as previously described (Khakimov et al. 2012). After the model validation, the PARAFAC2 resolved peak areas were extracted and put into final metabolite matrix that was then submitted to Principal Component Analysis.

### 4.5 **Correlation analysis and heat maps**

The analysis of the correlations is an alternative and/or additional tool useful for retrieving information from the metabolomics data. It is commonly based on pairwise correlation between concentration levels of the metabolites in a sample (Steuer 2006).

The most common correlation approach in metabolomics is *Pearson's correlation* (linear correlation between variables), however other options are also available (e.g. the non-linear Spearman correlation) (Camacho et al. 2005).

Even though calculating the correlations is quite straightforward, their link with biological functions or biochemical pathways is still poorly understood (Marcotte 2001). This is due to the complicated network of interactions among the different metabolic pathways that are often carried out simultaneously. However, the observed correlations can be considered as a global "fingerprint" of the observed system even though the full interpretation of the biological meaning is not immediate.

In *Paper IV* metabolite-metabolite Pearson's correlations have been calculated in order to better understand the metabolic pathways perturbed in the studied disease.

A commonly used tool for the representation of the calculated correlations is the so-called *Heat Map* where values, contained in a matrix, are usually represented using a colour scale. As showed in Figure 4.6 this representation can also be enriched with the corresponding p-value for each correlation, in order to distinguish among the significant and not significant data.

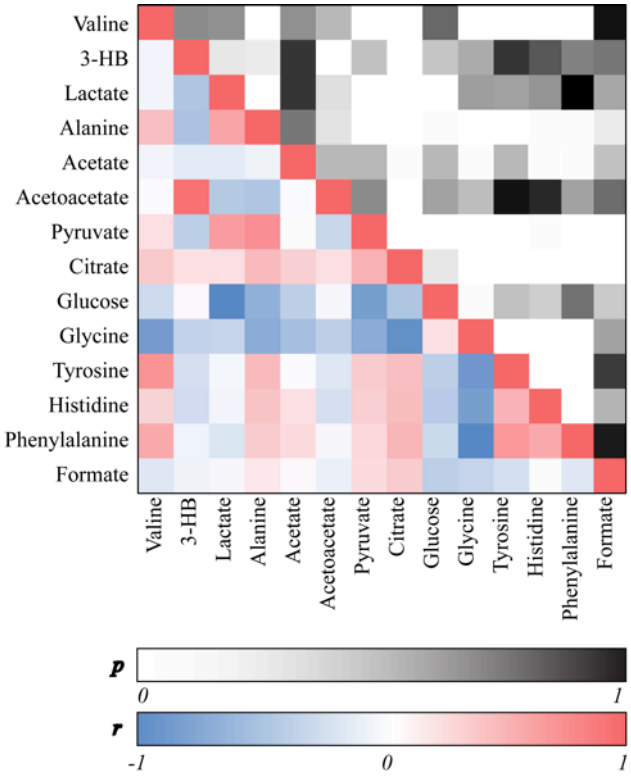


**Figure 4.6. Metabolite–metabolite correlation matrix (heat map) calculated using 41 follicular fluid samples *(Paper IV)*. The heat map is coloured according to Pearson's correlation coefficient (r) and the corresponding p-values (p). The red and blue colour of each cell depicts the Pearson's correlation coefficient value, with deeper colours indicating higher positive (red) or negative (blue) correlation coefficients. White cells stand for significant correlations ($p < 0.05$).**

### 4.5.1 *STOCSY*

Statistical Correlation Spectroscopy (STOCSY) (Cloarec et al. 2005; Holmes et al. 2007) is a widely used method to recover information from complex 1D-NMR metabolomics spectra. Basically, it calculates all the possible correlations among the NMR spectral variables, in a pairwise way. Therefore, the signals that arise from the same molecules, will also experience the same fluctuations in concentration among the samples of the dataset, thus showing a very high *r* value. The graphical representation on the STOCSY is comparable to that of a 2D correlation NMR experiment performed on one sample, namely TOCSY.

A real example of STOCSY is provided in Figure 4.7 from the study described in *Paper I.* The colour code has to be always present on the right of the plot since it explains the value, *r*, of the correlation. In this case red areas represent a correlation with $r > 0.99$ while blue regions are characterised by $r < 0.95$.

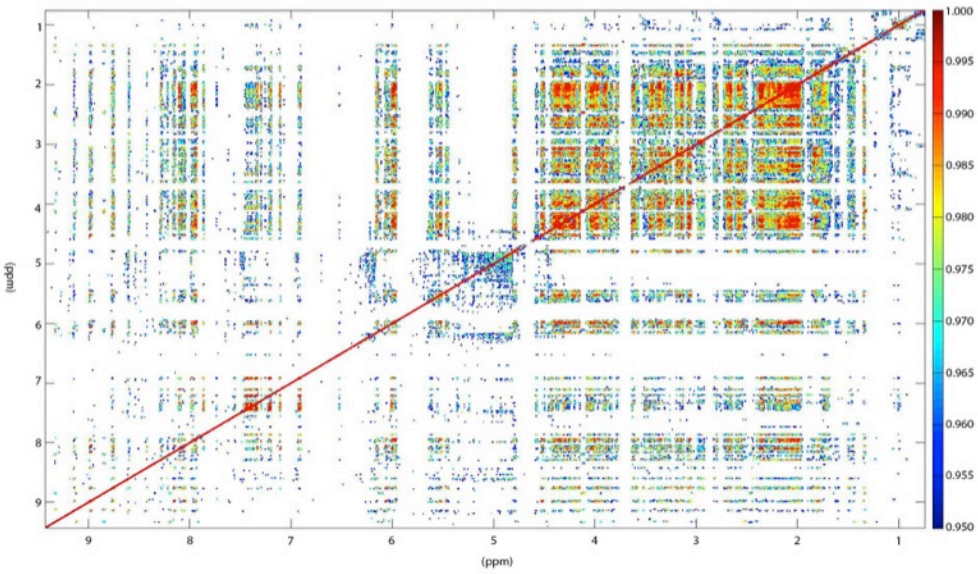


**Figure 4.7. STOCSY plot coloured according to the correlation coefficient (see coloured bar on the right).**

# Chapter 5
# NMR-BASED METABOLOMICS IN FOOD SCIENCE

The great potential of the NMR-based metabolomics has been widely recognized for qualitative and quantitative food analysis (Ramakrishnan & Luthria 2016). Minimum sample preparation, fast spectra acquisition, low cost per sample and good reproducibility are the main advantages of this technique. Indeed, NMR fingerprinting has been extensively exploited to obtain a 'holistic view' of the metabolome (foodome) of various kinds of beverages and foods, such as fruit juice (Belton et al. 1997; Sobolev et al. 2015), milk (Belloque 1999; Hu et al. 2007), wine (Godelmann et al. 2013) and olive oil (Mannina & Sobolev 2011). The possibility to differentiate cultivars, evaluate sensory properties, and investigate the influence of growing conditions and geographical origin of food crops, looking at a simple plot, has made NMR-based metabolomics one of the most used approaches in food science.

In this Chapter the results obtained from the food-related metabolomics project, conducted during this PhD, are presented.

## 5.1 **Characterization of monovarietal extra virgin olive oils from the province of Béjaïa (Algeria)** *(Paper II)*

Algeria is one of the countries where the olive oil production is particularly increased in the last ten years, thanks to two agricultural renewal programs, over the 2006–2008 and 2009–2014 periods, allowing producers to update their production tools. In these years, cultivation zones passed from 165,000 to 500,000 ha, and Algeria is nowadays considered as a new olive oil exporter. Despite the increased production, olive oils from this area are poorly studied compromising their exportation especially in the countries that are major producers of olive oils. Therefore, the aim of this work was to chemically characterize monovarietal Algerian EVOOs from different areas of the Béjaïa province, that is the area where the olive oil production is mostly increased, and, eventually, to suggest possible blends. The investigation was performed using traditional chemical analyses and $^{1}$H NMR spectroscopy,

as untargeted approach, to evaluate if geographical origin could influence the olive oil composition.

### 5.1.1 *Experimental design and results discussion*

Twenty-six monovarietal extra virgin olive oil (EVOO) samples from different areas of the province of Béjaïa (Algeria) (Figure 5.1) were analysed. The olive fruits coming from 19 different cultivars were randomly and manually picked from all parts of the selected fully-grown olive trees.

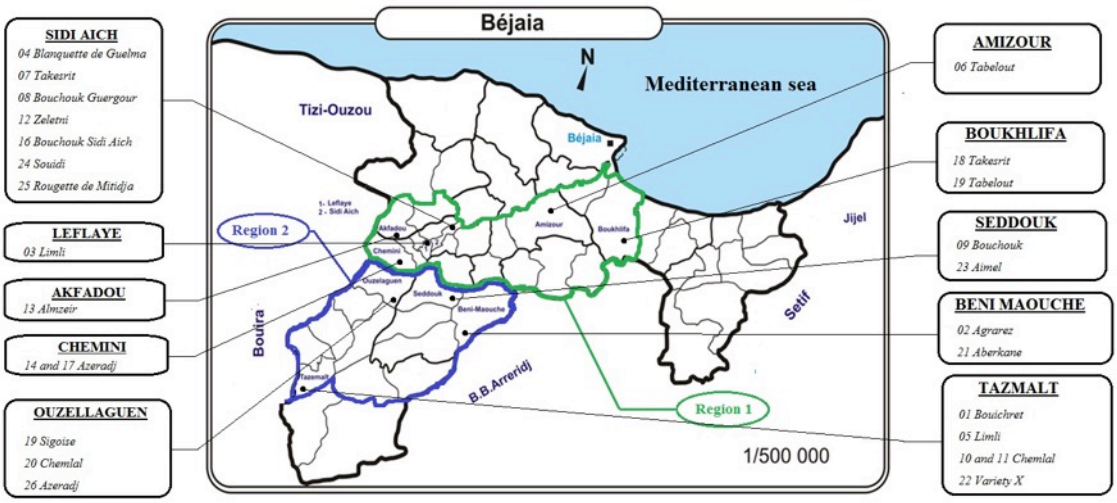


**Figure 5.1. Areas of the province of Béjaïa (Algeria) where the olives fruits were harvested. The province of Béjaïa is bordered by black bold line. Regions 1 and 2 are bordered by green and blue lines, respectively. Sample code, cultivar and area of origin of the samples are also reported.**

Then $^{1}$H NMR experiments were run both on whole EVOOs samples and their phenolic extracts. The spectra of whole olive oil showed the typical pattern of signals of the triglycerides. Nevertheless, a number of other signals belonging to minor components of the EVOO were also detectable. In particular, β-sitosterol, squalene, terpens, diacylglycerols and aldehydes were unambiguously identified (Figure 5.2).

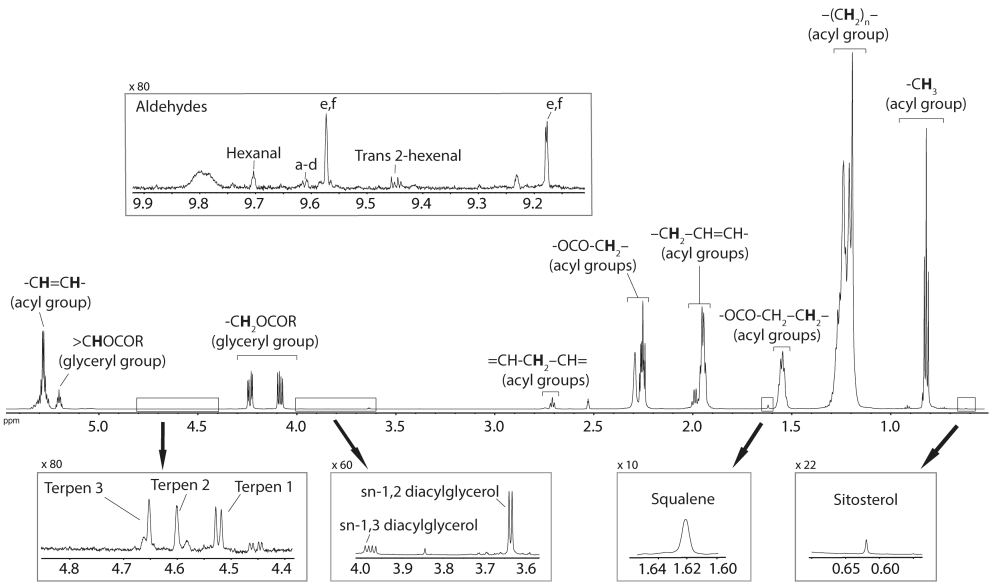


**Figure 5.2. $^{1}$H NMR spectrum of a representative EVOO sample. The expanded regions contain minor components of EVOO. The magnification value is indicated on the top left of each square. NMR signals in the left top frame correspond to dialdehydic form of ligstroside and oleuropein (*e* and *f* respectively) and aldehydic form of ligstroside (*a*, *b*) and oleuropein (*c*, *d*).**

In order to determine if the observable signals were able to give information related to the area of collection of the olives, a principal component analysis (PCA) was performed. The PCA showed that the spectra of whole olive oils are distributed according to their fatty acids compositions, however this information was not related to the geographical origin. On the other hand, the spectra of the phenolic extracts turned out to be very different among each other, and therefore particularly informative. In this case the PCA showed that the samples distribution was, in some ways, also correlated to the geographical origin of the samples. In particular, the oils produced from olives collected in the region closer to the *Mediterranean Sea* were richer in polyphenols than the samples collected in the inland region (Figure 5.3). As far as the loading plot of PC2 is concerned, it is interesting to note that the variables with the higher loading values (higher discriminating power) belong to the dialdehydic and monoaldehydic forms of ligstroside and oleuropein, indicating that the samples that lie in the bottom of the plot are richer in monoaldehydes, while the ones on the top have an higher content of dialdehydes. Interestingly, the dialdehydic form of decarboxymethyl ligstroside (indicated with letter *e* in Figure 5.1), also known as *oleocanthal,* has been investigated for its ibuprofen-like cyclooxygenase inhibiting activity (Beauchamp et al. 2005).

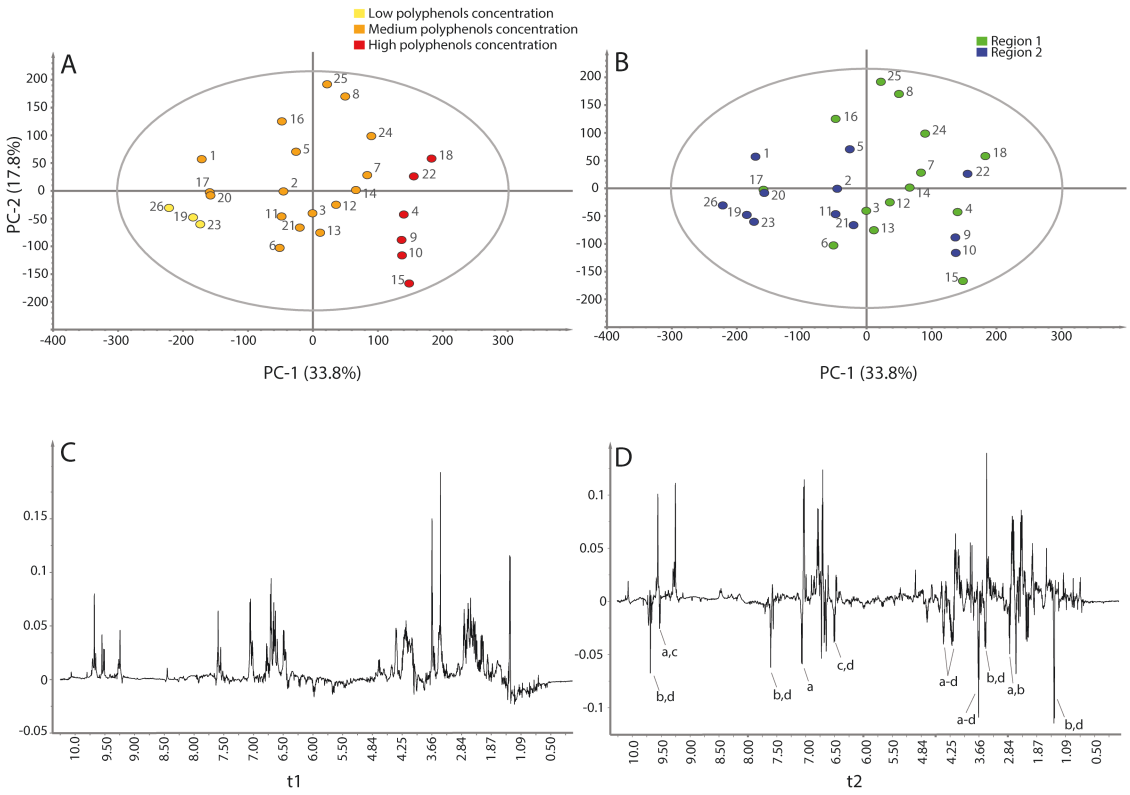


**Figure 5.3. A) PCA score plot coloured according to the content of polyphenols (yellow 50–160 mg kg$^{-1}$, orange 161–470 mg kg$^{-1}$ and red 471–1000 mg kg$^{-1}$). B) PCA score plot coloured according to the region of origin: region 1 (green); region 2 (blue). PC1 and PC2 loading plots are reported in panels C and D, respectively. Letters on the loading plots refer to aldehydic form of ligstroside (*a*, *b*) and oleuropein (*c*, *d*).**

This could suggest that these olive oils can be considered promising nutraceutical foods to be used for the treatment of inflammatory diseases (Iacono et al. 2010).

Generally, the results obtained in this study revealed that each olive oil seems to have strength and weakness points suggesting the potentiality of these cultivars to produce high quality blends that may compete with other Mediterranean products.

## 5.2 **Characterization of juices from ancient Danish apple cultivars**

This study was carried out at the University of Copenhagen under the supervision of Prof. Søren Balling Engelsen.

In this work a NMR-based metabolomics approach was applied, for the first time, to chemically characterize apple juices of about one hundred ancient Danish cultivars. The study was the result of collaboration with 'Pometet' (Figure 5.4), an experimental orchard and gene bank of the University of Copenhagen that hosts the national and international collection of

fruit genotypes. It is part of a project that aims at promoting the utilisation of ancient Danish apple cultivars for niche products since they may have unique flavour qualities that can be attractive in juices.

### 5.2.1 *Experimental design and results discussion*

A total of ninety-two cultivars from a local orchard (Taastrup, Denmark) were collected in 2010 during a period ranging from late August to middle October (Figure 5.4).

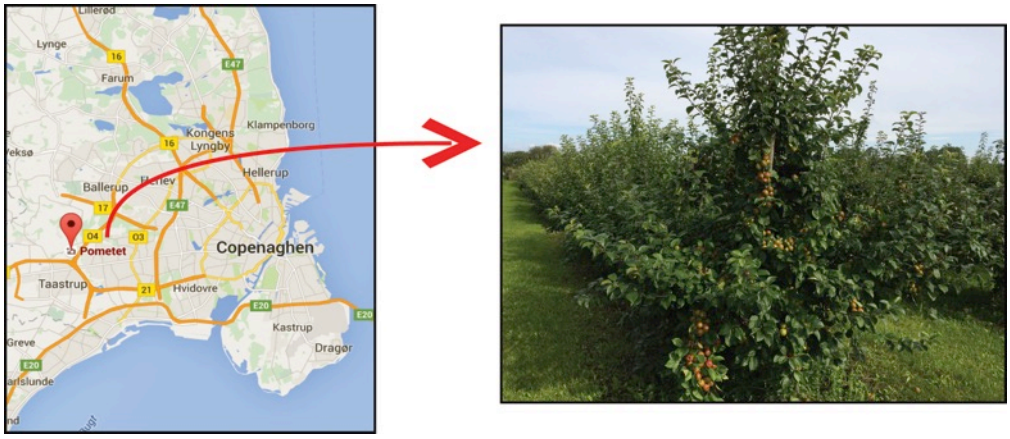


**Figure 5.4. The orchard "Pometet" is located in Taastrup, outskirts of Copenhagen (map on the left). A picture of the orchard, taken in September 2015, is showed on the right.**

High-field proton NMR spectroscopy was applied for samples characterization. 1D $^{1}H$ NMR spectra were acquired to determine the metabolic fingerprint of the juices, while 2D homonuclear experiments were acquired for assignments purposes. A total of 15 metabolites were clearly identified from the spectrum (Figure 5.5) following the references available in literature (Vandendriessche et al. 2013; Belton et al. 1997). In this study, the recently developed Bruker Spin Generated Fingerprint (SGF) Profiling – Juice Screener, was employed. It is a NMR-based screening method for the quality control of fruit juices that provides absolute quantification of around twenty compounds present in the apple juice, using a large reference database of more than 6000 samples of more than 50 different types of fruit juices from more than 50 countries (Spraul et al. 2009). A total of 26 metabolites were quantified in this study by the Bruker Profiling approach.

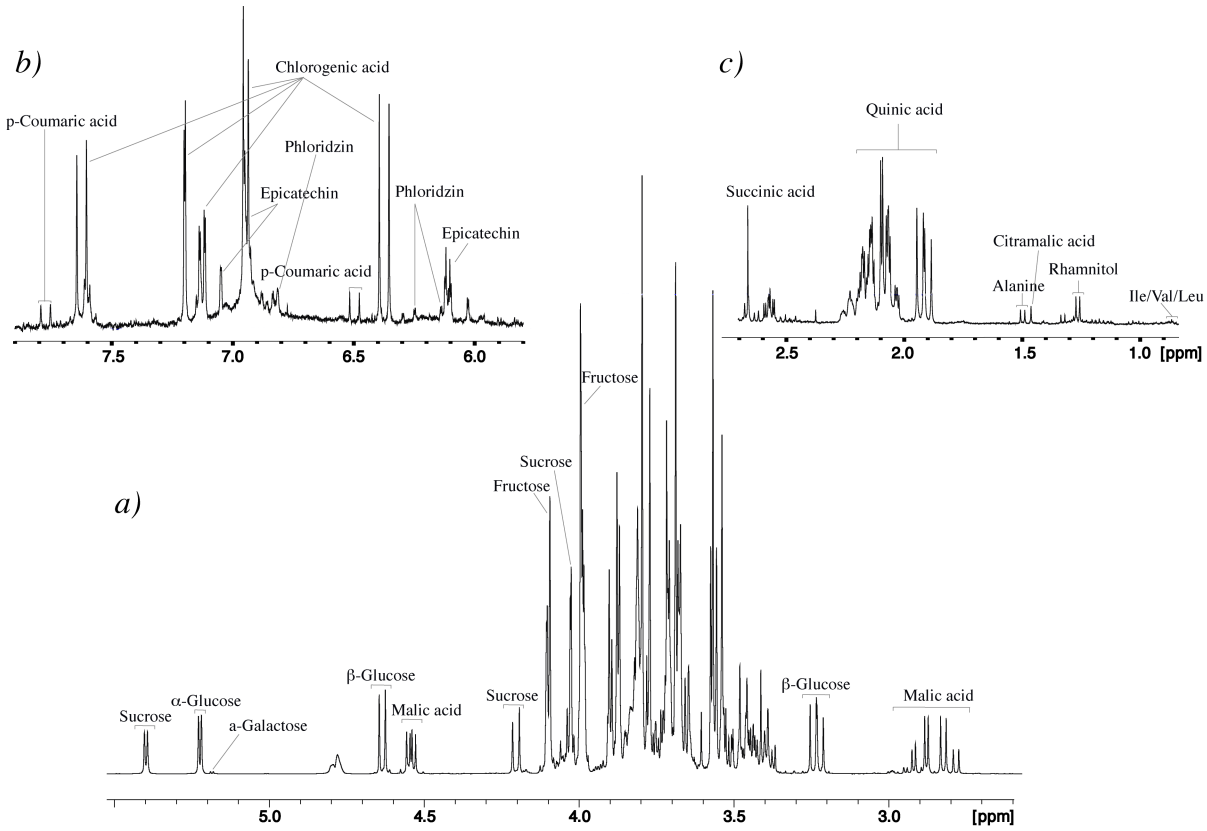


**Figure 5.5.** ***a)*** **$^{1}$H NMR spectrum of a representative apple juice sample (2.5 – 5.5 ppm).** ***b)*** **Aromatic region (5.8 – 7.9 ppm).** ***c)*** **Aliphatic region (0.9 – 2.7 ppm).**

A PCA was then performed in order to retrieve hidden information from the NMR spectra. The results have shown no particular clustering among the different cultivars, suggesting that each one allows the production of juices with peculiar chemical composition.

Since the spectra were dominated by sugars and malic acid signals, we decided to perform a PCA including only the aromatic region of each NMR spectrum (6.0 -8.0 ppm) (Figure 5.6). This last analysis showed that the cultivars *Bodil Neergaard, Barrittskoz Madæble and Gadeskovæblet* have an interesting polyphenolic composition.

The health related effects and bioavailability of polyphenols in human nutrition have been widely investigated (Dunlap et al. 2014; Visioli et al. 2011). The main health claims related to these compounds are based on their properties as scavengers of free radicals and reactive oxygen species (ROS). Phenolic compound characterization in whole apple fruit is well established, while their fate during transformation in juice has to be improved in order to better understand how to avoid the loss these precious compounds (Francini & Sebastiani 2013). In this frame, our results suggest the suitability of the identified cultivars for the production of juices with health benefits.

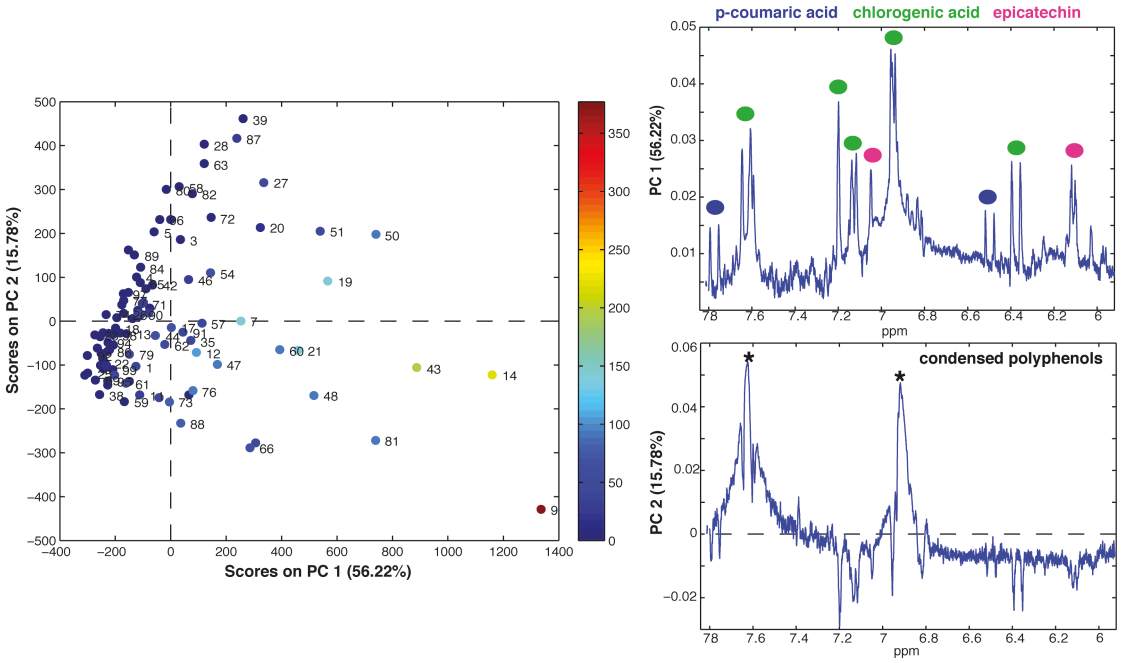


**Figure 5.6. PCA score plot (left) colored according to the content of chlorogenic acid. On the right side of the figure, PC1 (top) and PC2 (bottom) loading plots are showed. The metabolites responsible for the separation on the PC1 are indicated with coloured circles; the condensed polyphenols signals are responsible for the separation along the PC2.**

### 5.2.2 *Sensory study of apple juices (Paper III)*

Each apple juice sample of the study presented in 5.2.1 had also been submitted to a sensory evaluation prior the NMR analysis. Six different descriptors were evaluated: colour, overall odour, apple flavour, overall flavour, sweet taste and sour taste.

The sensory panellists were trained with a reference juice as well as with sucrose (11%) and malic acid (0.5%) water solutions for being able to properly recognize all the descriptors. The samples were evaluated using a continuous 0 (none) - 14 (very much) intensity scale and the scores of each sample were averaged over 5 assessors.

NMR and sensory data were employed to test a novel variable selection method (rPLS described in 4.2.1) in finding the spectral variables correlated with sweetness and sourness of the juices. Thus, the NMR dataset was used as *X* block while the sensory evaluation data represented the *y* vector.

In both cases (sweet and sour taste), the rPLS was able to develop a good regression model providing just a very limited set of variables that correlate with the *y* vector. In particular,

sucrose signals showed to be positively correlated with the sweet taste while malic acid signals with the acidity of the apple juice (Figure 5.7).

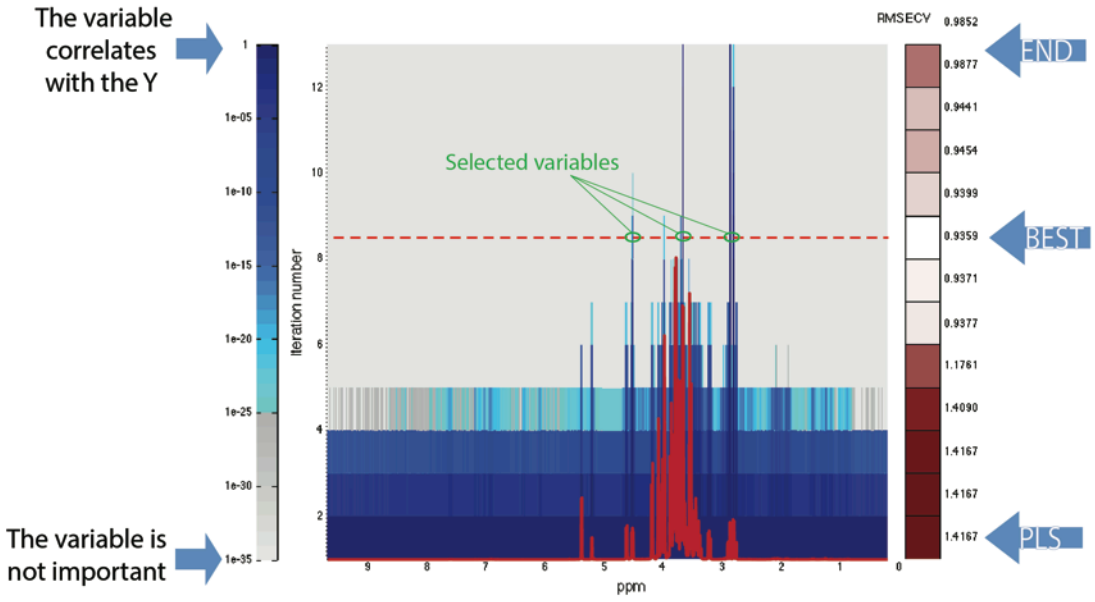


**Figure 5.7. The rPLS result for the prediction of sensory evaluated acidity. The model uses two latent variables. In each row the development of weights according to iterations is shown. The coloured scale on the right represents the RMSECV values, the white box indicates the row where the best model was built and its relative RMSECV value; the bar on the left shows the value of the weights. The value 1 means that the variable has a large weight and thus importance; 1e-35 means that it has not. The red dashed line shows the optimal rPLS model and the green circles indicate the variables selected by the algorithm. The thick red spectrum superimposed to the figure is the average of all the spectra in the dataset.**

The best iterative performance was obtained after eight iterations, as indicated also in (Figure 5.8 A). The global PLS model shows a predictive performance of RMSECV=1.70 while the rPLS global minimum shows a predictive performance of RMSECV=0.96. This result has two main advantages, (i) it performs clearly better than the global PLS model and (ii) it is three orders of magnitude more simple, as it contains only 25 variables instead of the 29149 spectral variables included in the global one, allowing the careful inspection of the single variables. Figure 5.8 B shows the regression vector of the global PLS model as well as the variables, counting for the NMR signals at 2.85 ppm (counting for twenty variables), 4.53 ppm (counting for only one variable), 3.69 ppm (counting for four variables), that have been identified by the rPLS as mainly responsible for the sour taste. The peaks can easily be

identified as the malic acid methylene (2.85 ppm) and methine (4.53 ppm) protons, while the peak at 3.69 ppm pertains to the glucose pyranose ring protons. These observations are in perfect agreement with the fact that the malic acid is the main acid in apple juice and therefore the main responsible for the sour taste of the samples. Moreover, the fact that glucose is also taken into account by the model, albeit with a numerically lower and negative regression coefficient, indicates an inverse correlation between malic acid content and glucose concentration. It is also interesting to notice that in the aromatic region, where polyphenols signals arise, the chlorogenic acid shows positive regression coefficients (Figure 5.8 B).

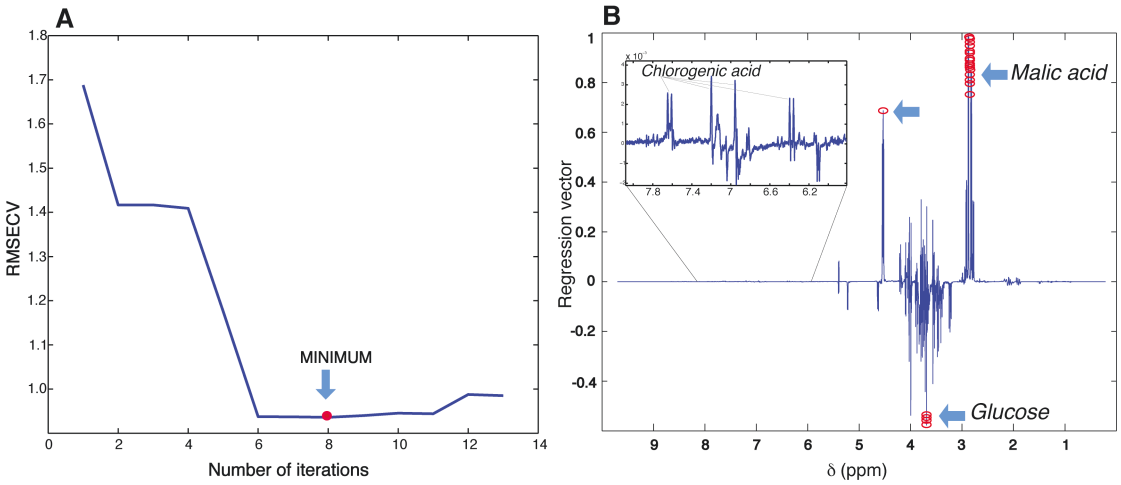


**Figure 5.8. A) The development in sourness prediction performance (RMSECV) during the rPLS iterations. B) The rPLS result for the prediction of sensory evaluated acidity. The regression coefficients for the full range PLS model (in blue) and for the rPLS reduced model (red circles).**

It is known that polyphenols can give bitterness and astringency to the apple juice (Berregi et al. 2003), however here the main polyphenol found in apple juice seems to have also a positive correlation with the sour taste. As far as the sweetness is concerned, the best rPLS result occurs after seven iterations. Also in this case the recursive approach brings a clear improvement when compared to the global PLS model, not only in terms of RMSECV, but also in terms of the number of variables to be inspected. Only five peaks have been selected by the rPLS. The signals around 3.81 and 3.67 ppm belong to sucrose and they are positively correlated to the sweetness, while the three glucose peaks (3.99, 3.79 and 3.68 ppm) are negatively correlated to this attribute. The inverse relation between sucrose and glucose content is already known from literature and it is likely due to their interconversion (Vermathen et al. 2011). Surprisingly, the sucrose turned out to be the main responsible for

the sweet taste even though the fructose is known to be the main sugar and thus sweetener in apple juice (Karadeniz & Ekşi 2002). This confirms the complexity in assigning the sweet taste to a specific chemical compound (Harker et al. 2002), since it should better be considered as the global result of the combination of several components. One of the advantages of the rPLS approach is that it does not only reduce the variable space and simplify the interpretation of the result, but it also includes the relevant covariation around the selected peaks. The latter information can be extremely useful for assignment purposes.

The results presented suggest a profitable use of the rPLS for the prediction of even more complex sensory features from different types of spectroscopic data.

# Chapter 6
# METABOLOMICS ON BIOLOGICAL SAMPLES: FROM CELLS TO HUMAN FLUIDS

Biological samples are by far the most commonly studied sample type in metabolic profiling studies. For instance, the analysis of the metabolome in body fluids has become an important tool to monitor the state of biological organisms as well as a diagnostic tool for disease (Zhang et al. 2012). Biological samples can include plasma, serum, urine, saliva, cerebrospinal fluid, synovial fluid, semen, tissue homogenates and cell extracts. Some of these fluids can be easily obtained with minimal invasion, thus allowing highly sampling frequency and, as a consequence, permitting a detailed characterisation of dynamic metabolic events. The absence of a specific sample pre-treatment in NMR-based metabolomics allow to obtain an unbiased "picture" of the sample chemical composition, thus reflecting the *in vivo* situation. NMR often needs specific settings to overcome potential issues caused by the high presence of proteins, lipids as well as the presence of water in aqueous samples (Keun & Athersuch 2011).
In this Chapter an overview of all the metabolomics studies conducted on biological samples is given.

## 6.1 **Study on Human Colon Cancer cells treated with anti-cancer drugs.** *(Paper I)*

Colon carcinoma is the third most commonly diagnosed cancer in the world and the second most common cause of death from cancer (Jemal et al. 2011). The analysis of metabolic profiles of this cell line provides a comprehensive assessment of the alterations in the metabolite levels in cells and can produce important information on in vitro actions of drugs towards their incorporation into novel therapeutic settings.

The aim of this study was to set up an optimized protocol for NMR metabolomics of adherent mammalian cell lines and perform a preliminary validation to drug-treated cancer

cells. Both traditional and more recent developed anti-cancer compounds were employed (Figure 6.1).

- *Compound 1*: one of the most promising G-quadruplex (G4) ligands discovered by virtual screening calculations (Di Leva et al. 2013)
- *Compound 2*: the pentacyclic acridine RHPS4 which is one of the most studied G4 ligands (Leonetti et al. 2004).
- *Compound 3*: the approved chemotherapeutic agent Adriamycin, also known as Doxorubicin (Young et al. 1981).

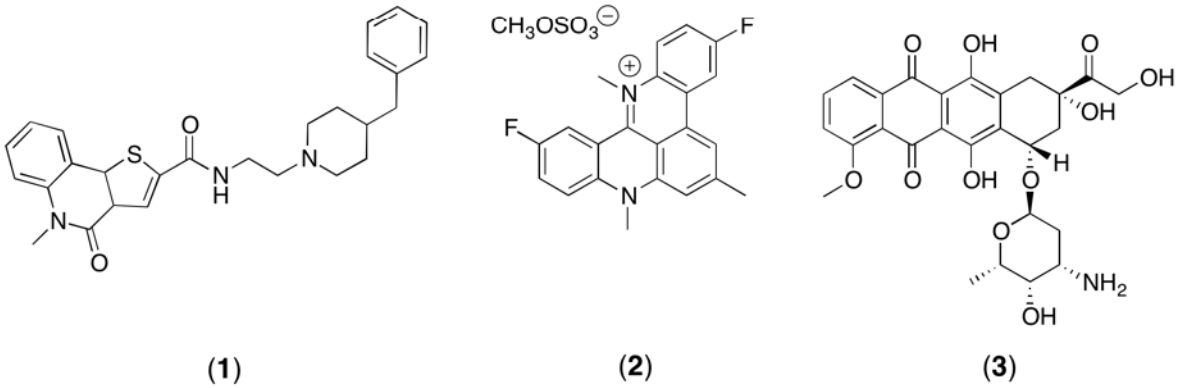


**Figure 6.1. The structure of compound (1), RHPS4 (2), and the structure of the traditional antitumor agent Adriamycin (3).**

### 6.1.1 *Experimental design and results discussion*

In order to reduce bias in the interpretation of the experiments, three biological replicates for each treatment were produced. Furthermore, three control samples (untreated cells) were also collected. Thus, a total of 12 samples were produced and studied by $^{1}H$ NMR spectroscopy. The whole design of experiment is summarized in Figure 6.2.

The dose and drug exposure duration time of cell culture for compounds 1 and 2 were established according to the literature ($IC_{50}$) (Salvati et al. 2007), while the optimal conditions for compound 3 were chosen on the basis of in-house unpublished results.

After cell treatment and washing, a quenching step with liquid $N_2$ was performed in order to stop the metabolism, hindering the metabolite degradation. Afterwards, a dual phase extraction (Bligh & Dyer 1959) was performed for the extraction of the intracellular metabolites. Only the aqueous upper phase contains water-soluble intracellular metabolites, was then analysed via NMR.

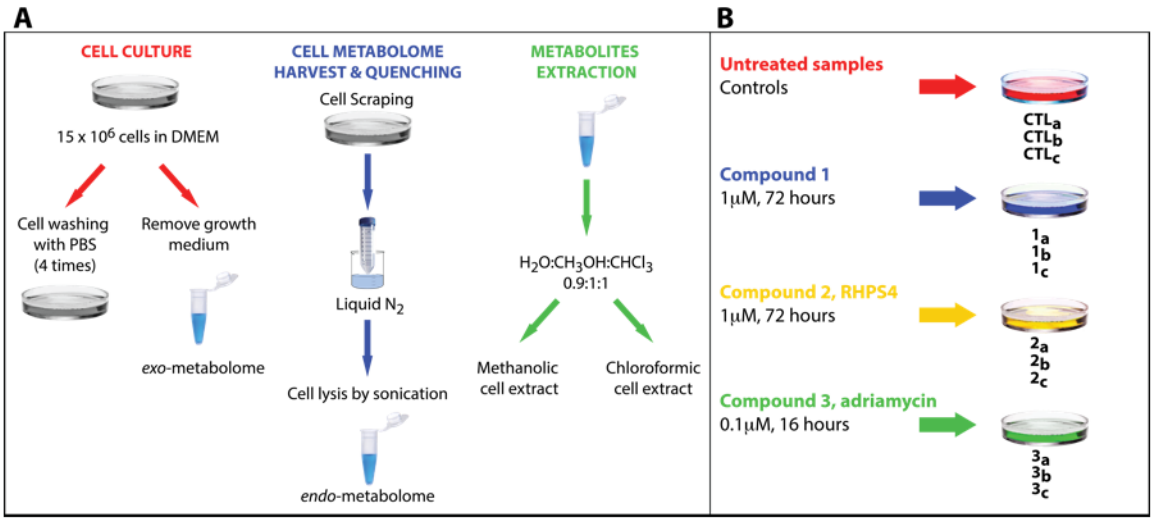


**Figure 6.2. A) General scheme describing the whole sample preparation protocol. B) Overview of the experimental design. Each compound has been tested in triplicate and three control samples (untreated ones) were also collected.**

1D NOESY-presat NMR experiments were carried out on the extracts, and the peak assignment was carried out (Figure 6.3).

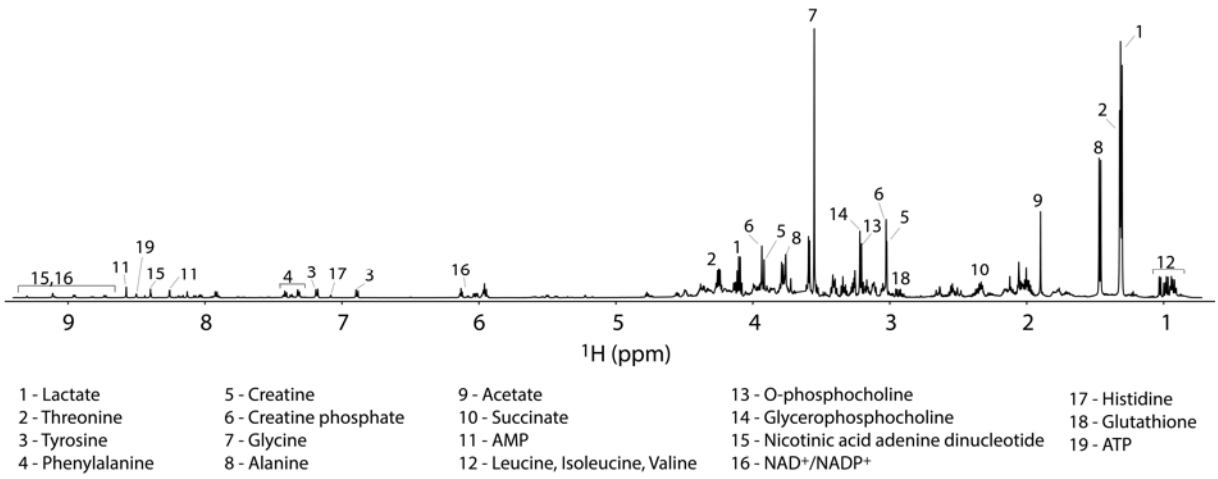


**Figure 6.3. $^{1}$H-NMR spectrum of a representative control sample along with the assignment of the signals.**

The 1D $^{1}$H NMR spectra were then processed and studied using a completely untargeted and unbiased multivariate data analytical approach. The aim was to identify the commonalities in the metabolic signatures associated with response to treatment for each

tested compound. For this reason, a principal component analysis (PCA) was performed on the NMR spectra (Figure 6.4).

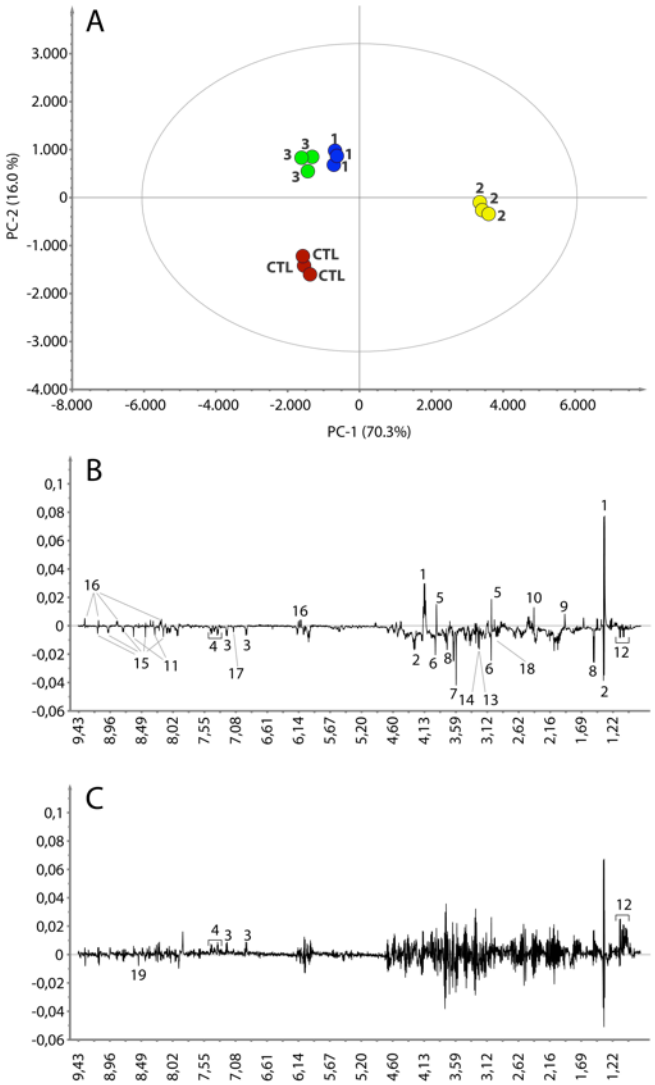


**Figure 6.4. A) PCA score plot. Control samples are coloured in red (CTL). Compound 1, 2 and 3 are coloured in blue, dark yellow and green, respectively B) PC-1 loading plot. C) PC-2 loading plot. Numbers on the loading plots refer to the NMR assignment reported in Figure 6.3.**

The PCA scores plot displaying the two main principal components (PCs) accounting for 86.3% of the variance (PC-1 70.3%, PC-2 16.0%) is shown in Figure 6.4A. The PCA scores plot shows that the samples of the cells treated with RHPS4 (2) are positioned on the extreme right side of the principal direction of variance PC1 and the samples of the cells treated with 1, 3 and controls are placed to the left. Along PC2 the treatments with 1 and 3, positioned in the up-left quadrant of the plot, differ from the control samples, which are found in the bottom-left quadrant. If we look at the PC1 loading plot (Figure 6.4B) it shows that the

samples treated with 2 are characterized by a higher content of lactate, creatine, acetate, succinate and $NAD^+$ /$NADP^+$ with 2 are characterized by a higher content of lactate, creatine, acetate, succinate and $NAD^+$/$NADP^+$, whereas the concentrations of threonine, glycine, alanine, tyrosine, phenylalanine, leucine, isoleucine, whereas the concentrations of threonine, glycine, alanine, tyrosine, phenylalanine, leucine, isoleucine, valine, histidine, creatine phosphate, glycerophosphocholine, O-phosphocholine, glutathione, NAAD valine, histidine, creatine phosphate, glycerophosphocholine, O-phosphocholine, glutathione, NAAD and AMP are lower with the respect of the samples that lie on the left of the plot. The loadings plot of the second principal component (Figure 6.4C) is much noisier than that observed for PC-1, thus it's difficult to draw reliable conclusions about the metabolome changes. In order to overcome this issue, a direct comparison of the average $^1H$ NMR spectra of the three replicates for each treatment and controls was performed. The detailed results of this approach are reported in *Paper I*.

In summary, the three tested compounds significantly altered the metabolism of the cells. The NMR data demonstrate that the treatments generally affect amino acid turnover or protein biosynthesis, tricarboxylic acid (TCA) cycle and mitochondrial activity (succinate, NAAD, NAD, ATP), urea cycle, anaerobic metabolism and protein and DNA biosynthesis and DNA repair. Furthermore, the specific alterations in the choline metabolism by compounds **1** and **2** indicate that cell death in HCT116 lines is induced interfering with DNA synthesis and DNA damaged repair and by inhibition of protein synthesis. The NMR data thus strongly suggest that treatments with compounds **1** and **2** slow down cellular metabolism, aggravate oxidative stress and reduces DNA synthesis and repair leading to cellular death and apoptosis in accordance with their anti-cancer activity. Compound **3** also drives cell death and apoptosis due to a general cytotoxicity in accordance with anti-cancer activity of Adriamycin (Cao et al. 2013).

In this study the preliminary insight into the biological behaviour of the three tested anti-cancer compounds was accomplished together with the implementation of a reliable NMR metabolomics analytical protocol for adherent mammalian cell lines.

## 6.2 **Study on plasma and faecal samples for the evaluation of effects of barley β-glucans in hypercholesterolemic rats.** *(Paper V - Manuscript draft)*

This study was carried out at the University of Copenhagen under the supervision of Prof. Søren Balling Engelsen.

In the last decade, barley (*Hordeum vulgare* L.), the world's fourth most produced cereal, normally employed for animal feed and beer production, has received increased nutritional attention due to its high content in soluble dietary fibre and its positive effects on lipid metabolism (Ho et al. 2016)**.**

In 2006 the consumption of barley products has been, indeed, associated by the Food and Drug Administration to the reduction of the Risk of Coronary Heart Disease for its cholesterol-lowering property (Food and Drug Administration 2006). Several studies have been conducted to prove the benefits of barley intake on plasma cholesterol on animal models such as rats (Yang et al. 2003; Kalra & Jood 2000) and hamsters (Tong et al. 2015; Wilson et al. 2004; Delaney et al. 2003), as well as humans (Behall et al. 2004a; Behall et al. 2004b; Keenan et al. 2007; Li et al. 2003; Rondanelli et al. 2011; Shimizu et al. 2008; Mikkelsen, Savorani, et al. 2014; Ibrugger et al. 2013). The component responsible for this healthy effect is called β-glucan (BG), a viscous soluble dietary fibre that can be found in cereals, mostly oat and barley, as well as yeast, bacteria, algae, and mushrooms (Theuwissen & Mensink 2008). It is a cell wall polysaccharide composed of glucose molecules that in oat and barley are joined by β-(1→4)- and β-(1→3)-glycosidic bonds (Figure 6.5).

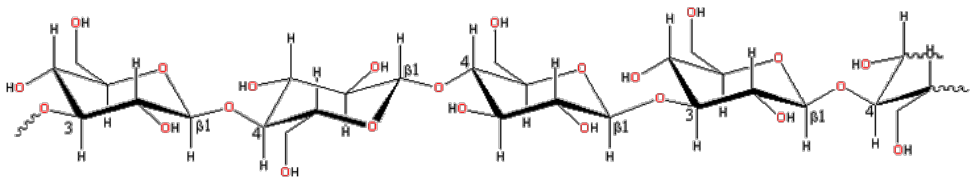

**Figure 6.5. Representative structure for β-glucans.** *[Adapted from www1.lsbu.ac.uk]*

The presence of this mixed linkage, which breaks up the regularity of the (cellulose) polymer, makes the molecule soluble and flexible (Lazaridou et al. 2004). The long chains of BG are characterized by repeating blocks of three (cellotriosyl, DP3) or four (cellotetraosyl, DP4) glucose units connected by β-D-(1,3) linkages. The so-called DP3/DP4 ratio as well as the numbers of β-(1→3)/ β-(1→4) linkages, are important functional parameters for β-glucan solubility which in turn is positively correlated with viscosity, a key feature for this fibre.

Interestingly, the mechanism of action of the BGs is based on the prevention of the intestinal reabsorption of bile acids and cholesterol (Figure 6.6). In particular, the hypotheses reported in literature that rationalize this mechanism include (i) the formation of a viscous barrier layer upon the enterocytes; (ii) the trapping of bile salts (BS) micelles containing cholesterol in a net made of fibres (Gunness & Gidley 2010; Wolever et al. 2010); and (iii) a direct hydrophobic interaction between the soluble fibre and the bile salts (Gunness et al. 2016).

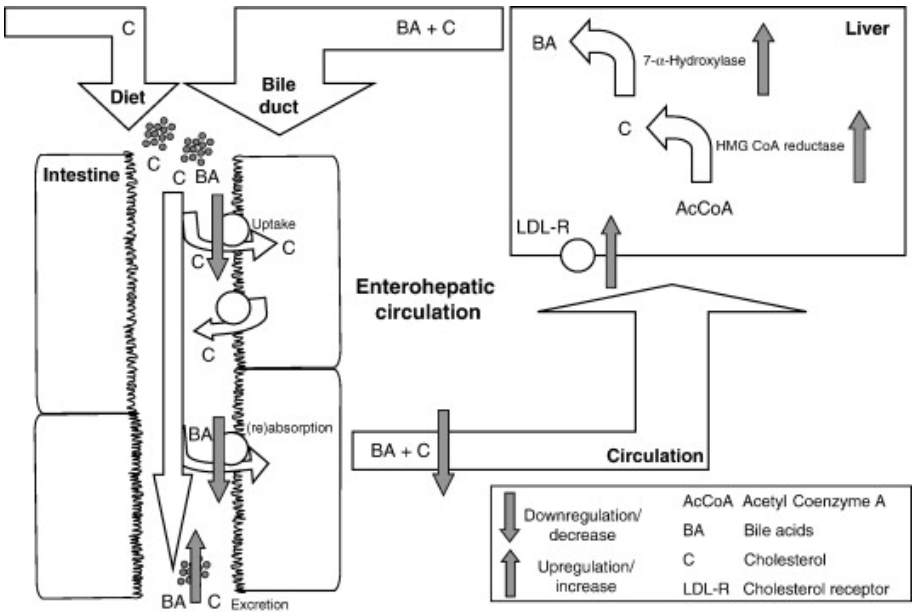


**Figure 6.6. Postulated hypocholesterolemic mechanism of water-soluble fibres.** *[Adapted from (Theuwissen & Mensink 2008)]*

As a consequence, faecal bile salts excretion increases dramatically and, in order to compensate for this loss, the endogenous cholesterol and *de novo* bile acid synthesis are promoted respectively via HMG-CoA and 7α-hydroxylase activation (CYP7A1) in the liver. Meanwhile, the receptors responsible for the uptake of the LDL cholesterol (LDL-R), also known as "bad cholesterol", are upregulated in order to provide the substrate for the bile acid synthesis cited before (Theuwissen & Mensink 2008). This leads to a further reduction of LDL cholesterol concentration in the blood. Therefore, the decrease in plasma cholesterol concentration is the indirect result of the BG intestinal action.

β-glucan preparations can vary in block structure, in MW, in purity and in MW distribution. So far, very few *in vivo* studies have compared the effect of BGs having different MW on plasma cholesterol. Immerstrand et al. have tested an entire MW range going from 10

to 2348 kDa, finding similar cholesterol-lowering effect in mice plasma (Immerstrand et al. 2010), whereas Wilson et al. have compared a reduced MW β-glucan (175 kDa) to a high MW one (1000 kDa), showing a slightly better effect of the latter in lowering the total and non-HDL cholesterol in hamsters (Wilson et al. 2004).

In this frame, Mikkelsen et al. decided to investigate a smaller range of molecular weight (from 150 to 530 kDa) BG concentrates on the plasma of hypercholesterolemic rats. In particular, an *in vivo* rat study was used to compare rat groups fed with medium (530 kDa) and low (150 kDa) MW β-glucan to the rats fed with commercially available Glucagel (100 kDa) (Mikkelsen et al. 2017). After four weeks of treatment, the authors observed a decreasing in total and LDL plasma cholesterol for all the studied groups. However, no significant difference among the groups fed with the three BG concentrates was found, suggesting that the BG molecular weight does not have any effect on the plasma lipids. Since such a targeted approach, that only takes into account the plasma lipids, might have hidden any other perturbation caused to the plasma by the ingestion of BGs, we decided, in collaboration with Mikkelsen and coworkers, to use an untargeted metabolomics approach to study the same plasma samples. Furthermore, in order to have a more complete picture of the function of BG, we decided also to study faecal samples of the treated rats by using GC-MS targeted towards bile acids.

### 6.2.1 *Experimental design*

A total of forty-eight male Wistar rats were used for the trial, conducted in triplicate (three blocks of sixteen rats per block), in order to increase the statistical power of the study. In each block, all rats were fed the control diet (CON) ad libitum for three weeks, in order to become hypercholesterolemic. After week 3, the rats were divided in four groups: control (CON) fed with the same diet of the first weeks, rats fed with the commercial glucan (GLU), rats fed with low molecular weight BG (LBG) and rats fed with medium molecular weight BG (MBG).

Four rats were housed in each cage, then, after 5.5 weeks, they were transferred to individual housing in metabolic cages for separate collection of faeces every second day during the last four days of week 7. Faecal samples were freeze-dried and stored at -20°C until analysis. Plasma was collected at the end of the study (week 7) and was kept at -80 °C until analysis (Mikkelsen et al. 2017). The experimental design is showed in Figure 6.7.

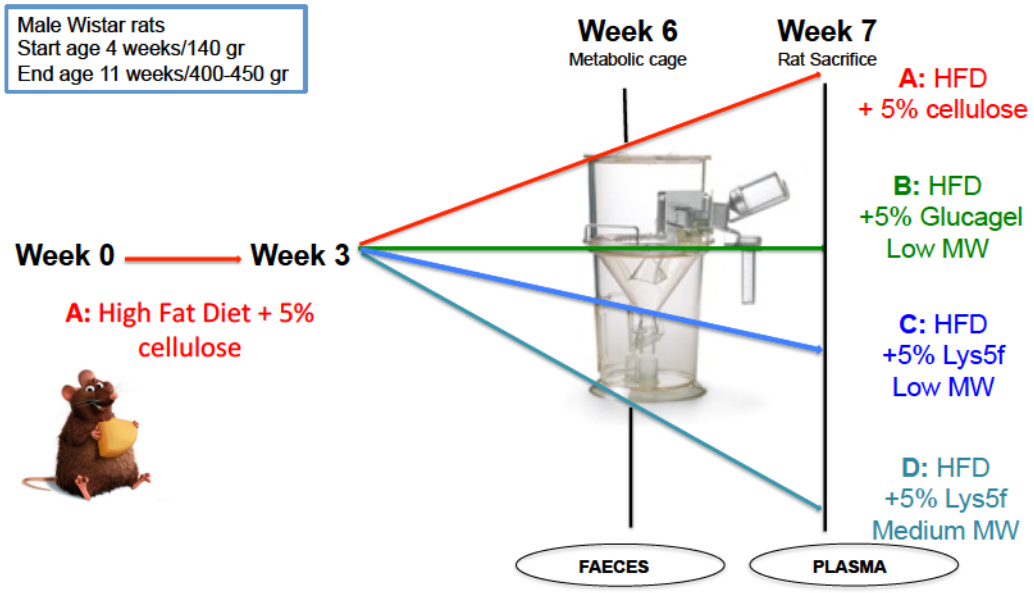


**Figure 6.7. Schematization of the rat study experimental design. From week 3, the rats were divided in four groups fed with different diets.**

### 6.2.2 *$^{1}$H NMR study on rat plasma*

Proton nuclear magnetic resonance spectroscopy was employed for the analysis of the rat plasma metabolome. Both CPMG and 1D NOESY NMR experiments were carried out, using a standardized protocol (Bruker GmbH, Rheinstetten, Germany) for the analysis of plasma. Spectra were aligned, normalized and pareto-scaled prior the multivariate data analysis.

A Principal Component Analysis was performed using both $^{1}$H CPMG and 1D-NOESY spectra. However, in both cases, no diet-related discrimination could be observed. The expected result for this analysis was the reduction of plasma lipids in BG treated groups compared to the CON group. Interestingly, our result was in agreement with Mikkelsen at al. (Mikkelsen et al. 2017) that measured, using a specific enzymatic kit, the concentration of plasma triglycerides, total cholesterol, low-density lipoprotein (LDL) and high-density lipoprotein (HDL) cholesterol in the four rat groups. No significant differences were observed in the lipids concentration between the CON and the BG treated groups, in line with a previous study conducted on humans (Ibrugger et al. 2013).

### 6.2.3 *GC-MS study on faecal samples*

Subsequently the faecal samples were investigated by GC-MS using a targeted method towards bile acids. Bile acids (BAs) and neutral sterols (NS) were extracted from the faecal samples using the protocol showed in Figure 6.8.

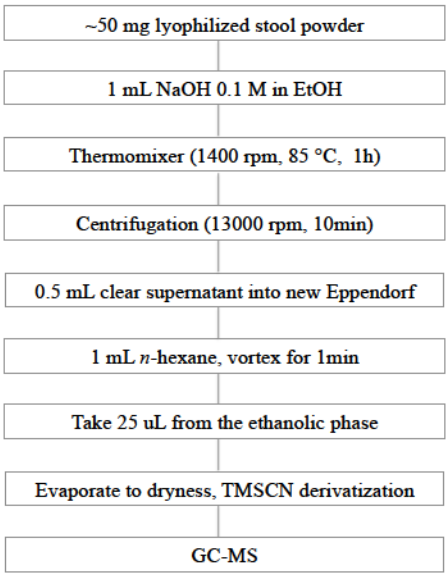


**Figure 6.8. Bile acid extraction protocol for GC-MS analysis.**

A total of 48 real samples with their replicate (96 samples in total) and 6 pooled samples (produced mixing the same amount of each real sample in the same tube), were submitted to GC-MS analysis.

The raw chromatograms were then manually divided into 33 intervals that were modelled by PARAFAC2 algorithm, using from one to seven component for each one, as described in section 4.2. After inspection of all the deconvoluted components (variables), some of them turned out to represent only baseline noise, reagent derived peaks or column bleed, and thus they were removed. Therefore, a final set of 54 variables was chosen, each having a specific EI mass values and Retention Index. Six out of 54 variables were safely identified as unconjugated bile acids, since authentic standards were used, namely: cholic acid (CA, variable 34 and 35), deoxycholic acid (DCA, variable 32 and 38), lithocholic acid (LCA, variable 29) and chenodeoxycholic acid (CDCA, variable 36). The fact that the same compound is represented by more than one variable is due to (i) the split of the peak belonging to the same molecule into two subsequent PARAFAC2 intervals (ii) different

trimethylsilyl derivatives having distinct retention times. Conjugated bile acids standards were also employed in the analysis, however, no corresponding peak was detected in the samples, in according with the results obtained by Dongowski and coworkers (Dongowski et al. 2002). The remaining metabolites were, as expected, neutral sterols, as well as fatty acids, fatty alcohols and disaccharides.

The final matrix (101x54) was then submitted to the Principal Component Analysis. The PCA, performed on the original dataset, lead to a fine discrimination between CON and BG groups, and also showed that the pooled samples were grouped all together as expected, thus indicating the soundness of the GC-MS analysis. After this evaluation, the pooled samples were removed from the data analysis, in order to just focus on the treatment-related variability. Furthermore, to be sure to exclude any interference from another potential source of variability, namely the block effect from the experimental design, an ANOVA-simultaneous component analysis (ASCA) was performed on the dataset.

Therefore, a new PCA was performed after ASCA. The PCA scores plot is reported in Figure 6.9. Inspection of PC1 versus PC3 produced an excellent sample separation. PC1 explained 20.08% of the total variance and separated the control group from the BG treated samples (LBG, MBG and GLU). The loadings for PC1 showed that all the identified bile acids are situated in the positive part of the plot, suggesting that the presence of 5% of BG in the test diets resulted in a higher BA excretion in the faeces. This is in agreement with previous studies that have reported an intestinal entrapment of bile salt micelles in a viscous network formed by the BGs, leading to an increased BA faecal excretion (Dongowski et al. 2002; Ellegård & Andersson 2007). Ellegard et al. showed an increased BS ileal content and hepatic BA synthesis after a BG-containing diet in ileostomic individuals, attributing this effect to the intestinal entrapment of BA caused by the fibre. In particular, they suggested that high viscosity oat-bran BG allows to entrap the entire BS-cholesterol mixed micelles in the intestinal lumen (Ellegård & Andersson 2007). Conversely, another study suggested that soluble dietary fibres form a barrier that increases the unstirred water layer lining the intestinal mucosal surface and hinder the formation of BA micelles (Theuwissen & Mensink 2008). Unfortunately, in this study, we were not able to study the effect of BGs on the cholesterol faecal excretion. This is due to the fact that the cholesterol signal in the GC-MS experiments turned out to be overloaded, and therefore not suitable for quantification purposes.

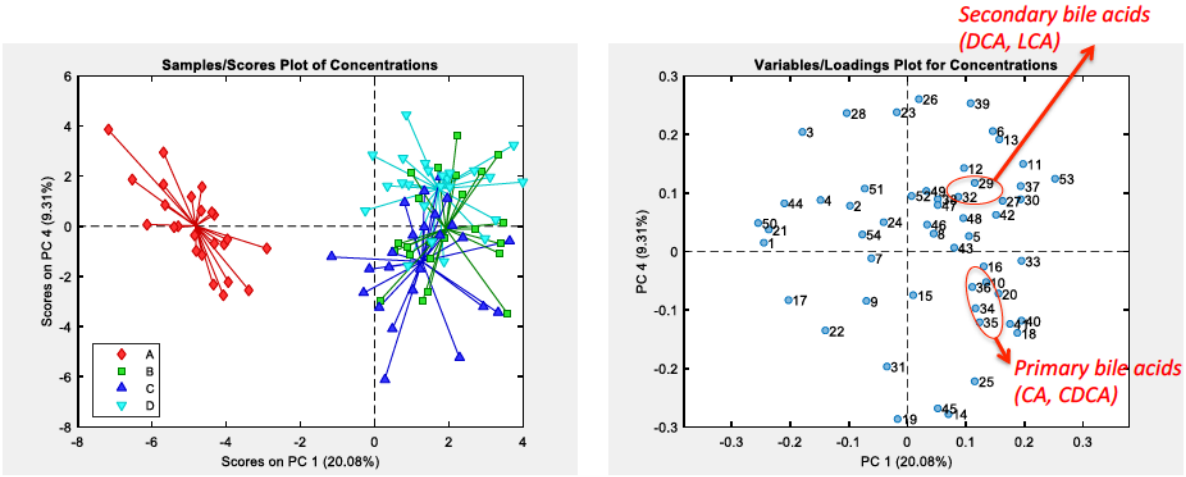


**Figure 6.9. PCA scores (left) and corresponding loadings plot (right) derived from the of PCA model performed on ASCA analysis of the GC-MS data of all the groups included in the experimental design (A= Control, B = Glucagel, C= Low MW BG, D = Medium MW BG). Variables shown on the loadings plot are numbered in the same order as in Table 2.**

Thereafter, in order to focus on the differences among the three BG treatments, the CON group was removed from the analysis, and a new PCA was computed (Figure 6.10). The new scores plot allows to better appreciate the separation among the LBG, MBG and GLU. The commercial β-glucan lies in the middle of the LBG and MBG groups, thus suggesting that low and medium molecular weight BG had a different influence on the faecal composition compared to Glucagel. In particular, the loadings plot reveals that the MBG group is characterized by higher levels of cholic acid while the LBG related samples show a higher amount of DCA and LCA. Even though a discrete number of unknown compounds is present in the loadings plot along the diagonal direction of separation of the three BG groups, the bile acids variables have the highest loadings values in the direction under consideration. Thus, they represent the main responsible for the observed separation indicating that the previous evaluations can be considered reliable. Cholic acid and chenodeoxycholic acid are defined as primary bile acids. They are derived from cholesterol by a sequence of enzymatic reactions occurring mainly in the liver. After the synthesis, they are first conjugated with glycine or taurine, and then excreted and stored in the gallbladder where they aggregate above a critical micellar concentration to form bile salts micelles.

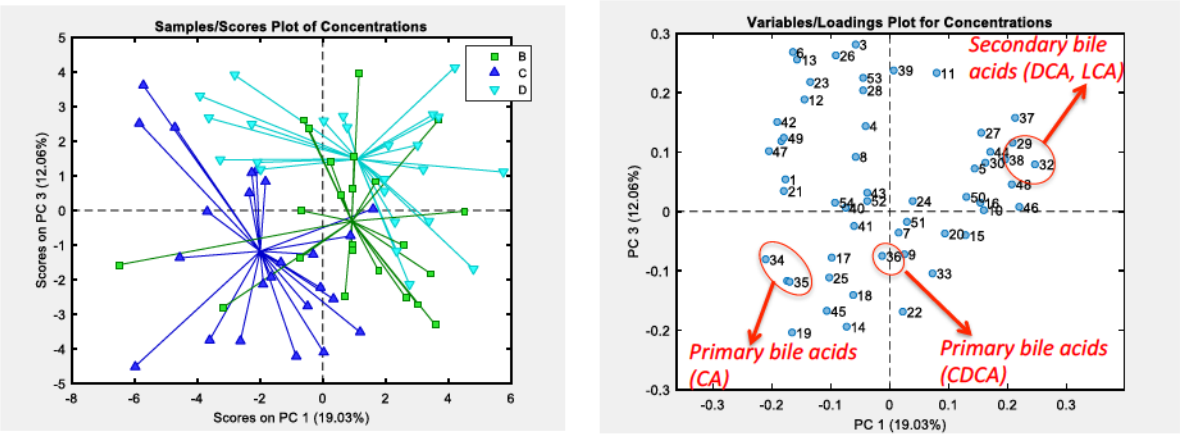


**Figure 6.10. PCA scores (left) and corresponding loadings plot (right) derived from the plots of PCA model performed on ASCA analysis of the GC-MS data of the BG-treated groups (B = Glucagel, C= Low MW BG, D = Medium MW BG). Variables shown on the loadings plot are numbered in the same order as in Table 2.**

These are then secreted into the duodenal lumen after meals in order to act as tensioactives and facilitate fat digestion. Once this action has been accomplished, BAs are largely reabsorbed (95%) in the terminal ileum by an active transport mechanism, in order to be sent back to the liver via portal blood thus taking part in the enterohepatic circulation (Hofmann et al. 1999). However, before going back to the liver, bile acids structure is altered on the side chain and on the nucleus by bacterial enzymes in the distal intestine. On the side chain, bile acids are deconjugated to form an unconjugated bile acid and glycine or taurine. Then, some of these unconjugated bile acids are absorbed, taking part to the enterohepatic cycle, and reconjugated in the hepatocytes. However, a small percentage (less than 5%) of unconjugated BAs remains in the intestine and enters in the colon (Nagengast 1995). Here, the steroidic nucleus undergoes to a 7α-dehydroxylation by anaerobic bacteria forming 7-deoxy bile acids. This process is responsible for the conversion of the primary into secondary bile acids. In particular, cholic acid is converted to deoxycholic acid, while chenodeoxycholic acid generates the lithocholic acid. Thus, DCA and LCA are called "secondary bile acids" because they are formed from primary bile acids. Generally, DCA is almost totally reabsorbed and enters the enterohepatic circulation, where it is conjugated in the liver and then secreted in the bile, whereas, LCA is almost insoluble and little of it is reabsorbed (Molino et al. 1986). However, the type and amount of BAs reabsorbed and excreted, are not constant but they depend on variations in diet, transit time, drugs, disease, etc. (Martínez-Augustin & de Medina 2008).

In this frame, our results confirmed that a diet containing BG increases the BAs excretion for the known viscous properties of this soluble fibre to entrap the bile acids/bile salts micelles. However, the most relevant outcome is that LBG and MBG seem to selectively increase the excretion of primary and secondary bile acids, respectively. To the best of our knowledge, this is the first time that such a modulation of faecal BAs excretion, depending on the molecular weight of the β-glucan, is reported, leading to the formulation of several hypotheses.

The direct interaction between β-glucans and bile salts micelles or monomers is one of the three hypothesis for the mechanism of action of the fibres (Gunness & Gidley 2010). Mikkelsen et al. (Mikkelsen et al. 2014) investigated the interaction between β-glucan and bile salts by Nuclear Magnetic Resonance assays, suggesting that the bile salts micelles are stabilized through transient, multivalent interactions with β-glucans, excluding, however, highly specific bindings. In this frame, we tend to exclude that the observed modulation may be due to a selective molecular interaction of LBG and MBG with primary and secondary BAs, respectively.

Another hypothesis includes the possibility that the conversion of primary into secondary BAs may be partially hindered in the LBG group. As described above, this conversion is performed by a group of anaerobic bacteria, thus it varies when colonic flora population, responsible for this biotransformation, changes (Ridlon et al. 2006). Mikkelsen et al. found significant alterations in the cecal microbiota of the BG groups compared to the CON group during their *in vivo* study. However, the three BG groups didn't show significant cecum microbiota alterations among them. Thus, suggesting that our result should not be ascribable to changes in the intestinal flora.

Furthermore, a crucial point in the BAs conversion is also represented by the colonic pH. Indeed, the 7α-deyhydroxylase, responsible for the BA transformation, is inhibited when the pH decreases (Van Munster 1993), thus reducing the amount of secondary bile acids produced. In this frame, the production of SCFA after the BGs fermentation process plays a key role, since it highly contributes to the pH lowering (Dongowski et al. 2002). In this study, the total amount of SCFA in the cecum was measured by Mikkelsen et al. resulting in higher levels of total organic acid pool in the MBG group. This would suggest that the pH is lower in the cecum of rats fed with MBG, as a consequence, the action of the enzyme would be prevented. Thus, we should expect to see less secondary bile acids in the faecal samples from this group compared to the GLU and LBG groups. In our case, the MGB samples showed,

instead, higher levels of secondary BAs. These considerations suggest that, probably, the main event that allows such a differentiation in the bile acid excretion has not to be found in an impaired BAs conversion, that however we cannot exclude, but in a "selective" entrapment of the BAs in the colon.

In this frame, viscosity could represent the key feature to explain the obtained results. The medium molecular weight BG has, indeed, 6-fold higher level of viscosity compared to the low molecular weight BG and Glucagel. Thus, a more viscous barrier may be able to prevent the reabsorption of more hydrophobic molecules as the secondary bile acids.

Indeed, when dietary fibres, as BG, are ingested, and partially fermented in the intestine, this leads to the formation of a layer of low density water that covers the thick layer of hydrophilic mucus (Chaplin 2003). The mucus and the fibre aqueous compartments can combine to produce a denser layer that delays or hinders the reabsorption of hydrophobic molecules that are potentially harmful.

Among the most common hydrophobic compounds that tend to accumulate in the distal colon, increasing the risk of cancer onset, there are polycyclic aromatic hydrocarbons from smoked foods, tryptophan metabolites from cooked red meat, fecapentaene-12 from anaerobic organisms, N-nitrosodimethylamine from protein and nitrates and secondary bile acids. Secondary BAs have been widely investigated for their role in cancer onset (Ajouz et al. 2014; Bernstein et al. 2011). In particular, lithocholic acid is considered a rare example of toxic endobiotics (Hofmann 2004) while deoxycholic acid was found to produce tumours in mice in 1939 (Cook 1940). The prolonged exposure of tissues to high physiological levels of BAs can lead to cause oxidative/nitrosative stress, DNA damage, mutation and cancer (Payne et al. 2008). Therefore, levels of secondary BA should be suppressed and the transit time of the gut contents should be decreased in order to reduce the risk of onset of carcinogenic process. In this frame the medium molecular weight BG could have a potential health benefit for the colonic mucosa.

All things considered, since secondary bile acids are more hydrophobic molecules than the primary BAs (Payne et al. 2008), this could suggest that they tend to be included in the high viscous barrier generated by the MBG more than in the layer generated from the LBG.

In summary, our findings demonstrate that medium (530 kDa) and low (150 kDa) MW BG as well as the commercially available Glucagel (100 kDa) concentrates from barley increase bile acids excretion in hypercholesterolemic rats. Interestingly, a modulation of the primary and secondary bile acid excretion occur, depending on the molecular weight (and thus the

viscosity) of the BGs employed, was observed. In particular, the viscosity of the BG is associated with greater excretion of the secondary bile acids. Despite the extensive literature on BGs, to the best of our knowledge, this is the first time that the MW of the β-glucan is reported to selectively influence the primary/secondary bile acids excretion.

Considering the growing interest in developing new nutraceutical barley-based products, the here reported results are extremely interesting demonstrating the considerable physiologic benefits of barley BG-enriched diet, with a particular focus on the molecular weight feature. Further studies are on going in our laboratory to further investigate the characteristics of the hydrophobic interactions between the bile acids and the tested BGs.

## 6.3 **Study on liver extracts to evaluate the effect of Phytosterols in murine colitis model**

Phytosterols are plant-derived sterols, structurally related to cholesterol. They have been classified into (i) *Sterols*, which have a double bond in the sterol ring (ii) *Stanols*, which lack a double bond in the sterol ring, so are saturated molecules (Figure 6.11). Phytosterols are known for their hypocholesterolemic effect, however they also present anti-inflammatory properties. Aldini and coworkers (Aldini et al. 2014), evaluated the effect of a mixture of phytosterols on prevention/induction/remission in a murine experimental model of colitis. Phytosterols were administered *per os* before, during and after colitis induction with Dextran Sodium Sulfate (DSS) in mice.

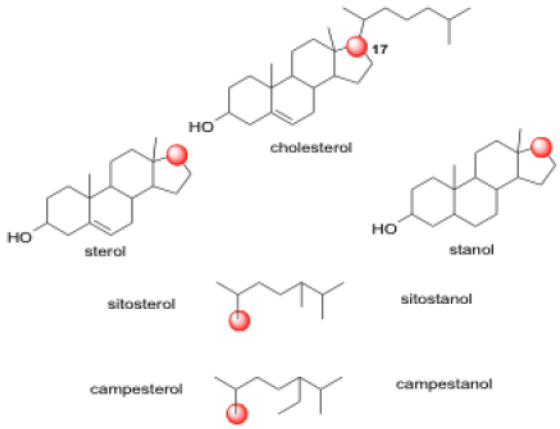


**Figure 6.11. Examples of phytosterols structures.**

The authors found that pre-treatment with phytosterols reduces the clinical symptoms and exerts a protective effect on DSS induced colonic inflammation decreasing infiltration of inflammatory cells and accelerating mucosal healing. These effects can be related to their antioxidant effects and to a regulation of the intestinal microflora. Phytosterols play also a role in the restoration of the intestinal motor pattern. These findings pave the way towards the role of phytosterols as potential nutraceutical tools in the management of Inflammatory Bowel Disease (IBD) and other intestinal inflammatory diseases.

This study has been developed in collaboration with the University of Bologna. The aim was to understand the metabolic pathways perturbed by the phytosterols in the liver as well as understand their role both on healthy and sick mice.

### 6.3.1 *Experimental design and results discussion*

A total of forty mice were divided in four groups in the first day of the study. Half of the animals (twenty mice) received the usual commercial control diet while the other half were fed the same diet enriched with a phytosterols (PH) preparation. For details about composition of the nutraceutical see (Aldini et al. 2014). After 14 days, the group fed the control diet was split in two more groups: controls (CT) that continued to receive the same diet and DSS treated group (DS) that received water containing DSS ad libitum in order to induce the colitis. In the same time, also the group fed the phytosterol-containing diet was split in two subgroups, exactly following the same procedure described above. Thus, the phytosterols treated group (PH) and the sick phytosterol treated group (PD) were generated. A schematization of the experimental design is given in Figure 6.12.

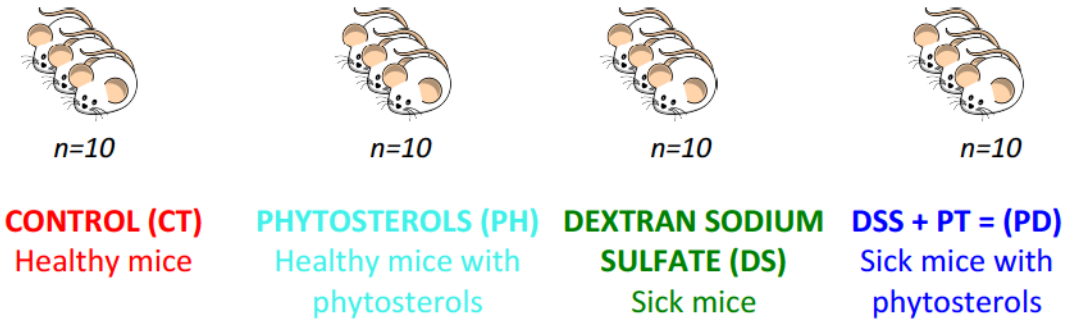


**Figure 6.12. Schematization of the experimental design for the phytosterols mice study.**

A dual phase extraction procedure was performed on the liver samples. Basically, a mixture of water, methanol and chloroform in the volume ratio of 1.8:2:2 was added to the lyophilised samples (Bligh & Dyer 1959; Wu et al. 2008).

About 40 mg of tissue were collected from each liver sample and mixed with about 960 µL of the extracting mixture. Then a homogenizing step was needed in order to break up the tissues and let the partition of the metabolites in the extracting solvents. A tissue homogenizer (FastPrep-24) with specialized Lysing Matrix beads was employed for this purpose. After this step, the supernatant was separated from the pellet. The latter was discarded. The supernatant was then centrifuged at 15000 rpm at 4 °C for 15 min at. This procedure generated a two-phase extract: the aqueous upper phase containing hydrophilic metabolites, while apolar metabolites as lipid molecules moved in the organic lower phase. Proteins and macromolecules are trapped, instead, in the thin skin-like layer between the two phases. The upper and lower phase were separated and transferred into different eppendorf tubes. Finally, solvents were completely removed from both fractions using a vacuum concentrator. Only the hydrophilic phase was taken into account in this study. All dried polar extracts were suspended in 540 µl of $D_2O$ together with 60 µl of a sodium phosphate buffer (1M, pH 7.2) to give a final buffer concentration of 0.1M. Samples were vortexed briefly and transferred into 5-mm NMR tubes for analysis. All 1D $^1$H-NMR spectra were acquired at 37 °C.

The assignment of the main hydrophilic metabolites was done according to the references available from literature (Feng et al. 2013; Shin et al. 2011). Chenomx NMR suite software and the Human Metabolome Database (HMDB) helped the metabolite identification.

A representative NMR spectrum of the liver extracts is showed in Figure 6.13.

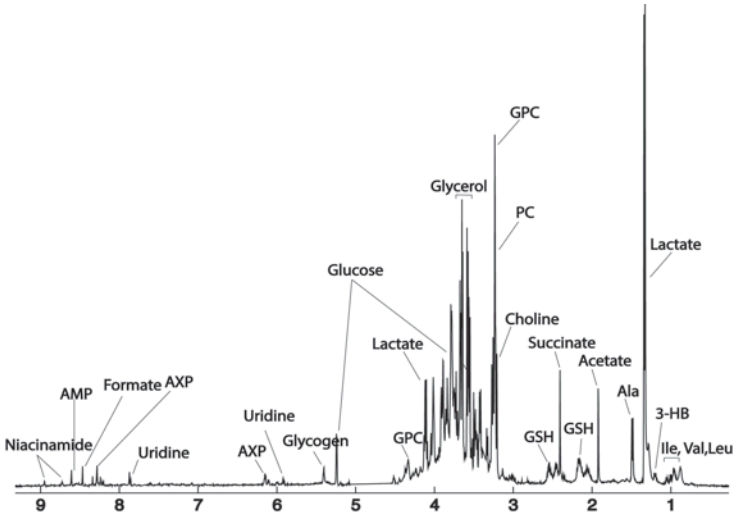


**Figure 6.13. $^{1}$H-NMR spectrum of a representative control sample along with the assignment of the signals. Keys: GPC, glycerophosphocholine; PC, phosphocholine; GSH, reduced glutathione.**

The NMR spectra were processed using iNMR software and then imported in Matlab. The NMR regions above 9.5 ppm and below 0 ppm were removed because contained only noise. Furthermore, regions between 7.67 and 7.72 ppm and between 3.35 and 3.38 ppm were discarded because of the residual signals of chloroform and methanol respectively. Finally, the region between 4.56 and 5.06 was discarded because of the residual signal of water. A global alignment step was carried out using icoshift: the acetate singlet at 1.92 ppm was chosen as reference signal. The NMR data matrix was then normalized to total area and pareto-scaled prior to multivariate data analysis. A unsupervised approach was chosen to explore the dataset, thus a PCA was computed (Figure 6.14). The PCA scores plot displaying the two main principal components (PCs) explained 46.2% of the variance (PC-1 31.4%, PC-2 14.8%). The score plot shows a nice separation among the samples belonging to the four different groups. Interestingly, a sort of trend that goes from the sick mice group to the healthy one is visible along the PC1. In particular, the livers extracted from the sick mice but treated with the phytosterols (PD, blue) are situated exactly in the middle of the controls (red) and the sick mice (green). This, at a glance, can suggest that the phytosterols do play an important role in bringing back the hepatic metabolome to the healthy condition.

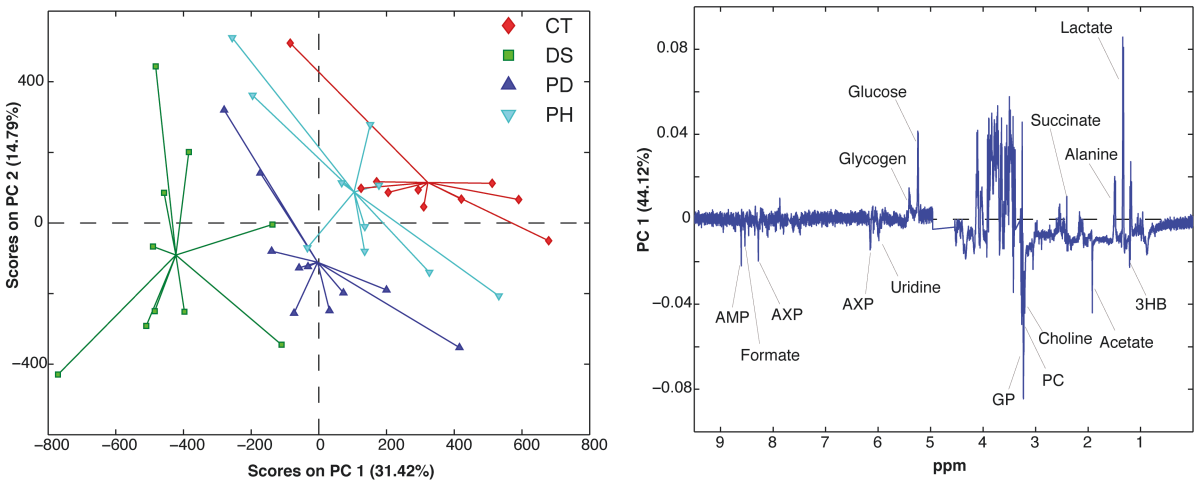


**Figure 6.14. PCA score plot (left) and loading plot (right) of the $^{1}$H NMR spectra of mice liver extracts. Keys: Controls (CT, red), Sick (DS, green), Sick with Phytosterols (PD, blue) and Healthy with Phytosterols (PH, turquoise).**

In order to understand the metabolic pathways perturbed by the presence of the intestinal disease caused by the DSS administration, it is just needed to look at the metabolite with higher loading values (both negative and positive values) in the loading plot (Figure 6.14).

Controls and sick mice are, indeed, at the very extreme of the PC1 in the score plot, thus meaning that loading plot of the corresponding component perfectly explains the difference between these two groups. However, two additional PCA were computed to allow a pairwise comparison of the studied groups. In particular, a PCA only with CT and DS was computed and it was compared to the PCA calculated including only PD (sick mice treated with phytosterols) and DS (sick mice). The idea was to evaluate if the administration of phytosterols before and after the disease, is able to bring the metabolome of a "sick" liver back to a healthy condition (Figure 6.15).

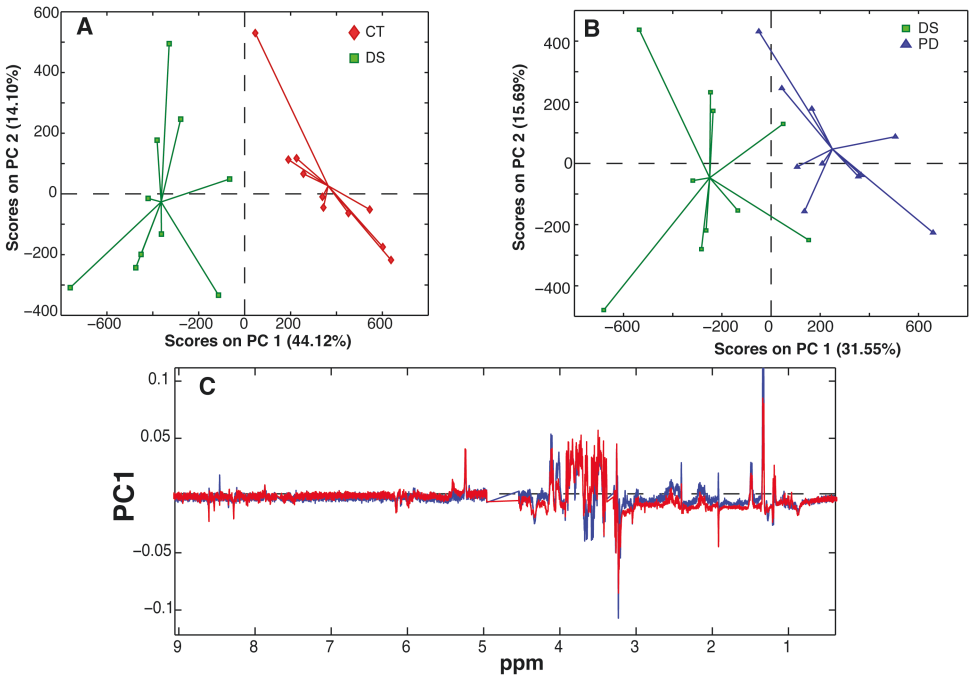


**Figure 6.15. A) PCA based on the pairwise comparison between controls (CT, red) and sick group (DS, green). B) PCA based on the pairwise comparison between sick mice fed the phytosterols (PD, blue) and sick group (DS, green). C) Superimposition of the PC1 loading plots of the PCA showed in A (red line) and in B (blue line).**

The superimposition of the two PC1 loadings plots showed that the phytosterols actually tend to bring the liver back to the healthy condition, since the loadings are very similar except for the acetate (1.92 ppm) and formate (8.34 ppm).

Interestingly, livers extracts from sick mice showed some dysfunctions in the nucleic acid synthesis/degradation pathways since UMP, UDP, UTP, CTP levels increase in this group. Moreover, it is well known that the intestinal damage caused by DSS reduces the absorption of nutrients. This could be a reason for the organism to move the glycolysis/ gluconeogenesis equilibrium towards the latter. Indeed, a reduction of alanine and lactate concentrations, major hepatic substrates for gluconeogenesis, is clearly visible in the sick mice livers. The negative loading value of the 3-hydroxybutyrate goes in the same direction of this hypothesis. 3-HB belongs to the ketonic bodies groups; an accumulation of these metabolites occurs when the lipids, and not the glucose, become the main source of energy for the organism.

Another typical colitis symptom is the dysregulation of lipid metabolism (Dong et al. 2013). Lipidic and phospholipidic metabolism are highly connected, thus the liver of the DS group shows an increase of glycerophosphocholine and phosphocholine.

These preliminary results show the strength of an NMR-based metabolomic study for the comprehension of the metabolic pathways perturbed in a living system, by the presence of a disease or a drug.

## 6.4 **Study on follicular fluid from patients with Polycystic Ovarian Syndrome (PCOS)** *(Paper IV – Submitted)*

Follicular fluid (FF) and cumulus cells (CC) have shown to play a key role in the oocyte health during its maturation. A highly coordinated network of interactions between the oocyte and the somatic cells influences the intrafollicular microenvironment allowing the folliculogenesis (Dumesic et al. 2015) (Figure 6.16).

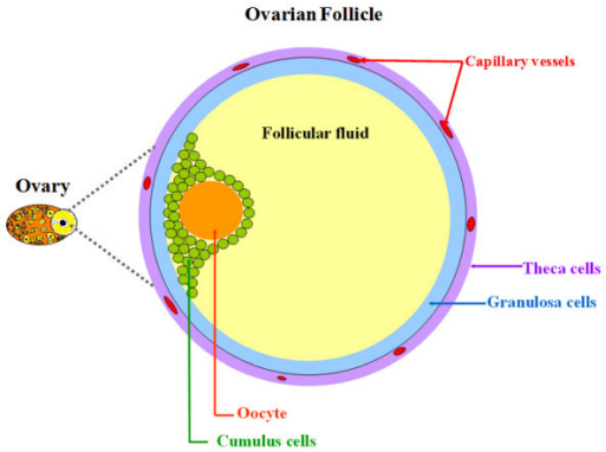


**Figure 6.16 Schematic representation of an ovarian antral follicle. Follicular fluid accumulates into the centre of the follicle and provides the micro-environment for growth, maturation and differentiation of follicular cells.** *[Adapted from (Fahiminiya et al. 2011)]*

Follicular fluid, also known as *liquor folliculi*, has been described for the first time in 1974 as an exudate from plasma enriched with secretions from the follicles (Edwards 1974). Thus, its composition varies when the plasma is altered by the presence of pathologies, and/or changes in the secretory processes of the granulosa and theca cells occur, reflecting the physiological state of the follicle. The variations in FF composition are strictly connected with both the maturation (Spitzer et al. 1996) and quality of the oocyte (Revelli et al. 2009). Indeed, Spitzer et al. observed that FF from immature follicles is characterized by a different protein pattern compared to the fluid collected from mature follicles, suggesting the potential

use of selected proteins as biomarkers for follicular maturity (Spitzer et al. 1996). The importance of finding biochemical predictors for the oocyte quality was investigated by Revelli et al. (Revelli et al. 2009). They analysed all the correlations reported in the literature between the embryo characteristics and several molecular markers in FF as hormones, growth factors, reactive oxygen species, anti-apoptotic factors, proteins, sugars and prostanoids. However, they were not able to identify substances as reliable markers for assessing the oocyte quality, most probably due to "univariate" scientific approach used. In fact, the same authors suggested to employ a metabolomic approach (based on a multivariate data analysis) that is more suitable to analyse complex biological mixture such as FF.

Metabolomics has been widely employed to analyse almost every kind of biological fluid using both nuclear magnetic resonance (NMR) and mass spectrometry. Compared to the latter, NMR has the advantage of being highly reproducible, requiring minimal sample handling and allowing the identification of a wide range of low-molecular-weight compounds. Pinero-Sagredo and coworkers were the first to perform a NMR study on FF, identifying the presence of at least 42 metabolites (Pinero-Sagredo et al. 2010). Significant correlations among glucose, β-hydroxybutyrate (3-HB), lactate, pyruvate, acetoacetate and acetate were found, thus suggesting the presence of an important anaerobic metabolism in overstimulated follicles. In addition, a statistically significant correlation was also observed between the glycolytic pathway and fatty acid metabolism in both young donors and the group with the higher fertilization rate. This study paved the way to the use of NMR-based metabolomics on FF for the discovery of biomarkers for the oocyte developmental competence (O'Gorman et al. 2013; Wallace et al. 2012).

Polycystic Ovary Syndrome (PCOS) represents the most commonly occurring metabolic and endocrinological disorder affecting 5–20% of women in their reproductive age worldwide (Goodarzi et al. 2011; Azziz et al. 2009). It is characterized by hyperandrogenism, ovulatory dysfunction and polycystic ovarian morphologic features. Being a syndrome, PCOS, is heterogeneous and also associated with other alterations such as repeated and quick gonadotropin-releasing hormone pulses, an excess of luteinizing hormone (LH), and insufficient follicle-stimulating hormone (FSH) secretion, which contributes to an excessive ovarian androgen production and ovulatory dysfunction. Furthermore, evidences of insulin resistance are common among women with PCOS, as a consequence of this, a compensatory hyperinsulinemia promotes adrenal and ovarian androgen production, thus contributing to the hyperandrogenism (McCartney & Marshall 2016).

One of the main features of PCOS is the increased number of antral follicles. In this frame, the growing interest in Anti-Müllerian Hormone (AMH) as follicular reserve marker has prompted many researchers to investigate the expression of this molecule in women with PCOS, in order to obtain information about the pathogenesis and the ability to calibrate the treatments in the case of ovulation induction or multiple follicular growth for medically assisted procreation cycles.

In order to understand the correlations among the FF composition, the hormones values and the hyper- and normoinsulemic conditions of PCOS women, we performed an exploratory NMR-based metabolomic study on 41 samples of FF provided from women that were undergoing an *in vitro* fertilization (IVF) therapy.

### 6.4.1 *Experimental design and results discussion*

FF samples from 41 women were provided by the "Centre for the Study and Treatment of Couple Sterility and Infertility" of Federico II University Hospital in Naples. Samples were collected by means of transvaginal ultrasound-guided puncture.

CPMG NMR experiments were performed in this study in order to reduce the broad resonances from high-molecular-weight compounds, allowing the observation of low-molecular-weight metabolites. A representative spectrum of FF sample is shown in Figure 6.17.

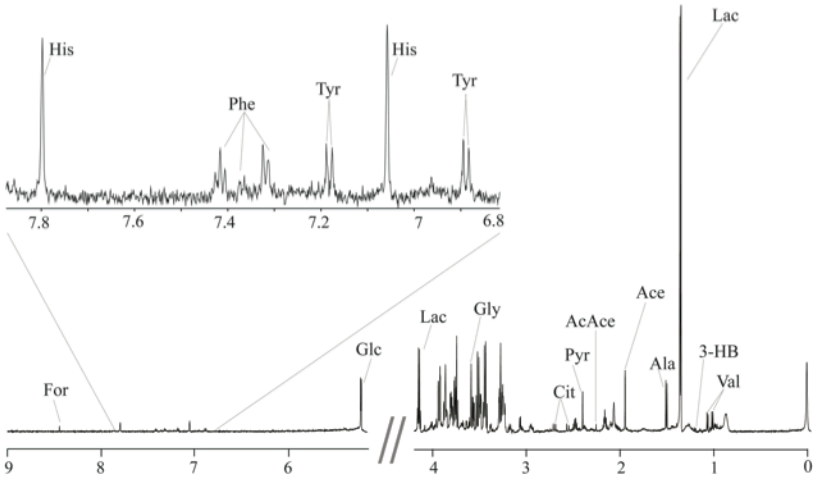


**Figure 6.17. $^{1}$H CPMG NMR spectrum of a representative follicular fluid sample measured at 700 MHz and 25 °C. Keys: Val = valine; 3-HB = 3-hydroxybutyrate; Lac = lactate; Ala = alanine; Ace = acetate; AcAce = acetoacetate; Pyr = pyruvate; Cit = citrate; Gly = glycine; Glc = glucose; Tyr = tyrosine; His = histidine; Phe = phenylalanine; For = formate.**

The FFs were characterized by low-molecular-weight metabolites including amino acids (valine, alanine, glycine, phenylalanine and tyrosine), organic acids (lactate, acetate, acetoacetate, pyruvate and formate) and glucose.

NMR spectra were then imported into MATLAB (R2012b; Mathworks, Natick, MA) where the spectral regions above 9 ppm and below 0 ppm were removed because containing only noise. Furthermore, the region between 4.16 and 5.20 ppm was discarded because it was dominated by the residual water signal. In order to correct for spectral misalignment, the entire dataset was globally aligned with respect to the acetate signal (1.91 ppm) using the icoshift algorithm (Savorani et al. 2010).

The data matrix was then submitted to the PLS toolbox version 8.1.1 (Eigenvector Research, Manson, USA) where the total area normalisation (1-norm) was performed prior to the pareto-scaling and mean-centering in order to compute the Principal Components Analysis (PCA).

An exploratory PCA was first performed on the data set consisting of the complete NMR spectra. The PCA scores plot (PC1/PC2, 48.8% of total variance explained) showed a slight separation of the samples according to the hyper-/normo-insulinemic condition of the PCOS patients along the diagonal direction of the plot. Unfortunately, the loading plot of PC1 (26.9% of the explained variance) is actually dominated by signals that cannot be assigned to traditional metabolites (except glycine), suggesting that they may be attributed to contaminant(s) that could affect the interpretation of the PCA. In order to analyse a PCA built by most reliable variables, all the peaks of the proton NMR spectrum that could be unambiguously assigned to known metabolites have been taken into consideration, and therefore integrated. Thus, 14 integral values were used to build a new data matrix that has been augmented by additional 12 clinical variables coming from the clinical measurements. This approach has the advantage to compare altogether a larger range of information, thus providing a more complete picture of the physiological condition of the studied population. So, a new PCA has been computed and it is reported in Figure 6.18.

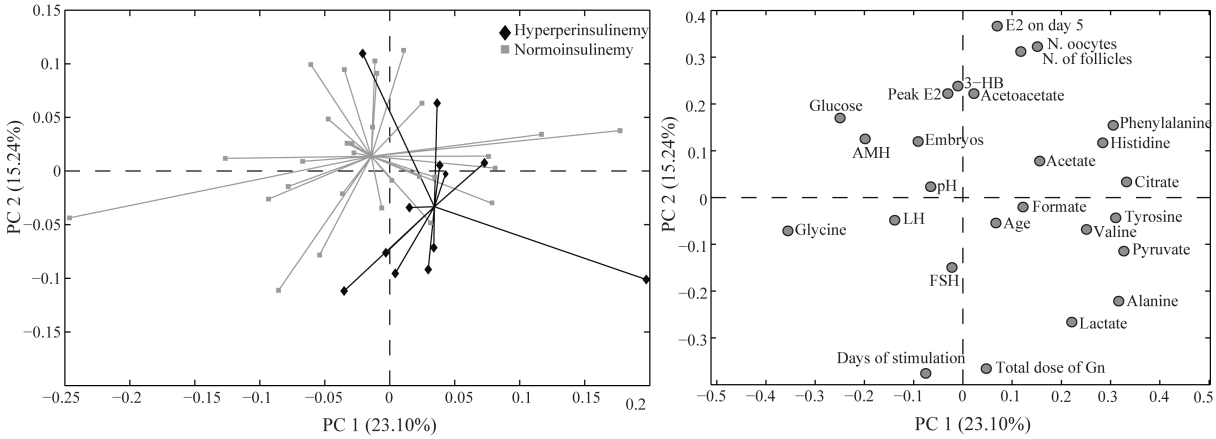


**Figure 6.18. PCA scores (left) and corresponding loadings plot (right) derived from the data analysis of the integrated peaks of the $^{1}$H NMR spectra of follicular fluids together with the clinical parameters measured among the PCOS women.**

As shown in the score plot, a better separation occurs between the hyperinsulinemic and normoinsulinemic PCOS women. The variables that better explain this separation are those having larger loading values along the direction of the sample separation (see the loadings plot in Figure 6.18). Therefore, the normoinsulinemic group is characterized by high value concentrations of glucose and AMH, while hyperinsulinemic women present higher concentration of lactate, alanine and pyruvate. From a metabolic point of view, glucose, lactate and pyruvate are strictly correlated to each other, being all metabolites of the glycolytic pathway. Interestingly, in normal physiological conditions, the follicles grow in an anaerobic environment since the avascular granulosa layer (that surround them) increases its thickness preventing the oxygen supply in the follicle (Gull et al. 1999). In particular, granulosa cells (GC) and cumulus cells (CC) are required to provide products of glycolysis for the development of the oocyte, which is unable to carry out this pathway (Sugiura et al. 2005). In these conditions, the pyruvate transformation into lactate is obtained due to the limited amounts of oxygen available. Pyruvate is the final key 3-carbon (3-C) intermediate transferred to the oocyte, however it is interchangeable with lactate which is the main 3-C compound produced by the GC. Likewise, alanine is linked to glycolysis pathway. In fact, alanine can be also produced by reductive amination of pyruvate. The increase of alanine concentration in the insulin-resistant PCOS women (compared to non insulin resistant patients) is in perfect agreement with a previous study reported by Zhang et al. (Zhang et al. 2014).

As mentioned above, among the variables that mostly contributed to the separation of the normoinsulinemic and hyperinsulinemic women there is also the anti-müllerian hormone (AMH). This hormone is a member of the transforming growth factor-β (TGF-β) superfamily and it is considered an important marker of ovarian reserve (van Rooij 2002). AMH values are important markers for PCOS, in fact, women affected with syndrome, have 2- to 3-fold higher levels of AMH than healthy women (Pigny et al. 2003). Interestingly, Takahashi et al. found that oocytes are more likely to be fertilized when follicles are able to make high concentrations of AMH in the follicular fluid. Thus, they speculated that AMH could be considered a prediction marker for fertilization (Takahashi et al. 2007). Lin et al. found also a negative correlation between AMH and BMI. This could be in line with our result, as normoinsulinemic women have a lower BMI than the insulin-resistant group. However, the authors didn't find any correlation between the hormone values and the insulin resistance parameters (Lin et al. 2011).

The study of the correlations among all the metabolites has been demonstrated to be a valuable analytical tool in systems biology (Steuer 2006). For this reason, the Pearson's correlation analysis of the metabolites was performed. At a threshold of $r \geq |0.5|$, 16 positive and 7 negative significant correlations were found, and the metabolites characterized by the highest correlation values turned out to be glucose, lactate, pyruvate, acetoacetate and 3-HB. We have found significant ($p < 0.05$) negative correlations between glucose and lactate (-0.78), and between glucose and pyruvate (-0.70). On the contrary, we found significant positive correlations between pyruvate and lactate (0.70), and between acetoacetate and 3-HB (0.95). Interestingly, these data are in agreement with the metabolite correlations found in FF of women not presenting the PCOS (Pinero-Sagredo et al. 2010). These correlations strengthens the hypothesis of an important anaerobic metabolism occurring in the hyperstimulated follicles (Gull et al. 1999), and the strong correlations among glucose, pyruvate and lactate confirm that FF provides lactate and pyruvate to the developing oocyte as a source of energy. Acetoacetate and 3-HB are synthesised from acetyl-CoA by fatty acid oxidation and their correlation can be explained by the fact that they are interconvertible by means of the β-hydroxybutyrate deidrogenase.

In summary, FF represents a suitable source of information since it is superfluous, abundant and easily available material during the IVF treatment. Metabolomics studies on this fluid can provide useful information about changes in the physiological state of patients, alterations of metabolic pathways, as well as biomarkers for oocyte quality and IVF success

rate. In the case reported here, the hyperinsulinemia in PCOS patients is responsible for a different FF metabolic profile. In particular, hyperinsulinemia seems to be also associated with impaired carbohydrate/glucose and lipid metabolism. Interestingly, AMH turned out to be positively correlated with glucose and negatively correlated with lactate, pyruvate and alanine. Although the data reported here are preliminary, this study paves the way to a better comprehension of the relationships among hormones and metabolites. To the best of our knowledge, this is the first study that attempts to correlate AMH values with the FF metabolites.

# Chapter 7
# CONCLUSIONS

The work presented in this Ph.D. thesis shows six examples of NMR-based metabolomics applications in the field of food, pharmaceutical, nutraceutical and medical sciences. In particular, the metabolomics approach was used to identify key metabolites as indicators of geographical origin and cultivar peculiarities (in food-related applications), as well as to identify, in all the other projects, candidate biomarkers of (i) *in vitro* anticancer activity, (ii) *in vivo* effects of nutraceuticals and (iii) disease related patterns.

NMR-based metabolomics has been widely used in recent years to obtain a 'holistic view' of the metabolome (foodome) of various kinds of beverages and foods. In Chapter 5, explorative studies on extra virgin olive oil and apple juice have been performed to retrieve information about chemical composition, cultivar-related variability, and influence of the growing condition. Algerian extra virgin olive oil composition turned out to be very variety-dependant. Each variety showed strength and weakness points in the chemical composition suggesting the potentiality of these cultivars to produce high quality blends that could be able to compete with other Mediterranean products. The metabolomics approach also allowed to correlate the chemical fingerprint of the cultivars to the geographical area of collection. In particular, the oils produced from olives collected in the region closer to the Mediterranean Sea were richer in polyphenols than the samples collected in the inland region. In the second study, a NMR-based metabolomics approach was applied, for the first time, to chemically characterize apple juices of about 100 ancient Danish cultivars. As in the case of the olive oil, also the apple juice chemical composition turned out to be very variety-dependant. In particular, the analysis of the aromatic region of the NMR spectrum showed that some juices have an interesting composition in polyphenolic content, suggesting the formulation of precisely characterized niche products. Moreover, this project also offered the possibility to demonstrate the application of a novel variable selection method (rPLS) for the prediction of sweet and sour taste of apple juice from Nuclear Magnetic Resonance (NMR) spectra.

In Chapter 6 the results obtained from four different metabolomics studies on samples of biological origin is provided. The first study, related to cell metabolomics, allowed to develop

a reliable experimental protocol for an efficient harvesting, quenching and extraction of cellular metabolites of human adherent cancer cell lines, while providing also a preliminary insight to the *in vitro* action of the three anticancer drugs tested. The second and the third studies concerned the evaluation of functional food ingredients, namely, β-glucans and phytosterols, using *in vivo* animal models. The β-glucans study demonstrated that medium and low molecular weight β-glucans, as well as the commercially available Glucagel concentrates from barley, increase bile acids excretion in hypercholesterolemic rats. Interestingly, a modulation of the primary and secondary bile acid excretion occurs, depending on the molecular weight (and thus the viscosity) of the BGs employed. Despite the extensive literature on BGs, this is the first time that the MW of the β-glucan is reported to selectively influence the primary/secondary bile acids excretion. Considering the growing interest in developing new nutraceutical barley-based products, the results reported here are extremely interesting, since demonstrate the considerable physiologic benefits of barley BG-enriched diet. Similar study has been performed to evaluate effect of phytosterols on the intestinal inflammation. The results of this explorative study showed the perturbation caused in the liver metabolome after the colitis induction, as well as the role of phytosterols in restoring the homeostatic equilibrium. The last study concerned the metabolomics study of follicular fluid from women affected with the Polycystic Ovary Syndrome. Although reported results are preliminary, they pave the way to a better comprehension of the relationships among serum hormones and follicular fluids metabolites.

The outcome of these studies further encourages the use of metabolomics to obtain a ‘holistic view’ of food and biological matrices. In particular, this thesis reports examples of the application of this approach and demonstrates, once again, how to unravel comprehensive biological questions and how to potentially improve products and quality of life.

# REFERENCES


Food and Administration and FDA Events & News (2006) FDA Finalizes Health Claim Associating Consumption of Barley Products with Reduction of Risk of Coronary Heart Disease. , pp.06–70. Available at: http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/2006/ucm108657.htm.

Ajouz, H., Mukherji, D. & Shamseddine, A., 2014. Secondary bile acids: an underrecognized cause of colon cancer. *World journal of surgical oncology*, 12(1), p.164.

Aldini, R. et al., 2014. Antiinflammatory effect of phytosterols in experimental murine colitis model: Prevention, induction, remission study. *PLoS ONE*, 9(9).

Amigo, J.M. et al., 2008. Solving GC-MS problems with PARAFAC2. *TrAC - Trends in Analytical Chemistry*, 27(8), pp.714–725.

Azziz, R. et al., 2009. The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report. *Fertility and Sterility*, 91(2), pp.456–488.

Barantin, L., Pape, A. Le & Akoka, S., 1997. A new method for absolute quantitation MRS metabolites. *Magnetic Resonance in Medicine*, 38(2), pp.179–182.

Beauchamp, G.K. et al., 2005. Phytochemistry: Ibuprofen-like activity in extra-virgin olive oil. *Nature*, 437(7055), pp.45–46.

Beckonert, O. et al., 2007. Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts. *Nature Protocols*, 2, pp.2692–2703.

Behall, K.M., Scholfield, D.J. & Hallfrisch, J., 2004a. Diets containing barley significantly reduce lipids in mildly hypercholesterolemic men and women. *Am J Clin Nutr*, 80(5), pp.1185–1193.

Behall, K.M., Scholfield, D.J. & Hallfrisch, J., 2004b. Lipids Significantly Reduced by Diets Containing Barley in Moderately Hypercholesterolemic Men. *Journal of the American College of Nutrition*, 23(1), pp.55–62.

Belloque, J., 1999. Application of NMR spectroscopy to milk and dairy products. *Trends in Food Science & Technology*, 10(10), pp.313–320.

Belton, P.S. et al., 1998. Application of chemometrics to the $^{1}$H NMR spectra of apple juices: Discrimination between apple varieties. *Food Chemistry*, 61(1–2), pp.207–213.

Belton, P.S. et al., 1997. High-field proton NMR studies of apple juices. *Magnetic Resonance in Chemistry*, 35(13), pp.S52–S60.

Bernstein, C. et al., 2011. Carcinogenicity of deoxycholate, a secondary bile acid. *Archives of Toxicology*, 85(8), pp.863–871.

Berregi, I. et al., 2003. Quantitative determination of (-)-epicatechin in cider apple juices by 1H NMR. *Talanta*, 61(2), pp.139–145.

Berry, D.P. et al., 2011. The integration of "omic" disciplines and systems biology in cattle breeding. *animal*, 5(4), pp.493–505.

Bligh, E.G. & Dyer, W.J., 1959. A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology*, 37(8), pp.911–917.

Bloch, F., Hansen, W.W. & Packard, M., 1946. The nuclear induction experiment. *Physical Review*, 70(7–8), pp.474–485.

Bro, R. et al., 2010. Mathematical chromatography solves the cocktail party effect in mixtures using 2D spectra and PARAFAC. *TrAC - Trends in Analytical Chemistry*, 29(4), pp.281–284.

Bro, R., 1997. PARAFAC. Tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, 38(2), pp.149–171.

Bro, R. et al., 1999. PARAFAC2—Part II. Modeling chromatographic data with retention time shifts. *Journal of Chemometrics*, 13(3–4), pp.295–309.

Bylund, D. et al., 2002. Chromatographic alignment by warping and dynamic programming as a pre-processing tool for PARAFAC modelling of liquid chromatography-mass spectrometry data. *Journal of Chromatography A*, 961(2), pp.237–244.

Camacho, D., de la Fuente, A. & Mendes, P., 2005. The origin of correlations in metabolomics data. *Metabolomics*, 1(1), pp.53–63.

Cao, B. et al., 2013. Metabolomic approach to evaluating adriamycin pharmacodynamics and resistance in breast cancer cells. *Metabolomics*, 9(5), pp.960–973.

Carr, H.Y. & Purcell, E.M., 1954. Effects of diffusion on free precession in nuclear magnetic resonance experiments. *Physical Review*, 94(3), pp.630–638.

Chaplin, M.F., 2003. Fibre and water binding. *The Proceedings of the Nutrition Society*, 62(1), pp.223–227.

Cloarec, O. et al., 2005. Statistical Total Correlation Spectroscopy: An Exploratory Approach

for Latent Biomarker Identification from Metabolic 1H NMR Data Sets. *Analytical Chemistry*, 77(5), pp.1282–1289.

Cook, J., 1940. Production of Tumours in mice by Deoxycholic acid. *Nature*.

Delaney, B. et al., 2003. Beta-glucan fractions from barley and oats are similarly antiatherogenic in hypercholesterolemic Syrian golden hamsters. *The Journal of nutrition*, 133(2), pp.468–475.

Dieterle, F. et al., 2006. Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in $^{1}$H NMR Metabonomics. *Analytical Chemistry*, 78(13), pp.4281–4290.

Dong, F. et al., 2013. Systemic responses of mice to dextran sulfate sodium-induced acute ulcerative colitis using 1H NMR spectroscopy. *Journal of Proteome Research*, 12(6), pp.2958–2966.

Dongowski, G. et al., 2002. Dietary fiber-rich barley products beneficially affect the intestinal tract of rats. *The Journal of nutrition*, 132(12), pp.3704–3714.

Dumesic, D.A. et al., 2015. Oocyte environment: Follicular fluid and cumulus cells are critical for oocyte health. *Fertility and Sterility*, 103(2), pp.303–316.

Dunlap, T. et al., 2014. NIH Public Access. , 6(12), pp.1771–1792.

Dunn, W.B. & Ellis, D.I., 2005. Metabolomics: Current analytical platforms and methodologies. *TrAC Trends in Analytical Chemistry*, 24(4), pp.285–294.

Edwards, R.G., 1974. Follicular fluid. *Journal of Reproduction and Fertility* , 37(1), pp.189–219.

Ellegård, L. & Andersson, H., 2007. Oat bran rapidly increases bile acid excretion and bile acid synthesis: an ileostomy study. *Eur J Clin Nutr*, 61(8), pp.938–945.

Emwas, A.H. et al., 2015. Standardizing the experimental conditions for using urine in NMR-based metabolomic studies with a particular focus on diagnostic studies: a review. *Metabolomics*, 11(4), pp.872–894.

Euceda, L.R., Giskeodegard, G.F. & Bathen, T.F., 2015. Preprocessing of NMR metabolomics data. *Scandinavian journal of clinical and laboratory investigation*, 75(3), pp.193–203.

Fahiminiya, S. et al., 2011. Proteomic analysis of mare follicular fluid during late follicle development. *Proteome Science*, 9(1), p.54.

Fan, T.W.-M. & Lane, A.N., 2016. Applications of NMR spectroscopy to systems biochemistry. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 92–93, pp.18–53.

Farrant, R.D. et al., 2010. NMR quantification using an artificial signal. *Magnetic Resonance in Chemistry*, 48(10), pp.753–762.

Feng, J. et al., 2013. Studies of Secondary Melanoma on C57BL/6J Mouse Liver Using 1H NMR Metabolomics. *Metabolites*, 3(4), pp.1011–1035.

Fiehn, O., 2001. Combining genomics , metabolome analysis , and biochemical modelling to understand metabolic networks. , pp.155–168.

Fiehn, O., 2002. Metabolomics – the link between genotypes and phenotypes. *Plant Molecular Biology*, 48(1), pp.155–171.

Flemming H. Larsen, 2006. An exploratory chemometric study of 1H NMR spectra of table wines. *Journal of Chemometrics*, (20), pp.198–208.

Forina, M., Casolino, C. & Millan, C.P., 1999. Iterative Predictor Weighting (IPW) PLS: a technique for the elimination of useless predictors in regression problems. *Journal of Chemometrics*, 184(November 1998), pp.165–184.

Förster, J. et al., 2003. Genome-Scale Reconstruction of the Saccharomyces cerevisiae Metabolic Network. *Genome Research* , 13(2), pp.244–253.

Francini, A. & Sebastiani, L., 2013. Phenolic Compounds in Apple (Malus x domestica Borkh.): Compounds Characterization and Stability during Postharvest and after Processing. *Antioxidants*, 2(3), pp.181–193.

Godelmann, R. et al., 2013. Targeted and nontargeted wine analysis by 1H NMR spectroscopy combined with multivariate statistical analysis. differentiation of important parameters: Grape variety, geographical origin, year of vintage. *Journal of Agricultural and Food Chemistry*, 61(23), pp.5610–5619.

Goodarzi, M.O. et al., 2011. Polycystic ovary syndrome: etiology, pathogenesis and diagnosis. *Nat Rev Endocrinol*, 7(4), pp.219–231.

Gull, I. et al., 1999. Anaerobic glycolysis: The metabolism of the preovulatory human oocyte. *European Journal of Obstetrics & Gynecology and Reproductive Biology*, 85(2), pp.225–228.

Gunness, P. et al., 2016. Molecular interactions of a model bile salt and porcine bile with (1,3:1,4)-β-glucans and arabinoxylans probed by 13C NMR and SAXS. *Food Chemistry*, 197, pp.676–685.

Gunness, P. & Gidley, M.J., 2010. Mechanisms underlying the cholesterol-lowering properties of soluble dietary fibre polysaccharides. *Food & function*, 1(2), pp.149–55.

Harker, F.R. et al., 2002. Sensory interpretation of instrumental measurements 2: Sweet and

acid taste of apple fruit. *Postharvest Biology and Technology*, 24(3), pp.241–250.

Harshman, R.A., 1972. PARAFAC2: Mathematical and technical notes. *UCLA Working Papers in Phonetics*, 22(10), pp.30–44.

Harshman, R. a, 1970. Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis. *UCLA Working Papers in Phonetics*, 16(10), pp.1–84.

Hendriks, M.M.W.B. et al., 2011. Data-processing strategies for metabolomics studies. *TrAC - Trends in Analytical Chemistry*, 30(10), pp.1685–1698.

Ho, H.V.T. et al., 2016. A systematic review and meta-analysis of randomized controlled trials of the effect of barley β-glucan on LDL-C, non-HDL-C and apoB for cardiovascular disease risk reductioni-iv. *European Journal of Clinical Nutrition*, (April), pp.1–7.

Hofmann, A.F., 2004. Detoxification of Lithocholic Acid, A Toxic Bile Acid: Relevance to Drug Hepatotoxicity. *Drug Metabolism Reviews*, 36(3–4), pp.703–722.

Hofmann, A.F. et al., 1999. The Continuing Importance of Bile Acids in Liver and Intestinal Disease. *Archives of Internal Medicine*, 159(22), p.2647.

Holmes, E. et al., 2007. Detection of Urinary Drug Metabolite (Xenometabolome) Signatures in Molecular Epidemiology Studies via Statistical Total Correlation (NMR) Spectroscopy. *Analytical Chemistry*, 79(7), pp.2629–2640.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. Baltimore: Warwick & York, p. 48.

Hu, F. et al., 2007. Nondestructive Quantification of Organic Compounds in Whole Milk without Pretreatment by Two-Dimensional NMR Spectroscopy. *Journal of Agricultural and Food Chemistry*, 55(11), pp.4307–4311.

Iacono, A. et al., 2010. Effect of oleocanthal and its derivatives on inflammatory response induced by lipopolysaccharide in a murine chondrocyte cell line. *Arthritis and Rheumatism*, 62(6), pp.1675–1682.

Ibrugger, S. et al., 2013. Extracted oat and barley beta-glucans do not affect cholesterol metabolism in young healthy adults. *The Journal of nutrition*, 143(10), pp.1579–1585.

Immerstrand, T. et al., 2010. Effects of oat bran, processed to different molecular weights of beta-glucan, on plasma lipids and caecal formation of SCFA in mice. *The British journal of nutrition*, 104(3), pp.364–373.

Jemal, A. et al., 2011. Global cancer statistics. *CA: A Cancer Journal for Clinicians*, 61(2),

pp.69–90.

Johnson, C.H., Ivanisevic, J. & Siuzdak, G., 2016. Metabolomics: beyond biomarkers and towards mechanisms. *Nature reviews. Molecular cell biology*, 17(7), pp.451–9.

Kalra, S. & Jood, S., 2000. Effect of Dietary Barley β-Glucan on Cholesterol and Lipoprotein Fractions in Rat. *Journal of Cereal Science*, 31(2), pp.141–145.

Karadeniz, F. & Ekşi, A., 2002. Sugar composition of apple juices. *European Food Research and Technology*, 215(2), pp.145–148.

Keenan, J.M. et al., 2007. The effects of concentrated barley β-glucan on blood lipids in a population of hypercholesterolaemic men and women. *British Journal of Nutrition*, 97(6), pp.1162–1168.

Keun, H.C. & Athersuch, T.J., 2011. Nuclear Magnetic Resonance (NMR)-Based Metabolomics. In T. O. Metz, ed. *Metabolic Profiling: Methods and Protocols*. Totowa, NJ: Humana Press, pp. 321–334.

Khakimov, B. et al., 2012. Plant metabolomics: Resolution and quantification of elusive peaks in liquid chromatography-mass spectrometry profiles of complex plant extracts using multi-way decomposition methods. *Journal of Chromatography A*, 1266, pp.84–94.

Kiers, H.A.L., Ten Berge, J.M.F. & Bro, R., 1999. PARAFAC2 — Part I. A Direct Fitting Algorithm for the PARAFAC2 Model. *J. Chemom.*, 13(July 1998), pp.275–294.

Lazaridou, A. et al., 2004. A comparative study on structure-function relations of mixed-linkage (1→3),(1→4)-β-D-glucan. *Food Hydrocolloids*, 18(5), pp.837–855.

Leonetti, C. et al., 2004. Biological Activity of the G-Quadruplex Ligand RHPS4 (3,11-Difluoro-6,8,13-trimethyl-8H-quino[4,3,2-kl]acridinium methosulfate) Is Associated with Telomere Capping Alteration. *Molecular Pharmacology* , 66(5), pp.1138–1146.

Di Leva, F.S. et al., 2013. Exploring the Chemical Space of G-Quadruplex Binders: Discovery of a Novel Chemotype Targeting the Human Telomeric Sequence. *Journal of Medicinal Chemistry*, 56(23), pp.9646–9654.

Li, J. et al., 2003. Effects of barley intake on glucose tolerance, lipid metabolism, and bowel function in women. *Nutrition*, 19(11–12), pp.926–929.

Lin, Y.-H. et al., 2011. Antimüllerian hormone and polycystic ovary syndrome. *Fertility and sterility*, 96(1), pp.230–235.

Ludwig, C. & Viant, M.R., 2010. Two-dimensional J-resolved NMR spectroscopy: Review of a key methodology in the metabolomics toolbox. *Phytochemical Analysis*, 21(1), pp.22–

32.

Mannina, L. & Sobolev, A.P., 2011. High resolution NMR characterization of olive oils in terms of quality, authenticity and geographical origin. *Magnetic Resonance in Chemistry*, 49(SUPPL. 1).

Marcotte, E.M., 2001. The path not taken. *Nat Biotech*, 19(7), pp.626–627.

Markley, J.L. et al., 2017. The future of NMR-based metabolomics. *Current Opinion in Biotechnology*, 43, pp.34–40.

Martínez-Augustin, O. & de Medina, F.S., 2008. Intestinal bile acid physiology and pathophysiology. *World Journal of Gastroenterology*, 14(37), pp.5630–5640.

McCartney, C.R. & Marshall, J.C., 2016. Polycystic Ovary Syndrome. *New England Journal of Medicine*, 375(1), pp.54–64.

Mikkelsen, M.S., Savorani, F., et al., 2014. New insights from a β-glucan human intervention study using NMR metabolomics. *Food Research International*, 63, pp.210–217.

Mikkelsen, M.S., Cornali, S.B., et al., 2014. Probing interactions between β-glucan and bile salts at atomic detail by $^{1}H$-$^{13}C$ nmr assays. *Journal of Agricultural and Food Chemistry*, 62(47), pp.11472–11478.

Mikkelsen, M.S., Jensen, M.G. & Mikkelsen, T.S., 2017. Effects of barley beta-glucans varying in molecular mass and oligomer structure on blood lipids and large intestinal fermentation in hypercholesterolemic rats. *Food & Function*, Submitted.

Molino, G. et al., 1986. Simulation of the metabolism and enterohepatic circulation of endogenous chenodeoxycholic acid in man using a physiological pharmacokinetic model. *European Journal of Clinical Investigation*, 16(5), pp.397–414.

Van Munster, I., 1993. The role of carbohydrate fermentation in colon cancer prevention. *Scandinavian Journal of gastroenterology*.

Nagana Gowda, G.A. & Raftery, D., 2015. Can NMR solve some significant challenges in metabolomics? *Journal of Magnetic Resonance*, 260, pp.144–160.

Nagengast, F., 1995. Role of Bile Acids in Colorectal Carcinogenesis. *European Journal of Cancer*.

Nicholson, J.K. & Lindon, J.C., 2008. Systems biology: Metabonomics. *Nature*, 455(7216), pp.1054–1056.

Nicholson, J.K., Lindon, J.C. & Holmes, E., 1999. "Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. *Xenobiotica*, 29(11), pp.1181–

1189.

Nielsen, N.P. V, Carstensen, J.M. & Smedsgaard, J., 1998. Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *Journal of Chromatography A*, 805(1–2), pp.17–35.

O'Gorman, A. et al., 2013. Metabolic profiling of human follicular fluid identifies potential biomarkers of oocyte developmental competence. *Reproduction*, 146(4), pp.389–395.

Pan, Z. & Raftery, D., 2007. Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics. *Analytical and Bioanalytical Chemistry*, 387(2), pp.525–527.

Patti, G.J., Yanes, O. & Siuzdak, G., 2012. Innovation: Metabolomics: the apogee of the omics trilogy. *Nat Rev Mol Cell Biol*, 13(4), pp.263–269.

Payne, C.M. et al., 2008. Hydrophobic bile acids, genomic instability, Darwinian selection, and colon carcinogenesis. *Clinical and experimental gastroenterology*, 1, pp.19–47.

Pearson, K., 1901. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(1), pp.559–572.

Pigny, P. et al., 2003. Elevated serum level of anti-mullerian hormone in patients with polycystic ovary syndrome: relationship to the ovarian follicle excess and to the follicular arrest. *The Journal of clinical endocrinology and metabolism*, 88(12), pp.5957–62.

Pinero-Sagredo, E. et al., 2010. NMR metabolic profile of human follicular fluid. *NMR in Biomedicine*, 23(5), pp.485–495.

Proctor, W.G. & Yu, F.C., 1950. The dependence of a nuclear magnetic resonance frequency upon chemical compound [6]. *Physical Review*, 77(5), p.717.

Purcell, E., Torrey, H. & Pound, R., 1946. Resonance Absorption by Nuclear Magnetic Moments in a Solid. *Physical Review*, 69(1–2), pp.37–38.

Ramakrishnan, V. & Luthria, D.L., 2016. Recent applications of NMR in food and dietary studies. *Journal of the Science of Food and Agriculture*, 97(1), pp.33–42.

Ramsden, J.J., 2009. Metabolomics and Metabonomics. In *Bioinformatics: An Introduction*. London: Springer London, pp. 1–6.

Revelli, A. et al., 2009. Follicular fluid content and oocyte quality: from single biochemical markers to metabolomics. *Reproductive biology and endocrinology : RB&E*, 7, p.40.

Ridlon, J.M., Kang, D.-J. & Hylemon, P.B., 2006. Bile salt biotransformations by human

intestinal bacteria. *Journal of lipid research*, 47(2), pp.241–259.

Rinnan, Å. et al., 2014. Recursive weighted partial least squares (rPLS): an efficient variable selection method using PLS. *Journal of Chemometrics*, 28(5), pp.439–447.

Rondanelli, M. et al., 2011. Beta-glucan- or rice bran-enriched foods: a comparative crossover clinical trial on lipidic pattern in mildly hypercholesterolemic men. *Eur J Clin Nutr*, 65(7), pp.864–871.

van Rooij, I.A.J., 2002. Serum anti-Mullerian hormone levels: a novel measure of ovarian reserve. *Human Reproduction*, 17(12), pp.3065–3071.

Salvati, E. et al., 2007. Telomere damage induced by the G-quadruplex ligand RHPS4 has an antitumor effect. *The Journal of Clinical Investigation*, 117(11), pp.3236–3247.

Savorani, F., Tomasi, G. & Engelsen, S.B., 2010. icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *Journal of Magnetic Resonance*, 202(2), pp.190–202.

Shimizu, C. et al., 2008. Effect of High β-Glucan Barley on Serum Cholesterol Concentrations and Visceral Fat Area in Japanese Men. A Randomized, Double-blinded, Placebo-controlled Trial. *Plant Foods for Human Nutrition*, 63(1), pp.21–25.

Shin, J.-H. et al., 2011. $^{1}$H NMR-based Metabolomic Profiling in Mice Infected with Mycobacterium tuberculosis. *Journal of Proteome Research*, 10(5), pp.2238–2247.

Smilde, A.K. et al., 2005. ANOVA-simultaneous component analysis (ASCA): A new tool for analyzing designed metabolomics data. *Bioinformatics*, 21(13), pp.3043–3048.

Smolinska, A. et al., 2012. NMR and pattern recognition methods in metabolomics: From data acquisition to biomarker discovery: A review. *Analytica Chimica Acta*, 750, pp.82–97.

Sobolev, A.P. et al., 2015. Untargeted NMR-based methodology in the study of fruit metabolites. *Molecules*, 20(3), pp.4088–4108.

Spitzer, D. et al., 1996. Different protein patterns derived from follicular fluid of mature and immature human follicles. *Human reproduction (Oxford, England)*, 11(4), pp.798–807.

Spraul, M. et al., 2009. NMR-based multi parametric quality control of fruit juices: SGF profiling. *Nutrients*, 1(2), pp.148–155.

Ståhle, L. & Wold, S., 1987. Partial least squares analysis with cross-validation for the two-class problem: A Monte Carlo study. *Journal of Chemometrics*, 1(3), pp.185–196.

Steuer, R., 2006. Review: On the analysis and interpretation of correlations in metabolomic data. *Briefings in Bioinformatics*, 7(2), pp.151–158.

Sugiura, K., Pendola, F.L. & Eppig, J.J., 2005. Oocyte control of metabolic cooperativity

between oocytes and companion granulosa cells: Energy metabolism. *Developmental Biology*, 279(1), pp.20–30.

Takahashi, C. et al., 2007. Anti-Müllerian hormone substance from follicular fluid is positively associated with success in oocyte fertilization during in vitro fertilization. *Fertility and Sterility*, 89(3), pp.586–591.

Tang, H. et al., 2004. Use of relaxation-edited one-dimensional and two dimensional nuclear magnetic resonance spectroscopy to improve detection of small metabolites in blood plasma. *Analytical Biochemistry*, 325(2), pp.260–272.

Theuwissen, E. & Mensink, R.P., 2008. Water-soluble dietary fibers and cardiovascular disease. *Physiology and Behavior*, 94(2), pp.285–292.

Tomasi, G., Van Den Berg, F. & Andersson, C., 2004. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *Journal of Chemometrics*, 18(5), pp.231–241.

Tong, L.T. et al., 2015. Effects of dietary hull-less barley β-glucan on the cholesterol metabolism of hypercholesterolemic hamsters. *Food Chemistry*, 169, pp.344–349.

Vandendriessche, T. et al., 2013. High-throughput NMR based metabolic profiling of Braeburn apple in relation to internal browning. *Postharvest Biology and Technology*, 80, pp.18–24.

Vermathen, M., 2011. Investigation of Different Apple Cultivars by High Resolution Magic Angle Spinning NMR. A Feasibility Study. *Journal of agricultural and food chemistryhemistry*, 59, pp.12784–12793.

Visioli, F. et al., 2011. Polyphenols and human health: a prospectus. *Critical reviews in food science and nutrition*, 51(6), pp.524–46.

Wallace, M. et al., 2012. An investigation into the relationship between the metabolic profile of follicular fluid, oocyte developmental potential, and implantation outcome. *Fertility and Sterility*, 97(5), pp.1078–84.

Want, E. J., Cravatt, B. F. and Siuzdak, G. 2005. The expanding role of mass spectrometry in metabolite profiling and characterization, ChemBioChem, 6(11), pp. 1941–1951.

Wilson, T.A. et al., 2004. Reduced and High Molecular Weight Barley β-Glucans Decrease Plasma Total and Non-HDL-Cholesterol in Hypercholesterolemic Syrian Golden Hamsters. *The Journal of Nutrition*, 134(10), pp.2617–2622.

Wishart, D.S. et al., 2013. HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic Acids Research*, 41(Database issue), pp.D801–D807.

Wold, S., Sjöström, M. & Eriksson, L., 2001. PLS-regression: A basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), pp.109–130.

Wolever, T.M.S. et al., 2010. Physicochemical properties of oat β-glucan influence its ability to reduce serum LDL cholesterol in humans: a randomized clinical trial. *The American journal of clinical nutrition*, 92(4), pp.723–32.

Wu, H. et al., 2008. High-throughput tissue extraction protocol for NMR- and MS-based metabolomics. *Analytical Biochemistry*, 372(2), pp.204–212.

Yang, J.-L. et al., 2003. Barley beta-glucan lowers serum cholesterol based on the up-regulation of cholesterol 7alpha-hydroxylase activity and mRNA abundance in cholesterol-fed rats. *Journal of nutritional science and vitaminology*, 49(6), pp.381–387.

Young, R.C., Ozols, R.F. & Myers, C.E., 1981. The Anthracycline Antineoplastic Drugs. *New England Journal of Medicine*, 305(3), pp.139–153.

Zhang, A. et al., 2012. Recent and potential developments of biofluid analyses in metabolomics. *Journal of Proteomics*, 75(4), pp.1079–1088.

Zhang, C.-M. et al., 2014. Metabolic heterogeneity of follicular amino acids in polycystic ovary syndrome is affected by obesity and related to pregnancy outcome. *BMC pregnancy and childbirth*, 14(1), p.11.

# APPENDIX

**Paper I**

"Development of an optimized protocol for NMR metabolomics studies of human colon cancer cell lines and first insight from testing of the protocol using DNA G-Quadruplex ligands as Novel Anti- Cancer Drugs". I. Lauri, F. Savorani, N. Iaccarino, P. Zizza, L.M. Pavone, E. Novellino, S.B. Engelsen, and A. Randazzo. *Metabolites*, 2016, 6 (1), pii: E4. doi: 10.3390/metabo6010004.

**Paper II**

"Characterization of monovarietal extra virgin olive oils from the province of Béjaïa (Algeria)". F. Laincer, N. Iaccarino, J. Amato, B. Pagano, A. Pagano, G. Tenore, A. Tamendjari, P. Rovellini, S. Venturini, G. Bellan, A. Ritieni, L. Mannina, E. Novellino, and A. Randazzo. *Food Research International*, 2016, 89, 1123-33.

**Paper III**

"Application of recursive partial least square regression for prediction of apple juice sensory attributes from NMR spectra". N. Iaccarino, C. Varming, M.A. Petersen, F. Savorani, A. Randazzo, B. Schütz, T.B Toldam-Andersen, and S.B. Engelsen. *Proceedings of the XIII International Conference on the Applications of Magnetic Resonance in Food Science,* 2016, 7-11.

**Paper IV**

"$^{1}$H NMR-based metabolomics study on follicular fluid from patients with PolyCystic Ovarian Syndrome (PCOS)". N. Iaccarino, J. Amato, B. Pagano, A. Pagano, L. D'Oriano, Sveva Pelliccia, Mariateresa Giustiniano, Diego Brancaccio, Francesco Merlino, Ettore Novellino, C. Alviggi, A. Randazzo. *Biochimica Clinica.* Submitted.

*metabolites*

MDPI

*Article*

# Development of an Optimized Protocol for NMR Metabolomics Studies of Human Colon Cancer Cell Lines and First Insight from Testing of the Protocol Using DNA G-Quadruplex Ligands as Novel Anti-Cancer Drugs

**Ilaria Lauri [1], Francesco Savorani [2,3,*], Nunzia Iaccarino [1], Pasquale Zizza [4], Luigi Michele Pavone [5], Ettore Novellino [1], Søren Balling Engelsen [2] and Antonio Randazzo [1,*]**

Received: 25 November 2015; Accepted: 11 January 2016; Published: 15 January 2016
Academic Editor: Peter Meikle

1 Department of Pharmacy, University of Naples "Federico II", via D. Montesano 49, 80131 Naples, Italy; ilaria.lauri@unina.it (I.L.); nunzia.iaccarino@unina.it (N.I.); ettore.novellino@unina.it (E.N.)
2 Spectroscopy & Chemometrics, Department of Food Science, Faculty of Science, University of Copenhagen, Rolighedsvej 26, 1958 Frederiksberg C, Denmark; se@food.ku.dk
3 Department of Applied Science and Technology (DISAT), Polytechnic University of Turin—Corso Duca degli Abruzzi 24, 10129 Torino, Italy
4 Experimental Chemotherapy Laboratory, Regina Elena National Cancer Institute, 00158 Rome, Italy; zizza@ifo.it
5 Department of Molecular Medicine and Medical Biotechnology, University of Naples "Federico II", via S. Pansini 5, 80131 Naples, Italy; luigimichele.pavone@unina.it
* Correspondence: frsa@food.ku.dk (F.S.); antonio.randazzo@unina.it (A.R.); Tel.: +45-3533-2565 (F.S.); Tel./Fax: +39-0816-78514 (A.R.)

**Abstract:** The study of cell lines by nuclear magnetic resonance (NMR) spectroscopy metabolomics represents a powerful tool to understand how the local metabolism and biochemical pathways are influenced by external or internal stimuli. In particular, the use of adherent mammalian cells is emerging in the metabolomics field in order to understand the molecular mechanism of disease progression or, for example, the cellular response to drug treatments. Hereto metabolomics investigations for this kind of cells have generally been limited to mass spectrometry studies. This study proposes an optimized protocol for the analysis of the *endo*-metabolome of human colon cancer cells (HCT116) by NMR. The protocol includes experimental conditions such as washing, quenching and extraction. In order to test the proposed protocol, it was applied to an exploratory study of cancer cells with and without treatment by anti-cancer drugs, such as DNA G-quadruplex binders and Adriamycin (a traditional anti-cancer drug). The exploratory NMR metabolomics analysis resulted in NMR assignment of all *endo*-metabolites that could be detected and provided preliminary insights about the biological behavior of the drugs tested.

**Keywords:** cell metabolomics; colon cancer; NMR spectroscopy; Multivariate statistical analysis; G-quadruplex ligands

## 1. Introduction

In the last decades, metabolomics studies have been performed on different biofluids (e.g., plasma, serum, urine, saliva, lymph and cerebrospinal fluid) with successful results, showing applications in many areas, such as biomarker discovery, clinical studies, nutritional studies, drug efficacy and toxicity evaluations and disease diagnosis [1–4]. However, recent developments in the use of metabolomics

involve the characterization and interpretation of the cell metabolome, starting from prokaryotes (especially *Escherichia coli*) to eukaryotes cell lines (yeast or mammalian cells) [5,6]. Complementary to the classic biofluid analyses, the metabolomic profiles of cells represent a powerful tool to understand how the local metabolism and biochemical pathways are influenced by pathologies and by external or internal stimuli. In particular, the metabolome analysis of cells grown *in vitro* provides important information for the development of models of biological pathways and networks. *In vitro* cell metabolomics analysis offers several advantages: experimental variables are easier to control, higher reproducibility, less expensive and easier to interpret than analysis of animal models and human subjects [7]. The use of mammalian cells is emerging in the metabolomics field in order to understand the molecular mechanism of disease progression, the cellular response to drug treatments [8] and the cell culture monitoring [9]. In particular, the identification and characterization of cancer cell metabolomic signature may play an important role in the early diagnosis as well as in the following therapeutic response, making it possible to map the drug action into metabolic pathways [10].

Colon carcinoma is the third most commonly diagnosed cancer in the world and the second most common cause of death from cancer [11]. Surprisingly, few metabolomic studies dealing with colon carcinoma cell lines are reported in the literature [12–16]. The analysis of metabolic profiles of this cell line provides a comprehensive assessment of the alterations in the metabolite levels in cells and can produce important information on *in vitro* actions of drugs towards their incorporation into novel therapeutic settings.

Recently, targeting of DNA secondary structures, for example G-quadruplexes, has been considered as an appealing opportunity for drug intervention in anti-cancer therapy [17]. G-quadruplex DNA (G4-DNA) structures are four-stranded helical DNA (or RNA) structures, comprising stacks of G-tetrads, which are the outcome of planar association of four guanines in a cyclic *Hoogsteen* hydrogen-bonding arrangement. From the biological point of view, G4-DNAs are widespread in the genome and they are present in the promoters of a wide range of genes, important in cell signaling, and recognized as hallmarks of cancer: c-Myc, c-Kit and K-Ras (self-sufficiency); pRb (insensitivity); Bcl-2 (evasion of apoptosis); VEGF-A (angiogenesis); hTERT (limitless replication); and PDGFA (metastasis) [18]. The G4-DNAs are also found in telomeric regions of the chromosome [19]. Telomeric DNA consists of tandem repeats of a simple short sequence, rich in guanine residues (TTGGGA). Telomeres protect the ends of the chromosome from damage and recombination, and their shortening is implicated in cellular senescence. The elongation of telomeric DNA, operated by the enzyme telomerase, leads cancer cells towards an infinite lifetime. The inhibition of telomerase, which is over-expressed in about 85% of tumors, represents the forefront of research for new effective anti-cancer drugs. Since this enzyme requires a single stranded telomeric primer, the formation of G-quadruplex complexes by telomeric DNA inhibits the telomerase activity. In this respect, it has been found that small molecules that stabilize G-quadruplex structures are effective telomerase inhibitors and can be considered as novel drugs candidates for anti-cancer therapy [20]. Recently, it has been discovered that a number of G-quadruplex ligands are exerting interesting antitumor activity *in vitro* [21,22].

Since the G-quadruplex ligands may be important for the development of new anti-cancer agents, this study is aimed to verify the feasibility of a NMR metabolomics study of HCT116 cells when treated with these agents. In particular, the treatment with compound **1**, which is one of the most promising ligands discovered by virtual screening calculations [23] (Figure 1), was compared to the treatment with pentacyclic acridine RHPS4 (2) (Figure 1), which is one of the most studied G4 ligands [24], and to treatment using the well-known antitumor agent Adriamycin (3) (Figure 1). Adriamycin is an approved chemotherapeutic agent with strong activity against a wide range of human malignant neoplasms including acute leukemia, non-Hodgkin lymphomas, breast cancer, Hodgkin's disease and sarcomas [25]. Thus, this study describes an optimized protocol for NMR metabolomics of adherent mammalian cell lines and the preliminary application and validation to treated cancer cells.

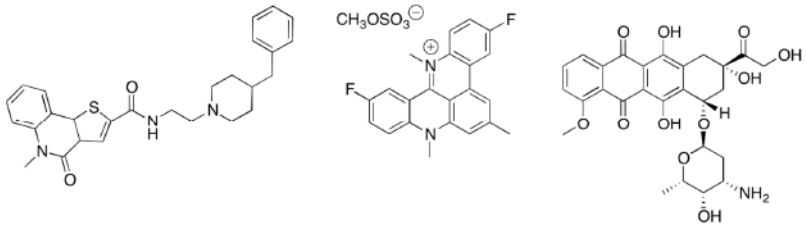


**Figure 1.** The structure of compound (**1**), RHPS4 (**2**), and the structure of the traditional antitumor agent Adriamycin (**3**).

## 2. Materials and Methods

### *2.1. Materials*

HCT116 cells were purchased from American Type Culture Collection (ATCC–Manassas, VA, USA). High glucose Dulbecco's Modified Eagle's medium (DMEM/HIGH Glucose) with L-Glutamine was purchased from Euroclone (MI, Italy), penicillin–streptomycin solution for cell culture was purchased from Gibco (NY, USA). Fetal bovine serum (FBS) was purchased from Thermo Scientific (HyCloneTM). Crystal phosphate buffer saline (PBS) (0.01 M Phosphate buffer, 0.0027 M KCl e 0.14 M NaCl, pH 7.4 at 25 °C) was purchased from Bioline (TR, Italy).

Deuterium oxide ($D_2O$, 99.8%D) was obtained from Sigma-Aldrich (St. Louis, MO, USA). All other reagents were of analytical grade.

### *2.2. Cell Culture*

The HTC116 cells were grown in high glucose (4.5 g/L) Dulbecco's Modified Eagle's Medium (DMEM/HIGH Glucose, Euroclone) supplemented with 10% FBS, L-Glutamine (2 mM), penicillin (100 U/mL) and streptomycin (1 mg/mL), at 37 °C in a humidified atmosphere of 5% $CO_2$. In order to obtain the final desired number of cells ($1.5 \times 10^6$) for each treatment, the cell growth was carried out in parallel in multiple (3×) 150 mm tissue culture dishes (Corning). Upon achievement of 90% cellular confluency, the culture medium (15 mL) was removed and the cells were processed for the endo-metabolomic analysis. In brief, the cells were extensively washed (4 times) with ice-cold phosphate-buffered saline (PBS 1X) in order to completely remove any residue of culture medium. Afterwards, 5.4 mL of PBS were added to each culture dish and cells were collected by scraping with a rubber policeman. Finally, the cells were counted, placed in Falcon tubes and the final PBS volumes were adjusted to obtain $15 \times 10^6$ cells into 5.4 mL PBS (pH 7.4). On the other side, the culture medium of each cell growth was collected and immediately stored at −80 °C to be used, in the close future, for the exo-metabolome analyses.

### *2.3. Anti-Cancer Drug Treatments*

The dose and drug exposure duration time of cell culture for compounds **1** and **2** were established according to the literature ($IC_{50}$) [26], while the optimal conditions for compound **3** were chosen on the basis of in-house unpublished results. In order to have a unique group of untreated control cells valid for all three treatments, compounds **1** and **2** were added to cell cultures 24 h after seeding. Cells were exposed to the drug treatment for 72 h with 1 μM final concentration; compound **3** was added to cell cultures 80 h after seeding. In this case, the drug exposure of cell cultures was for 16 h with 0.1 μM final concentration. Thus, all the cells (including controls) were detached from the plates after 96 h.

### *2.4. Cell Metabolome Quenching*

The Falcon tubes containing the detached cells were immersed into liquid nitrogen upon complete freezing of the samples and then slowly thawed in an ice bath. Finally, to destroy the cell membrane

favoring the release of the intracellular metabolites, the quenched cells were lysed by sonication (3 short-pulse cycles of 30 s each, at maximum power).

### *2.5. Metabolites Extraction for NMR Analysis*

Intracellular metabolites were extracted using a dual phase extraction procedure introduced by Bligh and Dyer in 1959 [27] with slight modifications. Adding 6 mL of cold methanol (−20 °C) and 6 mL of chloroform to the original solution (5.4 mL) containing quenched cells, briefly a mixture of water, methanol and chloroform in the volume ratio of 0.9:1:1 was obtained, corresponding to a total volume of 17.4 mL. Afterwards, this mixture containing quenched and lysed cells was incubated for 20 min on ice and vortexed frequently to facilitate the extraction. The cell extracts were centrifuged at 4000 *g* at 4 °C for 20 min. This extraction procedure generated a two-phase extract that can be described as follow: the aqueous upper phase contains water-soluble intracellular metabolites, while apolar metabolites as lipid molecules are in the organic lower phase. Proteins and macromolecules are trapped in the thin skin-like layer between the two phases. The upper and lower phase were separated and carefully transferred into different falcon tubes. Eventually, solvents were completely removed from both fraction using a vacuum concentrator (hydrophylic phase) and under a gentle flow of $N_2$ gas (organic phase). Only the hydrophilic phase has been taken into account in this study while the organic phase has been stored at −80 °C for future analysis.

### *2.6. Sample Preparation for NMR Analysis*

Each aqueous cell extract was dissolved in 540 µL of $D_2O$ together with 60 µL of a $D_2O$ solution containing the sodium salt of (trimethylsilyl) propanoic-2,2,3,3-d4 acid (TSP) (0.1% w/v), used as internal chemical shift reference ($\delta_H$ 0.00 ppm), to give a final concentration of 0.6 mM. Samples were vortexed briefly and transferred into 5-mm NMR tubes.

### *2.7. NMR Spectroscopy of Cell Extracts*

All one-dimensional $^1$H-NMR spectra were acquired at 300 K on a Bruker Avance III 600 MHz ultrashielded spectrometer (Bruker Biospin Gmbh, Rheinstetten, Germany) operating at 600.13 MHz for protons (14.09 Tesla) equipped with a double tuned cryo-probe (TCI) set for 5 mm sample tubes. $^1$H NMR spectra of hydrophilic cell extracts were acquired using a one-dimensional NOESY-presat pulse sequence (RD-90°-t-90°-tm-90°-ACQ). All the experiments were acquired with an acquisition time of 2.73 s, a relaxation delay of 4 s, mixing time of 10 ms, receiver gain of 181, 128 scans, 128 K data points and a spectral width of 18,029 Hz (30.041 ppm). All samples were automatically tuned, matched and shimmed.

Representative samples of treated cell extracts were examined by two-dimensional spectroscopy (JRES, COSY, TOCSY, HSQC and HMBC) to ensure the unambiguous assignment of the metabolites. A 700 MHz Varian Unity Inova spectrometer equipped with a 5 mm $^1$H{$^{13}$C/$^{15}$N} triple resonance probe was used for the acquisition of two-dimensional NMR experiments.

### *2.8. NMR Data Reduction and Processing*

Prior to Fourier transformation, each free induction decay (FID) was zero-filled to 128 K points and multiplied by an exponential function equivalent to a 1.0 Hz line broadening. The resulting spectra were phase and baseline corrected automatically using TOPSPIN™ (Bruker Biospin) and the ppm scale was referenced according to the TSP peak at 0.00 ppm.

The NMR regions above 9.43 ppm and below 0.8 ppm were removed because they only contain noise. Furthermore, the region between 4.75 and 4.62 ppm was also removed because containing the residual water signal. Since NMR spectra showed misalignments in chemical shift due to pH-sensitive peaks, the spectra were aligned using the interval correlation optimized shifting algorithm (icoshift) [28]. A normalization preprocessing step was carried out to correct variations of the overall concentrations of the samples. Since no quantitative internal standard was used, each spectrum of

the aligned NMR data matrix was normalized to unit area, obtained dividing every variable of each spectrum by the sum of the absolute value of all its variables. All preprocessing steps were performed using Matlab (2012b, The Mathworks Inc., Natick, MA, USA).

### *2.9. Multivariate Data Analysis*

The normalized data matrix was imported into Simca-P 13.0 (Umetrics, Umeå, Sweden) and Pareto-scaled [29]. The number of principal components (PCs) of the Principal Component Analyses (PCA) [30] was determined by leave one out cross-validation [31]. The quality of the models was described by the squared Pearson correlation coefficient $R^2$ and $Q^2$ values. $R^2$ is defined as the proportion of variance in the data explained by the models and indicates the goodness of fit. $Q^2$ is defined as the proportion of variance in the data predictable by the model and indicates predictability [29]. Both $R^2$ and $Q^2$ vary between 0 and 1: a good prediction model is indicated by $Q^2 > 0.5$, whereas a $Q^2 > 0.9$ means an excellent predictive ability of the model. In this study, all PCA models performed showed a $R^2 \geqslant 0.9$ and a $Q^2 \geqslant 0.8$, which means goodness of fit and goodness of prediction of the models.

### *2.10. Metabolite Identification*

Identification of hydrophilic metabolites was achieved by (i) comparison with the chemical shifts of the metabolites in the Human Metabolome Database (HMDB) [32]; (ii) peak fitting routine within the spectral database in Chenomx NMR Suite 5.0 software package (Chenomx, AB, Canada); (iii) analysis of literature data [33–35]; (iv) the interpretation of the bi-dimensional NMR spectra; and (v) the analysis of the Statistical Total Correlation Spectroscopy (STOCSY) [36].

### *2.11. Statistical Total Correlation Spectroscopy Analysis*

Statistical Total Correlation Spectroscopy (STOCSY) analysis (Figure S1) was performed on the binned (0.02 ppm) NMR (1D-NOESY) data set containing all samples, to obtain the correlations among the metabolite signals. The results were plotted using a threshold value of $R > 0.95$.

### *2.12. Metabolic Pathways Identification*

The impact of drug treatment of HCT116 colorectal carcinoma cell line on metabolic pathways was evaluated using a tool for metabolomic data analysis, which is available online [37]. The Pathway Analysis module combines results from powerful pathway enrichment analysis with the pathway topology analysis to help researchers identify the most relevant pathways involved in the conditions under study. By uploading the discriminatory compounds that were significantly influenced by drug treatment, the built-in *Homo sapiens* (human) pathway library for pathway analysis and hypergeometric test for over-representation analysis were employed. Results were then presented graphically as well as in a detailed table (Figure S2).

### *2.13. Statistics*

Values are presented as the mean $\pm$ SD. Differences between data sets were analyzed by a one-way ANOVA, and $p < 0.05$ was considered to be statistically significant.

## 3. Results and Discussion

### *3.1. Optimization of the Quenching and Extraction Procedures*

This investigation was aimed to verify the feasibility of the study of the metabolome of human colon cancer cell line (HCT116) by NMR when treated with anti-cancer drugs. Mammalian cell metabolomics is an emerging research field, however the number of studies concerning quenching and extraction methods for HCT116 cells is still limited and generally referred to studies performed by GC-MS and LC-MS.

In this study, several published protocols for NMR-based metabolomic analysis to recover the cell metabolome were tested and the best results were achieved by selecting and combining different steps described in the diverse procedures [38] (Figure 2A). By analyzing and investigating the different extraction protocols, a number of critical passages that required an extensive optimization were identified. For example, the effects of cell quenching in liquid nitrogen were thoroughly investigated. This commonly represents the first step in several extraction protocols (immediately after the growth medium removal) just before cell washing to remove the medium residues. However, in this study it was observed that washing the HCT116 cells after the quenching step turned in to a significant loss of the cell metabolites, presumably because the freezing step induces the cell wall breakage with consequent metabolite leakage. To overcome this problem, the order of quenching and washing was inverted. Moreover, as the HCT116 cells grow as a sub-confluent monolayer, it was found particularly difficult to completely remove the cell growth medium during the washing step. In particular, one of the most abundant components of the medium, glucose, challenged the spectral interpretation of the extracted metabolome due to its residual signals spread all over the central region of the NMR spectrum. In order to avoid this, the number of washing steps was increased to four. After this intense washing procedure, the cells could be detached from the dishes by mechanical scraping and the metabolic activity of the cells immediately quenched by liquid nitrogen. The optimized protocol is summarized in Figure 2A, and can be recapitulated in the following main steps: (i) growth of the cell culture; (ii) abundant washing; (iii) cell scraping; (iv) quenching in liquid nitrogen; (v) cell lysis by sonication; and (vi) dual phase extraction procedure of the metabolites. Experimental description of these steps is reported in Material and Methods Section.

### 3.2. Experimental Design

In order to reduce bias in the interpretation of the experiments, it was decided to produce three biological replicates for each treatment (namely with compounds **1**–**3**). Furthermore, three control samples (untreated ones) were also collected ($CTL_{a-c}$). Thus, a total of 12 samples were produced and studied by high-resolution $^1$H NMR. The whole design of experiment is summarized in Figure 2B. The most efficient dose and drug exposure duration time of cell culture were used for each compound.

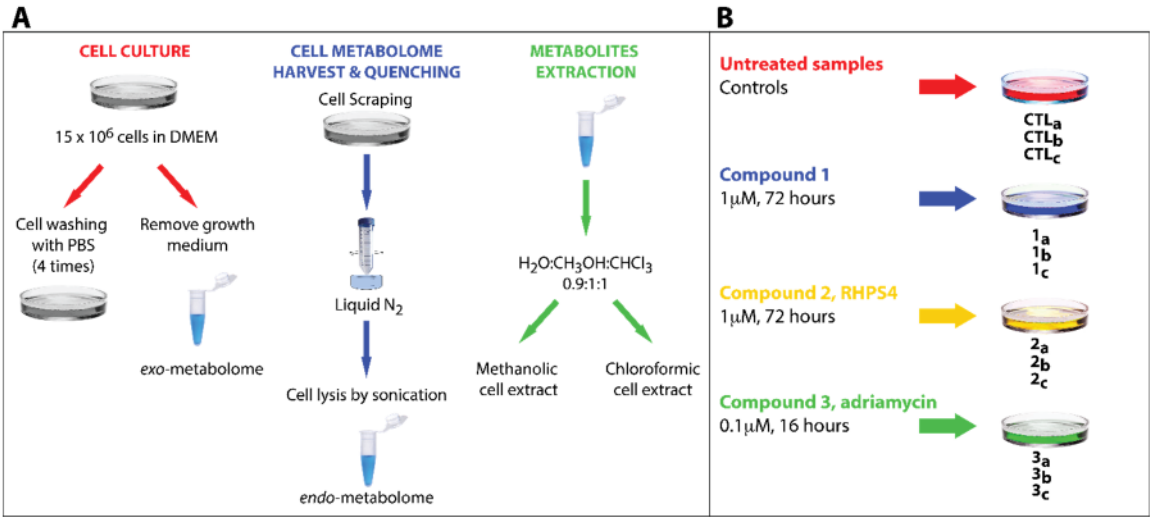


**Figure 2.** (**A**) General scheme describing the whole sample preparation protocol. (**B**) Overview of the experimental design. Each compound has been tested in triplicate and three control samples (untreated ones) were also collected ($CTL_{a-c}$).

**Table 1.** NMR assignment of the identified metabolites. The values indicate the percentage of increment or decrement in signal intensity of any given metabolite upon treatment with respect of the control. The values reported in talics are not statistically significant to be taken into account since the percentage of variation is less than three times the standard deviation (arbitrary threshold).

| Identification Number | Metabolites | Chemical Shifts (ppm) | Compound 1 | Compound 2 | Compound 3 |
|---|---|---|---|---|---|
| 1 | Lactate | 1.33(d)<br>4.13(q) | +19% ± 4% | +165% ± 18% | *+18 ± 10%* |
| 2 | Threonine | 1.34(d)<br>4.27(m) | +14% ± 4% | −46% ± 2% | *+17 ± 8%* |
| 3 | Tyrosine | 6.91(m)<br>7.21(m) | +28% ± 3% | −36% ± 3% | +17±5% |
| 4 | Phenylalanine | 7.34(d)<br>7.39(m)<br>7.44(m) | +23% ± 1% | −34% ± 2% | +13% ± 3% |
| 5 | Creatine | 3.04(s)<br>3.95(s) | +23% ± 2% | +49% ± 10% | +19% ± 5% |
| 6 | Creatine phosphate | 3.05(s)<br>3.96(s) | *−13% ± 5%* | −55% ± 2% | *0 ± 9%* |
| 7 | Glycine | 3.58(s) | *−8% ± 4%* | −43% ± 4% | *+6 ± 9%* |
| 8 | Alanine | 1.49(d)<br>3.81(q) | *+2% ± 3%* | −29 ± 5% | *+15 ± 9%* |
| 9 | Acetate | 1.92(s) | *0 ± 20%* | +14% ± 1% | *0% ± 50%* |
| 10 | Succinate | 2.39(s) | +7% ± 1% | +122% ± 122% | +13% ± 1% |
| 11 | AMP | 4.02(dd)<br>4.36(dd)<br>4.51(dd)<br>8.28(s)<br>8.59(s) | *+5 ± 4%* | −36 ± 5% | *+16% ± 8%* |
| 12 | Isoleucine, Leucine, Valine | 0.94(t)<br>1.02(d)<br>0.97(d)<br>0.99(d)<br>1.05(d) | +29% ± 4% | *−11% ± 5%* | *+15% ± 9%* |
| 13 | O-Phosphocholine | 3.23(s)<br>4.17(m) | −68% ± 1% | −61% ± 1% | +21% ± 7% |
| 14 | Glycerophosphocholine | 3.24(s) | −20% ± 2% | −33% ± 4% | *−1% ± 6%* |
| 15 | Nicotinic acid adenine dinucleotide (NAAD) | 8.06(t)<br>8.15(s)<br>8.42(s)<br>8.75(d)<br>8.95(d)<br>9.13(s) | −12% ± 2% | −48% ± 3% | +15% ± 2% |
| 16 | $NAD^+/NADP^+$ | 6.10(d)<br>8.18(m)<br>8.84(d)<br>9.12(d)<br>9.32(s) | −8% ± 2% | +62% ± 11% | *+6% ± 6%* |
| 17 | Histidine | 7.10(d)<br>7.86(d) | +24% ± 1% | −52% ± 3% | +14% ± 2% |
| 18 | Glutathione | 2.97(dd)<br>4.57(q)<br>2.58(m) | +9% ± 2% | −43% ± 9% | +13% ± 4% |
| 19 | ATP | 8.52(s) | −24% ± 7% | *−8% ± 4%* | −26% ± 6% |

The limited number of independent samples (12) used in this study is not sufficient to draw general conclusions, However, this represents a significant and necessary feasibility study before

setting up a much larger project. Indeed, the whole procedure described in Sections 2.2–5 from cell seeding to metabolites extraction, represents by far the most labor, cost and time intensive part of the whole study. In order to achieve a satisfactory and reliable result in terms of reproducibility and growth yield many trials were conducted to perfection the presented protocol. The purpose was to demonstrate the feasibility of the NMR metabolomics approach by developing a reliable protocol for cancer cell line metabolomics using a limited number of reliable sample results.

### *3.3. Metabolic Profile*

The 1D $^{1}$H NMR spectra were acquired to determine the metabolic fingerprints of the treated and untreated cancer cells, while 2D homo- and hetero-nuclear NMR experiments were acquired for the assignment of the metabolites. The metabolite assignment was accomplished by comparing data from literature, by peak fitting routine within the spectral database in Chenomx NMR software package, by the analysis of available chemical shifts databases (*i.e.*, HMDB) and by STOCSY correlation analysis (Figure S1). The results of the assignment are reported in the Table 1 and in Figure 3.

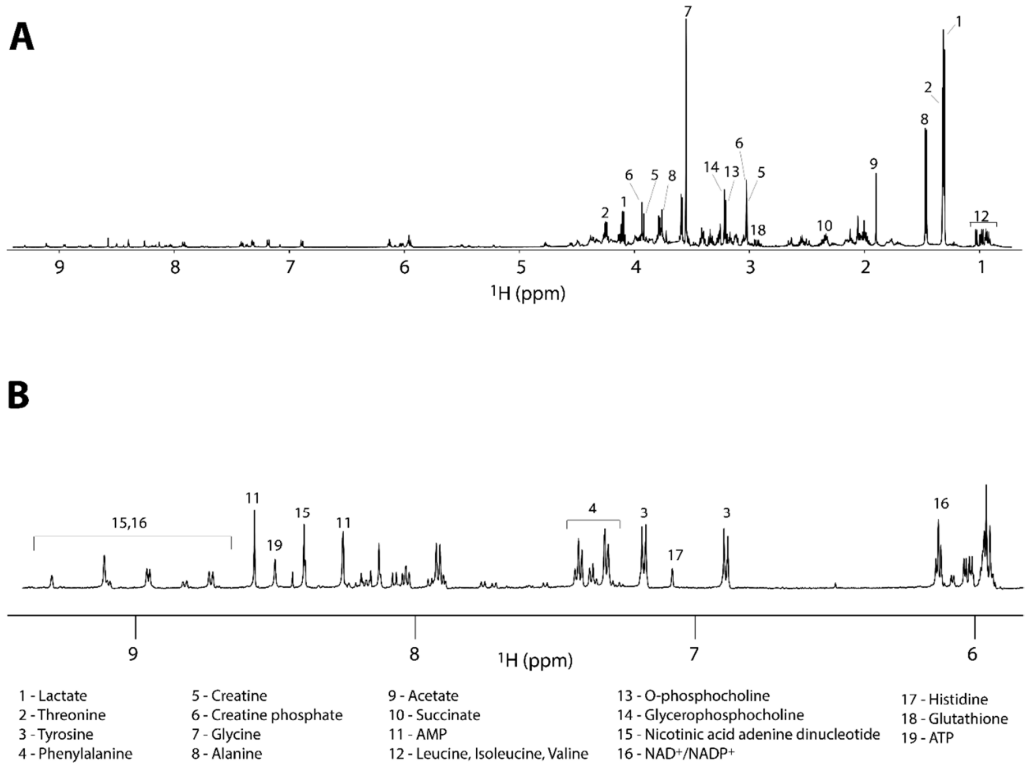


**Figure 3.** (**A**) Full $^{1}$H-NMR spectrum of a representative control sample along with the assignment of the most intense signals. (**B**) Expanded region of the spectrum reported in (**A**) with the assignment of the less intense metabolites.

The 1D $^{1}$H NMR spectra were processed and studied using a completely untargeted and unbiased multivariate data analytical approach. The aim was to identify the commonalities in the metabolic signatures associated with response to treatment for each tested compound. For this reason, a principal component analysis (PCA) was performed on the NMR spectra. The PCA scores plot displaying the two main principal components (PCs) accounting for 86.3% of the variance (PC-1 70.3%, PC-2 16.0%) is shown in Figure 4A. The PCA scores plot shows that the samples of the cells treated with RHPS4 (**2**) are positioned on the extreme right side of the principal direction of variance PC-1 and the samples of the cells treated with **1**, **3** and controls are placed to the left. Along with PC-2, the treatments with **1**

and **3**, positioned in the up-left quadrant of the plot, differ from the control samples, which are found in the bottom-left quadrant.

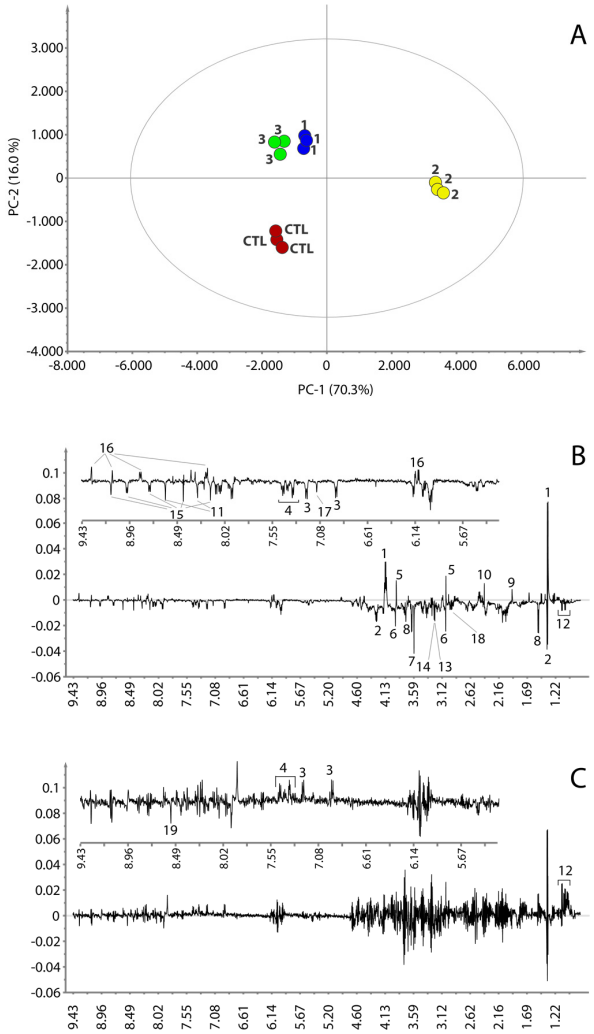


**Figure 4.** PCA score plot (**A**). PC-1 and PC-2 loading plots are reported in panel (**B**,**C**), respectively. Insets in (**B**,**C**) are expanded regions of the relative loading plots. Control samples are colored in red; compound **1**, **2** and **3** in blue, dark yellow and green, respectively. Numbers on the loading plots refer to the NMR assignment reported in Figure 3.

The loadings plot for the first principal component (Figure 4B) shows that the samples treated with **2** are characterized by a higher content of lactate, creatine, acetate, succinate and $NAD^+/NADP^+$, whereas the concentrations of threonine, glycine, alanine, tyrosine, phenylalanine, leucine, isoleucine, valine, histidine, creatine phosphate, glycerophosphocholine, O-phosphocholine, glutathione, NAAD and AMP are lower with the respect of the samples that lie on the left of the plot. The loadings plot of the second principal component (Figure 4C) is much noisier than that observed for PC-1. However, it appears that the samples treated with **1** and **3** differ from the control samples by having a higher content of leucine, isoleucine, valine, tyrosine, phenylalanine and a lower content of ATP. However, in order to better understand the effect of **1** and **3** on the metabolism of the cancer cells and to confirm the effect of **2**, a direct comparison of the average $^1$H NMR spectra of the three replicates for each treatment and controls was performed. The most interesting regions are reported in Figure 5. This comparison further corroborated the observation done for **2** and revealed that the treatments with **1** and **3** also caused variation in the content of lactate, threonine, glycine, creatine

phosphate, glycerophosphocholine, O-phosphocholine, histidine, $NAD^+/NADP^+$ and its precursor NAAD (Figure 5). Specifically, the concentration of lactate, threonine and creatine increases both in treatments with **1** and **3**. Glycine and creatine phosphate both decrease by treatment with **1**, whereas the treatment with **3** shows only a slight increment of creatine phosphate. O-phosphocholine and glycerophosphocholine were observed to decrease upon treatment with **1**, while samples treated with **3** showed only a slight increment of O-phosphocholine. The behavior of NAAD closely resembles that of creatine phosphate and glycerophosphocholine for all three treatments. On the other hand, histidine increased in the cell extracts by treatment with compounds **1** and **3**. On the contrary, the concentration of $NAD^+/NADP^+$ increased when the cell were treated with compound **3** and decreased by treatment with **1**. Furthermore, the content of acetate and succinate does not vary, while, concentration of ATP decreases in all three treatments. The behavior of all the cell metabolites is summarized in Table 1.

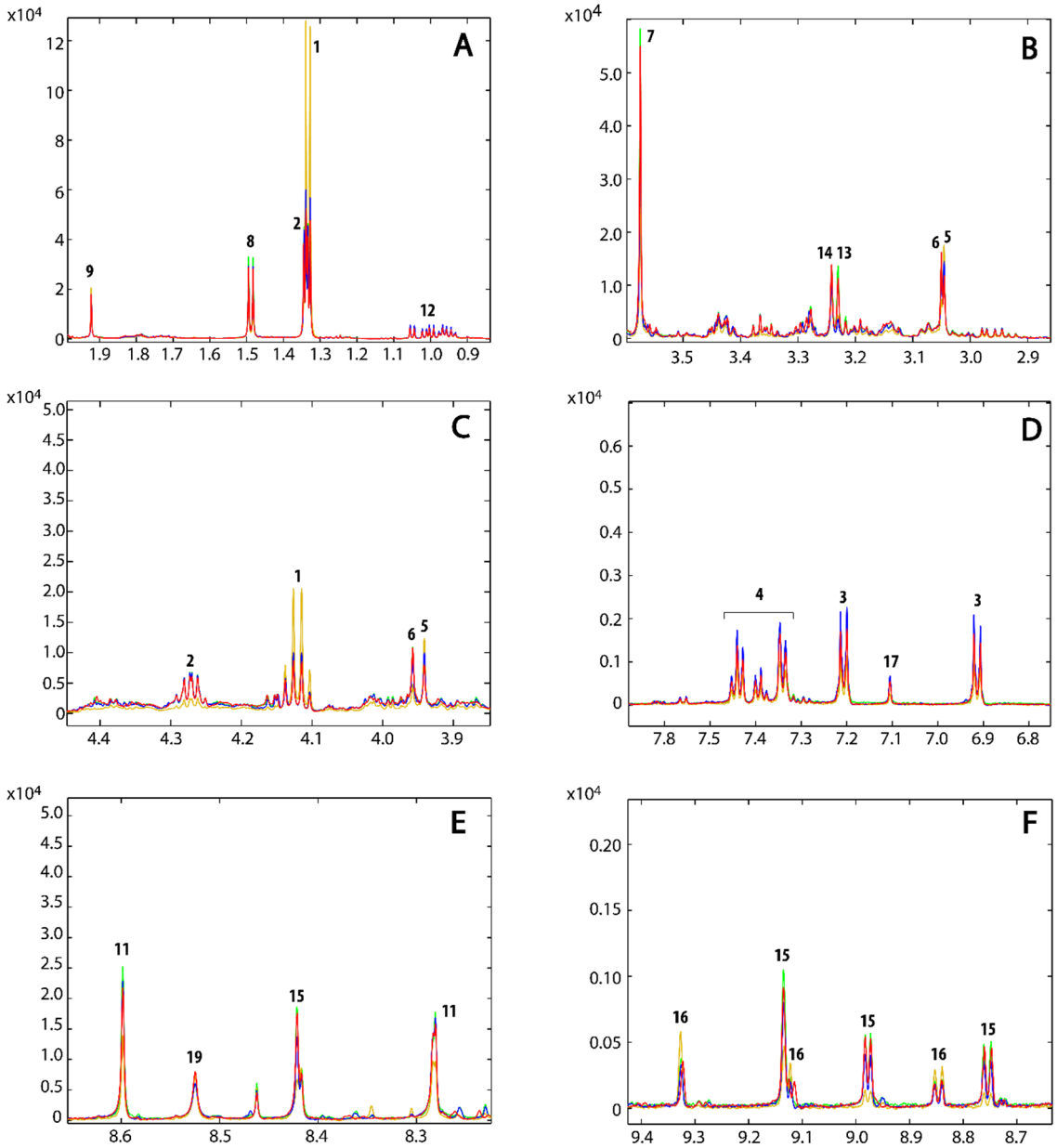


**Figure 5.** Expended regions of the superimposition of the mean NMR spectra of the untreated samples (**red**), and samples treated with compound **1** (**blue**), **2** (**dark yellow**) and **3** (**green**).

### *3.4. Metabolic Pathways Analysis*

As mentioned in the previous paragraph, many metabolites were affected by the treatment with compounds **1**, **2** and **3** (Table 1). In order to identify which metabolic pathways are involved, the MetaboAnalyst [37] web server was used. This tool suggests the most relevant pathways by uploading the discriminatory compounds that were significantly influenced by drug treatment. Results are provided in a so called "metabolic pathway analysis" and a "metabolite set enrichment overview" (Figure S2).

In particular, compound **2** significantly perturbs the levels of the metabolites that are involved in mitochondrial activities, compared to the untreated control cells. In fact, among the detected metabolites, the increased levels of succinate indicate inhibition of Complex I of respiratory chain of mitochondria useful to convert succinate in fumarate. This event thus impairs TCA cycle and production of ATP. Furthermore mitochondrion dysfunctions are shown by impaired conversion of creatine to creatine phosphate that results in further impairment of urea cycle and amino acid synthesis. Finally, decreased level of ATP and increased level of lactate and acetate are clear signs of apoptosis [39] and cell death in accordance with the down-regulated glutathione biosynthesis that suggest an increased reactive oxygen species (ROS) generation and a weakened ability to balance ROS. The cell death process was further supported by the reduction of choline metabolism that inhibits protein and DNA synthesis. Compound **1** behaves similarly to compound **2** because of the increase of lactate, creatine and decrease of creatine phosphate, ATP and glycine, as well as decrement of choline metabolism (glycerophosphocholine and O-phosphocholine). However, compounds **1** and **3** do not seem to interfere with TCA cycle, since succinate did not change. Compound **3**, similarly to **1** and **2**, drives cell death and apoptosis because of the increased lactate and creatine and decreased ATP.

In summary, the three tested compounds significantly altered the metabolism of the cells. The NMR data demonstrate that the treatments generally affect amino acid turnover or protein biosynthesis (alanine, glycine, isoleucine, leucine, valine, tyrosine, phenylalanine, threonine, histidine), tricarboxylic acid (TCA) cycle and mitochondrial activity (succinate, NAAD, NAD, ATP), urea cycle (creatine, creatine phosphate), anaerobic metabolism (lactate) and protein and DNA biosynthesis and DNA repair (choline and phosphocholine). Furthermore, the specific alterations in the choline metabolism by compounds **1** and **2** indicate that cell death in HCT116 lines is induced interfering with DNA synthesis and DNA damaged repair and by inhibition of protein synthesis. The NMR data thus strongly suggest that treatment with compounds **1** and **2** slow down cellular metabolism, aggravate oxidative stress and reduces DNA synthesis and repair leading to cellular death and apoptosis in accordance with their anti-cancer activity. Compound **3** also drives cell death and apoptosis due to a general cytotoxicity in accordance with anti-cancer activity of Adriamycin [40].

## 4. Conclusions

The implementation of a reliable NMR metabolomics analytical protocol has been quite challenging, owing to the number of critical steps along the way from cell culture to NMR tubes. This investigation was aimed at establishing a reliable protocol that describes how to handle the metabolome of the HCT116 human colon cancer cell line in order to perform a trustworthy metabolomic NMR analysis. This was pursued by simulating potential drug treatments using a limited number of "reliable" samples.

The best protocol was selected by combining different analytical procedures reported in the literature.

The optimized protocol can be summarized in the following main steps: (i) growth of the cell culture; (ii) abundant washing; (iii) cell scraping; (iv) quenching in liquid nitrogen; (v) cell lysis by sonication; and (vi) dual phase extraction procedure of the metabolites. It was demonstrated that the yield of the extraction and the quality of the extracted metabolome is of sufficiently high quality that the NMR assignment of detectable [41] metabolites could also be accomplished. Furthermore, preliminary insight into the biological behavior of the three tested anti-cancer compounds were accomplished.

**Supplementary Materials:** The following are available online at www.mdpi.com/2218-1989/6/1/4/s1, Figure S1: (A) Full STOCSY plot colored according to the correlation coefficient (see colored bar on the right); (B–D) Expanded regions of the STOCSY reporting the main correlations for NAD/NADP, NAAD and Lactate, respectively, Figure S2: Impact of the treatment with compounds 1-3 (panel A-C, respectively) on metabolic pathways of HCT116 cell, Table S1: CHEBI codes for the identified metabolites.

**Acknowledgments:** The authors acknowledge Xiaoyu Hu, Magnetic Resonance Center (CERM—University of Florence, Sesto Fiorentino, Italy) for providing useful information for the experimental protocol setting. This work has been financially supported by Associazione Italiana per la Ricerca sul Cancro AIRC IG 2013 # 14150 and by Italian Institute of Technology (IIT).

**Author Contributions:** Ilaria Lauri performed extraction of the metabolites, all the NMR experiments and preprocessing the data for the multivariate analysis. Ilaria Lauri, Nunzia Iaccarino and Antonio Randazzo took care of the metabolites' identification and NMR assignment. Cell cultures and treatments were performed by Pasquale Zizza. Luigi Michele Pavone provided the biochemical interpretation of the metabolite variations. Ilaria Lauri, Francesco Savorani, Nunzia Iaccarino and Antonio Randazzo wrote the manuscript. Francesco Savorani, Ettore Novellino, Søren Balling Engelsen and Antonio Randazzo supervised and coordinated most of this study, and edited the manuscript.

**Conflicts of Interest:** The authors declare that they have no conflict of interest.

## References

1. Griffin, J. Metabonomics: NMR spectroscopy and pattern recognition analysis of body fluids and tissues for characterisation of xenobiotic toxicity and disease diagnosis. *Curr. Opin. Chem. Biol.* **2003**, *7*, 648–654. [CrossRef] [PubMed]
2. Andrew Clayton, T.; Lindon, J.C.; Cloarec, O.; Antti, H.; Charuel, C.; Hanton, G.; Provost, J.-P.; Le Net, J.-L.; Baker, D.; Walley, R.J.; *et al*. Pharmaco-metabonomic phenotyping and personalized drug treatment. *Nature* **2006**, *440*, 1073–1077. [CrossRef] [PubMed]
3. Robertson, D.G. Metabonomics in Toxicology: A Review. *Toxicol. Sci.* **2005**, *85*, 809–822. [CrossRef] [PubMed]
4. Savorani, F.; Rasmussen, M.A.; Mikkelsen, M.S.; Engelsen, S.B. A primer to nutritional metabolomics by NMR spectroscopy and chemometrics. *Food Res. Int.* **2013**, *54*, 1131–1145. [CrossRef]
5. Tang, J. Microbial Metabolomics. *Curr. Genom.* **2011**, *12*, 391–403. [CrossRef] [PubMed]
6. Cuperlović-Culf, M.; Barnett, D.A.; Culf, A.S.; Chute, I. Cell culture metabolomics: Applications and future directions. *Drug Discov. Today* **2010**, *15*, 610–621. [CrossRef] [PubMed]
7. Zhang, A.; Sun, H.; Xu, H.; Qiu, S.; Wang, X. Cell Metabolomics. *OMICS* **2013**, *17*, 495–501. [CrossRef] [PubMed]
8. Bai, J.; Wang, M.X.; Chowbay, B.; Ching, C.B.; Chen, W.N. Metabolic profiling of HepG2 cells incubated with S(−) and R(+) enantiomers of anti-coagulating drug warfarin. *Metabolomics* **2011**, *7*, 353–362. [CrossRef] [PubMed]
9. Panopoulos, A.D.; Yanes, O.; Ruiz, S.; Kida, Y.S.; Diep, D.; Tautenhahn, R.; Herrerías, A.; Batchelder, E.M.; Plongthongkum, N.; Lutz, M.; *et al*. The metabolome of induced pluripotent stem cells reveals metabolic changes occurring in somatic cell reprogramming. *Cell. Res.* **2012**, *22*, 168–177. [CrossRef] [PubMed]
10. Serkova, N.; Glunde, K. Metabolomics of Cancer. In *Tumor Biomarker Discovery SE -20*; Tainsky, M.A., Ed.; Methods in Molecular Biology; Humana Press: New York, NY, USA, 2009; Volume 520, pp. 273–295.
11. Jemal, A.; Bray, F.; Center, M.M.; Ferlay, J.; Ward, E.; Forman, D. Global cancer statistics. *CA Cancer J. Clin.* **2011**, *61*, 69–90. [CrossRef] [PubMed]
12. Frezza, C.; Zheng, L.; Tennant, D.A.; Papkovsky, D.B.; Hedley, B.A.; Kalna, G.; Watson, D.G.; Gottlieb, E. Metabolic profiling of hypoxic cells revealed a catabolic signature required for cell survival. *PLoS ONE* **2011**. [CrossRef] [PubMed]
13. Kotze, H.L.; Armitage, E.G.; Sharkey, K.J.; Allwood, J.W.; Dunn, W.B.; Williams, K.J.; Goodacre, R. A novel untargeted metabolomics correlation-based network analysis incorporating human metabolic reconstructions. *BMC Syst. Biol.* **2013**, *7*, 107. [CrossRef] [PubMed]
14. Loftus, N.J.; Lai, L.; Wilkinson, R.W.; Odedra, R.; Wilson, I.D.; Barnes, A.J. Global Metabolite Profiling of Human Colorectal Cancer Xenografts in Mice Using HPLC–MS/MS. *J. Proteome Res.* **2013**, *12*, 2980–2986. [CrossRef] [PubMed]

15. Tolstikov, V.; Nikolayev, A.; Dong, S.; Zhao, G.; Kuo, M.-S. Metabolomics analysis of metabolic effects of nicotinamide phosphoribosyltransferase (NAMPT) inhibition on human cancer cells. *PLoS ONE* **2014**. [CrossRef] [PubMed]
16. Aftab, O.; Engskog, M.K.R.; Haglöf, J.; Elmsjö, A.; Arvidsson, T.; Pettersson, C.; Hammerling, U.; Gustafsson, M.G. NMR Spectroscopy-Based Metabolic Profiling of Drug-Induced Changes *in Vitro* Can Discriminate between Pharmacological Classes. *J. Chem. Inf. Model.* **2014**, *54*, 3251–3258. [CrossRef] [PubMed]
17. Balasubramanian, S.; Neidle, S. G-quadruplex nucleic acids as therapeutic targets. *Curr. Opin. Chem. Biol.* **2009**, *13*, 345–353. [CrossRef] [PubMed]
18. Brooks, T.A.; Kendrick, S.; Hurley, L. Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J.* **2010**, *277*, 3459–3469. [CrossRef] [PubMed]
19. Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative Visualization of DNA G-quadruplex Structures in Human Cells. *Nat. Chem.* **2013**, *5*, 182–186. [CrossRef] [PubMed]
20. Bidzinska, J.; Cimino-Reale, G.; Zaffaroni, N.; Folini, M. G-quadruplex structures in the human genome as novel therapeutic targets. *Molecules* **2013**, *18*, 12368–12395. [CrossRef] [PubMed]
21. Cosconati, S.; Rizzo, A.; Trotta, R.; Pagano, B.; Iachettini, S.; de Tito, S.; Lauri, I.; Fotticchia, I.; Giustiniano, M.; Marinelli, L.; *et al*. Shooting for Selective Druglike G-Quadruplex Binders: Evidence for Telomeric DNA Damage and Tumor Cell Death. *J. Med. Chem.* **2012**, *55*, 9785–9792. [CrossRef] [PubMed]
22. Read, M.; Harrison, R.J.; Romagnoli, B.; Tanious, F.A.; Gowan, S.H.; Reszka, A.P.; Wilson, W.D.; Kelland, L.R.; Neidle, S. Structure-based design of selective and potent G quadruplex-mediated telomerase inhibitors. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4844–4849. [CrossRef] [PubMed]
23. Di Leva, F.S.; Zizza, P.; Cingolani, C.; D'Angelo, C.; Pagano, B.; Amato, J.; Salvati, E.; Sissi, C.; Pinato, O.; Marinelli, L.; *et al*. Exploring the Chemical Space of G-Quadruplex Binders: Discovery of a Novel Chemotype Targeting the Human Telomeric Sequence. *J. Med. Chem.* **2013**, *56*, 9646–9654. [CrossRef] [PubMed]
24. Leonetti, C.; Amodei, S.; D'Angelo, C.; Rizzo, A.; Benassi, B.; Antonelli, A.; Elli, R.; Stevens, M.F.G.; D'Incalci, M.; Zupi, G.; *et al*. Biological Activity of the G-Quadruplex Ligand RHPS4 (3,11-Difluoro-6,8,13-trimethyl-8H-quino[4,3,2-kl]acridinium methosulfate) Is Associated with Telomere Capping Alteration. *Mol. Pharmacol.* **2004**, *66*, 1138–1146. [CrossRef] [PubMed]
25. Young, R.C.; Ozols, R.F.; Myers, C.E. The Anthracycline Antineoplastic Drugs. *N. Engl. J. Med.* **1981**, *305*, 139–153. [CrossRef] [PubMed]
26. Salvati, E.; Leonetti, C.; Rizzo, A.; Scarsella, M.; Mottolese, M.; Galati, R.; Sperduti, I.; Stevens, M.F.G.; D'Incalci, M.; Blasco, M.; *et al*. Telomere damage induced by the G-quadruplex ligand RHPS4 has an antitumor effect. *J. Clin. Investig.* **2007**, *117*, 3236–3247. [CrossRef] [PubMed]
27. Bligh, E.G.; Dyer, W.J. A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* **1959**, *37*, 911–917. [CrossRef] [PubMed]
28. Savorani, F.; Tomasi, G.; Engelsen, S.B. icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *J. Magn. Reson.* **2010**, *202*, 190–202. [CrossRef] [PubMed]
29. Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Wold, S. *Introduction to Multi- and Megavariate Data Anaysis Using Projection Methods (PCA & PLS)*; Umetrics: Umeå, Sweden, 1999.
30. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **1933**, *24*, 417–441. [CrossRef]
31. Martens, H.A.; Dardenne, P. Validation and verification of regression in small data sets. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 99–121. [CrossRef]
32. Wishart, D.S.; Jewison, T.; Guo, A.C.; Wilson, M.; Knox, C.; Liu, Y.; Djoumbou, Y.; Mandal, R.; Aziat, F.; Dong, E.; *et al*. HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic Acids Res.* **2013**, *41*, D801–D807. [CrossRef] [PubMed]
33. Bertini, I.; Hu, X.; Luchinat, C. Global metabolomics characterization of bacteria: Pre-analytical treatments and profiling. *Metabolomics* **2014**, *10*, 241–249. [CrossRef]
34. Duarte, I.F.; Marques, J.; Ladeirinha, A.F.; Rocha, C.; Lamego, I.; Calheiros, R.; Silva, T.M.; Marques, M.P.M.; Melo, J.B.; Carreira, I.M.; *et al*. Analytical Approaches toward Successful Human Cell Metabolome Studies by NMR Spectroscopy. *Anal. Chem.* **2009**, *81*, 5023–5032. [CrossRef] [PubMed]
35. Wang, X.; Hu, M.; Feng, J.; Liu, M.; Hu, J. $^{1}$H NMR metabolomics study of metastatic melanoma in C57BL/6J mouse spleen. *Metabolomics* **2014**, *10*, 1129–1144. [CrossRef] [PubMed]

36. Cloarec, O.; Dumas, M.-E.; Craig, A.; Barton, R.H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J.C.; Holmes, E.; *et al*. Statistical Total Correlation Spectroscopy: An Exploratory Approach for Latent Biomarker Identification from Metabolic 1H NMR Data Sets. *Anal. Chem.* **2005**, *77*, 1282–1289. [CrossRef] [PubMed]
37. Xia, J.; Psychogios, N.; Young, N.; Wishart, D.S. MetaboAnalyst: A web server for metabolomic data analysis and interpretation. *Nucleic Acids Res.* **2009**, *37*, W652–W660. [CrossRef] [PubMed]
38. Martineau, E.; Tea, I.; Loaëc, G.; Giraudeau, P.; Akoka, S. Strategy for choosing extraction procedures for NMR-based metabolomic analysis of mammalian cells. *Anal. Bioanal. Chem.* **2011**, *401*, 2133–2142. [CrossRef] [PubMed]
39. Wenzel, U.; Schoberl, K.; Lohner, K.; Daniel, H. Activation of mitochondrial lactate uptake by flavone induces apoptosis in human colon cancer cells. *J. Cell. Physiol.* **2005**, *202*, 379–390. [CrossRef] [PubMed]
40. Cao, B.; Li, M.; Zha, W.; Zhao, Q.; Gu, R.; Liu, L.; Shi, J.; Zhou, J.; Zhou, F.; Wu, X.; *et al*. Metabolomic approach to evaluating adriamycin pharmacodynamics and resistance in breast cancer cells. *Metabolomics* **2013**, *9*, 960–973. [CrossRef] [PubMed]
41. Krishnan, P.; Kruger, N.J.; Ratcliffe, R.G. Metabolic fingerprinting and profiling in plants using NMR. *J. Exp. Bot.* **2005**, *56*, 255–265. [CrossRef] [PubMed]


Food Research International 89 (2016) 1123–1133

Contents lists available at ScienceDirect

Food Research International

journal homepage: www.elsevier.com/locate/foodres

# Characterization of monovarietal extra virgin olive oils from the province of Béjaïa (Algeria)

Firdousse Laincer [a], Nunzia Iaccarino [b], Jussara Amato [b], Bruno Pagano [b], Alessia Pagano [b], Giancarlo Tenore [b], Abderezak Tamendjari [a], Pierangela Rovellini [c], Stefania Venturini [c], Giorgio Bellan [c], Alberto Ritieni [b], Luisa Mannina [d,e], Ettore Novellino [b], Antonio Randazzo [b,*]

[a] Laboratoire de Biochimie Appliquée, Faculté des Sciences de la Nature et de la Vie, Université de Béjaïa, 06000, Algérie
[b] Department of Pharmacy, University of Naples "Federico II", via D. Montesano 49, 80131 Naples, Italy
[c] INNOVHUB-SSI, Via Giuseppe Colombo 79, 20133 Milano, Italy
[d] Dipartimento di Chimica e Tecnologie del Farmaco, Sapienza Università di Roma, Piazzale Aldo Moro 5, I-00185 Rome, Italy
[e] Istituto di Metodologie Chimiche, Laboratorio di Risonanza Magnetica "Annalaura Segre", CNR, I-00015 Monterotondo, Rome, Italy

ARTICLE INFO

*Article history:*
Received 31 January 2016
Received in revised form 14 April 2016
Accepted 21 April 2016
Available online 23 April 2016

*Keywords:*
Olive oils
NMR
Algeria
Béjaïa
Statistical analysis

ABSTRACT

Olive fruits from 19 varieties and different areas of Béjaïa province (Algeria) were used to produce monovarietal olive oils in laboratory. The olive oils were analyzed using both traditional chemical analyses and nuclear magnetic resonance (NMR) methodology. The investigation involved pigment content determination, tocopherol analysis, fatty acid composition, and chromatographic determination of phenolic compounds. Chlorophyll, carotenoids, tocopherols and the content of oleic acid turned out to be variety dependent. The extra-virgin olive oils (EVOOs) were analyzed as a whole and as phenolic extract by NMR. The study gave general indication on olive oil quality and information about geographical origin of the samples. Overall, the results obtained in the present work reveal that Algerian monovarietal olive oils produced in Béjaïa province have the potential to produce blends that may compete with other Mediterranean products.



## 1. Introduction

Each olive oil has a peculiar fatty acid composition and also contains several micronutrients (Escrich, Moral, Grau, Costa, & Solanas, 2007) and a variety of minor components responsible for its particular sensory and nutritional characteristics (Boskou, 2006). Fatty acids (counting for about 98% of an olive oil weight) are mainly incorporated in triglycerides and diglycerides. Instead, minor components principally include pigments, tocopherols and phenolic compounds (Arslan & Schreiner, 2012; Manai-Djebali et al., 2012; Rigane, Boukhris, Bouaaziz, Sayadi, & Ben Salem, 2013; Vichi, Lazzez, Grati-Kamoun, & Caixach, 2012). Among all these compounds, monounsaturated fatty acids, pigments and antioxidants play an important role in extra-virgin olive oils (EVOOs) stability, as well as in the prevention of some diseases (Rotondi et al., 2004; Servili et al., 2009).

The absolute concentrations and relative proportions of olive oil's minor components are characteristic of each batch of olive oil, allowing for identification of the production area and possible adulterations (Angerosa, Campestre, & Giansante, 2006). The fine composition of an olive oil, besides being strongly dependent on the cultivar used for its production, is influenced by several other factors like climate, soil conditions and agricultural practices. More than 75% of the world olive oil production is concentrated in the Mediterranean area (Kavallari, Maas, & Schmitz, 2011) and it is constantly increasing. An in-depth chemical characterization of all Mediterranean olive oils is a key factor to increase the value for this specific production in order to preserve the unique landscape offered by the olive groves of the Mediterranean regions.

Algeria is one of the countries where the olive oil production is particularly increased in the last ten years, thanks to two agricultural renewal programs, over the 2006–2008 and 2009–2014 periods, allowing producers to update their production tools. In these years, cultivation zones passed from 165,000 to 500,000 ha, and Algeria is nowadays considered as a new olive oil exporter. Despite the increased production, olive oils from this area are poorly studied compromising their exportation especially in the countries that are major producers of olive oils.

Therefore, the aim of this work is to chemically characterize monovarietal Algerian EVOOs from different areas of the Béjaïa province, that is the area where the olive oil production is mostly increased, and, eventually, to suggest possible blends. The investigation was performed using traditional chemical analyses aimed to determine the content of fatty acids, pigments, tocopherols and phenolic compounds. These analyses were further supported by the use of NMR spectroscopy.

* Corresponding author.
*E-mail address:* antonio.randazzo@unina.it (A. Randazzo).

http://dx.doi.org/10.1016/j.foodres.2016.04.024
0963-9969/

Noteworthy, although NMR is nowadays recognized as an important tool for food characterization (Mannina, Sobolev, & Viel, 2012), this is the first NMR investigation of Algerian olive oils. In particular, NMR was used to characterize olive oils in terms of geographical origin, cultivar, and genuineness (Mannina et al., 2009; D'Imperio et al., 2007; Mannina, Marini, Gobbino, Sobolev, & Capitani, 2010; Mannina & Sobolev, 2011).

## 2. Materials and methods

### 2.1. Extra virgin olive oil sampling

Twenty-six monovarietal extra virgin olive oil (EVOO) samples from different areas of the province of Béjaïa (Algeria) (Fig. 1) were analyzed. The olive fruits coming from 19 different cultivars (Table S1) were randomly and manually picked from all parts of the selected fully-grown olive trees. In order to eliminate the influence of the maturation state on olive oil quality, the ripening degree was the same for all studied olive samples. After the harvest, olive fruits were immediately transported to the laboratory. The olives were washed and deleafed, and only healthy fruits, without any kind of infection or physical damage, were selected. Olive oil samples were obtained using a laboratory oil mill (Levi-Deleo-Lerogsame), consisting of three basic elements: a hammer crusher, a thermo-beater (mixer) and a pulp centrifuge. The olive fruits were milled in the hammer crusher, and then the olive paste was kneaded for 30 min with the addition of warm water (50 mL of water was added to 920 g of paste). After the vertical centrifugation, the olive oil was collected and left to stand. The olive oil samples were stored in amber-glass bottles, labeled with the laboratory code, without headspace and kept at 4 °C in the dark until analysis.

### 2.2. Pigment content determination

Chlorophyll and carotenoid pigments were determined by UV spectroscopy at 670 and 470 nm, respectively, in cyclohexane, using the specific extinction coefficients, according to the method proposed by Minguez-Mosquera, Rejano, Gandul, Sanchez, and Garrido (1991). The values of the specific extinction coefficients used were $\varepsilon_0 = 613$ for pheophytin, as a major component in the chlorophyll fraction, and $\varepsilon_0 = 2000$ for lutein as a major component in the carotenoid fraction. Thus, pigment content was calculated as follows:

$$\text{Chlorophyll}\left(\text{mg kg}^{-1}\right) = \left(A_{670} \times 10^6\right)/(613 \times 100 \times d)$$

$$\text{Carotenoid}\left(\text{mg kg}^{-1}\right) = \left(A_{470} \times 10^6\right)/(2000 \times 100 \times d)$$

where A is the absorbance and d is the spectrophotometer cell thickness (1 cm).

Chlorophyll and carotenoid contents were expressed as mg of pheophytin "a" and lutein per kg of oil, respectively.

### 2.3. Tocopherol analysis

Tocopherols were evaluated following the method developed by Rovellini, Azzolini, and Cortesi (1997). A solution of olive oil in acetone was analyzed by high performance liquid chromatography (HPLC) on a reversed phase silica column (Allsphere ODS2 Alltech 5 μm, 250 mm × 4.6 mm) and was eluted with acetonitrile/methanol (1:1) at a flow rate of 1.3 mL/min. An UV detector was used at 292 nm.

### 2.4. Fatty acid composition analysis

The fatty acid composition was determined as methyl ester derivatives by gas chromatography (GC) according to methods described in EC Regulation 796/, 2002. Fatty acid methyl esters were prepared by vigorous shaking of a solution of each olive oil sample in *n*-hexane (0.5 g in 5 mL) with 0.5 mL of 2 N methanolic potassium hydroxide solution. Chromatographic analysis was performed on a CHROMPACK C 9002 gas chromatograph equipped with a FID detector, using a capillary column DB 23 (30 m × 0.32 mm i.d. × 0.25 μm film thicknesses). The injector and detector temperatures were maintained at 250 °C; the oven

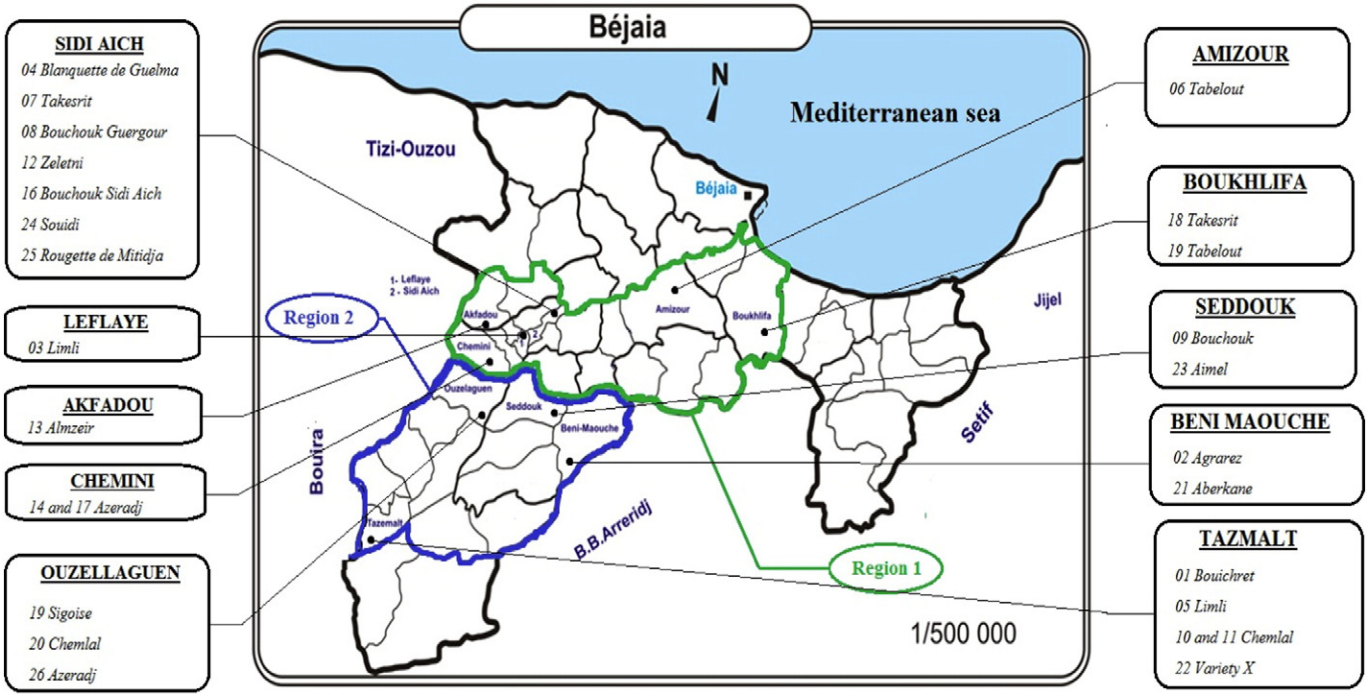


**Fig. 1.** Areas of the province of Béjaïa (Algeria) where the olives fruits were harvested. The province of Béjaïa is bordered by black bold line. Regions 1 and 2 are bordered by green and blue lines, respectively. Sample code, cultivar and area of origin of the samples are also reported.

temperature was set at 200 °C. Nitrogen was employed as a carrier gas with a flow rate of 1 mL/min.

### 2.5. Chromatographic determination of phenolic compounds

A solution of internal standard (1 mL of 0.015 mg $mL^{-1}$ of syringic acid in water/MeOH 20:80 v/v) was added to a sample of virgin olive oil (2 g). The mixture was shaken (30 s) and 5 mL of the extraction solution containing water and MeOH (20:80 v/v) was added. The obtained mixture was shaken for 1 min, extracted for 15 min in an ultrasonic bath and then centrifuged at 5000 rpm (2500 *g*) for 25 min at T = 20 °C (Norme Grassi e Derivati, NGD, 2010; IOC, 2010). The upper phase was filtered using a 0.45 μm PVDF syringe filter, and then 20 μL of the filtered solution was analyzed by HPLC with a UV detector at 280 nm (Fig. S1). The HPLC system consisted of a C18 Spherisorb ODS-2 reverse column (5 μm, 250 mm × 4.6 mm). Elution was performed at a flow rate of 1 mL $min^{-1}$ following a gradient composed of a mixture of water and orthophosphoric acid (99.8:0.2 v/v) (solvent A), MeOH (solvent B) and acetonitrile (solvent C): from 96% (A)–2% (B)–2% (C) to 0% (A)–50% (B)–50% (C) in 60 min. The last gradient was kept for 10 min. The successive gradient was: from 0% (A)–50% (B)–50% (C) to 96% (A)–2% (B)–2% (C) in 2 min and then kept for 10 min. The identification of phenolic compounds was performed by HPLC–MS. The main phenolic compounds were identified by comparison with relative retention times and UV spectra of pure standards.

### 2.6. Statistics

The chemical analysis data reported in the tables were subjected to analysis of variance using the Statistica 5.5 package (StatSoft 97 edition). Where statistical differences were noted, differences among packages were determined using the ANOVA/MANOVA following the Newman–Keuls test. Significance was defined at $p < 0.05$.

### 2.7. Phenolic compounds extraction for NMR experiments

Phenolic compounds were extracted from EVOOs according to Montedoro, Servilli, Baldioli, and Miniati (1992) and Christophoridou and Dais (2009) method with minor modifications: 8 mL of MeOH/water (80:20, v/v) solution was added to 8 g of EVOO; the mixture was vortexed for 30 s and the two phases were separated by centrifugation at 1400 *g* for 10 min. The extraction was repeated twice. Alcoholic extracts were then combined and concentrated under vacuum at T < 35 °C until a syrupy consistency was reached. Then 8 mL of acetonitrile was added to the extract, and it was partitioned with 8 mL of *n*-hexane. The two phases were then separated by centrifugation at 1400 *g* for 10 min and the acetonitrile phases were dried under a stream of nitrogen.

### 2.8. NMR experiments

The experiments were performed both on the EVOOs and the relative phenolic extracts. NMR samples were prepared by mixing 20 μL of the EVOO sample with 560 μL of deuterated chloroform and 20 μL of deuterated DMSO (DMSO-$d_6$). The addition of DMSO-$d_6$ has been proved to be useful for the dissolution of oil and fat minor components (Sacchi et al., 1996). The phenolic fractions were dissolved in DMSO-$d_6$ in order to get a final concentration of 10 mg $mL^{-1}$, finally 600 μL of this clear solution was transferred into the NMR tube. NMR experiments were conducted in 5 mm NMR tubes at 25 °C with a 700 MHz Varian Unity Inova spectrometer equipped with a 5 mm $^1H\{^{13}C/^{15}N\}$ triple resonance probe. $^1H$ NMR spectra of EVOOs and phenolic extracts were acquired using the same parameters: spectral width 11,193.507, acquisition time 2.92 s and relaxation delay 3 s. 90° pulses were used in both cases.

#### 2.8.1. NMR data reduction and preprocessing

*2.8.1.1. Olive oil.* The spectra were processed using iNMR software (www.inmr.net). $^1H$ NMR spectra were obtained by the Fourier Transformation (FT) of the free induction decay (FID), applying an exponential multiplication with a line-broadening factor of 0.5 Hz. The NMR regions above 11.7 ppm and below 0 ppm were removed because they contain only noise. Furthermore, regions between 7.1 and 7.5 ppm and between 2.49 and 2.58 ppm were discarded because of the residual signals of chloroform and DMSO respectively. Data reduction was obtained by dividing the spectra in evenly spaced windows of 0.002 ppm width. Then the normalization step was needed as in the case of extracts. Therefore each spectrum of the NMR data matrix was normalized referring to the area under the signal of the methyl protons at 0.821 ppm. After this preprocessing step of the spectra, a matrix of 26 rows (samples) and 5865 columns (ppm) was obtained and directly submitted to the multivariate statistic analysis.

*2.8.1.2. Phenolic extracts.* $^1H$ NMR spectra were segmented in buckets of fixed 0.002 ppm width and integrated. The NMR regions above 10.25 ppm and below 0 ppm were removed because of the absence of signals. The regions between 2.50 and 2.57 ppm and between 3.30 and 3.62 ppm were discarded because of the residual signals of DMSO and water, respectively. Furthermore, the regions of residual signals belonging to triglyceride esters still present in the extracted samples were also removed. Since NMR spectra showed slight misalignments in chemical shift due to pH-sensitive peaks, the spectra were aligned using the interval correlation optimized shifting algorithm (icoshift) (Savorani, Tomasi, & Engelsen, 2010). A normalization preprocessing step was carried out to correct variations of the overall concentrations of the samples. Since no quantitative internal standard was used, each spectrum of the aligned NMR data matrix was normalized to unit area, obtained by dividing every variable of each spectrum by the sum of the absolute value of all its variables. In the end, a total of 5115 variables were analyzed for each spectrum.

All preprocessing steps were performed using Matlab (The Mathworks Inc., Natick, MA).

#### 2.8.2. Statistical analysis

NMR data matrixes were imported into Simca-P 13.0 (Umetrics, Umeå, Sweden) and pareto-scaled (Eriksson, Johansson, Kettaneh-Wold, & Wold, 1999). A principal component analysis (PCA) was performed on the datasets. PCA is an unsupervised statistical approach that allows to identify patterns in a dataset highlighting hidden similarities and/or differences.

The goodness of the model is given by the degree of fit and by the predictive ability. The degree of fit tells how much of the variance from the dataset can be explained by the model; a quantitative measurement of the fit is given by the parameter $R^2$. The predictive power of the model is tested through the cross-validation (CV) procedure. The idea is to develop different models, keeping out, each time, some samples from the dataset. The model on the reduced dataset is then used to predict the omitted samples. $Q^2$ is the parameter that indicates the goodness of prediction of the model (Eriksson et al., 1999). In this case leave-one-out CV was employed given the limited number of samples, thus only one sample was removed each time in the CV process. Both $R^2$ and $Q^2$ can vary from 0 to 1, where 0 means no fitting and no prediction model. The PCA models performed in this study showed a $R^2 \geq 0.7$ and a $Q^2 \geq 0.5$, which indicate goodness both of fit and prediction. The optimal number of component was established looking at both $R^2$ and $Q^2$.

Statistical Total Correlation Spectroscopy (STOCSY) analysis was performed on the binned NMR data matrix of the phenolic extracts in order to obtain the correlations among the signals belonging to the same compounds. The results were plotted using a threshold value of $R \geq 0.9$ (Fig. S2).

*2.8.3. Metabolites identification*

*2.8.3.1. Olive oil.* The assignment of the major and minor components of olive oil was carried out according to the references Sacchi et al. (1996); Sacchi, Addeo, and Paolillo (1997); Segre and Mannina (1997) and Mannina, Patumi, Fiordiponti, Emanuele, and Segre (1999). The identified signals are shown in Fig. 2 and reported in Table S2.

*2.8.3.2. Phenolic extract.* Identification of phenolic compounds was obtained following the references in the literature (Christophoridou, Dais, Tseng, & Spraul, 2005; Christophoridou & Dais, 2009) and the analysis of the Statistical Total Correlation Spectroscopy (STOCSY) (Fig. S2). The complete assignment is reported in Table 1 and Fig. 3.

## 3. Results and discussion

*3.1. Chlorophyll and carotenoids*

Chlorophylls and carotenoids are the main pigments in vegetable oils. Their total content in olive oils is an important quality parameter, since it correlates with color, which is one of the first characteristics evaluated by consumers. These pigments act as pro-oxidants in the presence of light and as antioxidants in the darkness (Psomiadou & Tsimidou, 2002). Furthermore, they have biological and health properties (Ranalli & Modesti, 1999).

In the investigated EVOOs, chlorophyllic and carotenoid pigments were found to be at concentrations between 0.1 and 3.1 mg $kg^{-1}$ and between 0.1 and 1.6 mg $kg^{-1}$, respectively (Table 2). The data indicated a significant influence of the variety on the amount of chlorophylls and carotenoids ($p < 0.05$). For example, *Azeradj* (sample 26) and *Bouchret* (sample 1) olive oils contained the lowest level of chlorophyll (0.1 mg $kg^{-1}$) and carotenoids (0.1 mg $kg^{-1}$), respectively, whereas olive oil extracted from *Takesrit* (sample 7) contained the highest level of chlorophylls and carotenoid pigments (3.1 and 1.6 mg $kg^{-1}$, respectively).

*3.2. Tocopherol composition*

Four tocopherol isomers were detected and quantified in the analyzed samples: α-tocopherol (the most abundant), β-tocopherol, γ-tocopherol and δ-tocopherol (Fig. S3). As shown in Table 3, the amount of α-tocopherol showed significant differences between cultivars ($p < 0.05$), varying from 204.8 to 573.0 mg $kg^{-1}$. On the other hand, β-tocopherol and γ-tocopherol ranged from 1.3 to 9.2 mg $kg^{-1}$ and from 0.6 to 59.1 mg $kg^{-1}$, respectively. Concerning δ-tocopherol, the highest value was determined in the *Rougette de Mitidja* oil (sample 25) (1.6 mg $kg^{-1}$). Total tocopherols' content showed the same trend of α-tocopherol, since it was the most representative isomer in the tocopherol composition. The minimum and maximum contents (209.6 and 581.8 mg $kg^{-1}$) were observed in *Bouichret* (sample 1) and *Souidi* (sample 24) varieties, respectively. These results are in agreement with the literature, and indicated that tocopherol content is highly variety dependent (Deiana et al., 2002; Manai-Djebali et al., 2012; Dağdelen, Tümen, Ozcan, & Dündar, 2012). Interestingly, α-tocopherol, the active form of vitamin E, exerts, along with the polyphenols, an important antioxidant action that prevents the oxidation of low density lipoproteins (LDLs). All the analyzed samples were found to contain more than 200 mg $kg^{-1}$ of α-tocopherol, with an average content of total tocopherol of 300 mg $kg^{-1}$, with the exception of *Souidi* variety (sample 24), which is characterized by a 2-fold higher concentration. Thus, assuming a daily intake of only 10 g of olive oil from *Souidi* variety, it is possible to supply about 50% of daily vitamin E requirements.

*3.3. Fatty acid composition*

Differences among the cultivars were remarkable considering the relative concentration of saturated (SFA), monounsaturated (MUFA)

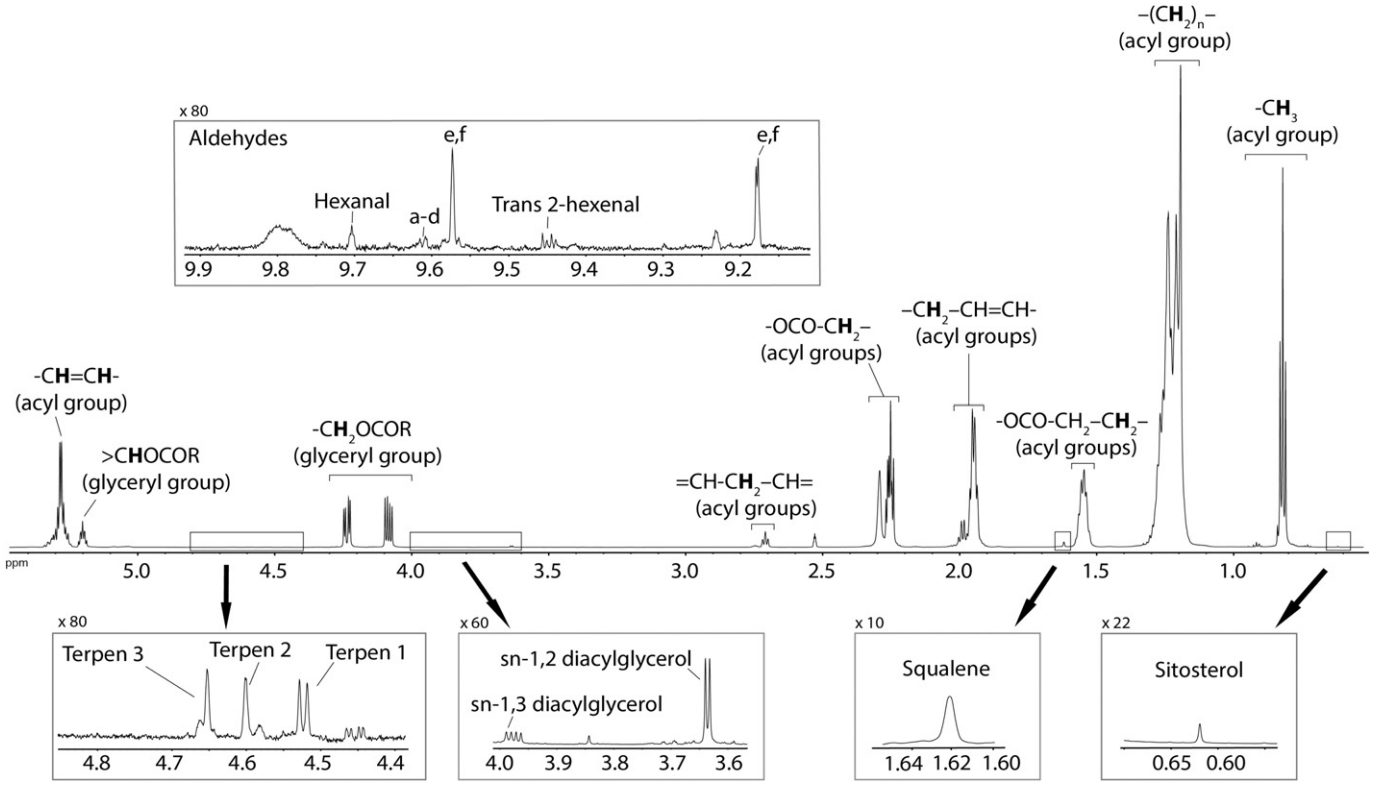


**Fig. 2.** $^1$H NMR spectrum of a representative EVOO sample. Some expanded regions containing minor components of EVOO are reported in the squared expansions. The magnification value is indicated on the top left of each square. Letters on the NMR signals in the left top frame correspond to the compounds reported in Table 1.

**Table 1**
Chemical shifts, multiplicity of phenolic compounds identified in EVOO extracts by high-resolution $^1$H NMR spectroscopy (700 MHz, DMSO-$d_6$).

| | Compound | Chemical shift δ (ppm) (multiplicity, J in Hz, assignment) |
|---|---|---|
| a | Aldehydic form of ligstroside (5S, 8R, 9S) | 1.34 (d, 6.6, H10), 2.81 (t, 6.8, H2′), 3.65 (s, $OCH_3$), 4.15–4.25 (H1a′, H1b′), 6.72 (d, 8.4, H5′, H7′), 7.06 (d, 8.4, H4′, H8′), 7.55 (s, H3), 9.52 (d, 1.5, H1) |
| b | Aldehydic form of ligstroside (5S, 8S, 9S) | 1.55 (d, 6.6, H10), 2.81 (t, 6.8, H2′), 3.23 (dd, 2.9, 11.0, H5), 3.66 (s, $OCH_3$), 4.15–4.25 (H1a, H1b, H8), 6.72 (d, 8.4, H5′, H7′), 7.06 (d, 8.4, H4′, H8′), 7.59 (s, H3), 9.67 (H1) |
| c | Aldehydic form of oleuropein (5S, 8R, 9S) | 1.34 (d, 6.6, H10), 3.65 (s, $OCH_3$), 4.15–4.25 (H1a, H1b, H8), 6.48 (dd, 1.8, 8.1, H8′), 6.69 (d, 1.8, H4′), 6.72 (d, 8.1, H7′), 9.50 (H1, d, 1.5) |
| d | Aldehydic form of oleuropein (5S, 8S, 9S) | 1.55 (d, 6.6, H10), 2.60 (H4b, H6a), 3.23 (dd, 2.9, 11.0, H5), 3.66 (s, $OCH_3$), 4.15–4.25 (H1a′, H1b′, H8), 6.70 (d, 1.8, H4′), 6.48 (dd, 1.8, 8.1, H8′), 6.72 (d, 8.4, H7′), 7.59 (s, H3), 9.67 (H1) |
| e | Dialdehydic form of ligstroside lacking a carboxymethyl group | 1.97 (d, 7.0, H10), 2.60 (H4b, H6a), 2.74 (ddd, 1.7, 8.5, 17.8, H4a), 4.00–4.20 (H1′a, H1′b), 6.77 (d, 8.4, H5′, H7′), 7.03 (H4′, H8′), 9.24 (d, 1.8, H1), 9.54 (s, br, H3) |
| f | Dialdehydic form of oleuropein lacking a carboxymethyl group | 1.97 (d, 7.0, H10), 2.74 (ddd, 1.3, 8.2, 17.8, H4a), 4.00–4.20 (H1′a, H1′b), 6.77 (d, 8.1, H7′), 9.24 (d, 1.8, H1), 9.54 (s, br, H3) |
| g | (+) 1-Acetoxypinoresinol | 1.69 (s, $COOCH_3$), 3.77 (s, 3′-$OCH_3$), 3.79 (s, 3″-$OCH_3$), 4.73 (d, 4.8, H6), 4.29 (d, 10.6, H8a), 4.92 (s, br, H2), 6.72 (d, 8.1, H5′), 6.81 (d, 8.1, H5″), 6.84 (dd, 1.8, 8.1 H6′), 6.89 (d, 1.8, H2′), 6.98 (d, 1.8, H2″) |
| h | (+) Pinoresinol | 3.78 (s, $OCH_3$), 4.61 (d, 4.0, H2, H6), 6.91 (d, 1.8, H2, H2) |
| i | Vanillin | 9.80 (s, H7) |
| l | Luteolin | 6.19 (d, 1.8, H6), 6.92 (d, 8.8, H5′), 7.40 (d, H2′), 7.42 (dd, 2.2, 8.8, H6′) |
| m | Apigenin | 6.22 (d, 1.8, H6), 7.94 (d, 8.8, H2′, H6′) |
| n | Free hydroxytyrosol | 6.46 (dd, 1.8, 8.1, H8′), 6.59 (d, H4′) |

**Table 2**
Chlorophyll and carotenoids content (mg $kg^{-1}$) of Algerian extra-virgin olive oils.

| Sample | Cultivar (cv.) | Chlorophyll | Carotenoids |
|---|---|---|---|
| 1 | *Bouichret* | $0.2^{a,b} \pm 0.1$ | $0.1^{a} \pm 0.0$ |
| 2 | *Agrarez* | $0.5^{f} \pm 0.0$ | $0.5^{g,h,i} \pm 0.0$ |
| 3 | *Limli* | $0.3^{b,c,d} \pm 0.1$ | $0.2^{a,b,c} \pm 0.0$ |
| 4 | *Blanquette de Guelma* | $0.8^{g} \pm 0.2$ | $0.4^{f} \pm 0.1$ |
| 5 | *Limli* | $0.4^{b,c,d,e} \pm 0.1$ | $0.2^{c,d} \pm 0.0$ |
| 6 | *Tabelout* | $0.2^{a,b} \pm 0.0$ | $0.1^{a,b} \pm 0.0$ |
| 7 | *Takesrit* | $3.1^{k} \pm 0.1$ | $1.6^{l} \pm 0.0$ |
| 8 | *Bouchouk Guergour* | $1.1^{i} \pm 0.1$ | $0.6^{i} \pm 0.2$ |
| 9 | *Bouchouk* | $0.6^{e,f} \pm 0.0$ | $0.3^{e} \pm 0.0$ |
| 10 | *Chemlal* | $0.8^{g,h} \pm 0.0$ | $0.5^{g,h} \pm 0.0$ |
| 11 | *Chemlal* | $0.4^{c,d,e,f} \pm 0.1$ | $0.2^{c,d} \pm 0.0$ |
| 12 | *Zeletni* | $0.8^{g,h} \pm 0.0$ | $0.5^{g,h,i} \pm 0.0$ |
| 13 | *Almzeir* | $0.5^{f} \pm 0.0$ | $0.3^{d,e} \pm 0.0$ |
| 14 | *Azeradj* | $1.0^{h} \pm 0.0$ | $0.6^{h,i} \pm 0.0$ |
| 15 | *Takesrit* | $0.9^{g,h} \pm 0.0$ | $0.5^{f,g} \pm 0.0$ |
| 16 | *Bouchouk Sidi Aich* | $0.5^{e,f} \pm 0.1$ | $0.7^{j} \pm 0.1$ |
| 17 | *Azeradj* | $0.3^{a,b,c} \pm 0.0$ | $0.1^{a,b,c} \pm 0.0$ |
| 18 | *Tabelout* | $0.5^{d,e,f} \pm 0.0$ | $0.2^{b,c,d} \pm 0.0$ |
| 19 | *Sigoise* | $0.3^{b,c} \pm 0.0$ | $0.2^{c,d} \pm 0.0$ |
| 20 | *Chemlal* | $0.5^{e,f} \pm 0.0$ | $0.3^{e} \pm 0.0$ |
| 21 | *Aberkane* | $0.3^{b,c} \pm 0.1$ | $0.1^{a} \pm 0.0$ |
| 22 | *Variety X* | $0.7^{g} \pm 0.0$ | $0.5^{g,h,i} \pm 0.0$ |
| 23 | *Aimel* | $0.3^{b,c,d} \pm 0.0$ | $0.2^{d,e} \pm 0.0$ |
| 24 | *Souidi* | $0.9^{g,h} \pm 0.0$ | $0.7^{j} \pm 0.0$ |
| 25 | *Rougette de Mitidja* | $1.6^{j} \pm 0.1$ | $0.9^{k} \pm 0.0$ |
| 26 | *Azeradj* | $0.1^{a} \pm 0.0$ | $0.2^{a,b,c} \pm 0.0$ |

Means ± standard deviation (n = 3).
[a–k]Means the column followed by different letters are significantly different ($p > 0.05$).

and polyunsaturated fatty acids (PUFA). Concerning the MUFA, *Rougette de Mitidja* olive oil (sample 25) was characterized by the highest percentage (79.2%), due to its high content in oleic acid. *Bouchouk* (sample 9) olive oil was rich in PUFA (18.8%), mainly due to the content in linoleic acid. *Taksreit* (sample 7) olive oil was rich in total saturated fatty acids (SFA) (21.8%), essentially due to its high content in palmitic acid.

More in detail, oleic acid (C18:1), the most abundant monounsaturated fatty acid, was dependent by the varieties (see Table 4). It reached a high percentage in *Limli* (sample 5) olive oil (78.2%), while it had lower levels in *Bouchouk* (sample 9) and *Souidi* (sample 24) oils (58.6% and 59.2% respectively). The remaining olive oil samples showed intermediate levels of oleic acid content.

As for linoleic acid (C18:2), sample 9 (*Bouchouk*) showed the highest mean value (18.0%), whereas the lowest one was found in *Rougette de Mitidja* (sample 25) (3.5%). The concentrations of palmitoleic (C16:1), stearic (C18:0) and linolenic (C18:3) acids were very low in all analyzed samples and changed from one olive oil to another. In any case, the

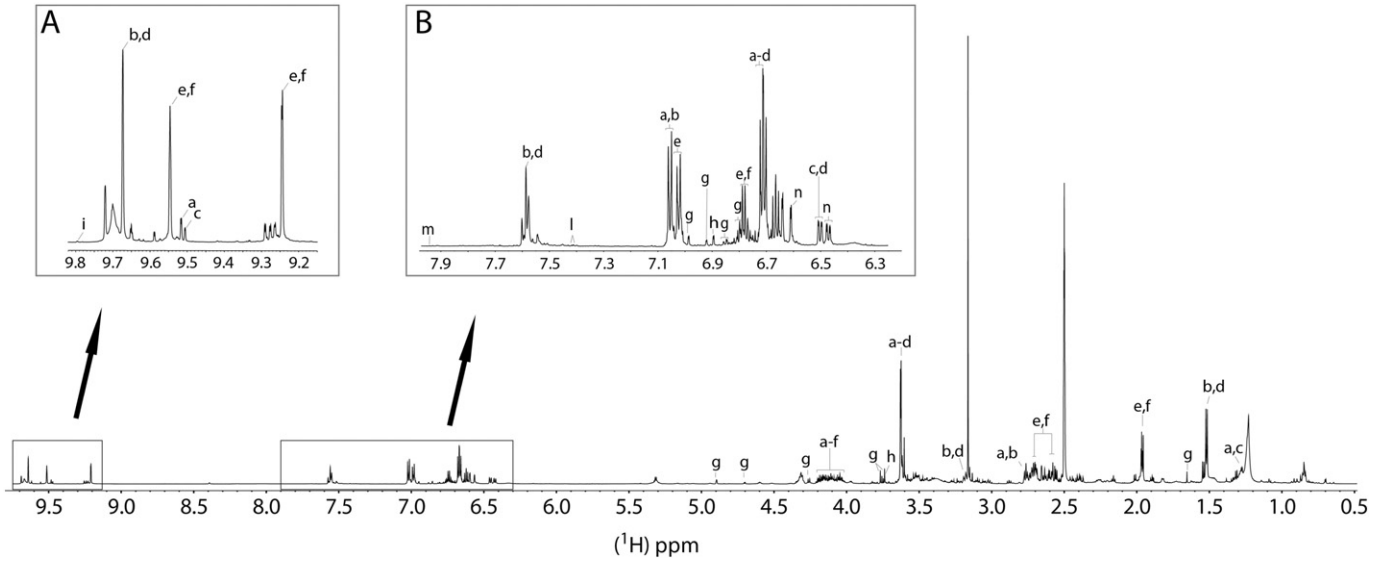


**Fig. 3.** $^1$H NMR spectrum of a representative phenolic extract of an EVOO sample. Letters on the NMR signals correspond to the compounds reported in Table 1.

**Table 3**
Tocopherol contents (mg $kg^{-1}$) of Algerian extra-virgin olive oils.

| Sample | Cultivar (cv.) | Tocopherol-δ | Tocopherol-γ | Tocopherol-β | Tocopherol-α | Tocopherol total |
|---|---|---|---|---|---|---|
| 1 | *Bouichret* | $0.2^{a,b} \pm 0.0$ | $0.7^{a} \pm 0.2$ | $1.3^{a} \pm 0.1$ | $207.5^{a} \pm 5.9$ | $209.6^{a} \pm 5.7$ |
| 2 | *Agrarez* | $0.1^{a,b} \pm 0.1$ | $2.2^{a,b} \pm 0.0$ | $1.8^{b,c} \pm 0.0$ | $220.4^{a,b} \pm 2.2$ | $224.5^{a,b} \pm 2.2$ |
| 3 | *Limli* | $1.1^{g} \pm 0.1$ | $47.3^{k} \pm 0.1$ | $2.3^{d,e,f} \pm 0.2$ | $300.0^{g} \pm 0.8$ | $350.8^{f,g} \pm 0.8$ |
| 4 | *Blanquette de Guelma* | $0.3^{a,b,c} \pm 0.0$ | $9.7^{g} \pm 0.0$ | $3.9^{k} \pm 0.1$ | $486.9^{j} \pm 7.5$ | $500.7^{j} \pm 7.6$ |
| 5 | *Limli* | $0.1^{a} \pm 0.0$ | $1.6^{a,b} \pm 0.1$ | $1.6^{b} \pm 0.1$ | $279.1^{c,d,e} \pm 4.7$ | $282.4^{d,e,f} \pm 4.7$ |
| 6 | *Tabelout* | $0.1^{a} \pm 0.0$ | $0.9^{a} \pm 0.0$ | $2.1^{d,e} \pm 0.0$ | $300.9^{d,e,f} \pm 4.1$ | $304.1^{f,g} \pm 4.1$ |
| 7 | *Takesrit* | $1.3^{h} \pm 0.2$ | $59.1^{l} \pm 0.1$ | $3.4^{j} \pm 0.4$ | $501.1^{k} \pm 3.3$ | $564.9^{j} \pm 3.6$ |
| 8 | *Bouchouk Guergour* | $0.4^{b,c,d} \pm 0.0$ | $13.3^{h} \pm 0.2$ | $2.5^{f,g} \pm 0.1$ | $233.9^{b,c} \pm 4.9$ | $249.9^{a,b,c} \pm 5.2$ |
| 9 | *Bouchouk* | $0.2^{a,b,c} \pm 0.1$ | $4.2^{c,d} \pm 0.1$ | $2.9^{h} \pm 0.0$ | $263.4^{c,d} \pm 4.4$ | $270.6^{c,d,e,f} \pm 4.5$ |
| 10 | *Chemlal* | $0.2^{a,b,c} \pm 0.0$ | $6.7^{e,f,g} \pm 0.3$ | $3.3^{i,j} \pm 0.1$ | $263.9^{d,e,f} \pm 4.8$ | $274.1^{e,f,g} \pm 5.0$ |
| 11 | *Chemlal* | $0.2^{a,b,c} \pm 0.1$ | $4.9^{d} \pm 0.0$ | $2.4^{e,f,g} \pm 0.0$ | $204.8^{a} \pm 0.2$ | $212.4^{a} \pm 0.1$ |
| 12 | *Zeletni* | $0.1^{a} \pm 0.0$ | $2.5^{a,b,c} \pm 0.1$ | $1.6^{b} \pm 0.0$ | $278.9^{c,d,e} \pm 4.1$ | $283.7^{d,e,f} \pm 4.1$ |
| 13 | *Almzeir* | $1.1^{g} \pm 0.0$ | $46.9^{k} \pm 0.1$ | $1.6^{b} \pm 0.2$ | $250.2^{d,e,f} \pm 2.6$ | $299.8^{b,c,d} \pm 2.9$ |
| 14 | *Azeradj* | $0.2^{a,b} \pm 0.0$ | $1.2^{ab} \pm 0.0$ | $1.9^{c,d} \pm 0.0$ | $322.8^{f,g} \pm 6.7$ | $326.7^{g} \pm 6.7$ |
| 15 | *Takesrit* | $0.7^{f} \pm 0.1$ | $42.2^{j} \pm 1.6$ | $2.1^{d} \pm 0.2$ | $271.9^{e,f,g} \pm 8.4$ | $316.8^{c,d,e,f} \pm 10.4$ |
| 16 | *Bouchouk Sidi Aich* | $0.1^{a,b} \pm 0.0$ | $1.8^{a,b} \pm 0.2$ | $3.4^{j} \pm 0.2$ | $377.9^{h} \pm 3.6$ | $383.2^{h} \pm 3.2$ |
| 17 | *Azeradj* | $0.4^{c,d} \pm 0.1$ | $7.3^{e,f} \pm 0.0$ | $3.4^{j} \pm 0.0$ | $257.7^{c,d} \pm 0.2$ | $268.7^{b,c,d,e} \pm 0.2$ |
| 18 | *Tabelout* | $0.4^{c,d} \pm 0.0$ | $7.9^{e,f,g} \pm 0.3$ | $4.1^{k} \pm 0.1$ | $262.3^{c,d,e} \pm 3.3$ | $274.7^{c,d,e,f} \pm 3.6$ |
| 19 | *Sigoise* | $0.5^{e} \pm 0.0$ | $13.8^{h} \pm 0.1$ | $3.0^{h,i} \pm 0.0$ | $384.9^{h} \pm 3.7$ | $402.2^{h} \pm 3.9$ |
| 20 | *Chemlal* | $0.2^{a,b,c} \pm 0.0$ | $8.9^{f,g} \pm 0.4$ | $3.5^{j} \pm 0.1$ | $317.4^{f,g} \pm 6.2$ | $330.1^{g} \pm 6.7$ |
| 21 | *Aberkane* | $0.5^{d,e} \pm 0.0$ | $7.2^{e,f} \pm 0.2$ | $2.3^{d,e,f} \pm 0.1$ | $220.2^{a,b} \pm 4.2$ | $230.2^{a,b} \pm 4.5$ |
| 22 | *Variety X* | $0.3^{a,b,c} \pm 0.1$ | $9.6^{g} \pm 0.2$ | $3.3^{j} \pm 0.1$ | $416.8^{i} \pm 0.1$ | $430.1^{i} \pm 0.4$ |
| 23 | *Aimel* | $1.0^{g} \pm 0.1$ | $46.9^{k} \pm 0.1$ | $2.1^{d,e} \pm 0.2$ | $252.2^{d,e,f} \pm 3.2$ | $302.3^{b,c,d} \pm 3.5$ |
| 24 | *Souidi* | $0.1^{a} \pm 0.0$ | $3.2^{b,c,d} \pm 0.1$ | $5.5^{l} \pm 0.1$ | $573.1^{k} \pm 16.8$ | $581.8^{l} \pm 17.0$ |
| 25 | *Rougette de Mitidja* | $1.6^{i} \pm 0.0$ | $18.5^{i} \pm 0.3$ | $9.2^{m} \pm 0.1$ | $540.6^{k} \pm 4.0$ | $569.8^{k} \pm 4.4$ |
| 26 | *Azeradj* | $0.2^{a,b} \pm 0.1$ | $6.7^{e} \pm 0.0$ | $2.6^{g} \pm 0.1$ | $259.6^{c,d} \pm 3.6$ | $269.0^{c,d,e} \pm 3.7$ |

Means ± standard deviation (n = 3).
[a–m]Means the column followed by different letters are significantly different ($p < 0.05$).

compositions of the fatty acid were in the expected range of EVOOs (Manai-Djebali et al., 2012; Ramos-Escudero, Teresa Morales, & Asuero, 2015).

Palmitic acid (C16:0) was the most abundant saturated fatty acid in the investigated olive oils. Its content ranged between 10.6% and 18.5%, according to cultivars. The highest percentage was observed in *Chemlal* olive oil (sample 20) (18.5%), whereas the lowest levels were observed in *Azeradj* (sample 26) and *Limli* (sample 5) olive oils (10.6 and 10.8%, respectively).

### 3.4. Phenolic composition

Quantitative data for the phenolic content of the twenty-six samples are reported in Tables 5 and 6. Five main phenolic groups were

**Table 4**
Fatty acid compositions (%) of Algerian extra-virgin olive oils.

| Sample | Cultivar (cv.) | C16:0 | C16:1 | C18:0 | C18:1 | C18:2 | C18:3 | SFA | PUFA | MUFA | C18:1/C18:2 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | *Bouichret* | $13.6^{g} \pm 0.1$ | $0.9^{e} \pm 0.0$ | $3.1^{n} \pm 0.0$ | $68.7^{l} \pm 0.1$ | $12.9^{o} \pm 0.0$ | $0.8^{l} \pm 0.0$ | $16.7^{f} \pm 0.1$ | $13.7^{o} \pm 0.0$ | $69.7^{m} \pm 0.1$ | $5,3^{h} \pm 0,0$ |
| 2 | *Agrarez* | $12.7^{d} \pm 0.0$ | $1.2^{h} \pm 0.0$ | $3.1^{n} \pm 0.0$ | $73.4^{o} \pm 0.1$ | $8.9^{f} \pm 0.1$ | $0.7^{e} \pm 0.0$ | $15.8^{d} \pm 0.0$ | $9.6^{e} \pm 0.1$ | $74.7^{s} \pm 0.2$ | $8,2^{o} \pm 0,1$ |
| 3 | *Limli* | $17.5^{l} \pm 0.1$ | $2.0^{o} \pm 0.0$ | $2.9^{i} \pm 0.0$ | $65.4^{f} \pm 0.1$ | $11.4^{k} \pm 0.0$ | $0.7^{h} \pm 0.0$ | $20.4^{l} \pm 0.1$ | $12.1^{j} \pm 0.4$ | $67.5^{h} \pm 0.1$ | $5,8^{j} \pm 0,0$ |
| 4 | *Blanquette de Guelma* | $12.7^{d} \pm 0.1$ | $0.5^{a} \pm 0.0$ | $3.0^{j,k} \pm 0.0$ | $66.4^{g} \pm 0.1$ | $16.7^{u} \pm 0.0$ | $0.8^{i} \pm 0.0$ | $15.6^{d} \pm 0.1$ | $17.5^{u} \pm 0.0$ | $66.9^{g} \pm 0.1$ | $4,0^{d} \pm 0,0$ |
| 5 | *Limli* | $10.8^{b} \pm 0.1$ | $0.7^{b} \pm 0.0$ | $3.3^{p} \pm 0.0$ | $78.2^{s} \pm 0.1$ | $6.4^{b} \pm 0.0$ | $0.6^{d} \pm 0.0$ | $14.1^{b} \pm 0.1$ | $7.1^{b} \pm 0.0$ | $78.9^{v} \pm 0.1$ | $12,2^{q} \pm 0,1$ |
| 6 | *Tabelout* | $11.3^{c} \pm 0.1$ | $0.8^{c} \pm 0.0$ | $3.2^{o} \pm 0.0$ | $71.1^{m} \pm 0.0$ | $12.8^{n} \pm 0.0$ | $0.8^{j} \pm 0.0$ | $14.5^{c} \pm 0.1$ | $13.6^{n} \pm 0.0$ | $71.9^{q} \pm 0.1$ | $5,6^{i} \pm 0,0$ |
| 7 | *Takesrit* | $18.3^{n} \pm 0.0$ | $1.8^{l} \pm 0.0$ | $3.5^{r} \pm 0.0$ | $63.5^{d} \pm 0.0$ | $12.0^{l} \pm 0.0$ | $0.9^{o} \pm 0.0$ | $21.8^{o} \pm 0.0$ | $12.9^{l} \pm 0.0$ | $65.3^{d} \pm 0.0$ | $5,3^{h} \pm 0,0$ |
| 8 | *Bouchouk Guergour* | $14.4^{h} \pm 0.1$ | $0.9^{d} \pm 0.0$ | $2.6^{g} \pm 0.0$ | $66.9^{h} \pm 0.2$ | $14.3^{s} \pm 0.0$ | $0.9^{l} \pm 0.0$ | $17.0^{g} \pm 0.1$ | $15.2^{s} \pm 0.0$ | $67.8^{i} \pm 0.2$ | $4,7^{f} \pm 0,0$ |
| 9 | *Bouchouk* | $17.9^{m} \pm 0.1$ | $2.5^{r} \pm 0.0$ | $2.2^{e} \pm 0.0$ | $58.6^{a} \pm 0.1$ | $18.0^{x} \pm 0.0$ | $0.7^{g} \pm 0.0$ | $20.1^{k} \pm 0.1$ | $18.8^{x} \pm 0.0$ | $61.1^{a} \pm 0.1$ | $3,3^{a} \pm 0,0$ |
| 10 | *Chemlal* | $17.0^{k} \pm 0.1$ | $2.5^{r} \pm 0.0$ | $2.0^{d} \pm 0.0$ | $64.1^{e} \pm 0.1$ | $13.8^{r} \pm 0.0$ | $0.6^{a} \pm 0.0$ | $19.0^{i} \pm 0.2$ | $14.4^{r} \pm 0.1$ | $66.6^{f} \pm 0.1$ | $4,7^{f} \pm 0,0$ |
| 11 | *Chemlal* | $15.3^{j} \pm 0.1$ | $1.7^{k} \pm 0.0$ | $1.8^{a} \pm 0.0$ | $66.9^{h} \pm 0.0$ | $13.7^{q} \pm 0.1$ | $0.6^{a} \pm 0.0$ | $17.2^{g} \pm 0.1$ | $14.3^{q} \pm 0.1$ | $68.6^{l} \pm 0.0$ | $4,9^{g} \pm 0,0$ |
| 12 | *Zeletni* | $17.1^{k} \pm 0.1$ | $1.6^{j} \pm 0.0$ | $2.9^{i} \pm 0.0$ | $69.3^{j} \pm 0.1$ | $8.4^{e} \pm 0.0$ | $0.7^{g} \pm 0.0$ | $20.0^{j} \pm 0.1$ | $9.1^{e} \pm 0.0$ | $71.0^{o} \pm 0.1$ | $8,3^{o} \pm 0,0$ |
| 13 | *Almzeir* | $17.5^{l} \pm 0.0$ | $1.9^{n} \pm 0.0$ | $3.0^{l,m} \pm 0.0$ | $63.6^{d} \pm 0.0$ | $13.3^{p} \pm 0.0$ | $0.7^{f,g} \pm 0.0$ | $20.5^{l} \pm 0.1$ | $14.0^{p} \pm 0.0$ | $65.5^{e} \pm 0.0$ | $4,8^{g} \pm 0,1$ |
| 14 | *Azeradj* | $13.3^{f} \pm 0.1$ | $0.8^{c} \pm 0.0$ | $3.7^{s} \pm 0.0$ | $71.7^{n} \pm 0.1$ | $9.8^{h} \pm 0.0$ | $0.8^{h} \pm 0.0$ | $17.0^{g} \pm 0.1$ | $10.5^{g} \pm 0.0$ | $72.5^{r} \pm 0.1$ | $7,4^{m} \pm 0,0$ |
| 15 | *Takesrit* | $17.9^{m} \pm 0.1$ | $1.9^{m} \pm 0.0$ | $3.0^{k,l} \pm 0.0$ | $66.4^{g} \pm 0.1$ | $10.3^{i} \pm 0.0$ | $0.6^{b} \pm 0.0$ | $20.8^{m,n} \pm 0.1$ | $10.9^{h} \pm 0.0$ | $68.2^{k} \pm 0.2$ | $6,4^{k} \pm 0,0$ |
| 16 | *Bouchouk Sidi Aich* | $15.0^{i} \pm 0.1$ | $1.3^{i} \pm 0.0$ | $3.0^{j} \pm 0.0$ | $70.6^{l} \pm 0.2$ | $9.1^{g} \pm 0.0$ | $1.1^{p} \pm 0.0$ | $17.9^{h} \pm 0.2$ | $10.2^{f} \pm 0.0$ | $71.9^{q} \pm 0.2$ | $7,8^{n} \pm 0,0$ |
| 17 | *Azeradj* | $13.0^{e} \pm 0.1$ | $1.1^{g} \pm 0.0$ | $2.7^{h} \pm 0.0$ | $70.1^{k} \pm 0.1$ | $12.2^{m} \pm 0.0$ | $0.9^{n} \pm 0.0$ | $15.7^{d} \pm 0.1$ | $13.1^{m} \pm 0.0$ | $71.2^{p} \pm 0.1$ | $5,8^{j} \pm 0,0$ |
| 18 | *Tabelout* | $17.8^{m} \pm 0.1$ | $2.1^{p} \pm 0.0$ | $2.5^{f} \pm 0.0$ | $68.6^{i} \pm 0.1$ | $8.3^{d} \pm 0.0$ | $0.7^{g} \pm 0.0$ | $20.3^{k,l} \pm 0.1$ | $9.0^{d} \pm 0.0$ | $70.8^{n} \pm 0.1$ | $8,3^{o} \pm 0,0$ |
| 19 | *Sigoise* | $15.2^{j} \pm 0.1$ | $2.7^{s} \pm 0.0$ | $1.9^{b} \pm 0.0$ | $62.0^{c} \pm 0.1$ | $17.3^{w} \pm 0.0$ | $0.9^{m} \pm 0.0$ | $17.1^{g} \pm 0.1$ | $18.2^{w} \pm 0.0$ | $64.7^{c} \pm 0.1$ | $3,6^{c} \pm 0,0$ |
| 20 | *Chemlal* | $18.5^{o} \pm 0.1$ | $2.2^{q} \pm 0.0$ | $1.9^{c} \pm 0.0$ | $61.8^{c} \pm 0.1$ | $14.7^{t} \pm 0.0$ | $0.9^{l} \pm 0.0$ | $20.4^{l} \pm 0.12$ | $15.6^{t} \pm 0.1$ | $64.0^{b} \pm 0.1$ | $4,2^{e} \pm 0,0$ |
| 21 | *Aberkane* | $11.2^{c} \pm 0.0$ | $0.8^{d} \pm 0.0$ | $2.9^{i} \pm 0.0$ | $77.3^{q} \pm 0.0$ | $7.1^{c} \pm 0.0$ | $0.7^{f} \pm 0.0$ | $14.1^{b} \pm 0.0$ | $7.7^{c} \pm 0.1$ | $78.2^{u} \pm 0.0$ | $11,0^{p} \pm 0,1$ |
| 22 | *Variety X* | $17.4^{l} \pm 0.1$ | $2.0^{n} \pm 0.0$ | $3.0^{m} \pm 0.0$ | $63.7^{d} \pm 0.1$ | $13.3^{p} \pm 0.0$ | $0.7^{g} \pm 0.0$ | $20.5^{l} \pm 0.1$ | $14.0^{o} \pm 0.0$ | $65.6^{e} \pm 0.1$ | $4,9^{g} \pm 0,0$ |
| 23 | *Aimel* | $17.5^{l} \pm 0.1$ | $1.7^{k} \pm 0.0$ | $3.5^{q} \pm 0.0$ | $66.3^{g} \pm 0.0$ | $10.4^{i} \pm 0.08$ | $0.6^{c} \pm 0.0$ | $21.0^{n} \pm 0.1$ | $11.0^{i} \pm 0.1$ | $68.0^{j} \pm 0.1$ | $6,4^{k} \pm 0,1$ |
| 24 | *Souidi* | $17.8^{m} \pm 0.1$ | $2.0^{n} \pm 0.1$ | $2.9^{i} \pm 0.0$ | $59.2^{b} \pm 0.2$ | $17.2^{v} \pm 0.03$ | $0.9^{o} \pm 0.0$ | $20.7^{m} \pm 0.2$ | $18.2^{v} \pm 0.0$ | $61.1^{a} \pm 0.3$ | $3,4^{b} \pm 0,0$ |
| 25 | *Rougette de Mitidja* | $13.5^{f,g} \pm 0.0$ | $1.6^{j} \pm 0.0$ | $3.0^{k} \pm 0.0$ | $77.6^{r} \pm 0.0$ | $3.5^{a} \pm 0.04$ | $0.9^{k} \pm 0.0$ | $16.4^{e} \pm 0.04$ | $4.4^{a} \pm 0.0$ | $79.2^{w} \pm 0.0$ | $22,2^{r} \pm 0,2$ |
| 26 | *Azeradj* | $10.6^{a} \pm 0.1$ | $1.0^{f} \pm 0.0$ | $2.2^{e} \pm 0.0$ | $74.0^{p} \pm 0.2$ | $11.0^{j} \pm 0.01$ | $1.2^{q} \pm 0.0$ | $12.8^{a} \pm 0.2$ | $12.2^{k} \pm 0.0$ | $75.0^{t} \pm 0.2$ | $6,7^{l} \pm 0,0$ |
| | EVOO(EEC.2003) | 7.5–20.0 | 0.3–3.5 | 0.5–5.0 | 55.0–83.0 | 3.5–21.0 | ≤1.0 | | | | |

Means ± standard deviation (n = 3).
Fatty acid nomenclature: 16:0 (palmitic acid); 16:1 (palmitoleic acid); 18:0 (stearic acid); 18:1 (oleic acid); 18:2 (linoleic acid); 18:3 (linolenic acid).
SFA: saturated fatty acids; MUFA: monounsaturated fatty acids; PUFA: polyunsaturated fatty acids.
[a–x]Means the column followed by different letters are significantly different ($p < 0.05$).

**Table 5**
Phenolic compound composition (mg $kg^{-1}$) of Algerian extra-virgin olive oils.

| Sample | Cultivar (cv.) | Total phenolic compounds | Oleuropein | Oleuropein derivatives | Ligstroside derivatives | Secoiridoid acids | Decarboxymethylelenolic acid | Elenolic acid | Oleocanthal | Lignans |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | *Bouichret* | $236.9^{g} \pm 1.5$ | $0.0^{a} \pm 0.0$ | $76.7^{h} \pm 0.4$ | $55.9^{e} \pm 1.8$ | $80.6^{i} \pm 2.5$ | $9.7^{h} \pm 0.1$ | $71.0^{j} \pm 2.6$ | $22.6^{g} \pm 0.8$ | $56.4^{o} \pm 0.2$ |
| 2 | *Agrarez* | $169.9^{d} \pm 1.2$ | $0.0^{a} \pm 0.0$ | $41.9^{d} \pm 0.3$ | $83.9^{j} \pm 0.4$ | $131.8^{n} \pm 1.4$ | $29.0^{n} \pm 0.2$ | $102.9^{o} \pm 1.2$ | $45.6^{n} \pm 0.2$ | $28.8^{g} \pm 0.4$ |
| 3 | *Limli* | $273.0^{l} \pm 0.2$ | $0.0^{a} \pm 0.0$ | $117.7^{m} \pm 0.4$ | $72.7^{g} \pm 0.1$ | $90.4^{k} \pm 0.5$ | $7.0^{e,f} \pm 0.0$ | $83.3^{m} \pm 0.6$ | $22.1^{g} \pm 0.0$ | $36.8^{i} \pm 0.1$ |
| 4 | *Blanquette de Guelma* | $1006.7^{t} \pm 3.0$ | $0.0^{a} \pm 0.0$ | $732.4^{v} \pm 1.3$ | $195.2^{o} \pm 1.2$ | $344.6^{t} \pm 0.5$ | $92.0^{p} \pm 0.0$ | $252.6^{u} \pm 0.66$ | $40.3^{l} \pm 0.3$ | $26.2^{f} \pm 0.2$ |
| 5 | *Limli* | $264.3^{k} \pm 1.5$ | $0.0^{a} \pm 0.0$ | $104.0^{k} \pm 0.8$ | $83.0^{i} \pm 0.9$ | $146.6^{p} \pm 0.7$ | $39.5^{o} \pm 0.5$ | $107.1^{p} \pm 1.2$ | $46.9^{o} \pm 0.6$ | $40.7^{k} \pm 0.3$ |
| 6 | *Tabelout* | $181.0^{e} \pm 0.6$ | $0.0^{a} \pm 0.0$ | $75.9^{h} \pm 04$ | $66.8^{f} \pm 0.0$ | $23.8^{a} \pm 0.1$ | $0.7^{a} \pm 0.0$ | $23.1^{b} \pm 0.1$ | $16.3^{e} \pm 0.0$ | $26.9^{f} \pm 0.1$ |
| 7 | *Takesrit* | $259.8^{j} \pm 0.2$ | $0.0^{a} \pm 0.0$ | $83.8^{i} \pm 0.3$ | $115.4^{j} \pm 0.1$ | $172.4^{r} \pm 0.7$ | $7.5^{f,g} \pm 0.1$ | $164.9^{s} \pm 0.8$ | $44.5^{m} \pm 0.1$ | $32.7^{h} \pm 0.1$ |
| 8 | *Bouchouk Guergour* | $338.2^{m} \pm 1.8$ | $0.3^{b} \pm 0.0$ | $78.0^{h} \pm 0.7$ | $158.5^{m} \pm 0.9$ | $71.6^{g} \pm 0.5$ | $17.9^{l} \pm 0.2$ | $53.7^{e} \pm 0.4$ | $135.6^{u} \pm 0.7$ | $38.9^{j} \pm 0.1$ |
| 9 | *Bouchouk* | $491.1^{o} \pm 1.3$ | $0.0^{a} \pm 0.0$ | $253.4^{q} \pm 0.9$ | $131.5^{k} \pm 0.2$ | $72.9^{g} \pm 0.4$ | $0.9^{a,b} \pm 0.0$ | $72.0^{j} \pm 0.4$ | $48.5^{p} \pm 0.4$ | $56.2^{o} \pm 0.1$ |
| 10 | *Chemlal* | $537.2^{p} \pm 2.6$ | $0.0^{a} \pm 0.0$ | $211.2^{p} \pm 0.6$ | $204.8^{p} \pm 0.4$ | $178.2^{s} \pm 1.5$ | $12.0^{j} \pm 0.1$ | $166.2^{t} \pm 1.4$ | $71.3^{q} \pm 0.4$ | $70.7^{p} \pm 0.5$ |
| 11 | *Chemlal* | $223.3^{f} \pm 2.2$ | $0.2^{b} \pm 0.0$ | $73.2^{g} \pm 0.3$ | $65.5^{f} \pm 3.3$ | $61.4^{e} \pm 0.4$ | $4.0^{c} \pm 0.0$ | $57.4^{f} \pm 0.3$ | $22.1^{g} \pm 0.0$ | $36.4^{i} \pm 0.3$ |
| 12 | *Zeletni* | $224.1^{f} \pm 0.1$ | $0.0^{a} \pm 0.0$ | $110.4^{l} \pm 0.2$ | $61.8^{f} \pm 0.3$ | $137.4^{o} \pm 0.2$ | $5.2^{d} \pm 0.0$ | $132.2^{q} \pm 0.2$ | $33.4^{j} \pm 0.1$ | $20.2^{e} \pm 0.0$ |
| 13 | *Almzeir* | $251.1^{i} \pm 0.9$ | $0.0^{a} \pm 0.0$ | $100.5^{j} \pm 0.2$ | $77.7^{g,h} \pm 0.5$ | $66.3^{f} \pm 0.4$ | $3.8^{c} \pm 0.0$ | $62.4^{h} \pm 0.4$ | $32.1^{i} \pm 0.3$ | $32.2^{h} \pm 0.0$ |
| 14 | *Azeradj* | $337.6^{m} \pm 0.1$ | $0.0^{a} \pm 0.0$ | $142.7^{n} \pm 0.3$ | $113.5^{j} \pm 0.1$ | $100.7^{m} \pm 0.3$ | $12.5^{j} \pm 0.0$ | $88.3^{n} \pm 0.3$ | $71.0^{q} \pm 0.0$ | $43.5^{l} \pm 0.3$ |
| 15 | *Takesrit* | $598.0^{q} \pm 2.6$ | $0.0^{a} \pm 0.0$ | $343.3^{s} \pm 1.7$ | $150.6^{l} \pm 1.0$ | $154.3^{q} \pm 0.4$ | $1.1^{a,b} \pm 0.0$ | $153.2^{r} \pm 0.4$ | $39.3^{k} \pm 0.2$ | $52.3^{n} \pm 0.3$ |
| 16 | *Bouchouk Sidi Aich* | $169.0^{d} \pm 2.3$ | $1.2^{e} \pm 0.1$ | $65.1^{f} \pm 0.6$ | $44.5^{d} \pm 0.7$ | $33.4^{b} \pm 0.2$ | $3.4^{c} \pm 0.0$ | $30.0^{c} \pm 0.2$ | $26.6^{h} \pm 0.2$ | $16.6^{d} \pm 0.2$ |
| 17 | *Azeradj* | $172.0^{d} \pm 1.1$ | $0.0^{a} \pm 0.0$ | $83.9^{i} \pm 0.2$ | $34.9^{c} \pm 0.6$ | $38.5^{c} \pm 0.2$ | $16.1^{k} \pm 0.1$ | $22.4^{b} \pm 0.1$ | $19.5^{f} \pm 0.1$ | $15.6^{c} \pm 0.1$ |
| 18 | *Tabelout* | $782.3^{r} \pm 5.5$ | $0.6^{c} \pm 0.3$ | $458.9^{t} \pm 3.5$ | $239.8^{q} \pm 15.3$ | $80.5^{i} \pm 0.9$ | $1.8^{b} \pm 0.0$ | $78.7^{l} \pm 0.9$ | $144.3^{v} \pm 0.8$ | $46.6^{m} \pm 1.2$ |
| 19 | *Sigoise* | $55.0^{a} \pm 0.1$ | $0.0^{a} \pm 0.0$ | $12.3^{a} \pm 0.1$ | $16.7^{a} \pm 0.1$ | $59.1^{d} \pm 0.3$ | $7.6^{f,g} \pm 0.0$ | $51.5^{d} \pm 0.2$ | $2.8^{a} \pm 0.0$ | $9.6^{a} \pm 0.0$ |
| 20 | *Chemlal* | $172.1^{d} \pm 0.4$ | $0.3^{b} \pm 0.1$ | $39.6^{c} \pm 0.7$ | $37.4^{c} \pm 0.4$ | $76.1^{h} \pm 0.1$ | $10.5^{j} \pm 0.1$ | $65.6^{i} \pm 0.1$ | $14.1^{d} \pm 0.4$ | $36.4^{i} \pm 0.0$ |
| 21 | *Aberkane* | $337.9^{m} \pm 3.4$ | $0.0^{a} \pm 0.0$ | $198.4^{o} \pm 2.1$ | $41.0^{c,d} \pm 0.2$ | $85.7^{j} \pm 0.5$ | $8.3^{g} \pm 0.0$ | $77.5^{k,l} \pm 0.5$ | $16.9^{e} \pm 0.8$ | $21.0^{e} \pm 0.7$ |
| 22 | *Variety X* | $803.4^{s} \pm 1.5$ | $0.0^{a} \pm 0.0$ | $583.4^{u} \pm 1.4$ | $185.4^{n} \pm 0.7$ | $94.9^{l} \pm 0.4$ | $18.6^{m} \pm 0.0$ | $76.3^{k} \pm 0.3$ | $117.9^{t} \pm 0.3$ | $20.6^{e} \pm 0.4$ |
| 23 | *Aimel* | $141.9^{c} \pm 0.5$ | $1.0^{d} \pm 0.2$ | $60.9^{e} \pm 0.3$ | $50.7^{e} \pm 2.9$ | $79.9^{j} \pm 0.2$ | $9.1^{h} \pm 0.0$ | $70.8^{j} \pm 0.2$ | $10.4^{c} \pm 0.1$ | $10.7^{b} \pm 2.5$ |
| 24 | *Souidi* | $468.9^{n} \pm 1.6$ | $0.0^{a} \pm 0.0$ | $265.6^{r} \pm 1.3$ | $134.0^{k} \pm 0.0$ | $65.0^{f} \pm 0.6$ | $1.4^{a,b} \pm 0.0$ | $63.6^{h} \pm 0.6$ | $91.4^{r} \pm 0.2$ | $43.2^{l} \pm 0.2$ |
| 25 | *Rougette de Mitidja* | $244.7^{h} \pm 0.3$ | $0.0^{a} \pm 0.0$ | $12.8^{a} \pm 0.8$ | $136.8^{k} \pm 0.3$ | $177.3^{s} \pm 3.0$ | $116.8^{q} \pm 2.2$ | $60.5^{g} \pm 0.8$ | $99.0^{s} \pm 0.5$ | $51.3^{n} \pm 0.0$ |
| 26 | *Azeradj* | $76.9^{b} \pm 0.4$ | $0.0^{a} \pm 0.0$ | $18.0^{b} \pm 0.4$ | $23.7^{b} \pm 0.3$ | $23.9^{a} \pm 0.1$ | $6.6^{e} \pm 0.1$ | $17.3^{a} \pm 0.0$ | $4.1^{b} \pm 0.1$ | $8.8^{a} \pm 0.1$ |

Means ± standard deviation (n = 3).
[a–v]Means the column followed by different letters are significantly different ($p < 0.05$).

detected: phenolic alcohols (hydroxytyrosol and tyrosol), secoiridoids (mainly derivatives of oleuropein and ligstroside and elenolic acid), lignans, flavonoids (luteolin and apigenin), and phenolic acids. The total phenolic content (oleuropein derivatives, ligstroside derivatives, flavonoids, phenolic acids and biophenols oxidized) showed large variation, ranging from 55.0 mg $kg^{-1}$ for *Sigoise* variety (sample 19) to 1006.7 mg $kg^{-1}$ for *Blanquette de Guelma* (sample 4). Significant differences ($p < 0.05$) among the different varieties were observed. Our data are in agreement with the values reported in the literature, where the total phenolic content of olive oils varies from 50 to 1000 mg $kg^{-1}$ (Montedoro et al., 1992).

**Phenolic alcohols** amount varied between 1.7 mg $kg^{-1}$ (*Sigoise*) and 300.6 mg $kg^{-1}$ (*Blanquette de Guelma*). In this class, hydroxytyrosol and tyrosol were among the most representative compounds, and, in

**Table 6**
Phenolic compound composition (mg $kg^{-1}$) of Algerian extra-virgin olive oils.

| Sample | Cultivar (cv.) | Phenolic alcohols | Hydroxytyrosol | Tyrosol | Flavonoids | Luteolin | Apigenin | Phenolic acids | Biophenols oxidized |
|---|---|---|---|---|---|---|---|---|---|
| 1 | *Bouichret* | $4.9^{c,d} \pm 0.2$ | $1.7^{c} \pm 0.1$ | $3.2^{c} \pm 0.1$ | $21.8^{k} \pm 1.2$ | $16.6^{n} \pm 0.5$ | $5.2^{j} \pm 0.7$ | $14.7^{p} \pm 0.2$ | $11.4^{e} \pm 0.4$ |
| 2 | *Agrarez* | $5.2^{d} \pm 0.0$ | $0.7^{a,b} \pm 0.0$ | $4.4^{f} \pm 0.0$ | $6.4^{e} \pm 0.2$ | $4.6^{d} \pm 0.1$ | $1.8^{d} \pm 0.1$ | $3.5^{f} \pm 0.0$ | $5.5^{a} \pm 0.0$ |
| 3 | *Limli* | $20.8^{m} \pm 0.6$ | $4.0^{e,f} \pm 0.6$ | $16.9^{p} \pm 0.0$ | $22.9^{l} \pm 0.1$ | $17.1^{n} \pm 0.1$ | $5.8^{k} \pm 0.0$ | $5.5^{i} \pm 0.0$ | $17.5^{i} \pm 0.1$ |
| 4 | *Blanquette de Guelma* | $300.6^{q} \pm 1.5$ | $252.4^{n} \pm 1.7$ | $48.3^{r} \pm 0.2$ | $22.6^{l} \pm 0.5$ | $15.5^{m} \pm 0.4$ | $7.1^{m} \pm 0.1$ | $15.3^{q} \pm 0.0$ | $15.0^{g} \pm 0.2$ |
| 5 | *Limli* | $6.6^{f} \pm 0.2$ | $3.5^{e} \pm 0.2$ | $3.1^{c} \pm 0.0$ | $14.8^{i} \pm 0.2$ | $10.2^{i} \pm 0.1$ | $4.6^{i} \pm 0.1$ | $12.5^{n} \pm 0.0$ | $9.3^{c} \pm 0.1$ |
| 6 | *Tabelout* | $8.3^{g} \pm 0.2$ | $3.9^{e,f} \pm 0.2$ | $4.4^{f} \pm 0.0$ | $0.5^{a} \pm 0.0$ | $0.4^{a} \pm 0.0$ | $0.1^{a} \pm 0.0$ | $1.5^{a} \pm 0.0$ | $9.3^{c} \pm 0.1$ |
| 7 | *Takesrit* | $11.4^{i} \pm 0.0$ | $1.0^{a,b,c} \pm 0.0$ | $10.4^{n} \pm 0.0$ | $5.4^{d} \pm 0.0$ | $4.2^{d} \pm 0.1$ | $1.2^{c} \pm 0.0$ | $4.4^{g} \pm 0.1$ | $18.0^{j} \pm 0.2$ |
| 8 | *Bouchouk Guergour* | $6.0^{e} \pm 0.1$ | $1.5^{b,c} \pm 0.1$ | $4.5^{f,g} \pm 0.0$ | $15.4^{i} \pm 0.1$ | $12.7^{j} \pm 0.1$ | $2.8^{e} \pm 0.0$ | $32.5^{s} \pm 0.2$ | $14.9^{g} \pm 0.2$ |
| 9 | *Bouchouk* | $13.4^{k} \pm 0.1$ | $8.8^{j} \pm 0.1$ | $4.6^{f,g} \pm 0.0$ | $13.5^{h} \pm 0.1$ | $8.0^{f} \pm 0.0$ | $5.5^{j,k} \pm 0.0$ | $2.0^{b,c} \pm 0.1$ | $34.6^{o} \pm 0.2$ |
| 10 | *Chemlal* | $13.0^{j,k} \pm 0.0$ | $5.1^{g,h} \pm 0.0$ | $7.9^{j} \pm 0.0$ | $19.1^{j} \pm 0.8$ | $12.5^{j} \pm 0.4$ | $6.6^{l} \pm 0.4$ | $2.5^{e} \pm 0.2$ | $29.0^{m} \pm 0.0$ |
| 11 | *Chemlal* | $7.1^{f} \pm 0.1$ | $2.9^{d} \pm 0.0$ | $4.2^{e} \pm 0.0$ | $14.7^{i} \pm 0.1$ | $9.9^{h,l} \pm 0.1$ | $4.7^{i} \pm 0.1$ | $3.4^{f} \pm 0.0$ | $30.2^{n} \pm 0.4$ |
| 12 | *Zeletni* | $9.6^{h} \pm 0.2$ | $4.9^{g,h} \pm 0.2$ | $4.6^{f,g} \pm 0.0$ | $12.2^{g} \pm 0.1$ | $8.7^{g} \pm 0.1$ | $3.5^{f} \pm 0.0$ | $4.9^{h} \pm 0.1$ | $14.5^{g} \pm 0.1$ |
| 13 | *Almzeir* | $12.9^{j,k} \pm 0.0$ | $4.4^{f,g} \pm 0.0$ | $8.5^{l} \pm 0.0$ | $18.5^{j} \pm 0.1$ | $14.4^{l} \pm 0.0$ | $4.2^{h} \pm 0.1$ | $2.6^{e} \pm 0.0$ | $19.6^{k} \pm 0.1$ |
| 14 | *Azeradj* | $8.9^{g} \pm 0.0$ | $5.3^{h} \pm 0.0$ | $3.6^{d} \pm 0.0$ | $13.0^{h} \pm 0.0$ | $10.2^{i} \pm 0.0$ | $2.8^{e} \pm 0.0$ | $14.5^{o} \pm 0.0$ | $10.4^{d} \pm 0.0$ |
| 15 | *Takesrit* | $29.5^{p} \pm 0.4$ | $13.8^{l} \pm 0.2$ | $15.8^{o} \pm 0.1$ | $26.4^{o} \pm 0.3$ | $16.9^{n} \pm 0.3$ | $9.5^{n} \pm 0.0$ | $3.5^{f} \pm 0.1$ | $21.6^{l} \pm 0.4$ |
| 16 | *Bouchouk Sidi Aich* | $4.4^{c} \pm 0.1$ | $0.8^{a,b} \pm 0.0$ | $3.6^{d} \pm 0.0$ | $13.5^{h} \pm 0.3$ | $9.6^{h} \pm 0.3$ | $4.0^{g,h} \pm 0.1$ | $17.4^{r} \pm 0.1$ | $11.9^{e} \pm 0.5$ |
| 17 | *Azeradj* | $14.2^{l} \pm 0.0$ | $10.1^{k} \pm 0.1$ | $4.1^{e} \pm 0.0$ | $23.8^{m} \pm 0.0$ | $18.4^{o} \pm 0.3$ | $5.4^{j,k} \pm 0.3$ | $7.2^{k} \pm 0.0$ | $6.7^{b} \pm 0.1$ |
| 18 | *Tabelout* | $23.2^{n} \pm 0.1$ | $13.6^{l} \pm 0.1$ | $9.6^{m} \pm 0.0$ | $25.1^{n} \pm 0.0$ | $13.5^{k} \pm 0.1$ | $11.6^{p} \pm 0.1$ | $8.7^{m} \pm 0.0$ | $18.3^{j} \pm 0.6$ |
| 19 | *Sigoise* | $1.7^{a} \pm 0.0$ | $0.5^{a} \pm 0.0$ | $1.2^{a} \pm 0.0$ | $7.6^{f} \pm 0.1$ | $5.6^{e} \pm 0.1$ | $2.0^{d} \pm 0.0$ | $2.3^{d} \pm 0.1$ | $6.6^{b} \pm 0.0$ |
| 20 | *Chemlal* | $7.1^{f} \pm 0.1$ | $2.3^{d} \pm 0.1$ | $4.8^{h} \pm 0.1$ | $18.4^{j} \pm 0.0$ | $12.8^{j} \pm 0.1$ | $5.6^{k} \pm 0.1$ | $2.1^{c,d} \pm 0.0$ | $37.8^{p} \pm 0.2$ |
| 21 | *Aberkane* | $12.8^{j,k} \pm 0.3$ | $9.7^{k} \pm 0.2$ | $3.2^{c} \pm 0.0$ | $51.5^{p} \pm 1.1$ | $41.4^{p} \pm 1.0$ | $10.1^{o} \pm 0.1$ | $7.8^{l} \pm 0.1$ | $18.1^{j} \pm 0.4$ |
| 22 | *Variety X* | $22.8^{n} \pm 0.0$ | $17.5^{m} \pm 0.0$ | $5.3^{i} \pm 0.0$ | $2.0^{b} \pm 0.0$ | $1.8^{b} \pm 0.1$ | $0.2^{a} \pm 0.0$ | $1.9^{b} \pm 0.1$ | $10.0^{d} \pm 0.2$ |
| 23 | *Aimel* | $12.5^{j} \pm 0.2$ | $4.4^{f,g} \pm 0.2$ | $8.1^{k} \pm 0.1$ | $4.2^{c} \pm 0.1$ | $3.0^{c} \pm 0.1$ | $1.2^{b,c} \pm 0.0$ | $2.6^{e} \pm 0.1$ | $12.8^{f} \pm 0.2$ |
| 24 | *Souidi* | $10.8^{i} \pm 0.1$ | $6.2^{i} \pm 0.1$ | $4.7^{g} \pm 0.0$ | $11.7^{g} \pm 0.5$ | $8.0^{f} \pm 0.5$ | $3.7^{f,g} \pm 0.0$ | $4.3^{g} \pm 0.0$ | $10.1^{d} \pm 0.1$ |
| 25 | *Rougette de Mitidja* | $26.6^{o} \pm 0.3$ | $1.4^{b,c} \pm 0.1$ | $25.2^{q} \pm 0.2$ | $7.8^{f} \pm 0.2$ | $4.4^{d} \pm 0.3$ | $3.5^{f} \pm 0.2$ | $17.5^{r} \pm 0.2$ | $18.5^{j} \pm 0.7$ |
| 26 | *Azeradj* | $3.3^{b} \pm 0.2$ | $0.4^{a} \pm 0.1$ | $2.9^{b} \pm 0.1$ | $5.1^{d} \pm 0.3$ | $4.2^{d} \pm 0.2$ | $0.9^{b} \pm 0.1$ | $5.8^{j} \pm 0.2$ | $15.6^{h} \pm 0.8$ |

Means ± standard deviation (n = 3).
[a–v]Means the column followed by different letters are significantly different ($p < 0.05$).

agreement with the literature (Cerretani et al., 2006; Romero, Saavedra, Tapia, Sepúlveda, & Aparicio, 2015), they were found at levels below 17.5 and 25.2 mg $kg^{-1}$, respectively. However, only the variety *Blanquette de Guelma* provided higher levels of these phenols (252.4 and 48.3 mg $kg^{-1}$ for hydroxytyrosol and tyrosol, respectively). Such high values are not unique for olive oils, in fact Gilbert-López et al. (2014) found values of hydroxytyrosol concentration above 50 mg $kg^{-1}$ in 78 samples of extra virgin olive oil collected from ten different countries, and detected higher levels of both phenols in *Arauco* VOO. Interestingly, it should be noted that no Algerian olive oils have showed such high values before.

**Secoiridoid derivatives**, mainly represented by oleuropein and ligstroside derivatives, were by far the most abundant group of phenolic compounds in all the analyzed samples regardless of geographic origin and variety. As reported by different authors, these compounds are widely associated with the sensory properties of EVOO as they are the main contributors to its bitterness and pungency (Bendini et al., 2007; Lauri et al., 2013). In our case, the highest concentrations of oleuropein derivatives were observed in the olive oil from *Blanquette de Guelma* (sample 4) (732.4 mg $kg^{-1}$), whereas the lowest concentrations were detected in *Sigoise* (sample 19) and *Rougette de Mitidja* (sample 25) olive oils (12.3 and 12.8 mg $kg^{-1}$ respectively). As for the ligstroside derivatives, extra-virgin olive oil from *Sigoise* (sample 19) showed the lowest value (16.7 mg $kg^{-1}$), while the highest value was observed in *Tabelout* (sample 18) (239.8 mg $kg^{-1}$). As far as the oleuropein is concerned, it was detected only in olive oil from *Takesrit* (sample 7), *Bouchouk Guergour* (sample 8), *Chemlal* (sample 11), *Bouchouk Sidi Aich* (sample 16), *Tabelout* (sample 18), *Chemlal* (sample 20) and *Aimel* (sample 23). In any case, oleuropein's content, as expected, did not exceed 1.2 mg $kg^{-1}$ (Krichene et al., 2009; Ballus et al., 2014).

Similarly, the content of elenolic acid (a secoiridoid acid) ranged from 17.3 (*Azeradj* — sample 26) to 252.6 mg $kg^{-1}$ (*Blanquette de Guelma* — sample 4). Moreover, another secoiridoid acid, namely decarboxymethylelenolic acid, ranged from 0.8 (*Tabelout* — sample 6) to 116.8 mg $kg^{-1}$ (*Rougette de Mitidja* — sample 25). These data are in agreement with those reported in literature (Pinelli, Galardi, Mulinacci, & Romani, 2003).

Newly pressed EVOOs contain also (−)-decarboxymethyl ligstroside aglycone, also known as oleocanthal. Oleocanthal has been shown to mimic the pharmacology of ibuprofen (Cicerale, Lucas, & Keast, 2012). Oleocanthal concentration in EVOO is highly variable, ranging from 0.2 mg $kg^{-1}$ to 498 mg $kg^{-1}$ (Gómez-Rico, Salvador, La Greca, & Fregapane, 2006). Beauchamp et al. (2005) demonstrated that EVOOs produced in different countries had variable oleocanthal concentrations. For instance, EVOO produced in the U.S.A. contained a low concentration of oleocanthal (22.6 ± 0.6 mg $kg^{-1}$), while EVOOs produced in Italy contained very high quantities of this compound (up to 191.8 ± 2.7 mg $kg^{-1}$). All the twenty-six studied EVOOs contained oleocanthal, which was found in considerable amount in *Tabelout* (sample 18), *Bouchouk Guergour* (sample 8) and *Variety X* (sample 22) oils (144.3, 135.6 and 117.9 mg $kg^{-1}$, respectively). *Sigoise* (sample 19) variety was characterized by the lowest value (2.8 mg $kg^{-1}$).

**Lignans** represent an important group of phenolic compounds that characterize EVOOs. In our case, they varied between 70.7 mg $kg^{-1}$ in *Chemlal* (sample 10) and 8.8 mg $kg^{-1}$ in *Azeradj* (sample 26) variety. However, these compounds are not indicative of the variety of the cultivar, but rather of the process used to obtain the olive oil. In fact, as observed by Owen et al. (2000), these compounds are the main components of the phenolic fraction of the olive seed and are practically absent from the pulp, leaves, and limbs, and therefore their presence in the oil must be due to the breaking of the pits when the olives are crushed. These could be used as an index of the crushing conditions and of the fruit pulp/seed ratio during olive production.

Regarding the **flavonoids**, they were found within the range of 0.5–51.5 mg $kg^{-1}$ detected in *Tabelout* (sample 6) and *Aberkane* (sample 21) samples, respectively. Luteolin and apigenin were the most relevant compounds within this group. Luteolin, the most abundant flavonoid occurring in the analyzed samples, ranged from 0.4 mg $kg^{-1}$ in *Tabelout* (sample 6) to 41.4 mg $kg^{-1}$ in *Aberkane* (sample 21), while apigenin concentration ranged between 0.1 and 11.6 mg $kg^{-1}$ in two different samples of *Tabelout* (samples 6 and 18, cultivated in Amizour and Boukhlifa, respectively). These data suggest the importance of environmental factors on the content of these compounds. Interestingly, the levels of luteolin were higher than in all the others Mediterranean cultivars analyzed so far (Pinelli et al., 2003; Abaza et al., 2005; García-Villalba et al., 2010), therefore this could be potentially considered as a discriminative characteristic of Algerian EVOOs produced in the Béjaïa province.

**Phenolic acids.** Phenolic acids have already been associated with color and sensory qualities, as well as with the health-related antioxidant properties of foods (Cartoni, Coccioli, Jasionowska, & Ramires, 2000). All the twenty-six Algerian EVOOs showed low concentration of this class of compounds, as the content ranged from 1.5 to 17.5 mg $kg^{-1}$, with the only exception of the sample 8 (*Bouchouk Guergour*), that had almost 2-fold levels of phenolic acids (32.5 mg $kg^{-1}$). Interestingly, as for the oxidized biophenols, they were detected in all the twenty-six samples. Recently, the interest in oxidized form of olive oil phenols has significantly increased, especially in relation to determination of freshness/aging status (Rovellini et al., 1997; Tovar, Motilva, & Paz Romero, 2001; Owen et al., 2000). In this context, it is interesting to note that the highest amount of oxidized biophenols was found in *Chemlal* variety (sample 20) (37.8 mg $kg^{-1}$).

### 3.5. NMR spectroscopy

The twenty-six EVOO samples were also investigated by $^1$H NMR. In particular, the experiments were performed both on the intact EVOOs (Fig. 2) and the phenolic extracts (Fig. 3).

The assignment of the major and minor components of the whole olive oil was carried out according to the literature (Sacchi et al., 1996, 1997; Segre & Mannina, 1997 and Mannina et al., 1999). The identified signals are shown in Fig. 2 and reported in Table S2. The spectra of whole olive oil showed the typical pattern of signals of the triglycerides. Nevertheless, a number of other signals belonging to minor components of the EVOO were also detectable. In particular, β-sitosterol (0.62 ppm), squalene (1.62 ppm), terpens (4.53, 4.60, 4.66 ppm), diacylglycerols (3.66, 3.99 ppm) and aldehydes were unambiguously identified (see below).

In order to determine if the observable signals were able to give information on the analyzed olive oils, a principal component analysis (PCA) was performed. Particularly, the PCA (Fig. S4) performed on the data of the spectra of whole olive oils showed that the samples are distributed according to their fatty acid compositions. Olive oils with high content of oleic acid and low content of linoleic acid and saturated fatty acids are in the left part of PCA plot. Among these samples, *Limli*, *Aberkane*, and *Rougette de Mitidja*, (samples 25, 21, and 5) turned out to be particularly interesting for their composition (Table 4), that is similar to the fatty acid composition of some prestigious varieties grown in other Mediterranean areas.

On the other hand, the spectra of the phenolic extracts turned out to be particularly informative. Fig. 3 shows the $^1$H NMR spectrum (700 MHz) of a representative phenolic extract; three main regions could be easily distinguished: aldehydic (9.0–9.8), aromatic (8.0–6.2) and aliphatic (5.0–1.0) regions. Identification of phenolic compounds was made following the references in the literature (Christophoridou et al., 2005; Christophoridou & Dais, 2009) and the analysis of the Statistical Total Correlation Spectroscopy (STOCSY) (Fig. S2). The assignment is reported in Table 1 and Fig. 3. In particular, the more intense aldehydic signals between 9.20 and 9.75 ppm (Fig. 3) were assigned to the protons of the hydrolysis products (**a–f,** Table 1) of the monoaldehydic and dialdehydic forms of ligstroside and oleuropein. Furthermore, the signal at 9.80 ppm was assigned to vanillin (**i**). The aromatic region, instead,

was characterized by the presence of the signals of the free hydroxytyrosol (**n**) (6.4–6.6 ppm). The presence of (+)-1-acetoxypinoresinol (**g**) and (+)-pinoresinol (**h**) was detected by the assignment of the signals at 6.91 and 6.89 ppm, respectively, while the signal at 7.94 and the set of signals at around 7.42 could be attributed to apigenin (**m**) and luteolin (**l**), respectively.

PCA was also performed on the dataset obtained by the NMR spectra of the phenolic extract. Three principal components (PCs) accounting for 61.5% (PC1 33.8%, PC2 17.8%, PC3 9.9%) of the variation were identified. The plots of PC1 and PC2 scores (Fig. 4A and B) showed the positioning of the samples according to their chemical characteristics and allowed the identification of the most important NMR signals for sample differentiation (loading plot in Fig. 4C and D). The PCA shows that the samples 19, 23 and 26 (that are those containing the lowest amount of polyphenols) are placed in the very left side of the score plot, opposite to the samples 4, 9, 10, 15, 18 and 22 (having the highest values of polyphenols) that lie on the very right side of the plot. The loading plot of PC1 (Fig. 4C) closely resembles the NMR spectra of the polyphenolic extract, thus indicating that this distribution is due to the overall content of polyphenols. In order to determine if this distribution is, in some ways, also correlated to the geographical origin of the samples, they were labeled according to two geographical areas of Béjaïa province (Fig. 1): region 1, closer to the Mediterranean Sea, and region 2, much more in-land. Surprisingly, in spite of the two geographical areas that are not so different between each other, they turned out to be well separated in the score plot. Only four samples, namely 9, 10, 17 and 22, were not properly grouped, suggesting that other factors certainly influence this separation. Therefore, overall these data suggest that the content in polyphenols is influenced also by the origin of the samples.

As far as the loading plot of PC2 is concerned (Fig. 4D), it is interesting to note that the variables with the higher loading values (higher discriminating power) belong to the dialdehydic and monoaldehydic forms of ligstroside and oleuropein, indicating that the samples that lie in the bottom of the plot are richer in monoaldehydes, while the ones on the top have an higher content of dialdehydes. Unfortunately, the limited number of samples and the lack of replicates prevented us to ascribe with certainty this peculiar separation to a defined cause. For the time being, it can be highlighted that the varieties *Rougette de Mitidja* (sample 26) and *Bouchouk Guergour* (sample 8), cultivated in the Sidi Aich region, offer a very high content of oleocanthal, known for its ibuprofen-like cyclooxygenase inhibiting activity (Beauchamp et al., 2005), suggesting that these olive oils can be considered promising nutraceutical foods to be used for the treatment of inflammatory diseases (Iacono et al., 2010).

This NMR study gave us also the opportunity to confirm the assignment of the signal of some minor compounds in the NMR spectra of whole oils. Particularly, the signals at 9.70 and 9.45 ppm were confirmed to belong to hexanal and (E)-2-hexenal, respectively. As far as the peaks at 9.62 and 9.57 ppm are concerned, they are generally assigned, respectively, to generic branched alkanal and alkenal (Mannina et al., 1999). Recently, thanks to the use of separation techniques, these peaks were tentatively assigned to monoaldehydes and dialdehydes forms of ligstroside and oleuropein (Mannina et al., 1999). Moreover, in another investigation, the signal at 9.18 ppm was tentatively attributed to dialdehydic derivatives of ligstroside and oleuropein (Dugo et al., 2015). In order to confirm those assignments, a comparison between the NMR spectra of the phenolic extracts (where the signals were unambiguously assigned) and of the whole

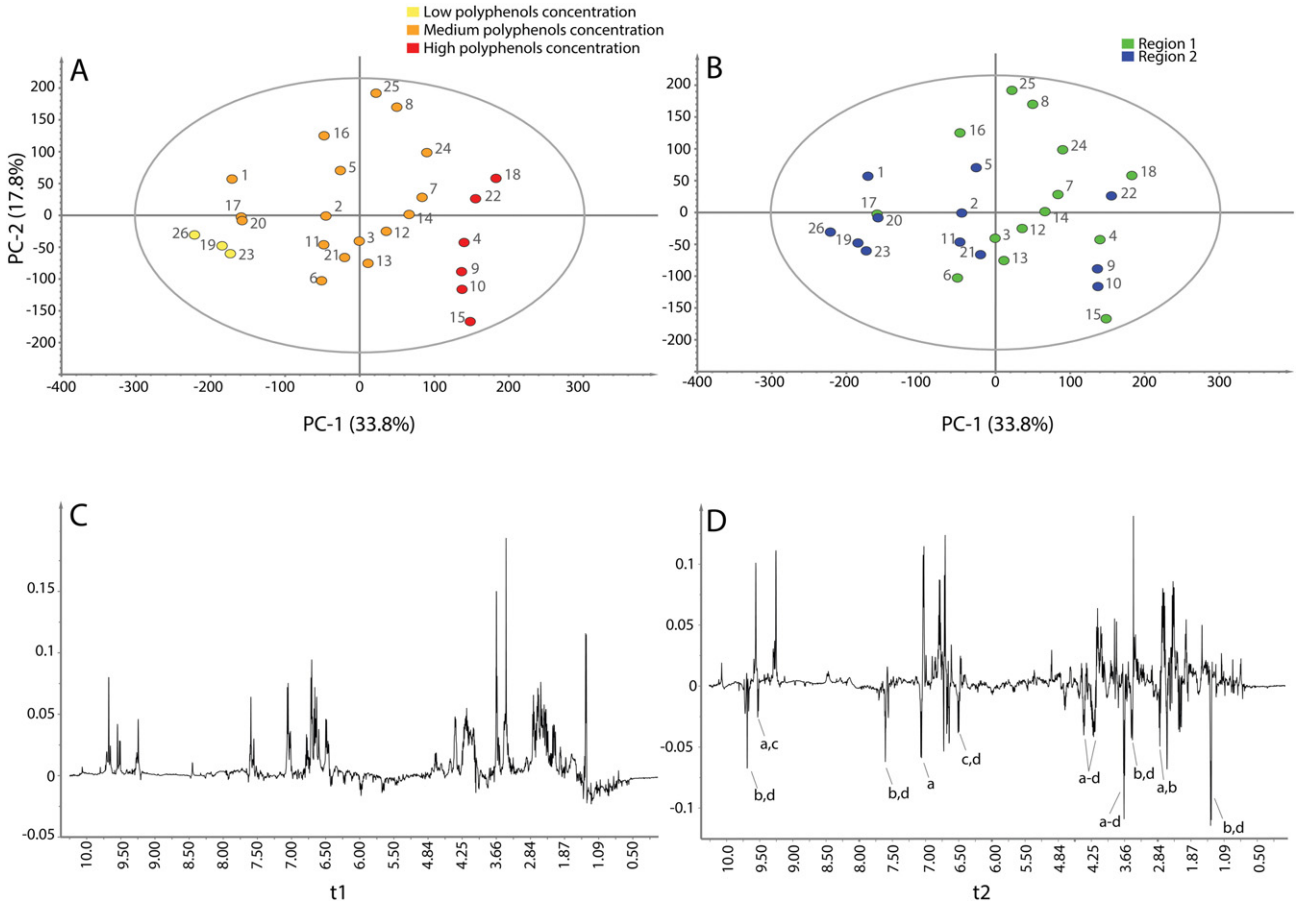


**Fig. 4.** (A) PCA score plot colored according to the content of polyphenols (yellow 50–160 mg $kg^{-1}$, orange 161–470 mg $kg^{-1}$ and red 471–1000 mg $kg^{-1}$). (B) PCA score plot colored according to the region of origin: region 1 (green); region 2 (blue). PC1 and PC2 loading plots are reported in panels C and D, respectively. Letters on the loading plots refer to the NMR assignment reported in Table 1.

oils was performed. The basic idea is that if the molecules generating those signals in the spectra of whole oils are also present in the NMR spectra of the phenolic extracts, the intensities of their signals should be correlated. Thus we computed a matrix of correlation coefficients calculated between the intensities of the signals in the aldehydic region of the spectra of whole olive oils and the signals present in the same region of the spectra of the phenolic extracts. Very interestingly, a correlation coefficient higher than 0.9 was found between the peaks unambiguously assigned to the dialdehydic forms of ligstroside and oleuropein in the phenolic extracts spectra (**e**–**f**) and the signals at 9.57 and 9.18 ppm in the spectra of whole oils. Similarly, a correlation coefficient higher than 0.8 was found between the peaks attributed to the aldehydic compounds in the phenolic extracts (**a**–**d**) and the signal at 9.62 ppm in the intact oil spectra. Therefore, the signals at 9.57 and 9.18 ppm can be unambiguously assigned to the dialdehydic forms of ligstroside and oleuropein, respectively, while the signal at 9.62 ppm can be unambiguously assigned to the their monoaldehydes derivatives.

## 4. Conclusions

Since consumers are more and more oriented toward consumption of food products with certified authenticity and geographical origin, the quality control and authentication of olive oil are of primary importance. For this reason, here we reported an in-depth characterization of twenty-six monovarietal olive oils produced in the Béjaïa province, which is the Algerian area where the olive oil production is mostly increased in the last years. The investigation was first performed using traditional chemical analysis. Each olive oil turned out to have strength and weakness points. For instance, *Souidi* and *Sigoise* varieties had high levels of tocopherols, important for their antioxidant activity, but also a low amount of oleic acid and high amount of palmitic acid. *Blanquette de Guelma* variety was characterized by high contents of phenolic compounds and tocopherols, and by not high amounts of oleic acid. Sample 5 (*Limli* variety of Tazmalt area) had a high value of oleic acid and a low level of saturated fatty acid, even if the tocopherol and phenolic compounds contents were lower with respect to the other olive oils. *Rougette de Mitidja* variety had a high value of oleic acid, a low level of saturated fatty acid, a high level of tocopherols, but a low, although acceptable, level of phenolic compounds. Overall, the results of this study suggest that the analyzed olive oils could be blended in order to obtain oils of better quality. For instance, the blend of *Blanquette de Guelma* and *Rougette de Mitidja* varieties in the suitable proportion could lead to an olive oil with interesting properties.

Furthermore, a NMR study was also performed on the intact olive oils and on the phenolic extracts of the twenty-six samples. This, to the best of our knowledge, represents the first NMR study of Algerian EVOOs. While the spectra of whole olive oils were useful to have an idea of the quality of monovarietal olive oils, the NMR spectra of phenolic extract provided very interesting information about geographical origin of the samples. Of course, these findings need further validation, based on substantially larger sets of olive oil samples, in order to investigate the possibility to extend this method to discrimination of different cultivars, year of harvest, age of oil sample, etc. The data acquired in this investigation could be used, along with those coming from other studies, to establish a bank of original data concerning the chemical composition of Algerian monovarietal olive oils and providing information that can be used by Algerian olive growers for large-scale plantation.

## Conflict of interest

The authors declare no competing financial interest.

## Acknowledgments

The authors are grateful to the staff of ITAFV (Institut d'Arboriculture Fruitière et de la Vigne) Takeriets (Béjaïa Algeria).

## Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.foodres.2016.04.024.

## References

Abaza, L., Taamalli, W., Ben Temime, S., Ben Miled Daoud, D., Gutierrez, F., & Zarrouk, M. (2005). Natural antioxidant composition as correlated to stability of some Tunisian virgin olive oils. *Rivista Italiana delle Sostanze Grasse*, *82*(1), 12–18.

Angerosa, F., Campestre, C., & Giansante, L. (2006). In B. Dimitrios (Ed.), *Analysis authentication, olive oil chemistry and technology* (2nd ed.). AOCS Publishing.

Arslan, D., & Schreiner, M. (2012). Chemical characteristics and antioxidant activity of olive oils from Turkish varieties grown in Hatay province. *Scientia Horticulturae*, *144*, 141–152.

Ballus, C. A., Meinhart, A. D., Campos, F., Jr., da Silva, L., de Oliveira, A., & Godoy, H. (2014). A quantitative study on the phenolic compound, tocopherol and fatty acid contents of monovarietal virgin olive oils produced in the southeast region of Brazil. *Food Research International*, *62*, 74–83.

Beauchamp, G. K., Keast, R. S., Morel, D., Lin, J., Pika, J., Han, Q., ... Breslin, P. A. (2005). Phytochemistry: Ibuprofen-like activity in extra-virgin olive oil. *Nature*, *437*, 45–46.

Bendini, A., Cerretani, L., Carrasco-Pancorbo, A., Gomez-Caravaco, A. M., Segura-Cerretano, A., Fernandez-Gutierrez, A., & Lercker, G. (2007). Phenolic molecules in virgin olive oils; a survey of their sensory properties, health effects, antioxidant activity and analytical methods. An overview of the last decade. *Molecule*, *12*(8), 1679–1719.

Boskou, D. (2006). Sources of natural phenolic antioxidants. *Trends Food Science Technology*, *17*, 505–512.

Cartoni, G. P., Coccioli, F., Jasionowska, R., & Ramires, D. (2000). HPLC analysis of the benzoic and cinnamic acids in edible vegetable oils. *Italian Journal of Food Science*, *12*(2), 163–173.

Cerretani, L., Bendini, A., Del Caro, A., Piga, A., Vacca, V., & Caboni, M. F. (2006). Preliminary characterisation of virgin olive oils obtained from different cultivars in Sardinia. *European Food Research and Technology*, *222*(3), 354–361.

Christophoridou, S., & Dais, P. (2009). Detection and quantification of phenolic compounds in olive oil by high resolution 1H nuclear magnetic resonance spectroscopy. *Analytica Chimica Acta*, *633*(2), 283–292.

Christophoridou, S., Dais, P., Tseng, L. H., & Spraul, M. (2005). Separation and identification of phenolic compounds in olive oil by coupling high-performance liquid chromatography with postcolumn solid-phase extraction to nuclear magnetic resonance spectroscopy (LC-SPE-NMR). *Journal of Agricultural and Food Chemistry*, *53*(12), 4667–4679.

Cicerale, S., Lucas, L., & Keast, R. S. (2012). Antimicrobial, antioxidant and antiinflammatory phenolic activities in extra virgin olive oil. *Current Opinion in Biotechnology*, *23*(2), 129–135.

D'Imperio, M., Mannina, L., Capitani, D., Bidet, O., Rossi, E., Bucarelli, F. M., & Segre, A. L. (2007). NMR and statistical study of olive oils from Lazio: A geographical, ecological and agronomic characterization. *Food Chemistry*, *105*(3), 1256–1267.

Dağdelen, A., Tümen, G., Ozcan, M. M., & Dündar, E. (2012). Determination of tocopherol contents of some olive varieties harvested at different ripening periods. *Natural Product Research*, *26*(15), 1454–1457.

Deiana, M., Rosa, A., Cao, C. F., Pirisi, F. M., Bandino, G., & Dessi, M. A. (2002). Novel approach to study oxidative stability of extra virgin olive oils: Importance of alpha-tocopherol concentration. *Journal of Agricultural and Food Chemistry*, *50*(15), 4342–4346.

Dugo, G., Rotondo, A., Mallamace, D., Cicero, N., Salvo, A., Rotondo, E., & Corsaro, C. (2015). Enhanced detection of aldehydes in extra-virgin olive oil by means of band selective NMR spectroscopy. *Pyshica A*, *420*, 258–264.

E.C. (2002). Regulation n° 796 of 6 May 2006 on changes EC-regulation. 2568/91. *Official J.L.128/815/05/02. 2002* (Bruxelles (Belgium)).

Eriksson, L., Johansson, E., Kettaneh-Wold, N., & Wold, S. (1999). *Introduction to multi- and megavariate data analysis using projection methods (PCA & PLS)*. Umeå, Sweden: Umetrics, 213–225.

Escrich, E., Moral, R., Grau, L., Costa, I., & Solanas, M. (2007). Molecular mechanisms of the effects of olive oil and other dietary lipids on cancer. *Molecular Nutrition and Food Research*, *51*(10), 1279–1292.

García-Villalba, R., Carrasco-Pancorbo, A., Oliveras-Ferraros, C., Vázquez-Martín, A., Menéndez, J.A., Segura-Carretero, A., & Fernández-Gutiérrez, A. (2010). Characterization and quantification of phenolic compounds of extra-virgin olive oils with anticancer properties by a rapid and resolutive LC-ESI-TOF MS method. Journal of Pharmaceutical and Biomedical Analysis, 51(2), 416–429.

Gilbert-López, B., Valencia-Reyes, Z. L., Yufra-Picardo, V. M., García-Reyes, J. F., Ramos-Martos, N., & Molina-Díaz, A. (2014). Determination of polyphenols in commercial extra virgin olive oils from different origin Mediterranean. *Food Analytical Methods*, *7*(9), 1824–1833.

Gómez-Rico, A., Salvador, M. D., La Greca, M., & Fregapane, G. (2006). Phenolic and volatile compounds of extra virgin olive oil (*Olea europaea* L. Cv. Cornicabra) with regard to fruit ripening and irrigation management. *Journal of Agricultural and Food Chemistry*, *54*(19), 7130–7136.

Iacono, A., Goḿez, R., Sperry, J., Conde, J., Bianco, G., Meli, R., ... Gualillo, O. (2010). Effect of oleocanthal and its derivatives on inflammatory response induced by lipopolysaccharide in a murine chondrocyte cell line. *Arthritis Rheumatism*, *62*(6), 1675–1682.

IOC (2010). International Olive Council. Determination of the content of waxes, fatty acid methyl esters and fatty acid ethyl esters by capillary gas chromatography. *COI/T. 20/NC No 28/Rev. 1* (pp. 1–17).

Kavallari, A., Maas, S., & Schmitz, M. (2011). Examining the determinants of olive oil demand in nonproducing countries: Evidence from Germany and the UK. *Journal of Food Products Marketing*, *17*(2), 355–372.

Krichene, D., Allalout, A., Baccouri, B., Fregapane, G. Q., Salvador, M. D., & Zarrouk, M. (2009). Territorial investigation based on the chemical composition of Chemlali virgin olive oils. *Asian Journal of Biochemistry*, *4*(1), 1–12.

Lauri, I., Pagano, B., Malmendal, A., Sacchi, R., Novellino, E., & Randazzo, A. (2013). Application of "magnetic tongue" to the sensory evaluation of extra virgin olive oil. *Food Chemistry*, *140*(4), 692–6929.

Manai-Djebali, H., Krichene, D., Ouni, Y., Gallardo, L., Sanchez, J., Osorio, E., ... Zarrouk, M. (2012). Chemical profiles of five minor olive oil varieties grown in central Tunisia. *Journal of Food Composition and Analysis*, *27*(2), 109–119.

Mannina, L., & Sobolev, A. P. (2011). High resolution NMR characterization of olive oils in terms of quality, authenticity and geographical origin. *Magnetic Resonance in Chemistry*, *49*, 3–11.

Mannina, L., D'Imperio, M., Capitani, D., Rezzi, S., Guillou, C., Mavromoustakos, T., ... Aparicio, R. (2009). $^{1}$H NMR-based protocol for the detection of adulterations of refined olive oils with refined hazelnut oil. *Journal of Agricultural and Food Chemistry*, *57*(24), 11550–11556.

Mannina, L., Marini, F., Gobbino, M., Sobolev, A. P., & Capitani, D. (2010). NMR and chemometrics in tracing European olive oils: The case study of Ligurian samples. *Talanta*, *80*(5), 2141–2148.

Mannina, L., Patumi, M., Fiordiponti, P., Emanuele, M. C., & Segre, A. L. (1999). Olive and hazelnut oils: A study by high-field $^{1}$H NMR and gas chromatography. *Italian Journal of Food Science*, *2*, 139–149.

Mannina, L., Sobolev, A. P., & Viel, S. (2012). Liquid state 1H high field NMR in food analysis. *Progress in Nuclear Magnetic Resonance Spectroscopy*, *66*, 1–39.

Minguez-Mosquera, M. I., Rejano, L., Gandul, B., Sanchez, A. H., & Garrido, J. (1991). Color-pigment correlation in virgin olive oil. *Journal of the American Oil Chemists' Society*, *68*(5), 322–337.

Montedoro, G. F., Servilli, M., Baldioli, M., & Miniati, E. (1992). Simple and hydrolyzable phenolic compounds in olive oil. 1. Their extraction, separation, and quantitative and semi quantitative separation and evaluation by HPLC. *Journal of Agricultural and Food Chemistry*, *40*(9), 1571–1576.

Norme Grassi e Derivati (NGD) (2010). *Divisione SSOG — Innovhub SSI, Milano, NGD C89 and COI/T20 Doc. N. 29*.

Owen, R. W., Mier, W., Giacosa, A., Hull, W. E., Spiegelhalder, B., & Bartsch, H. (2000). Olive-oil consumption and health: The possible role of antioxidants. *The Lancet Oncology*, *1*, 107–112.

Pinelli, P., Galardi, C., Mulinacci, N., & Romani, A. (2003). Minor polar compound and fatty acid analyses in monocultivar virgin olive oils from Tuscany. *Food Chemistry*, *80*(3), 331–336.

Psomiadou, E., & Tsimidou, M. (2002). Stability of virgin olive oil. 1. Autoxidation studies. *Journal of Agricultural and Food Chemistry*, *50*(4), 716–721.

Ramos-Escudero, F., Teresa Morales, M., & Asuero, A. G. (2015). Characterization of bioactive compounds from monovarietal virgin olive oils: Relationship between phenolic compounds-antioxidant capacities. *International Journal of Food Properties*, *18*(2), 348–358.

Ranalli, A., Modesti, G., Jaren Galan, M., & Hornero Mendez, D. (1999). Processing technologies and biotechnologies affect the composition of green and yellow lipochromes and the chromatic features of virgin olive oil. In M. I. Mınguez-Mosquera (Ed.), *Proceedings of the 1st international congress on pigments in food technology, Sevilla* (pp. 239).

Rigane, G., Boukhris, M., Bouaaziz, M., Sayadi, S., & Ben Salem, R. (2013). Analytical evaluation of two monovarietal of Tunisia: Jemri-Bouchouka and Chemlali-Tataouin cultivars. *Journal of the Science of Food and Agriculture*, *93*(5), 1242–1248.

Romero, N., Saavedra, J., Tapia, F., Sepúlveda, B., & Aparicio, R. (2015). Influence of agroclimatic parameters on phenolic and volatile compounds of Chilean virgin olive oils and characterization based on geographical origin, cultivar and ripening stage. *Journal of the Science of Food and Agriculture*, *96*(2), 583–592.

Rotondi, A., Bendini, A., Cerretani, L., Mari, M., Lercker, G., & Toschi, T. G. (2004). Effect of olive ripening degree on the oxidative stability and organoleptic properties of Cv. Nostrana di Brisighella extra virgin olive oil. *Journal of Agricultural and Food Chemistry*, *52*(11), 3649–3654.

Rovellini, P., Azzolini, M., & Cortesi, N. (1997). Tocoferoli e tocotrienoli in olie grassi vegetali mediante HPLC. *Rivista Italiana delle Sostanze Grasse*, *74*, 1–5.

Sacchi, R., Addeo, F., & Paolillo, L. (1997). $^{1}$H and $^{13}$C NMR of virgin olive oil. An overview. *Magnetic Resonance Chemistry*, *35*(13), 5133–5145.

Sacchi, R., Patumi, M., Fontanazza, G., Barone, P., Fiordiponti, P., Mannina, L., ... Segre, A. (1996). A high-field 1H nuclear magnetic resonance study of the minor components in virgin olive oils. *Journal of the American Oil Chemists' Society*, *73*(6), 747–758.

Savorani, F., Tomasi, G., & Engelsen, S. B. (2010). Icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. *Journal of Magnetic Resonance*, *202*(2), 190–202.

Segre, A., & Mannina, L. (1997). $^{1}$H-NMR study of edible oils. *Recent Research Development Oil Chemistry*, *1*, 297–308.

Servili, M., Esposto, S., Fabiani, R., Urbani, S., Taticchi, A., Mariucci, F., ... Montedoro, G. F. (2009). Phenolic compounds in olive oil: Antioxidant, health and organoleptic activities according to their chemical structure. *Inflammopharmacology*, *17*(2), 76–84.

Tovar, M. J., Motilva, M. J., & Paz Romero, M. (2001). Changes in the phenolic composition of virgin olive oil from young trees (*Olea europaea* L. cv. Arbequina) grown under linear irrigation strategies. *Journal of Agricultural Food Chemistry*, *49*(11), 5502–5508.

Vichi, S., Lazzez, A., Grati-Kamoun, N., & Caixach, J. (2012). Modifications in virgin olive oil glycerolipid fingerprint during olive ripening by MALDI-TOF MS analysis. *LWT - Food Science and Technology*, *48*(1), 24–29.

doi: 10.1255/mrfs.2

# Application of recursive partial least square regression for prediction of apple juice sensory attributes from NMR spectra

N. Iaccarino[a], C. Varming[b], M.A. Petersen[b], F. Savorani[b,c], A. Randazzo[a], B. Schütz[d], T.B. Toldam-Andersen[e] and S.B. Engelsen[b]*
[a]Department of Pharmacy, University of Naples "Federico II", Via D. Montesano 49, 80131 Naples, Italy.
E-mail: nunzia.iaccarino@unina.it
[b]Department of Food Science, University of Copenhagen, Rolighedsvej 26, 1958 Frederiksberg C, Denmark.
Corresponding Author: se@food.ku.dk
[c]Department of Applied Science and Technology, Polytechnic University of Turin, Corso Duca degli Abruzzi 24, 10129 Turin, Italy.
[d]Bruker BioSpin, Silberstreifen 4, 76287 Rheinstetten, Germany.
[e]Department of Plant and Environmental Sciences, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark.

This study demonstrates the application of a novel variable selection method here employed for the prediction of sweet and sour taste of apple juice from Nuclear Magnetic Resonance (NMR) spectra. The method is called recursive weighted Partial Least Square (rPLS). It operates by iteratively re-weighting the spectral variables using the regression coefficients calculated by PLS. The only parameter to be estimated by the operator is the number of latent factors to be used in the model. This approach provides an easier model interpretation than a regular PLS model, since it converges towards a very limited number of variables and therefore the assignment effort is drastically reduced. These properties suggest a profitable use of the rPLS for the prediction of even more complex sensory features from different types of spectroscopic data.

## Introduction

Nuclear Magnetic Resonance (NMR) spectroscopy has been widely applied to food systems in order to obtain a 'holistic view' of the metabolome (foodome) of various kinds of beverages and foods, such as fruit juice[1,2] milk[3,4], wine[5] and olive oil[6]. In recent years, some studies have focused on the correlation between the NMR metabolomic fingerprint of the food samples and the sensory features evaluated by a panel test. These studies include sour cherry juice[7], tomatoes[8], olive oil[9] and coffee beans extracts[10] and some even suggest that NMR spectroscopy can be considered as a "magnetic tongue" for a better characterisation as well as prediction of the taste of food products. In this context Multivariate Data Analysis plays a fundamental role in the understanding of the correlation between the spectral dataset (**X**) and the response parameters from the sensory evaluation (**y**). For this purpose the Partial Least Square regression[11] is the most widely used algorithm. It first calculates a set of loading weights, **W**, which finds the combination between **X** and **y** and then calculates the regression coefficients, **b**, that provide an estimation of **y** when it is multiplied by the **X** matrix. In this study an advanced version of PLS is performed for the prediction of sweet and sour taste of apple juice from NMR spectra (Figure 1). These attributes are considered important drivers of the market preferences[12], therefore their evaluation is crucial and trained sensory panels are employed for this purpose. Finding a method that is able to predict these features avoiding the employment of the panellists and/or reducing chemical analysis to the minimum would help both companies and researchers in selecting only the best cultivars. As far as apples are concerned, titratable acidity and °Brix values were found to be quite good descriptors respectively for the acid and sweet taste[13]. In this study we propose an approach that could be useful when only spectroscopic data are available and a prediction of sensory attributes is needed. This approach is based on the so-called recursive PLS, or just rPLS[14], which is a recently developed variable selection method where the regression coefficients are recursively used as weights on the original data matrix. This concept is based on the fact that the regression vector reflects the importance of the variables: weights around 0 indicate variables not correlated with **y**, and weights with large absolute values indicate important variables. Repeating the weighting, the rPLS model has the property to converge to a limited number of variables (equal to the number

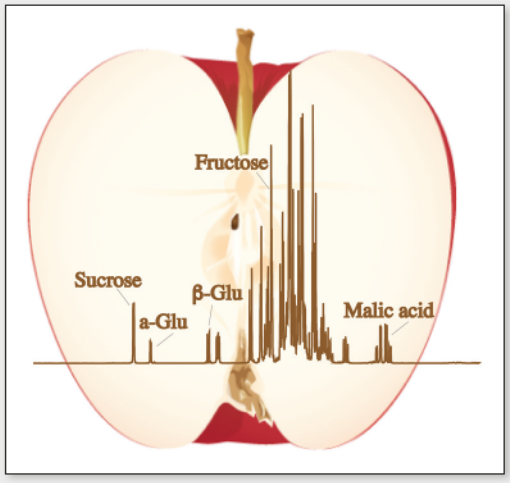


**Figure 1.** The average NMR spectrum of apple juice.

of Latent Variables/PLS components), facilitating the interpretation and reducing the time consuming step of a thorough signal assignment.

## Materials and methods

### Samples

The dataset consists of ninety-two apple juice samples obtained from ancient Danish apple cultivars. Each apple juice was then submitted both to NMR analysis and sensory evaluation. Six different descriptors were evaluated: colour, overall odour, apple flavour, overall flavour, sweet taste and sour taste. In order to test the efficacy of the rPLS algorithm, sweetness and acidity, considered easier to interpret, were taken into account in this preliminary study. The sensory panelists were trained with a reference juice as well as with sucrose (11%) and malic acid (0.5%) water solutions for being able to properly recognize all the descriptors. The samples were evaluated using a continuous 0 (none) - 14 (very much) intensity scale and the scores of each sample were averaged over 5 assessors.

### Nuclear magnetic resonance data and processing

Bruker Spin Generated Fingerprint (SGF) profiling[15] was employed for the NMR analysis. Each sample required minimal preparation effort consisting of 90% juice with 10% buffer containing 0.1% of TSP (sodium salt of 3-trimethylsilyl-propionate acid-d4) and 0.013% of sodium azide to suppress microorganism activity. This NMR-based screening method is based on an Avance 400 NMR spectrometer with a 9.4-T Ultrashield™ Plus magnet and utilizes flow-injection NMR (BEST™ NMR) with a 4-mm flow-cell probe with Z-gradient and a Gilson liquids handler for sample storage, preparation and transfer. Samples are provided in bar-coded cryovials placed in a Gilson cooling rack that keeps the temperature low (about 4 °C) prior to injection. Then a heated transfer line from the Gilson unit to the probe allows the pre-equilibration of the sample to the desired temperature (300 K) during the transfer. The overall experimental procedure is fully controlled by Bruker's SampleTrack software including temperature adjustment, tuning and matching, locking, shimming and the optimization of the pulses and presaturation power for each sample. The resulting spectra do not need any manual processing step, as they are automatically phase corrected and referenced by the Bruker procedure. Thus, they are ready to be imported in MATLAB (The Mathworks Inc., Massachusetts, USA) where they are at first aligned in the horizontal direction using the icoshift tool developed by Savorani et al.[16] and subsequently mean centered prior to any further chemometrics calculation.

### Chemometrics

The rPLS procedure starts by performing a classical PLS regression model between **X** and **y**, while in the following steps, it recursively re-weights the **X** by multiplying it by the regression vector **b** calculated during the previous iteration. This re-weighting is iteratively repeated until no further progress in the regression coefficients occurs. The idea of using the regression vector variables as weights is based on the fact that they reflect the importance of the original spectral variables. Regression vector weights near 0 indicate variables not involved in the correlation with **y**, and weights with large absolute values indicate important variables. The Root Mean Square Error of Cross Validation (RMSECV) is calculated by using venetian blind cross validation (5 groups). The rPLS algorithm is implemented in MATLAB (The MathWorks, Inc.) and made freely available for non-commercial use at www.models.life.ku.dk.

## Results and discussion

This apple juice NMR dataset was selected to show the ability of the rPLS algorithm to find the useful information, correlated with the sourness and sweetness in this case, among thousands of spectral variables. The result for the prediction of the sour taste is shown in Figure 2. This plot contains most of the information calculated by the algorithm and it is a direct output of the rPLS script. As observed, the number of variables considered "important" become lower and lower when the iterations increase, until they converge to

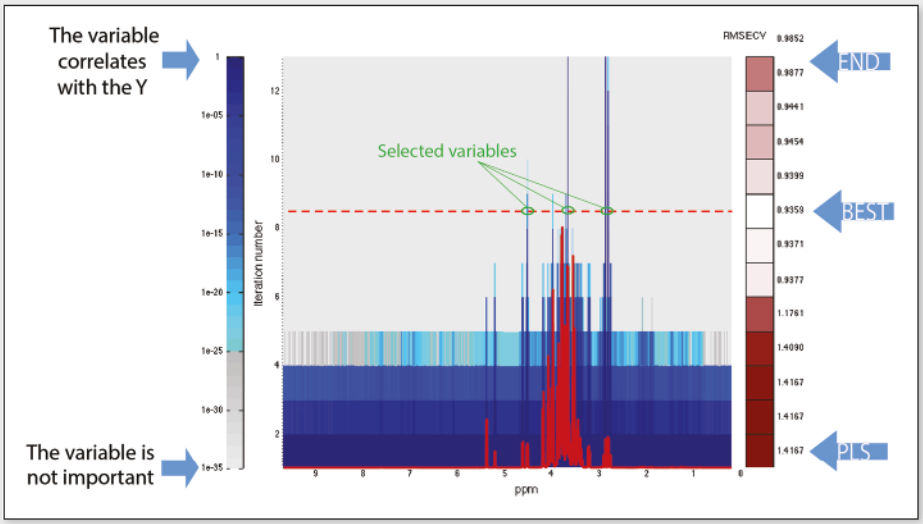


**Figure 2.** The rPLS result for the prediction of sensory evaluated acidity. The model uses two latent variables. In each row the development of weights according to iterations is shown. The coloured scale on the right represents the RMSECV values, the white box indicates the row where the best model was built and its relative RMSECV value; the bar on the left shows the value of the weights. The value 1 means that the variable has a large weight and thus importance; 1e-35 means that it has not. The red dashed line shows the optimal rPLS model and the green circles indicate the variables selected by the algorithm. The thick red spectrum superimposed to the figure is the average of all the spectra in the dataset.

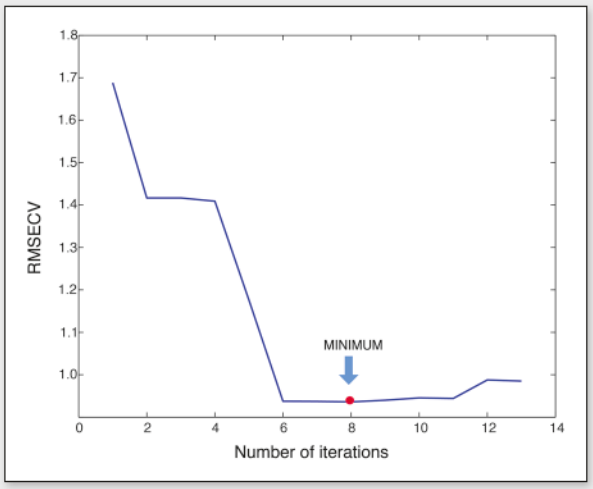


**Figure 3.** The development in sourness prediction performance (RMSECV) during the rPLS iterations.

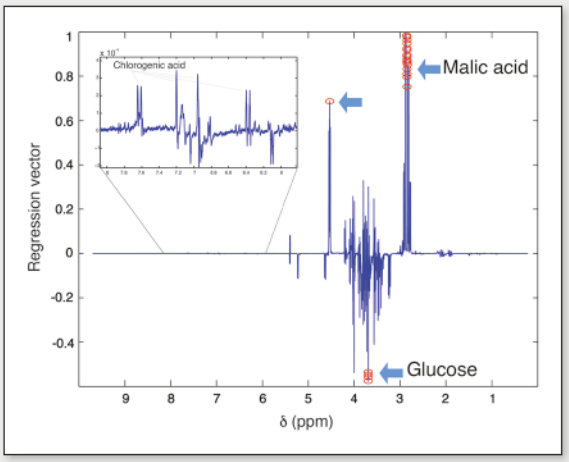


**Figure 4.** The rPLS result for the prediction of sensory evaluated acidity. The regression coefficients for the full range PLS model (in blue) and for the rPLS reduced model (red circles).

the same number of Latent Variables chosen for the optimal model of the very first step (two in this case). On the left part of the figure, a coloured bar represents the weights used for the variables. The dark blue colour means that the variable has weight equals to 1 and thus it is very important for the regression model. In contrast, the variables situated in the grey region can be considered useless in terms of predictivity since their weight is very close to 0. On the right hand side of the rPLS plot, the RMSECV of each iteration step is shown. The best iterative performance was obtained after eight iterations, as indicated also in Figure 3.

The global PLS model shows a predictive performance of RMSECV=1.70 while the rPLS global minimum shows a predictive performance of RMSECV=0.96. This result has two main advantages, (i) it performs clearly better than the global PLS model and (ii) it is three orders of magnitude more simple, as it contains only 25 variables instead of the 29149 spectral variables included in the global one, allowing the careful inspection of the single variables. Figure 4 shows the regression vector of the global PLS model as well as the variables, counting for the NMR signals at 2.85 ppm (counting for twenty variables), 4.53 ppm (counting for only one variable), 3.69 ppm (counting for four variables), that have been identified by the rPLS as mainly responsible for the sour taste. The peaks can easily be identified as the malic acid methylene (2.85 ppm) and methine (4.53 ppm) protons, while the peak at 3.69 ppm pertains to the glucose pyranose ring protons. These observations are in perfect agreement with the fact that the malic acid is the main acid in apple juice and therefore the main responsible for the sour taste of the samples. Moreover, the fact that glucose is also taken into account by the model, albeit with a numerically lower and negative regression coefficient, indicates an inverse correlation between malic acid content and glucose concentration.

It's also interesting to notice that in the aromatic region, where polyphenols signals arise, the chlorogenic acid shows positive regression coefficients (Figure 4). It is known that polyphenols can give bitterness and astringency to the apple juice[17], however here the main polyphenol found in apple juice seems to have also a positive correlation with the sour taste.

As far as the sweetness is concerned, the best rPLS result occurs after seven iterations (Figure 5). Also in this case the recursive approach brings a clear improvement when compared to the global PLS model, not only in terms of RMSECV, but also in terms of the number of variables to be inspected. Only five peaks have been selected by the rPLS. The signals around 3.81 and 3.67 ppm belong to sucrose and they are positively correlated to the sweetness, while the three glucose peaks (3.99, 3.79 and 3.68 ppm) are negatively correlated to this attribute. The inverse relation between sucrose and glucose content is already known from literature and it is likely due to their interconversion[18]. Surprisingly, the sucrose turned out to be the main responsible for the sweet taste even though the fructose is known to be the main sugar and thus sweetener in apple juice[19]. This confirms the complexity in assigning the sweet taste to a specific chemical compound[13], since it should better be considered as the global result of the combination of several components. One of the advantages of the rPLS approach is that it does not only reduce the variable space and simplify the interpretation of the result, but it also includes the relevant covariation around the selected peaks. The latter information can be extremely useful for assignment purposes.

## Conclusion

In this work, we have shown the utility of the rPLS method for the prediction of sensory attributes from Nuclear Magnetic Resonance data of apple juices. Two of the six sensory descriptors available were used for this purpose, namely sweet and sour taste.

In both cases the rPLS was able to develop a good regression model providing just a very limited set of variables that correlate with the y vector. The advantage of this technique is that no parameter must be set by the operator, apart from the optimal number of latent variables required for the initial PLS model. Indeed, the strength of this approach lies in the strong variable selection iteratively performed by the algorithm which not only improves the prediction performance, but it also makes the interpretation of the result tremendously easier for the operator. Thus the aim of this work is to demonstrate the validity of the rPLS method for predicting simple sensory parameters and suggest that it can be a useful tool for the prediction of even more complex sensory features from NMR data avoiding the need of any further chemical analysis. However, it is important to note that when a correlation is obtained from the data,

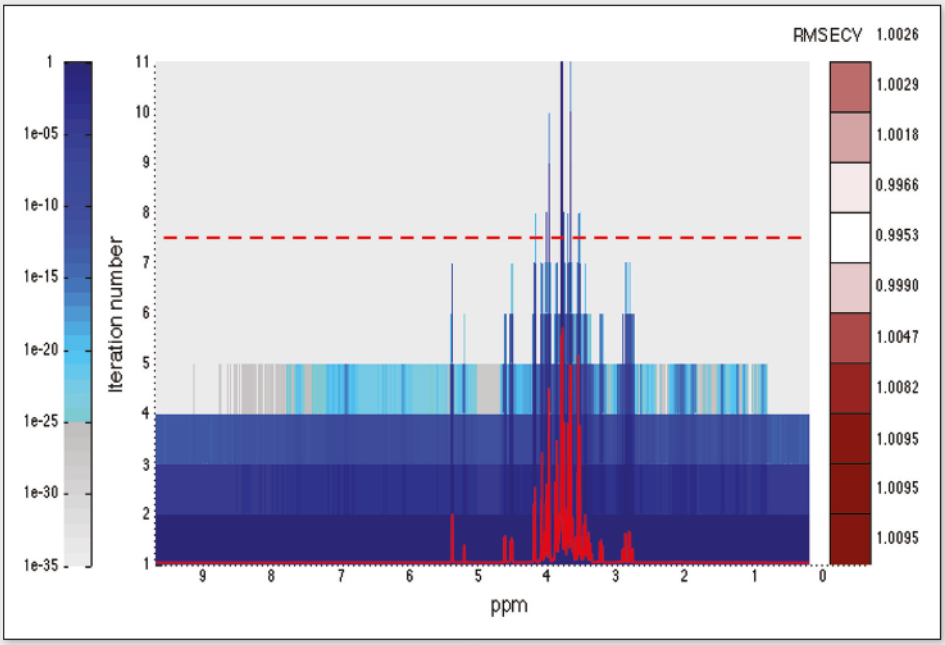


**Figure 5.** The rPLS result for the prediction of sensory evaluated sweetness.

this does not necessarily mean that there is a direct cause-effect between the correlated parts. An indirect causality is often present in any kind of dataset and this can potentially be misleading for the interpretation. Thus, the results of the model always need to be carefully verified.

The use of the rPLS is also promising for applications to the metabolomics field, as showed by Rinnan et al.[14] since it is able to extract only the useful information from a highly complex metabolomics dataset. Finally, it should be emphasised that in this preliminary research we have been using rPLS to augment the interpretation of the data. If improved prediction models instead are the target, then model validation with an independent test set is obligatory when regression and variable selection is mixed as it is in the rPLS model.

## Acknowledgements

The apples were provided by the experimental orchard "Pometum" (Høje Taastrup, Denmark) in the frame of the YDUN project, in collaboration with the Nordic Genetic Resource Center (NordGen). The Ministry of Food, Agriculture and Fisheries financially supported the project through the Danish Food Industry Agency.

## References

1. P.S. Belton, I. Delgadillo, A.M. Gil, P. Roma, F. Casuscelli, I.J. Colquhoun, M.J. Dennis and M. Spraul, "High-field proton NMR studies of apple juices", *Magn Reson Chem*. **35**(13), S52-S60 (1997). doi: 10.1002/(SICI)1097-458X(199712)35:13<S52::AID-OMR212>3.0.CO;2-D
2. A.P. Sobolev, L. Mannina, N. Proietti, S. Carradori, M. Daglia, A.M. Giusti, R. Antiochia and D. Capitani, "Untargeted NMR-based methodology in the study of fruit metabolites", *Molecules*. **20**(3), 4088-4108 (2015). doi: 10.3390/molecules20034088
3. J. Belloque and M. Ramos, "Application of NMR spectroscopy to milk and dairy products", *Trends Food Sci Technol*. **10**(10), 313-320 (1999). doi: 10.1016/S0924-2244(00)00012-1
4. F. Hu, K. Furihata, Y. Kato and M. Tanokura, "Nondestructive quantification of organic compounds in whole milk without pretreatment by two-dimensional NMR spectroscopy", *J. Agric. Food. Chem*. **55**, 4307-4311(2007). doi: 10.1021/jf062803x
5. R. Godelmann, F. Fang, E. Humpfer, B. Schütz, M.Bansbach, H. Schäfer and M. Spraul, "Targeted and nontargeted wine analysis by 1H NMR spectroscopy combined with multivariate statistical analysis. Differentiation of important parameters: Grape variety, geographical origin, year of vintage", *J Agric Food Chem*. **61**(23), 5610-5619 (2013). doi: 10.1021/jf400800d
6. R. Sacchi, F. Addeo and L. Paolillo, "1H and 13C NMR of virgin olive oil. An overview", *Magn. Reson. Chem.* **35**, S133-145 (1997). doi: 10.1002/(SICI)1097-458X(199712)35:13<S133::AID-OMR213>3.0.CO;2-K
7. M.R. Clausen, B.H. Pedersen, H.C. Bertram and U. Kidmose, "Quality of sour cherry juice of different clones and cultivars (Prunus cerasus L.) determined by a combined sensory and NMR spectroscopic approach", *J Agric Food Chem*. **59**(22), 12124-12130 (2011). doi: 10.1021/jf202813r
8. A. Malmendal, C. Amoresano, R. Trotta, I. Lauri, S. De Tito, E. Novellino and A. Randazzo, "NMR spectrometers as "magnetic tongues": Prediction of sensory descriptors in canned tomatoes", *J Agric Food Chem*. **59**(20),10831-10838 (2011). doi: 10.1021/jf203803q
9. I. Lauri, B. Pagano, A. Malmendal, R. Sacchi, E. Novellino and A. Randazzo, "Application of "magnetic tongue" to the sensory evaluation of extra virgin olive oil", *Food Chem*. **140**(4), 692-699 (2013). doi: 10.1016/j.foodchem.2012.10.135
10. F. Wei, K. Furihata, T. Miyakawa and M. Tanokura, "A pilot study of NMR-based sensory prediction of roasted coffee bean extracts", *Food Chem*. **152**, 363-369 (2014). doi: 10.1016/j.foodchem.2013.11.161
11. S. Wold, M. Sjöström and L. Eriksson, "PLS-regression: A basic tool of chemometrics", *Chemom Intell Lab Syst*. **58**(2), 109-130 (2001). doi: 10.1016/S0169-7439(01)00155-1
12. S.R. Jaeger, Z. Andani, I.N Wakeling and H.J. MacFie, "Consumer preferences for fresh and aged apples: a cross-cultural compari-

son", *Food Qual Prefer*. **9**(5), 355-366 (1998). doi: 10.1016/S0950-3293(98)00031-7

13. F.R. Harker, K.B. Marsh, H. Young, S.H. Murray, F.A. Gunson, S.B. Walker, "Sensory interpretation of instrumental measurements 2: Sweet and acid taste of apple fruit", *Postharvest Biol Technol*. **24**(3), 241-250 (2002). doi: 10.1016/S0925-5214(01)00157-0
14. Å. Rinnan, M. Andersson, C. Ridder and S.B. Engelsen, "Recursive weighted partial least squares (rPLS): an efficient variable selection method using PLS", *J Chemom*. **28**(5), 439-447 (2014). doi: 10.1002/cem.2582
15. M. Spraul, B. Schütz, P. Rinke, S. Koswig, E. Humpfer, H. Schäfer, M. Mörtter, F. Fang, U.C. Marx and A. Minoja, "NMR-based multi parametric quality control of fruit juices: SGF profiling", *Nutrients*. **1**(2),148-155 (2009). doi: 10.3390/nu1020148
16. F. Savorani, G. Tomasi, S.B. Engelsen, "icoshift: A versatile tool for the rapid alignment of 1D NMR spectra", *J Magn Reson*. **202**(2), 190-202 (2010). doi: 10.1016/j.jmr.2009.11.012
17. I. Berregi, J.I. Santos, G. Del Campo, J.I. Miranda, J.M. Aizpurua, "Quantitation determination of chlorogenic acid in cider apple juices by 1H NMR spectrometry", *Anal Chim Acta*. **486**(2), 269-274 (2003). doi: 10.1016/S0003-2670(03)00496-3
18. M. Vermathen, M. Marzorati, D. Baumgartner, C. Good, P. Vermathen, "Investigation of Different Apple Cultivars by High Resolution Magic Angle Spinning NMR. A Feasibility Study", *J Agric food Chem*. **59**, 12784-12793 (2011). doi: 10.1021/jf203733u
19. F. Karadeniz and A. Ekşi, "Sugar composition of apple juices", *Eur Food Res Technol*. **215**(2), 145-148 (2002). doi: 10.1007/s00217-002-0505-2

# $^{1}$H NMR-based metabolomics study on follicular fluid from patients with PolyCystic Ovary Syndrome (PCOS)

**Nunzia Iaccarino[1], Jussara Amato[1], Bruno Pagano[1], Alessia Pagano[1], Laura D'Oriano[2], Sveva Pelliccia[1], Mariateresa Giustiniano[1], Diego Brancaccio[1], Francesco Merlino[1], Ettore Novellino[1], Carlo Alviggi[2], Antonio Randazzo[1]***

[1]Department of Pharmacy, University of Naples Federico II, Via D. Montesano 49, 80131 Naples, Italy.

[2]Department of Neuroscience, Reproductive Medicine, Odontostomatology, University of Naples Federico II, Via S. Pansini 5, 80131 Naples, Italy.

*Corresponding author

E-mail: antonio.randazzo@unina.it

TEL/FAX: +39 081 678514

**Abstract**

Among reproductive-age women, polycystic ovary syndrome (PCOS) is a common endocrine disorder often associated with infertility and insulin resistance. Metabolomics studies on follicular fluid (FF) have shown to provide information about changes in the physiological state of patients as well as biomarkers for oocyte quality. In this context, we performed an exploratory NMR-based metabolomic study on FF samples provided from PCOS women that were undergoing an *In Vitro* Fertilization (IVF) therapy. We found that the hyperinsulinemia in PCOS patients is responsible for a different FF metabolic profile. In particular, hyperinsulinemia seems to be also associated with impaired carbohydrate/glucose and lipid metabolism. Interestingly, anti-müllerian hormone (AMH) turned out to be positively correlated with glucose and negatively correlated with lactate, pyruvate and alanine. To the best of our knowledge, this is the first study that attempts to correlate AMH values with the FF metabolites. Although the data reported here are preliminary, this study paves the way to a better comprehension of the relationships among hormones and metabolites.

## Introduction

Follicular fluid (FF) and cumulus cells (CC) have shown to play a key role in the oocyte health during its maturation. A highly coordinated network of interactions between the oocyte and the somatic cells influences the intrafollicular microenvironment allowing the folliculogenesis (1).

Follicular fluid, also known as *liquor folliculi*, has been described for the first time in 1974 as an exudate from plasma enriched with secretions from the follicles (2). Thus, its composition varies when the plasma is altered by the presence of pathologies, and/or changes in the secretory processes of the granulosa and theca cells occur, reflecting the physiological state of the follicle. The variations in FF composition are strictly connected with both the maturation (3) and quality of the oocyte (4). Indeed, Spitzer et al. observed that FF from immature follicles is characterized by a different protein pattern compared to the fluid collected from mature follicles, suggesting the potential use of selected proteins as biomarkers for follicular maturity (3). The importance of finding biochemical predictors for the oocyte quality was investigated by Revelli et al. (4). They analysed all the correlations reported in the literature between the embryo characteristics and several molecular markers in FF as hormones, growth factors, reactive oxygen species, anti-apoptotic factors, proteins, sugars and prostanoids. However, they were not able to identify substances as reliable markers for assessing the oocyte quality, most probably due to "univariate" scientific approach used. In fact, the same authors suggested to employ a metabolomic approach (based on a multivariate data analysis) that is more suitable to analyse complex biological mixture such as FF.

Metabolomics has been widely employed to analyse almost every kind of biological fluid using both Nuclear Magnetic Resonance (NMR) and mass spectrometry. Compared to the latter, NMR has the advantage of being highly reproducible, requiring minimal sample handling and allowing the identification of a wide range of low-molecular-weight compounds. Pinero-Sagredo and coworkers were the first to perform a NMR study on FF, identifying the presence of at least 42 metabolites (5). Significant correlations among glucose, β-hydroxybutyrate (3-HB), lactate, pyruvate, acetoacetate and acetate were found, thus suggesting the presence of an important anaerobic metabolism in overstimulated follicles. In addition, a statistically significant correlation was also observed between the glycolytic pathway and fatty acid metabolism in both young donors and the group with the higher fertilization rate. This study paved the way to the use of NMR-based metabolomics on FF for the discovery of biomarkers for the oocyte developmental competence (6,7).

Polycystic Ovary Syndrome (PCOS) represents the most commonly occurring metabolic and endocrinological disorder affecting 5–20% of women in their reproductive age worldwide (8,9). It

is characterized by hyperandrogenism, ovulatory dysfunction and polycystic ovarian morphologic features. Being a syndrome, PCOS, is heterogeneous and also associated with other alterations such as repeated and quick gonadotropin-releasing hormone pulses, an excess of luteinizing hormone (LH), and insufficient follicle-stimulating hormone (FSH) secretion, which contributes to an excessive ovarian androgen production and ovulatory dysfunction. Furthermore, evidences of insulin resistance are common among women with PCOS, as a consequence of this, a compensatory hyperinsulinemia promotes adrenal and ovarian androgen production, thus contributing to the hyperandrogenism (10).

One of the main features of PCOS is the increased number of antral follicles. In this frame, the growing interest in Anti-Müllerian Hormone (AMH) as follicular reserve marker has prompted many researchers to investigate the expression of this molecule in women with PCOS, in order to obtain information about the pathogenesis and the ability to calibrate the treatments in the case of ovulation induction or multiple follicular growth for medically assisted procreation cycles.

In order to understand the correlations among the FF composition, the hormones values and the hyper- and normoinsulemic conditions of PCOS women, we performed an exploratory NMR-based metabolomic study on 41 samples of FF provided from women that were undergoing an IVF therapy.

## Materials and methods

*Experimental design*

The women included in the study, ranging from 23 to 38 years old, were diagnosed with PCOS (in both ovaries) according to Rotterdam criteria (11,12). All women were characterized by the following parameters: number of follicles ($n > 12$), diameter of follicles (2 - 9 mm) and/or ovarian volume ($V > 0.10$ $cm^3$). The selected donors showed no genetic anomalies, and no chronic or inflammation diseases. Moreover, they never had ovarian surgery and were not under hormonal therapy during the 6 months prior to the study.

The patients were divided in two subgroups according to specific features that were considered valid to determine a hyper- or normoinsulinemic condition. In particular, three parameters were taken into account to assess the hyperinsulinemic state: the body mass index (BMI > 27 $kg/m^2$), the waist to hip ratio (WHR > 0.85) and the homeostatic model assessment for insulin resistance (HOMA-IR > 2.5). Only patients with positivity to all three parameters where considered belonging to the group of hyperinsulinemic. Conversely, the patients showing a HOMA-IR < 2.5, WHR < 0.85 and BMI < 27 $kg/m^2$ were assigned to the group of normoinsulinemic. For each patient, a pelvic ultrasound with vaginal probe (6.5 MHz) was then performed.

All the individuals involved in the study were undergoing to a GnRH antagonist type therapy for the induction of multiple follicular growth as well as stimulation with recombinant FSH (rFSH). In addition, a folic acid-based treatment and inositol therapy were employed for 5 months. For the final oocyte maturation, 10,000 IU of human chorionic gonadotropin (hCG) were administered, in order to stimulate the ovulation.

*Sample collection*

FF samples from 41 women were provided by the "IVF Center, Department of Neuroscience, Reproductive Science and Odontostomatology", Federico II University Hospital in Naples. Samples were collected by means of transvaginal ultrasound-guided puncture. The fluid was then separated from the oocytes, and stored at -80 °C until analysis, as detailed in previous study (13).

*Serum analysis*

In order to assess the insulin resistance (IR) among the patients, an Homeostatic Model Assessment was performed as described by Matthews et al. (14). Furthermore, the serum concentration of AMH was also measured, being a key factor in the PCOS pathology. The complete serum analysis and the characteristics of the studied population are reported in **Table 1**.

*NMR spectroscopy*

FF samples were thawed at room temperature and then centrifuged at 10,000 rpm for 5 min. NMR samples were prepared mixing 175 $\mu$L of the supernatant with 35 $\mu$L of deuterium oxide ($D_2O$) and 140 $\mu$L of phosphate buffer in order to obtain a pH value of 7.4. The samples were then transferred to 3 mm NMR tubes for the analysis. The spectra were recorded at 25 °C with a 700 MHz Varian Unity Inova spectrometer equipped with a 5 mm $^{1}H\{^{13}C/^{15}N\}$ triple resonance probe.

A Carr-Purcell-Meiboom-Gill (CPMG) pulse sequence was employed in this study in order to reduce the broad resonances from high-molecular-weight compounds, allowing the observation of low-molecular-weight metabolites. A total of 128 scans and 16K data points were used to perform each measurement. The spectra were then processed using iNMR (www.inmr.net). An exponential line broadening of 0.5 Hz was applied to the free-induction decay prior to Fourier transformation.

All acquired NMR spectra were phase- and baseline-corrected, then referenced to the singlet at 1.91 ppm (acetate). Spectral $^{1}H$ assignments were made based on both the literature values (5) and the library provided by the Chenomx database (Chenomx NMR Suite 8.1, Chenomx Inc., Canada).

*Data processing and Multivariate data analysis*

NMR spectra were then imported into MATLAB (R2012b; Mathworks, Natick, MA) where the spectral regions above 9 ppm and below 0 ppm were removed because containing only noise. Furthermore, the region between 4.16 and 5.20 ppm was discarded because it was dominated by the residual water signal. In order to correct for spectral misalignment, the entire dataset was globally aligned with respect to the acetate signal (1.91 ppm) using the icoshift algorithm (15).

The data matrix was then submitted to the PLS toolbox version 8.1.1 (Eigenvector Research, Manson, USA) where the total area normalisation (1-norm) was performed prior to the pareto-scaling and mean-centering in order to compute the Principal Components Analysis (PCA).

PCA is an unsupervised pattern recognition method that allows the reduction of the dimensionality of a data set consisting of a large number of interrelated variables, providing a visual representation of the major variance in the data (16). Thus, the original variables are transformed into a smaller set of new uncorrelated variables, called principal components (PCs), which are ordered according to the explained variance that they are able to retain. Most of the relevant systematic information is usually calculated by the first few PCs, while the following ones are often computed considering chance variation and noise. The outcome of this analysis consists of two plots: a scores plot, where the samples are displayed as scores, and a loadings plot that shows the variables.

The scores are the projection of the data onto the new coordinate system defined by the PCs, whereas the loadings define the size of the contribution of each original variable to the component.

Thus, samples having similar scores will cluster together as well as variables with comparable loadings will be close, meaning that they are highly correlated.

The multivariate analysis described above have been used to study the NMR spectra of the sampled FFs. The NMR dataset was taken into consideration using both the entire set of variables and, in order to reduce model complexity, using only the NMR relative quantification of 14 selected metabolites. In particular, the peak areas of the well-separated resonances were manually integrated and submitted to the data analysis.

Finally, a metabolite-metabolite correlation analysis was conducted on all sampled FFs using the selected integrated NMR peaks. The Pearson correlation analysis generates a coefficient indicated as $r$, which value can range from $-1$ for a perfect negative linear relationship to $+1$ for a perfect positive linear relationship. A value around 0 (zero) indicates no relationship between two variables. Furthermore, to assess the statistical significance of a correlation, the p-value calculation was also performed.

## Results and discussions

A total of 41 FF samples were analysed by $^{1}$H NMR spectroscopy. A representative spectrum of FF sample is shown in **Figure 1**. The FFs were characterized by low-molecular-weight metabolites including amino acids (valine, alanine, glycine, phenylalanine and tyrosine), organic acids (lactate, acetate, acetoacetate, pyruvate and formate) and glucose.

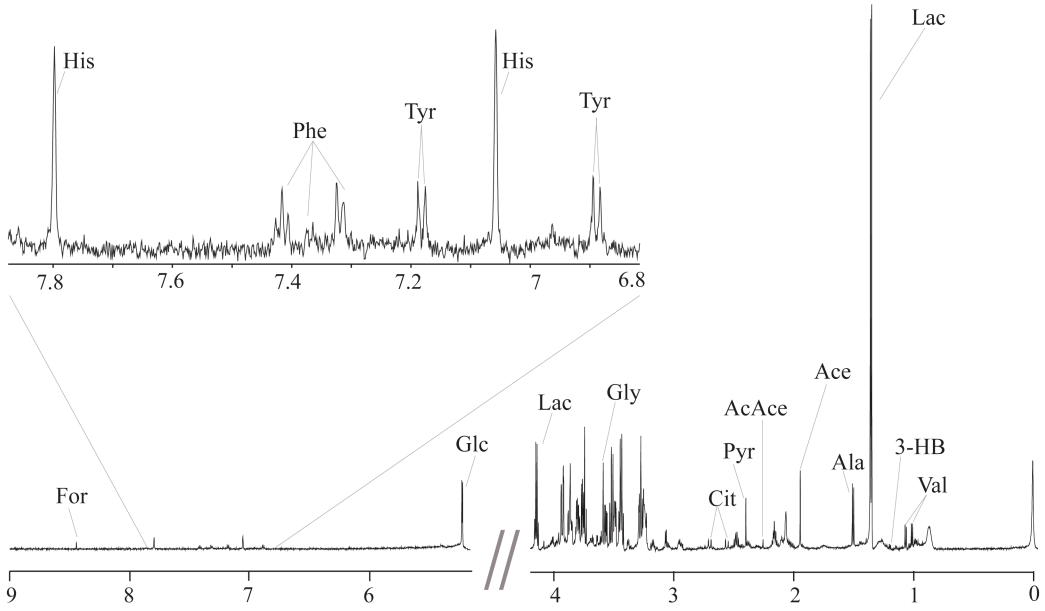


**Figure 1**. $^{1}$H CPMG NMR spectrum of a representative follicular fluid sample measured at 700 MHz and 25 °C. Keys: Val = valine; 3-HB = 3-hydroxybutyrate; Lac = lactate; Ala = alanine; Ace = acetate; AcAce = acetoacetate; Pyr = pyruvate; Cit = citrate; Gly = glycine; Glc = glucose; Tyr = tyrosine; His = histidine; Phe = phenylalanine; For = formate.

An exploratory PCA was first performed on the data set consisting of the complete NMR spectra. The PCA scores plot (PC1/PC2, 48.8% of total variance explained) showed a slight separation of the samples according to the hyper-/normo-insulinemic condition of the PCOS patients along the diagonal direction of the plot (data not shown). Unfortunately, the loading plot of PC1 (26.9% of the explained variance) is actually dominated by signals that cannot be assigned to traditional metabolites (except glycine), suggesting that they may be attributed to contaminant(s) that could affect the interpretation of the PCA. In order to analyze a PCA built by most reliable variables, all the peaks of the proton NMR spectrum that could be unambiguously assigned to known metabolites have been taken into consideration, and therefore integrated. Thus, 14 integral values were used to build a new data matrix that has been augmented by additional 12 clinical variables coming from the clinical measurements (see paragraph *Serum analysis* in Materials and methods section). This approach has the advantage to compare altogether a larger range of information, thus providing a more complete picture of the physiological condition of the studied population. So, a new PCA has

been computed and it is reported in **Figure 2**. As shown in the score plot, a better separation occurs between the hyperinsulinemic and normoinsulinemic PCOS women. The variables that better explain this separation are those having larger loading values along the direction of the sample separation (see the loadings plot in **Fig. 2**). Therefore, the normoinsulinemic group is characterized by high value concentrations of glucose and AMH, while hyperinsulinemic women present higher concentration of lactate, alanine and pyruvate. From a metabolic point of view, glucose, lactate and pyruvate are strictly correlated to each other, being all metabolites of the glycolytic pathway. Interestingly, in normal physiological conditions, the follicles grow in an anaerobic environment since the avascular granulosa layer (that surround them) increases its thickness preventing the oxygen supply in the follicle (17). In particular, granulosa cells (GC) and cumulus cells (CC) are required to provide products of glycolysis for the development of the oocyte, which is unable to carry out this pathway (18). In these conditions, the pyruvate transformation into lactate is obtained due to the limited amounts of oxygen available. Pyruvate is the final key 3-carbon (3-C) intermediate transferred to the oocyte, however it is interchangeable with lactate which is the main 3-C compound produced by the GC.

Likewise, alanine is linked to glycolysis pathway. In fact, alanine can be also produced by reductive amination of pyruvate. The increase of alanine concentration in the insulin-resistant PCOS women (compared to non-insulin resistant patients) is in perfect agreement with a previous study reported by Zhang et al. (19).

As mentioned above, among the variables that mostly contributed to the separation of the normoinsulinemic and hyperinsulinemic women there is also the anti-müllerian hormone (AMH). This hormone is a member of the transforming growth factor-β (TGF-β) superfamily and it is considered an important marker of ovarian reserve (20). AMH values are important markers for PCOS, in fact, women affected with syndrome, have 2- to 3-fold higher levels of AMH than healthy women (21). Interestingly, Takahashi et al. found that oocytes are more likely to be fertilized when follicles are able to make high concentrations of AMH in the follicular fluid. Thus, they speculated that AMH could be considered a prediction marker for fertilization (22). Lin et al. found also a negative correlation between AMH and BMI. This could be in line with our result, as normoinsulinemic women have a lower BMI than the insulin-resistant group. However, the authors didn’t find any correlation between the hormone values and the insulin resistance parameters (23).

The study of the correlations among all the metabolites has been demonstrated to be a valuable analytical tool in systems biology (24). For this reason, the Pearson's correlation analysis of the metabolites was performed. At a threshold of $r \geq |0.5|$, 16 positive and 7 negative significant correlations were found, and the metabolites characterized by the highest correlation values turned

out to be glucose, lactate, pyruvate, acetoacetate and 3-HB. We have found significant ($p < 0.05$) negative correlations between glucose and lactate (-0.78), and between glucose and pyruvate (-0.70). On the contrary, we found significant positive correlations between pyruvate and lactate (0.70), and between acetoacetate and 3-HB (0.95). Interestingly, these data are in agreement with the metabolite correlations found in FF of women not presenting the PCOS (5). These correlations strengthens the hypothesis of an important anaerobic metabolism occurring in the hyperstimulated follicles (17), and the strong correlations among glucose, pyruvate and lactate confirm that FF provides lactate and pyruvate to the developing oocyte as a source of energy. Acetoacetate and 3-HB are synthesised from acetyl-CoA by fatty acid oxidation and their correlation can be explained by the fact that they are interconvertible by means of the β-hydroxybutyrate deidrogenase.

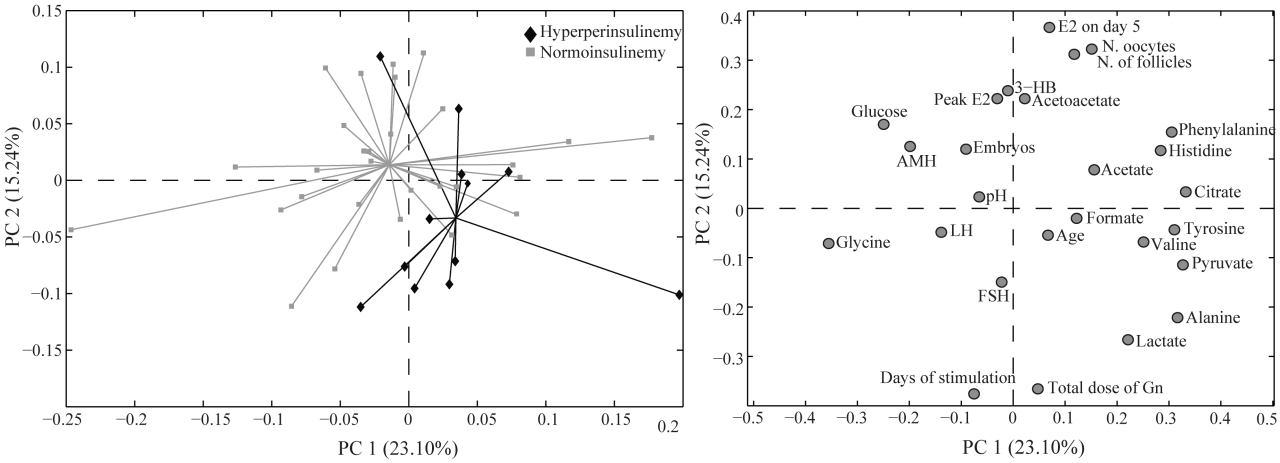


**Figure 2**. PCA scores (left) and corresponding loadings plot (right) derived from the data analysis of the integrated peaks of the $^{1}$H NMR spectra of follicular fluids together with the clinical parameters measured among the PCOS women.

In summary, FF represents a suitable source of information since it is superfluous, abundant and easily available material during the IVF treatment. Metabolomics studies on this fluid can provide useful information about changes in the physiological state of patients, alterations of metabolic pathways, as well as biomarkers for oocyte quality and IVF success rate. In the case reported here, the hyperinsulinemia in PCOS patients is responsible for a different FF metabolic profile. In particular, hyperinsulinemia seems to be also associated with impaired carbohydrate/glucose and lipid metabolism. Interestingly, AMH turned out to be positively correlated with glucose and negatively correlated with lactate, pyruvate and alanine. Although the data reported here are preliminary, this study paves the way to a better comprehension of the relationships among hormones and metabolites. To the best of our knowledge, this is the first study that attempts to correlate AMH values with the FF metabolites.

**Table 1.** Baseline characteristics of the studied population.

| **Patient characteristics** | **Normoinsulinemic** n = 29 samples | **Hyperinsulinemic** n = 12 samples |
|---|---|---|
| Age (years) | 31.3 ± 2.6 | 31.4 ± 3 |
| BMI (Kg/m2) | 22.6 ± 2.3 | 31.4 ± 2.7 |
| HOMA | 1.4 ± 0.6 | 3.4 ± 0.7 |
| FSH (IU/L) | 5.6 ± 2.0 | 5.7 ± 1.4 |
| LH (IU/L) | 6.4 ± 3.9 | 5.9 ± 2.6 |
| AMH (IU/L) | 5.7 ± 4.1 | 2.9 ± 1.3 |
| E2 on day 5 (pmol/L) | 432.0 ± 310.4 | 173.4 ± 179.1 |
| E2 peak (pmol/L) | 1111.5 ± 682.6 | 1363.7 ± 657.8 |
| No. of follicles | 10.7 ± 4.7 | 8.4 ± 3.6 |
| No. of MII oocytes | 8.8 ± 4.1 | 7.4 ± 3.8 |
| Total dose of Gn (IU/L) | 1100.6 ± 471.6 | 1657.5 ± 707.1 |
| Days of Gn stimulation | 9.8 ± 3.3 | 11.2 ± 2.4 |

Data are reported as mean ± standard deviation.

BMI: body max index; HOMA: homeostasis model assessment; FSH: follicle stimulating hormone; LH: luteinizing hormone; AMH: antimullerian hormone; E2: Estradiol.

**References:**


1. Dumesic DA, Meldrum DR, Katz-Jaffe MG, et al. Oocyte environment: Follicular fluid and cumulus cells are critical for oocyte health. Fertil Steril 2015;103(2):303–16.
2. Edwards RG. Follicular Fluid. J Reprod Fertil 1974;37(1):189–219.
3. Spitzer D, Murach KF, Lottspeich F,et al. Different protein patterns derived from follicular fluid of mature and immature human follicles. Hum Reprod 1996;11(4):798–807.
4. Revelli A, Delle Piane L, Casano S, et al. Follicular fluid content and oocyte quality: from single biochemical markers to metabolomics. Reprod Biol Endocrinol 2009;7:40.
5. Pinero-Sagredo E, Nunes S, de Los Santos MJ, et al. NMR metabolic profile of human follicular fluid. NMR Biomed. 2010;23(5):485–95.
6. O'Gorman A, Wallace M, Cottell E, et al. Metabolic profiling of human follicular fluid identifies potential biomarkers of oocyte developmental competence. Reproduction. 2013;146(4):389–95.
7. Wallace M, Cottell E, Gibney MJ, et al. An investigation into the relationship between the metabolic profile of follicular fluid, oocyte developmental potential, and implantation outcome. Fertil Steril. 2012;97(5).
8. Goodarzi MO, Dumesic DA, Chazenbalk G, et al. Polycystic ovary syndrome: etiology, pathogenesis and diagnosis. Nat Rev Endocrinol 2011;7(4):219–31.
9. Azziz R, Carmina E, Dewailly D, et al. The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report. Fertil Steril 2009;91(2):456–88.
10. McCartney CR, Marshall JC. Polycystic Ovary Syndrome. N Engl J Med 2016;54–64.
11. Azziz R. Diagnosis of polycystic ovarian syndrome: The Rotterdam criteria are premature. J Clin Endocrinol Metab 2006;91(3):781–5.
12. Azziz R, Carmina E, Dewailly D, et al. Position statement: Criteria for defining polycystic ovary syndrome as a predominantly hyperandrogenic syndrome: An androgen excess society guideline. J Clin Endocrinol Metab 2006;91(11):4237–45.
13. Alviggi C, Cariati F, Conforti A, et al. The effect of FT500 Plus ® on ovarian stimulation in PCOS women. Reprod Toxicol. 2016;59:40–4.
14. Matthews DR, Hosker JP, Rudenski AS, et al. Homeostasis model assessment: insulin resistance and β-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia 1985;28(7):412–9.
15. Savorani F, Tomasi G, Engelsen SB. icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. J Magn Reson 2010;202(2):190–202.

16. Hotelling H. Analysis of a complex of statistical variables into principal components. J. Educ. Psychol. 1933; 24: 417–441.
17. Gull I, Geva E, Lerner-Geva L, et al. Anaerobic glycolysis: The metabolism of the preovulatory human oocyte. Eur J Obstet Gynecol Reprod Biol. 1999;85(2):225–8.
18. Sugiura K, Pendola FL, Eppig JJ. Oocyte control of metabolic cooperativity between oocytes and companion granulosa cells: Energy metabolism. Dev Biol. 2005;279(1):20–30.
19. Zhang C-M, Zhao Y, Li R, et al. Metabolic heterogeneity of follicular amino acids in polycystic ovary syndrome is affected by obesity and related to pregnancy outcome. BMC Pregnancy Childbirth 2014;14(1):11.
20. Van Rooij IAJ. Serum anti-Mullerian hormone levels: a novel measure of ovarian reserve. Hum Reprod 2002;17(12):3065–71.
21. Pigny P, Merlen E, Robert Y, et al. Elevated serum level of anti-mullerian hormone in patients with polycystic ovary syndrome: relationship to the ovarian follicle excess and to the follicular arrest. J Clin Endocrinol Metab 2003;88(12):5957–62.
22. Takahashi C, Fujito A, Kazuka M, et al. Anti-Müllerian hormone substance from follicular fluid is positively associated with success in oocyte fertilization during in vitro fertilization. Fertil Steril 2007;89(3):586–91.
23. Lin Y-H, Chiu W-C, Wu C-H, et al. Antimüllerian hormone and polycystic ovary syndrome. Fertil Steril. 2011;96(1):230–5.
24. Steuer R. Review: On the analysis and interpretation of correlations in metabolomic data. Brief Bioinform. 2006;7(2):151–8.