Alma Mater Studiorum – Università di Bologna

in cotutela con Università di Copenhagen

DOTTORATO DI RICERCA IN

SCIENZE E TECNOLOGIE AGRARIE, AMBIENTALI E ALIMENTARI

Ciclo XXIX

Settore Concorsuale di afferenza: 03/B1

Settore Scientifico disciplinare: CHIM/03

A 'FOODOMIC' APPROACH FOR THE EVALUATION OF FOOD QUALITY AND ITS IMPACT ON THE HUMAN METABOLOME

Presentata da: TRIMIGNO Alessia

Coordinatore Dottorato
Giovanni DINELLI

Relatore
Francesco CAPOZZI

Relatore
Søren Balling ENGELSEN

Esame finale anno 2017

A 'FOODOMIC' APPROACH FOR THE EVALUATION OF FOOD QUALITY AND ITS IMPACT ON THE HUMAN METABOLOME

ABSTRACT

In recent years, omic sciences have been increasingly employed in a multitude of research fields thanks to their high-throughput capabilities and holistic approach.
Among the omic sciences, metabolomics and foodomics have recently emerged in the investigation of food and nutrition and their relation to the individual health and wellness status (Chapter 1).
The analytical platforms used are ideal for non-targeted analysis, due to their capability of detecting and identifying a large set of variables (or metabolites) in complex biological samples. The most employed metabolomics techniques are mass spectrometry and nuclear magnetic resonance spectroscopy (Chapters 2.1, 2.2 and 2.3), which can reach high analytical power in the non-targeted realm thanks to the advent and advancement of multivariate data analysis (Chapter 2.4).
This thesis, which is the result of three years of PhD studies carried out both in Italy and in Denmark, outlines the analytical pipeline of the foodomic approach and highlights the current challenges in the field (Chapter 2.3). The thesis traces the path of modern foodomics and metabolomics from the definition and description of food quality (Chapters 3 to 6), to the profiling of the metabolome (Chapters 7 to 8.5), and finally the investigation of the impact of food on the human health, the prevention of diseases, and the identification of biomarkers of health status (Chapters 8.6 and 8.7).
The impact of factors such as genetic modification or farming method was investigated in grapes, defining the specific changes in the fruit metabolism correlated with these factors (Chapter 4). For example, a linear decrement in the content of malic acid in GM grapes in relation to the number of transgene copies inserted was observed, showing a shift in the plant metabolism due to the increased fruit yield. Moreover, the impact of the food matrix and digestion processes on the stability and bioaccessibility of specific, beneficial molecules was assessed (Chapters 5 and 6).
The animal metabolome was also investigated as a model of the human body. In one of the studies reported, the effect of antibiotic treatment on necrotizing enterocolitis was investigated (Chapter 7.2), proving that the employment of the selected drugs successfully treated the condition and impacted on the gut microbiota. This model can be of great help in the investigation of the

condition in human newborns, too.
Many studies have also been carried out on the human metabolome (plasma, serum, urine). These studies have helped in the development of specific algorithms for the search of dietary biomarkers in observational studies (Chapter 8.6.1). Moreover, food intake biomarkers have been discovered in an intervention study (i.e. galactose for milk intake) and will be further investigated (Chapter 8.6.2). Research was also carried out to investigate on specific disease biomarkers and to discover possible trajectories from a disease state to a healthier condition (Chapter 8.7).

INDEX

Introduction

Chapter 1 WHAT IS FOODOMICS?
1.1 The pathway from metabolomics to foodomics to nutrimetabolomics
REFERENCES CHAPTER 1

Chapter 2 MATERIALS AND METHODS
2.1 ANALYTICAL TECHNIQUES: Mass Spectrometry
2.2 ANALYTICAL TECHNIQUES: NMR Spectrometry
2.2.1 Principles
2.2.2 The NMR Spectrometer
2.2.3 NMR spectra characteristics
2.2.4 NMR techniques
2.3 NMR Data Processing
2.3.1 Phase and Baseline correction
2.3.2 Referencing, Alignment and Binning
2.3.3 Removal of unwanted regions
2.3.4 Normalization
2.3.5 Scaling
2.4 MULTIVARIATE DATA ANALYSIS
2.4.1 Multivariate Statistical Analysis
2.4.2 Principal Component Analysis (PCA)
2.4.3 Partial Least Squares Regression (PLS)
2.4.4 ASCA
2.4.5 MultiLevel Analysis
2.5 PEAK ASSIGNMENT
2.5.1 Assignment principles
2.5.2 Assignment through online tools
2.5.3 Assignment through 2D-NMR spectroscopy
2.5.4 Assignment through softwares
REFERENCES CHAPTER 2

Chapter 3 DEFINING FOOD QUALITY
REFERENCES CHAPTER 3 (INTRODUCTION)
3.1 Describing food quality through metabolomics
REFERENCES CHAPTER 3.1
3.2 Definition of food quality by NMR-based foodomics
REFERENCES CHAPTER 3.2
3.3 Definition of meat quality through metabolomics
REFERENCES CHAPTER 3.3

Chapter 4 FOODOMICS FOR THE INVESTIGATION OF FOOD QUALITY IN PLANT-DERIVED PRODUCTS
4.1 Investigation on the impact of farming system
REFERENCES CHAPTER 4.1
4.2 Foodomics to investigate genetic modification
REFERENCES CHAPTER 4.2

Chapter 5 FOODOMICS TO INVESTIGATE THE “MATRIX EFFECT”
5.1 Investigation on the stability of three different food matrices
5.2 Investigation on the stability of enriched pancakes
5.3 Final conclusions
REFERENCES CHAPTER 5

Chapter 6 FOODOMICS TO INVESTIGATE THE DIGESTION PROCESS
6.1 Genetic impact on digestibility
6.2 Processing impact on digestibility

Index

6.3 Cooking impact on digestibility
6.4 Digestibility of food supplements
REFERENCES CHAPTER 6

Chapter 7 FOODOMICS IN THE INVESTIGATION OF THE ANIMAL METABOLOME
REFERENCES CHAPTER 7 (INTRODUCTION)
7.1 Foodomics to investigate pig olostrums
REFERENCES CHAPTER 7.1
7.2 Foodomics for the investigation of the pig metabolome
REFERENCES CHAPTER 7.2

Chapter 8 THE HUMAN METABOLOME
8.1 Urine
8.2 Plasma and serum
8.3 Faeces
8.4 Research work overview
8.4 Foodomics for the Investigation of the Human Urinary Metabolome – The CHANCE Project
8.6 Foodomics for the Search of Food Intake Biomarkers
8.6.1 The quest for dietary biomarkers in observational studies
8.6.2 Search of food intake biomarkers in an intervention study
8.7 Foodomics for the definition of the health status
8.7.1 Foodomics for the search of disease specific biomarkers
8.7.2 Foodomics for the investigation of the dietary changes in breast cancer patients
REFERENCES CHAPTER 8

Chapter 9 Conclusions

Annex 1 – Materials and Methods (SOPs)
- Equipment
- Software
- Main SOPs employed

REFERENCES ANNEX 1

Annex 2 – List of Abbreviations

CHAPTER 1: WHAT IS FOODOMICS?

1.1 The pathway from metabolomics to foodomics to nutrimetabolomics

The post-genomic or “omic” sciences are considered incredibly useful tools for the investigation of the underlying mechanisms in human biochemistry and health [Lenz et al., 2003] and to describe normal and shifted biological functions since they study genes (genomics), gene expression (transcriptomics), protein profiles (proteomics) and, finally, metabolic profiles (metabolomics). [Kochhar et al., 2006]. Many definitions have been given for the term **metabolomics**. It has been called the science capable of providing a “functional readout of the physiological state of an organism” [Roessner & Bowne, 2009], “bridging the gap between genotype and phenotype” [Roessner & Bowne, 2009] or a “comprehensive strategy to address the health challenges facing the modern world” [German et al., 2005]. In any case, though, this new omic science targets the investigation and profiling of the metabolome, the set of all the low-molecular weight molecules (metabolites) in a cell, tissue or organism, taking part in all the key metabolic reactions. The complete set of the metabolome, which depends on the species and organism considered, can vary from 600 metabolites in yeast to over 20 thousands in humans [Fiehn, 2002; Dunn et al., 2005; Bernini et al., 2009]. Being founded on a holistic approach, metabolomics is not just helpful for the study of particular organism, but it has given a great boost also to food science. The field of food science, in effect, has been increasingly correlated to other research fields such as medicine, biology, veterinary and agricultural sciences or genetics. For this reason the metabolomic approach could successfully help in the investigation of the relations between these fields. Metabolomics can in fact help in the study of the effects of food quality on health, together with the assessment of the shifts in food qualities and characteristics caused by many external factors, such as human or environmental. Therefore, **foodomics** has been developed as the application of metabolomics to the investigation of the relation between food quality and nutrition. Thus, an increasing number of research studies has been and is now focused on foodomics, assessing for example the impact on food quality and nutritional value of genetic modification, geographical origin or new processing techniques and technologies (see Chapters 3-4), since factors like these have an expanding importance in consumers' choice and opinion. Another important characteristic that is often investigated is digestibility (see Chapter 6), since many nutrients, especially the ones added in enriched foods, have to be correctly assimilated by our bodies. For all these reasons, the capability of describe food quality throroughly is a paramount tool for both food producers and consumers

[Capozzi & Trimigno, 2015]. The power of metabolomics and its potential in many applied sciences has also been explored, as aforementioned, in the assessment of the link between nutrition, food and human health. Due to the aging of population and the increasing number of chronic diseases correlated to disfunctional metabolic functions and nutritional habits, new approaches for this kind of investigation were developed. Traditional nutritional research, based on typical measurements on single or a few nutrients has now been replaced by more high-throughput approaches, founded on the study of genes, gene expression, proteins and metabolites. The so called **nutritional metabolomics** (or nutrimetabolomics) is thus the application of metabolomics to nutrition sciences for the determination of the metabolic phenotype (metabotype) and its variations due to diet and its correlated metabolic and biological impairments (See Chapter 8). This approach could help in the future to develop targeted diets to match specific needs for each individual in order to improve or mantain a good health status. Up to now, though, much research has been done and still needs to be carried out, on the separation of all the effects of intrinsic and extrinsic factors playing a role in the characterization of metabolomes. These factors, apart from diet, can be genetics, use of specific drugs, lifestyle, gut microbiota, presence of particular diseases, etc. It is therefore important to understand the role of each stressor and their combinate effects [Rezzi et al., 2013] and to determine the stable part of the metabolome, the so-called metabolic fingerprint or metabotype, to then be able to better understand the shifts from this condition. The complex interaction between all the described factors and the human health status is depicted in Figure 1.

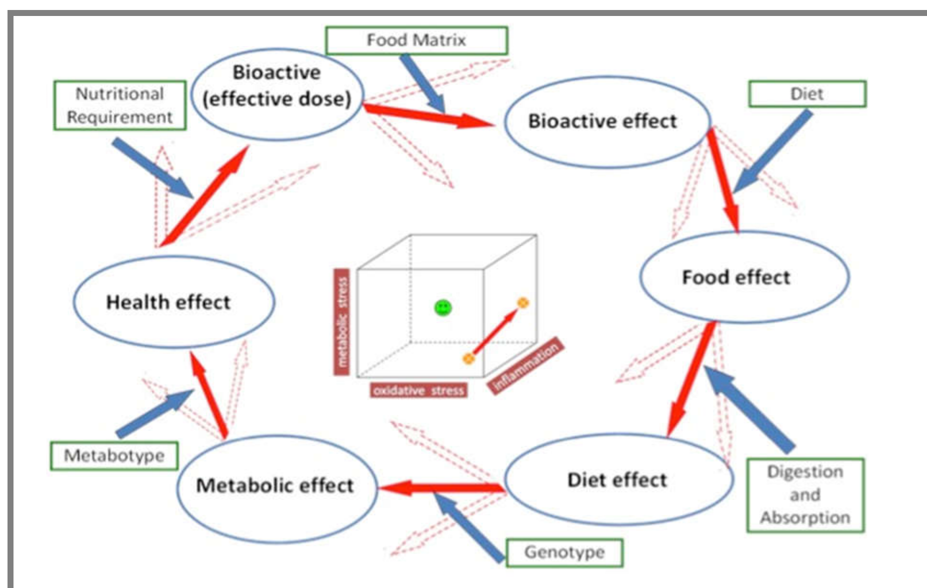


Fig. 1: Definition of the health status of a human being and the complex factors and effects impacting on it and causing its shift from the ideal state [Bordoni & Capozzi, 2014].

The wellbeing of a subject can in fact be defined in a three-dimensional space set by the

coordinates of: oxidative stress, inflammation status and metabolic stress. In each moment, each individual will be in a defined position, which can shift towards better or worse conditions in relation to the impact of many internal and external factors, which are, in turn, correlated to each other and interdependent. It is thus of great interest to dig into this intricate net of topics and characteristics in order to shed a light into the better understanding of human health. Each of the factors influencing the various effects outlined above are investigated in this research work, highlighting the power of the foodomic approach in modern research. The techniques mainly employed in the foodomic research field are described in Chapter 2.

REFERENCES CHAPTER 1

Bernini, P., Bertini, I., Luchinat, C., Nepi, S., Saccenti, E., Schäfer, H., Schütz, B., Spraul, M., & Tenori , L. (2009) “Individual Human Phenotypes in Metabolic Space and Time” Journal of Proteome Research, 8(9), 4264–4271.

Bordoni, A., & Capozzi, F. (2014) “Foodomics for healthy nutrition” Current Opinion in Clinical Nutrition & Metabolic Care, 17(5), 418-424.

Capozzi, F., & Trimigno, A. (2015). “Using metabolomics to describe food in detail” Metabolomics as a Tool in Nutrition Research. Edited by Sebedio JL, Brennan L. Woodhead Publishing, 203-228.

Dunn, W. B., Bailey, N.J.C., & Johnson, H.E. (2005) “Measuring the metabolome: current analytical technologies” - Analyst, 130(5), 606–625.

Fiehn O. (2002)“Metabolomics – the link between genotypes and phenotypes.” Plant Molecular Biology, 48(1-2), 155 – 171.

German, J.B., Hammock, B.D., & Watkins, S.M. (2005) “Metabolomics: building on a century of biochemistry to guide human health.” Metabolomics, 1(1), 3-9.

Kochhar S., Jacobs D. M., Ramadan Z., Berruex F., Fuerholz A. & Fay L B. (2006) “Probing gender-specific metabolism differences in humans by nuclear magnetic resonance-based metabonomics.” Analytical Biochemistry, 352(2), 274–281.

Lenz, E.M., Bright, J., Wilson, I.D., Morgan, S.R. , & Nash, A.F.P. (2003)“A 1H NMR-based metabonomic study of urine and plasma samples obtained from healthy human subjects” Journal of Pharmaceutical and Biomedical Analysis, 33(5), 1103-1115.

Rezzi, S., Collino, S., Goulet, L., & Martin, F.P. (2013) “Metabonomic approaches to nutrient metabolism and future molecular nutrition” TrAC Trends in Analytical Chemistry, 52, 112-119.

Roessner, U. & Bowne, J. (2009) “What is metabolomics all about?” BioTechniques, 46(5), 363-365.

CHAPTER 2: MATERIALS AND METHODS

2.1 ANALYTICAL TECHNIQUES: Mass Spectrometry

One of the most employed metabolomics techniques is mass spectrometry. Mass spectrometry (MS) is based on molecular ionization and on the subsequent fragmentation of the molecule in ions in relation to the different mass/charge ratio (m/z). Since molecules are ionized in gas phase through the expulsion of an electron, MS is a destructive technique. This ionization forms the so-called "molecular ion", a radical cation, which in turn will be fragmented, generating both neutral molecules and/or radicals (not detectable) and cations and/or radical cations (detectable). The spectrometer will thus discriminate between the molecular ion and its fragments in relation to the M/z ratio and these will be observed by a final detector. The output of MS analysis is a mass spectrum, which represents the relative abundance of the different ions in relation with the M/z value. The mass spectrum is composed by an x axis, corresponding to values of mass, and a y axis, reporting the intensity of the specific mass value of a determined fragment. Thus, the maximum mass value will represent the molecular ion.

In order to interpret the mass spectrum, the steps to follow will be:

- identification of the molecular ion
- identification of the most characteristic ions
- recognition of the fragmentation pattern
- identification of the molecule

Through MS it is thus possible to identify metabolites from their fragmentation profiles. Usually, separation techniques such as gas chromatography (GC) or high performance liquid chromatography (HPLC) are employed in association with MS, in order to help the separation of metabolites and the later identification.

2.2 ANALYTICAL TECHNIQUES: NMR Spectroscopy

2.2.1 Principles

In a NMR experiment, the sample is inserted into a probe, which, in turn, is immersed into a strong magnetic field generated by a permanent magnet. The nuclei contained in the sample spin with an angular momentum P. Since they are also charged, they will produce a magnetic moment μ. At first, all the magnetic moments of the nuclei are randomly oriented, but later become aligned in the magnetic field and their sum gives a weak magnetization (M) to the sample. Nuclei magnetic moments also have precessions at high frequency (MHz – radio frequencies) in the magnetic field.

This precession frequency (ωL) is proportional to the static field strength (B0) by a constant called gyromagnetic ratio (γ). An alternating current at radio frequency (RF) is run through the inductor (wire coil) arond the sample to produce an oscillating field (B1(t)). The precessing nuclei can absorb the energy of the oscillating magnetic field if the energy is on resonance, thus at the same frequency of the nuclei's precession. The circuit of the coil will both excite the nuclei and detect the RF. The magnetic moments can be aligned parallelly or anti-parallelly to the static magnetic field B0 due to quantization. The subset of nuclei aligned paralelly to the field is slightly bigger and the energy difference between the parallel and anti-parallel population is called ΔE. If an energy equal to is provided by B1(t), which could have a frequency ν so that $\Delta E = h\nu$, magnetic moments can re-orient. This frequency ν is the precession frequency $\nu L = \omega L / 2\pi$. B1(t) is actually the sum of two fields rotating in opposite directions, and one of them is capable of causing resonance. As said, when this oscillating field is on, the magnetic moments will get in phase and generate a vector of magnetization M, which follows B1(t) in the rotating frame. The excitation field is applied in short pulses and sample magnetization will precess about it, with an angle of rotaton $\Delta\theta$. The time in which it rotates is called pulse width (Δt) and if the angle $\Delta\theta$ is 90°, then the magnetization will be on the x' axis and the detection on the transverse plane (x,y). After each pulse, the magnetization will continue to precess in the transverse plane, responding to B0 and inducing an electromotive force in the coil. Pulse will therefore be detected by electromagnetic induction. The precession frequency of the nuclei is modified by the magnetic shielding of electron clouds in the molecule (so-called “chemical shift effect”) and the magnetic coupling with adjacent nuclei (effects such as the J-coupling, the spin-spin or the multiplet). For this reason, the signal measured will have a frequency related to chemical structure information on the nuclei and can give a great insight on the sample's characteristics. Nuclear relaxation will make the magnetization die. The magneization has a component parallel to B0 (longitudinal) and one perpendicular to B0 (transverse). The induction signal, as aforementioned, is detected in the transverse plane, thus the signal will reflect the decay of magnetization of the transverse component. Another relaxation will be witnessed in the experiment, since also the longitudinal component will return to its equilibrium. This is the longitudinal or spin-lattice relaxation, characterized by the time constant T1, though, not being in the transverse plane, it will not be measured directly, but only with inversion or saturation recovery experiments. The Free Induction Decay (FID) will give two main information:

- the number of resonant nuclei, thus the concentration, since the amplitude at t=0 is proportional

- physical information on the sample since the decay rate is modified by molecuar motions (i.e. slow in liquid phases due to fast motion).

The NMR signal will thus have a decay rate in the transverse component magnetization called spin-spin (or transverse) relaxation, composed by both the nuclear relaxation and the inhomogeneity of B0. The obtained FID curve will be described by a time constant called T2*, whilst T2 is the true nuclear relaxation time. The NMR signal in time-domain will thus give info on the physical state of the sample but usualley what is important is the chemical information. For this reason, a Fourier Transform (FT) must be applied to obtain the magnetic resonance spectrum we usually employ. If the FT is applied to the decay of solid sample, the spectrum will have a single and broad resonances, not giving out chemical information, whilst with liquids narrow resonances are obtained, with line widths smaller than the separation between the different resonances, so chemical information can be acquired. When a sample has many components, each will have its own resonance, so the spectrum will be a resulting pattern of resonances. The higher the field strength, the better the separation among the resonances and greater sensitivity. The main characteristics of the peaks which appear in a spectrum and give information on the sample are:

- peak area, related to the amount of substance
- frequency or chemical shift, linked to the chemical type
- line shape, giving information on the state of the sample and on its molecular structure

The line width is related to T2* by the equation $\Delta\nu=1\pi/T2^*$.

2.2.2 The NMR Spectrometer

The NMR instrument will constitute of the following parts:

- a superconducting stable magnet
- a probe
- a high-power RF transmitter delivering short pulses
- a receiver amplyfying NMR signals
- an Analogue to Digital Converter (ADC) to convert signals into a storable form
- a pulse programmer
- a computer

To generate the B0 field, as stated, a stable and permanent superconducting magnet is employed. This magnet is made by a wire coil through which current flows and generates the magentic field. This wire is superconducting, so at very low temperatures the resistance will be equal to 0, thus the current, once set, will go on forever and no electrical power is needed for the magnetic field.

The superconducting state of the wire is maintained by having it immersed in a liquid helium bath. Outside of this helium layer, stands a heat shield, kept at constant temperature (77K) by a liquid nitrogen bath. All of this is contained in a vacuum flask so that the heat flow is reduced. The sample, though, will have to be around room temperature, so there is a zone called bore tube, built in the magnet, where room temperature is kept. The bore tube consists of a vertical tube going through the magnet and the magnetic field will be in its direction. The probe consists in a metal cylindrical tube. It is inserted into the bore tube. On top of the probe, the small coil which excites and detects the signal is hold in a way that the sample can drop into it from the top of the magnet. The coil needs to be as close as possible to the sample, though allowing it to drop down. This coil will formed a tuned circuit with the coil and a capacitor so that the inductance and the capacitance they form resonates at Larmor frequency. The capacitor needs to be adjusted so that the tuned circuit resonates at the Larmor frequency. This operation is called tuning and it is very sensitive to solvent or ion concentration changes. In addition, a matching operation needs to be carried out, for which the circuit design is adjusted to maximize power transfer between the probe and the transmitter/receiver. The probe might also contain arrangements for the cooling or heating of the sample. The part of the instruments generating pulses is the RF transmitter. The RF source produces a stable frequency that can be precisely defined by the operator, so that the transmitter can be moved to various part of the spectrum (i.e. In selective excitation experiments). The RF source is usually a frequency synthetizer, since it can generate a stable frequency and this can be set by the operator. In addition such RF source is readily controlled by a computer interface.

The synthetizer output is gated so that it creates the pulse only for the short time that it is needed. The gate will be controlled by a computer to define pulse length and time. The RF source is usually at the level of few mW so, before being applied to the probe, it will need to be increased to provide a useful oscillating field. The B1(t) will be more intense the stronger the power applied to the probe is and the resulting 90° pulse length will be shorter. Nonetheless, there is a maximum power, stated by the manufacturer, that can be applied, after which a discharge will be generated and this can destroy coil and capacitor and give out erratic oscillating fields. The NMR signal generated from the probe is of the order of µV, thus it will be amplified before being digitized. Modern amplifiers are low-noise, designed so that only minimum extra noise is introduced. A first amplifier, the pre-amp, is as close to the probe as possible, generally at the foot of the magnet, to boost the signal before sending it to the spectrometer console. A diplexer is also present, as a fast switch is needed to prevent the high-power pulse to go into the sensitive receiver. This switch

works so that when the pulse is off, the receiver is connected to the probe and the transmitter not, whilst when the pulse is on, RF flows to the probe and the receiver is disconnected. Diplexers can require external power, while others are passive. One of the last parts of the instrument is the ADC, converting the NMR signal from voltage to binary, so that it can be stored in computer memory. The signal is sampled by the ADC at regular intervals and the FID is thus represented in data points. The number of bits used by the ADC will define the largest output number. The larger the bits number, the more output levels present. Generally ADC use between 16 and 32 bits. Since the ADC uses approximation of the FID, small sidebands (digitization sidebands) near the base of peaks in the spectrum, which can be a problem when a they belong to a strong peak and can hide a close weak peak. These sidebands can be reduced by improving the resolution.

2.2.3 NMR spectra characteristics

The main type of information retrieved by a NMR spectra derive from the number of signals, their position and intensity and their splittings. Protons that are chemically equivalent, thus in the same magnetic environment, will absorb at the same frequency, whilst protons in a compound with different magnetic environment will five separate signals. Protons in a CH_3 or in a CH_2 group will be generally equivalant, as protons in symmetrical compounds (i.e. benzene). Though, if protons differ in any possible way (also stereochemically), they are not equivalent and show separate signals in the spectrum. This happens very often since most compounds are not symmetrical. The position were signals arise along the x-axis of frequency is the so-called chemical shift and it is measured in ppm (parts per million). The chemical shift will depend solely on the changing local magnetic fields from adjacent protons and not from the Earth or the NMR magnetic field, since those affect similarly all protons in the compound. The chemical shift can be influenced by various effects. Firstly, when a magnetic field is applied, the electrons in the electron clouds will produce a small but opposing magnetic field that will lower the effective one perceived by protons and will shift the signal to the right of the spectrum (upfield). This is the local diamagnetic shielding, since the electron cloud shieds the proton from the applied magnetic field. If a proton is attached to electronegative atoms, its signal will be more downfield. Moreover, when electronegative atoms are present, they will remove electrons from the electron cloud, resulting in a lower shielding effect, moving signals lowfield (higher chemical shift). The deshielding effects is inversely proportional to the distance between the proton and the electronegative atom. If a molecule is charachterized by anisotropy, since the orbitals will not be spherically symmentrical, the proton will experience different shielding effects to the applied magnetic field. Similar functional groups

will have similar chemical shift, though the variability of molecules will make it difficult to assign with 100% certainty a specific chemical shift to a specific type of proton. The integral of a signal will be related to the numbers of protons responsible for it, thus the number of proton equivalents. This means that the area under the signal will give information on the number of protons and on the concentration of that molecule. NMR signals, though, will not be simply single signals, but are usually split in doublets, triplets, quartets or multiplets. When signals are splitted the distance between the resulting peaks is called J, or coupling constant. Nearby protons can interact and orient in a parallel or antiparallel way to the applied magnetic field, producing different spin filp energies. When a proton is so close to another than can affect its spin, we have spin-spin coupling. The splitting will always be returned, so if Ha is splitting Hb, Hb will in turn split Ha. For this reason, splitting will give important information on the closeby protons of a proton in a molecule. The general rule is if that a proton has n neighbours, the signal will be split into n+1 peaks. Other rules that apply for the spin-spin coupling are:

- protons must not be equivalent to couple: if nuclei are isochronous, thus in the same chemical shift, they will not couple.
- If protons are separated by three bonds (vicinal) they can couple, if they are separated by more than three bonds they cannot couple, since they feel a very weak signal from their neighbours. An exception to this is allylic coupling, where coupling occurs through 4 bonds if one is a pi-bond; this happens because the pi-bond will have a higher electron density. In addition, all hydrogens on a benzene will couple with each other, though they cannot couple with hydrogens attached to other atoms even within the limit of three bonds.
- If a hydrogen is linked to a nitrogen or oxygen it generally will not couple and give out singlets in the spectra.

In addition to first-order splitting, producing normal splitting patterns with equal J constants, non-first order splitting will not be predicted by the N+1 rule. This happens when a proton couples with two or more sets of neighbouring protons with different values of J.

2.2.4 NMR techniques

The main spectroscopical techniques employed in this thesis are NOESY and Carr-Purcell Meiboom-Gill (CPMG). A NOESY experiment will show correlation of protons through dipolar coupling, observed only in solid or partially oriented phases. The correlation shown is among protons close in space, no matter if that is because of bonds or not. The Nuclear Overhauser Effect (NOE) will show also a change in the intensity of a signal if the signal of a couple proton is saturated. To wit,

NOE occurs when the integrated absorption intensity of a spin changes if another nearby spin is saturated. The NO effect is based on the behavior in relaxation and can bring both reduction and enhancement of signals. The NOESY pulse sequence, using the NOE principle, consists of three pulses of $\pi/2$ of which:

- the first one creates the transverse spin magnetization
- the second one produces a longitudinal magnetization equal to the transverse component perpendicular to the pulse direction
- the third pulse generates a transverse magnetization from the remaining longitudinal one

After the third pulse, acquisition starts, observing the transverse magnetization as a function of t2. NOESY experiments also have a mixing period τm, initialised during the second pulse and kept constant throughout all the 2D experiment. The CPMG sequence allows the measurement of spin-spin T2 relaxation times. The half-height linewidth measured in the experiment (d) at a defined resonance is related to T2* in this way: $d=1/(\pi T2^*)$

The pulse sequence, based on the spin-echo pulse one, procedes this way:

- a 90° pulse generates transverse magnetization
- a spin-echo period will determine the decay of the transverse magnetization. This period (delay-180°-delay block) is repeated for a n number of times
- acquisition

The CPMG experiment can measure conventional 1D spectra, where signal intensity are dependent on the spin-echo length. If the spin-echo period is short, all signals will have positive intensities, though when the period is increased, intensity will decrease, until the period is long that the intensities of signals become null. Due to this, CPMG experiments are particularly useful when broad signals have to be neutralized, such in the case of samples with protein and large polypeptides, like plasma or serum (Figure 1). In these cases, CPMG is applied with parameters that will prevent the total suppression of informative and sharp signals, while removing the broad bands which can interfere with later analysis. After acquisition, NMR spectra need to undergo a set of pre-processing steps which can greatly impact on the results of the following data analysis and thus need to be carried out carefully. These steps are reported in the following chapter (Chapter 2.3). Both NMR and MS have recently reached great analytical power due to the improvement of multivariate statistical analysis, which is described in chapter 2.4. Dealing in fact with large datasets characterized by the many different factors, it is fundamental to have tools capable of dealing with this type of matrices and of carrying out complex data mining procedures.

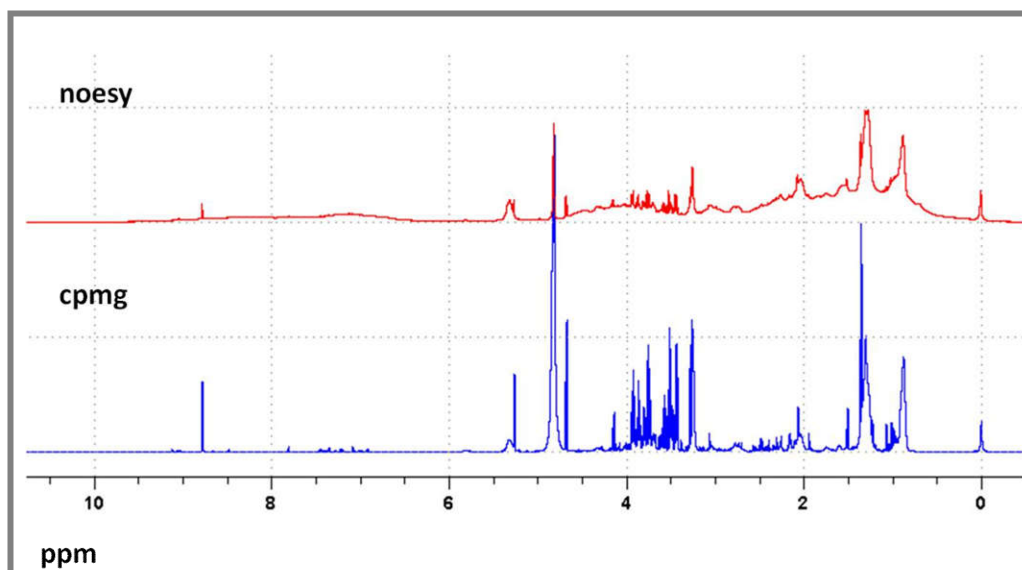


Fig. 1: Serum samples analyzed by NMR spectroscopy employing either a 1D NOESY (red) or a 1D CPMG (blue) pulse sequence. Being samples characterized by a large amount of proteins, the CPMG sequence prevents the spectrum to be dominated by broad signals and allows the better identification of metabolites through the assignment of their sharp signals.

2.3 NMR Data Processing

Before both univariate and multivariate statistical analysis, pre-processing steps are necessary and fundamental for the preparation of data for the investigation.

2.3.1 Phase and baseline correction

Phase correction is a necessary step carried out after Fourier Transform. In effect, after FT, spectra tend to look dispersive and "out of phase", thus they need to be adjusted until they look in pure absorption mode. This operation is generally done on specific softwares working with the NMR instrument itself, such as TopSpin, by a "click and drag" operation or automatically. There are two kinds of phase correction: the zero-order (PH0) and the first-order (PH1). PH0 applies a change of phase to all signals equally, whilst PH1 applies a phase change positively correlated to the distance from the reference signal. Baseline correction is usually also carried out, since there might be a problem in the calculation of integrals and in signal recognition if baselines are not flat. In addition, is hard to phase spectra with bad baselines. The distortion of the baseline is usually caused by a distortion in the very first FID-points. In all the studies here presented baseline and phase correction were applied on FT specta through the software TopSpin.

2.3.2 Referencing, Alignment and Binning

One of the first step in spectral processing is the referencing to the internal standard. In case of all the studies analysed in this thesis, the reference signal was the one for trimethyl-sylil-propionic acid (TSP), which will be put at 0.0 ppm and all other signals will be referenced to that.

Spectral data, though, as aforementioned, can be affected by numerous external variation and this could invalidate statistical analysis and its results. One of the first visible effects of external factors is in fact the variation in peak position from sample o sample even after the referencing all all signals to the chemical shift of a standard compound. The main factor causing this variation is the difference in pH between samples, since the chemical shift is dependent on the ionization state of the molecules in the sample. In addition, the concentration of metal ions, binding of metabolites with proteins and chemical exchanges can also cause variation in chemical shift values of specific metabolites, like citrate [Cloarec et al., 2005]. These differences in chemical shifts between samples can lead to false clustering or artificial discriminations after multivariate data analysis, especially in the case of urine samples. In fact, urine generally has a pH varying from 5.5 to 6.5, but this range can expand as far as 4.6-8.0 due to the health status, specific diets or drug treatments. In addition, the carboxylic acids, organic amines and amino acids that mainly compone human urine (citrate, hippurate, dimethylamine and trimethylamine, histidine, taurine and glycine) have chemical shifts highly dependent on the pH of the sample since the carboxyl and amino groups can be ionized, as described by the Henderson-Hasselbach equation. Another factor that can create inter-sample variations in chemical shifts is the variable amount of ionic species present in urine. This can affect ionic strength and molecular bindings and, thus, the value of chemical shifts for specific molecules, such as citrate, which can vary due to different divalent cation concentrations. Though adding specifical chemicals to bind the metal ions and avoid chemical shifts problems will create other issues, such as the formation of overlapping signals on the metabolites peaks, as for examples happens after the addition of EDTA to form comples with divalent cations (Ca2+ or Mg2+). This will make it hard to have the right information to be compared among spectra, since it would ruin the necessary bilinearity for the analytical approach and the model that can be calculated. The principle of bilinearity that is essential for the creation of a valid model: it is fundamental that every variable column contains information of the same specific signal, generated by the identical metabolite in each sample to have the statistical analysis work and give out real results. Recent studies have proved that the chemical shift values are positively correlated to ionic strength and negatively correlated to pH and that variations can be reduced to less than 0.04 ppm if the pH is controlled and kept between 7.1 and 7.7 and the concentration of salt is less than 0.15 M. For this reasons, some researchers suggest to control this conditions by the employment of specific buffers [Xiao et al., 2009]. The main methods employed to reduce this external factors-induced variations still involve the integration of signals within so call bins or

buckets, spectral regions with a general width between 0.01 and 0.05 ppm (in most cases of 0.04 ppm) [Cloarec et al., 2005]. As stated, these intervals will contain the spectral integral and information contained in that chemical shift region. The new variable contained in each bin could be the integral as the sum or average of spectral intensities or the maximum. Generally the average is chosen. One of the major drawbacks of these methods is the fact that spectral resolution is lost and, together with that, information on signals shape. In fact, the traditional dimension of bins (0.04 ppm) was considered to reduce matrices dimensions and to increase the tolerance for the chemical shifts intersample variations, though such bin width will combine in many cases different signals for separate metabolites that could change in accordingly in their concentration, overshadowing therefore real metabolic changes. In some cases, therefore, smaller buckets (0.004 ppm) were considered, though, variations of chemical shifts will still be visible and generating problems in subsequent data analysis [Xiao et al., 2009]. More advanced binning techniques are now also available, such as Adaptive-Intelligent binning or Gaussian binning, focused on a better peak definition identifying conserved minima throughout the spectra and applied those as bin edges. In this way it is hoped that signals are not divided into separate bins. Other techniques that work in this sense are peak-picking algorithms, capable, sometimes using libraries, of identifying peaks – generally after alignment – and dividing them into bins. Also in this case, each bin will contain the average, summed or maximum intensity [De Meyer et al., 2010]. In any way, though, the problem of loss of spectral information still remains in most of these cases. In order to avoid this, more advanced methods such as "dynamic time warping" (DTW) or "correlation optimized warping" (COW) were developed. These methods are though based on really complex computer calculations and could give out unrealistic relations between shift and width of signals, due to the fact that they employ local compressions or expansions and this could ruin one of the major character for signal identification. Other techniques find the most relevant signals in spectra, transforming each specturm into a list of signals and their characteristics, so to reduce the dimensions of the original dataset. In this case, the disadvantage is that some information about signal shape can again be lost and, in addition, it is difficult to define the specific parameters to be applied for the choice of signals and their alignment. Most modern method are instead based on the creation of specific algorithms capable of automatic alignment of spectral signals, with the least possible human effect and keeping most of the spectral information. These techniques started by using the genetic algorithm as the base for the alignment of spectral regions, partial linear adaptation for spectral alignment and a PCA-based algorithm to find the spectral

regions which were not correctly aligned. These methods were though not too efficient and lengthy calculus-wise, therefore Wong solve the problem of their speed by using a correlation engine employing a Fast Fourier Transform (PAFFT), using regular width spectral intervals, in order to align both individually and rigidly the regions. This method was later improved, though it still holds its intrinsic complexity, since it is always necessary to find the most relevant spectra, whilst a simple automated research could be easier [Savorani et al., 2010]. In most cases elucidated in this PhD thesis, binning with intervals of 100 points was employed, since it would fix the major shifts evident among samples. In some studies, though, the algorithm of icoshift was employed for signal alignment. This algorithm, developed by Savorani et al. [Savorani et al., 2010] employs an FFT engine to concurrently align all spectra and it is based on correlation shifting, as the name suggests. It is a fast algorithm for full-resolution alignment even of large datasets and in this way it is possible to keep spectral resolution and all signal information. It independently aligns each signal in the NMR spectra with reference to a target that could be a real signal or a calculated signal (i.e. Average or median) maximizing the cross-correlation in the spectral intervals defined, which could be of regular width or user-defined). Three steps form the icoshift algorithm: firstly the intervals are defined, then the cross-correlation is maximized through the employment of the FFT engine and finally signals are reconstructed into the aligned spectrum. Many elements in the algorithm can be chosen by the user, making this pre-processing step more valuable [Savorani et al., 2010]. Still, this method can have some drawbacks, especially in the case of complex spectra, with many signals overlapping and misaligned, such as in the case of urine. This can generate spectral artifacts or force the alignment of peaks from different metabolites, therefore ruining statistical analysis. It is thus necessary to evaluate which kind of technique to employ for misalignment problems in each specific case.

2.3.3 Removal of unwanted regions

Another fundamental step before data analysis is data reduction. Generally the region between 0.5 and 9.5 ppm is considered for the analysis, in order to remove noise from the most upfield and downfield regions and the internal reference signal. In addition, the residual water signal (generally between 4.9 and 4.6 ppm) and the broad urea signal (5.5-6.1 ppm) are excluded from the analysis in order to focus on the real information of the spectra and the interesting metabolites.It is in fact desirable to remove both the uninteresting signals and the ones who could corrupt further analysis and retain quantitative accuracy. In effect, urea signals in urine can exchange with the solvent hydrogen, interfering with signal intensities and corrupting quantitation. Signals with similar

behaviour should be eliminated from the data matrix, too. Data reduction should be always carried out trying not to meddle with overlapping and nearby signals, which could actually be relevant [Dai & Eads, 2010].

2.3.4 Normalization

Many variables from the type of experiment, the instrument and from biological characteristics can have an influence on NMR spectra and on signal intensity. One of the main factors of variation in this sense is sample dilution, which could change to experimental error in buffer addition or to different water content [Torgrip et al., 2008]. For example, urine will generally show aspecific concentration differences amongst all samples. Usually, the concentration varies in a range as big as half an order of magnitude, though in extreme cases, such in starving conditions or when some particular medicines are taken, the difference in dilution from normal urine could be as big as ten-fold [Dieterle et al., 2006]. In this case, though the information on real differences among samples is masked by the dilution range and the statistical power of the analysis is greatly decreased. Specific metabolic changes due to metabonomic responses to determined factors will in effect influence only some metabolites and their signals and not the whole spectra, as dilution does. And these are the changes that are the object of metabonomics research studies. In order to have the real information on signal intensities differences among samples it is thus necessary to procede with a normalization step, in order to compensate for dilution changes and highlight the real shifts in metabolites' concentration [Dieterle et al., 2006]. The standard method employed for spectral normalization is the so-called integral normalization or constant/total sum normalization (CSN). This technique assumes that spectral integrals are mostly related to the total sample concentration: a linear series of concentrations in urine will result in a linear series of spectral integrals. This presumes that the shifts in individual metabolite concentration is way smaller than the global concentration shift. This type of normalization divides each bin or spectral point by the integral of the whole spectrum or one of its region, thus normalizing each spectra for the same integral. The cons of this method are mostly related to accuracy and robustness, since, as aforementioned, it is not always the case that the global integral is simply a function of dilution. Most of the normalization methods employ this general equation:

$$I(i) = \frac{I_i(i)}{\sum_{i}^{k} \int_{j_{ik}}^{j_{uk}} \left(I(x)^n \, dx\right)^{\frac{1}{n}}}$$

where $I_{i(i)}$ and $I_{(i)}$ are the intensities for variable i (bin, chemical shift point, wavelength, etc.) before and after normalization, respectively. k is an index of the spectral regions employed for normalization and j_{ik}

and j_{uk} are the upper and lower limits for k region for which the nth power of $I_{(x)}$ intensity is integrated. Integral normalization sets n equal to 1. Generally, metabonomics 1H-NMR urine studies will regard the spectral region between 0 and 1 ppm for integration, after the removal, as stated, of water and urea signals. Finally, each variable is usually multiplied by 100 in order to have a total integral for each spectrum equal to 100. Another method of normalization employed is the one on the creatinine signal. This technique derives from clinical chemistry, which generally normalizes the concentration of urine analites on creatinine concentration in human and animal studies. The hypothesis behind this method is that the creatinine clearance is constant, thus that this analite is an indicator of urine dilution. It will employ creatinine signals as references, thus in the general equation j_{ik} and j_{uk} will be for example the left limit of one of the creatinine signal (3.05 or 4.05 ppm). In the field of metabonomics, some studies employed this normalization, though some biological and technical issues arise. The first thing that can occur is that metabolites having overlapping signals with those of creatinine will interfere in the correct determination of creatinine concentration. In addition, the chemical shift of creatinine at 4.05 ppm, especially, will change with pH values, thus a complex and robust algorithm for peak selection is needed. For what concerns biology, instead, it was proved that creatinine can actually change in concentration due to different metabonomic responses. It can in effect change due to the protein intake, but mostly in correlation to the skeletal muscle mass, having therefore generally a higher concentration in male urine. For all these reasons, this method is currently less and less employed in metabonomics. Similar "partial integrals" normalization techniques, though, can be employed and give good results. If it is known that some spectral regions are very stable, these could be taken into account for integral normalization. Other normalization techniques employ n equal to 2. In this case spectra are seen as vectors, so to correct different concentrations, the length of each vector is set to 1. This approach is employed in pattern recognition techniques, which assume that the concentration of a sample will influence the vector length, whilst a sample content will determine its direction. Many drawbacks, though, have been found in all types of integral normalization. For example, it was demonstrated that when samples contain a great concentration of one particular metabolite, either from normal metabolism or from the intake of specific xenobiotics, this normalization fails. Integral normalization could also negatively influence multivariate analysis and data correlation, for the reasions mentioned above, thus it was necessary to find new and alternative algorithms. One of the proposed methods is the so-called Probabilistic Quotient Normalization (PQN). This tehcnique normalizes spectra using the most probable dilution factors, by estimating them through

the analysis of the distribution of quotients between signal amplitude in the analysed spectrum and in a reference spectrum [Dieterle et al., 2006]. In this way, PQN does not create the artifacts that other techniques generated [Torgrip et al., 2008]. PQN assumes that most of the signals will have intensities correlated by dilution, but not every signal present in the spectrum. For this reason, the most probable quotient present between signals in a spectrum and a chosen reference spectrum is calculated and employed as a normalizing factor. This most probable quotient can be calculated in many ways. Generally, the most employed and robust way is to employ the median spectrum as a reference for the calculation of the quotient. Other ways are to use a spectrum for the database, an average spectrum, or the average/median spectrum calculated on a subset of spectra. It was though demonstrated that this choice is not as critical as expected, though the median is still the most advisable reference, especially in studies with few samples, which might greatly be affected by outliers. In addition, it is advised to employ an integral normalization before quotient normalization. For this reason, the general algorithm of PQN is thus divided:

- Step 1: integral normalization, generally with a constant integral of 100
- Step 2: choice or calculation of reference spectrum (generally median)
- Step 3: calculation of all the quotients for each interesting variable in the test spectrum in relation to the reference spectrum
- Step 4: calculation of the median of these quotients
- Step 5: each spectral variable in the test spectrum is divided by this median

PQN can be applied both on raw or on binned spectra, though spectral intervals without signals should be excluded from the calculation [Dieterle et al., 2006]. Probabilist Quotient Normalization was proved very useful for urine samples, though, in this thesis it will be shown that in different samples it is advisable not to employ it. This is especially the case of samples in which particular metabolites will arise or disappear during time, such in food shelf-life/stability studies. In these cases it is advisable to find molecules that are considered stable during storage and normalized all samples with reference to the signals of those metabolites, in order to find the real variations occurring in analites concentration during the selected shelf-life. Other methods of normalization employ an external or internal chemical reference or an electronic one. An external chemical reference employs either coaxial inserts with the reference or separate tubes for the compound and the reference. The assumption is that the signal area S is directly proportional to the number of nuclei (N) resonating at that frequency, correlated to the moles and thus the concentration of that molecule. Therefore S is proportional to the concentration [C] by a constant k' through the

eqaution S=k'[C]. This constant depends on various measurement factors such as the specific spectrometer or the temperature, though in this normalization it is assumed that the k' for the reference is equal to that of the compound, by employing the same conditions and, when possible, by measuring at the same time. In any case, the accuracy of the measure of k' cannot be better than 5% in most cases, though in particular cases where many precautions are taken, a precision equal or lower than 1% can be possible and achieved. New and more detailed equations can be now used to avoid these errors and employ more robustly external referencing, such as the Henderson equation. Anyhow, the advantage of external referencing is that there is no contact between the standard and the compound, so that there are no problems in reactivity, solubility, polarity or homogeneity between the compound and the reference chemical, as instead happens with internal referencing. Internal referencing, instead, consists in mixing a homogeneous solution of the reference and the test compound. The reference and the matrix are weighted and selected signals from both are integrated in the spectrum. The two, though, should have no chemical interaction between them or with the solvent, should dissolve completely, shoud give no slow exchange and their signals should not overlap (at least one per each). These conditions are hard to meet and that is way there is still no official quantitative NMR reference substance. Dimethylsulfone was proposed as an universal standard, since it is very useful for quantitation os water or polar soluble samples. In some cases, the residual proton signal from the deuterated solvent was proposed as a reference and assumed to be constant in dilute solution. Thus, this signal will be set to 100 arbitrarily and will be the internal standard, so other analytes will have their concentrations measured in relation to this. Others have proposed the used of the solvent signal as a reference for concentration. one of the problems in this approach is radiation damping, especially when the signal from the tested compound is much smaller than the solvent. Alternative strategies were now introduced and these methods propose a reference signal produced electronically. This method, also called ERETIC (Electronic Reference To access In vivo Concentrations) started as a way to avoid the introduction of a reference substance in in vivo NMR. In the ERETIC method, a pseudo-FID, is generated by an electronic device and transmitted into the probe during the reception of the signal. This electronic signal, having the same shape and frequency of a FID, will give out a specific peak in the spectra after Fourier Transform. The ERETIC peak then needs to be calibrated through this equation:

$$[ERETIC]=k[R]SERETIC/SR$$

in which [ERETIC] is the resulting concentration of the ERETIC, [R] is the concentration of a

calibration solution, SR is the area of the calibration peak, SERETIC the area of the ERETIC and k a value which considers the number of protons per chemical group. Thus, for this calibration, a sample with known concentration and similar to the test sample should be used. After the calibration of the ERETIC peak, any compound can have its concentration determined by the equation:

$$[C]=k[ERETIC]SC/SERETIC$$

in which [C] is the concentration of the analyte to be determed and SC is the area of the peak of the molecule to be quantified. To inject the ERETIC signals there are three ways: transmitting it through an additional antenna in the probe, through the secondary changel via a directional couple and finally by a gradient coil using capacitive coupling. It is clear though, that anyhow, it is necessary to have a particular spectrometer, equipped with three RF channels, one gradient output and a dual probe. In addition, this procedure is usually lengthy, especially when many samples are analyzed. For these reasons, it was recently proposed to generate the ERETIC mathematically and subsequently merging it into the spectrym during data processing in order to use it as a quantitative reference. The first step is scaling the artificial signal to a reference sample with a known concentration in order to attribute it a nominal concentration. The artificial signal is added then to the analyzed spectra at a chosen chemical shift. Its intensity is chosen so to consider the differences between acquisitions on the reference and on the analyzed samples, such as the number of scans, the receiver gain or the solvent. Originally this method was referred to as QUANTAS, QUANTification by Artificial Signal and it is now included in Bruker softwares called "ERETIC 2". This method was proved to have similar accuracy (circa 5%) of the internal reference method [Giraudeau et al., 2014]. Recently, a work from Hochrein et al. [Hochrein et al., 2015] proposed a different approach for the normalization of NMR spectra of samples such as human plasma or urine. The new normalization pipeline described surged from the fact that in many cases there is unbiased regulation of metabolites in the commonly analysed human biofluids. This study proposed firstly to normalized spectra on the internal standard (i.e. TSP) in order to correct for eventual variations due to the performance of the NMR spectrometer in time, then unbalanced regulation is detected through Shapiro-Wilk test and manual data inspection. In case this occurs, a step of reference feature selection based on variance should be applied before employing normalization. This study also once again highlights that PQN is generally a good normalization procedure for human biofluid spectra, especially if the further step described is applied when data shows unbalanced regulation. As described, therefore normalization is a pre-processing step that

can be done in many different ways. The choice, though, of the normalization method can have a great impact on the outcomes of later statistical analysis. A lot of investigation and trial was done in this sense in the projects described in this thesis and in the years of my PhD study. A simple example of this trial and the impact of the normalization technique is shown in Figure 2.

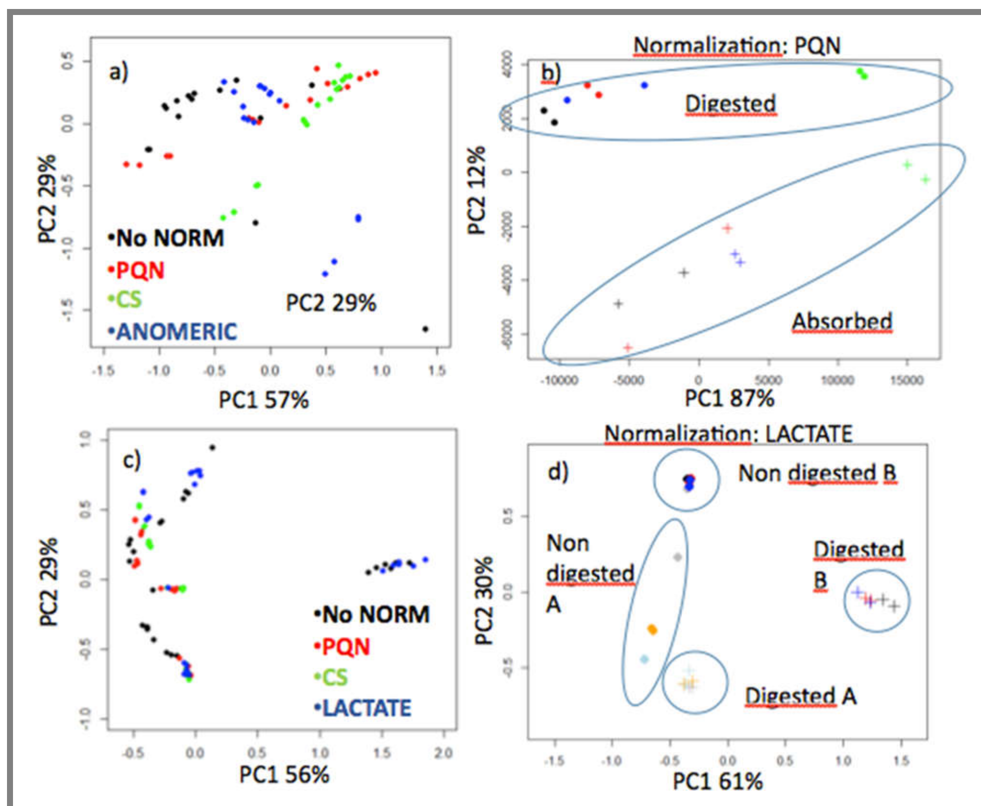


Fig. 2: The impact of normalization on the results of statistical analysis, all represented data were from studies carried out during the PhD. a) PCA calculated on a dataset normalized in four different ways: no normalization in black, PQN in red, constant sum normalizaiton in green and normalization on the area of the anomeric signal of a sugar in blue. The sample show very different values in the scoreplot depending on the method of choice. In this case, it was seen that the best method was PQN, as shown in subfigure b). c) Shows a PCA calculated on another dataset again normalized in four different ways; in this case the fourth method was to normalize on the signal of lactate, which was actually the best method for this dataset as shown in subfigure d).

This proves how the choice of normalization technique to apply is a crucial step in data analysis and should be carefully thought of for each individual research study. Knowledge of the spectral features and on the sample characteristics are of great importance in this choice. For example, being sure that a concentration of a specific metabolite should stay constant in every sample can help, since the best option might be to normalize on the signal(s) of that metabolite. Moreover, knowing that some molecules should display higher or lower concentration in specific samples subgroups can help in validating the normalization procedure of choice. For all these reason, it is fundamental to consider all these aspects when normalizing the NMR dataset and not to stick with a one-size-fits-all approach.

2.3.5 Scaling

While normalization is a row operation, meaning that it is done spectrum by spectrum, scaling is a column operation, performed on each spectral variable across all samples. Each variable in the matrix, thus each column, will undergo a certain operation. For example, mean-centering can be performed: the mean of each column can be set to zero, substracting the column mean by each column value. This a method commonly used in Principal Component Analysis (PCA) so that all the calculated component origin from the data centroid. In mean-centering data weighting will reflect variable covariance. In other cases, colimns can be divided by the standard deviation of the column, thus scaled to unit variance. In this way, the weighting reflects data correlation. Another form of scaling is Pareto scaling, dividing each variable by the square root of the standard deviation of the column. In this way values, will relate to their scores of order of magnitude [Craig et al., 2006].

2.4 Multivariate Data Analysis

2.4.1 Multivariate Statistical Analysis

The non-target power of metabolomics techniques such as mass spectrometry or NMR spectroscopy is strongly relying on the development and improvement of multivariate statistics.
Since MS and NMR generate large datasets, usually with a greater number of variables than that of individuals, more powerful and specific statistical tools were needed. Multivariate statistics is exactly the definition of this, since it is the method employed when many measures are done on a single object or individual (Fig. 3). These measurements are called variables, whilst the objects/individuals are refered to as observations. Thanks to the capability of modern computers to perform quick and difficult calculations, multivariate analysis (MVA) has been more and more employed. The origin of MVA relies in biological and behavioural studies and has then spreaded to all kinds of applications and research. Usually, as in all spectroscopical analysis, all variables studied with MVA are measured simultaneously in each sample and are correlated. This correlation will overlap with the underlying structure that needs to be studied, thus it needs to be untangled and simplified by MVA. Thus, exploratory MVA techniques exist, to reduce dataset dimensions and generate hypothesis on the analysed sets. Two main area of statistics are usually studied with MVA: the descriptive and inferential realm. In descriptive statistics, linear combination of variables are seeked and obtained; in interferential statistics MVA techniques, usually deriving from univariate ones, try to make inference from data to more general conditions and are useful because the significance level can be set by the researcher.

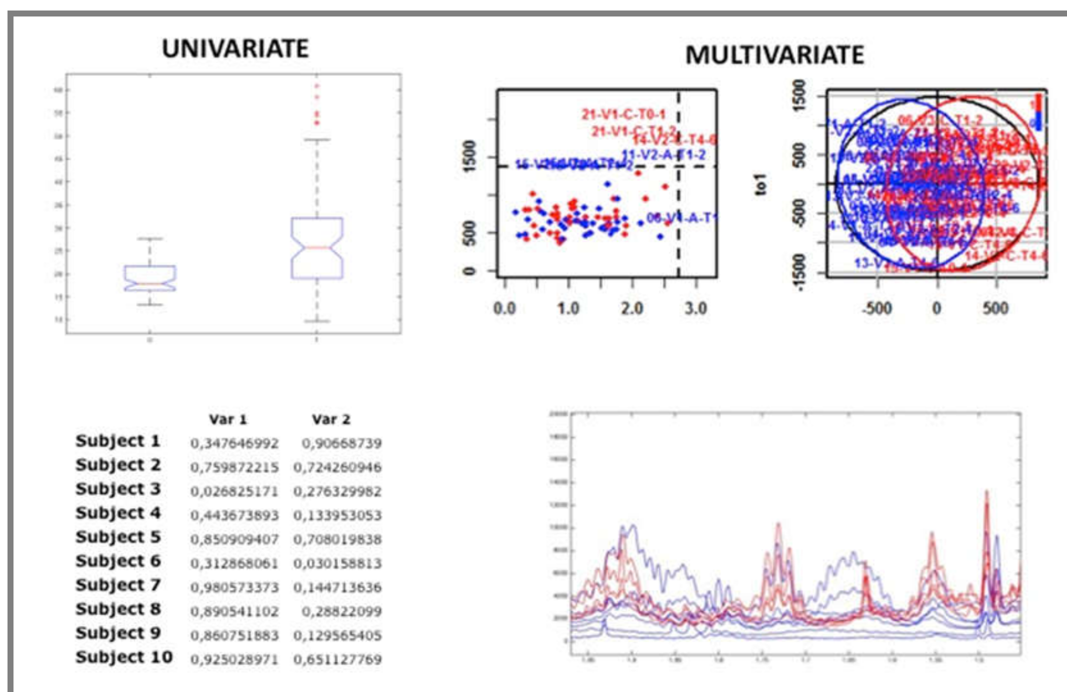


Fig. 3: Difference between univariate and multivariate statistical analysis. The first one is helpful when the number of measured variables is small and lower than the number of investigated individuals, whilst multivariate statistical analysis is useful for spectroscopical techniques, such as NMR, were many variables are simultaneously measured for each individual.

Anyhow, the choice of the methods to use depends on the objectives of the investigation, which could be:

- matrix simplification/data reduction: the need to make interpretation easier by simplifying the sample set without sacrificing valuable information to
- grouping of similar objects or variables
- study of the relationship among variables
- prediction of variable values after the identification of their relationships
- construction and testing of hypothesis

The most common techniques applied can be divided into two main approaches:

- UNSUPERVISED ANALYSIS: datasets are investigated without prior information on groupings and classes
- SUPERVISED ANALYSIS: the model is generated on the basis of known classes

One of the main techniques employed belonging to unsupervised analysis is PCA, whilst PLS or PLS-DA are techniques in which known classes are fed into the algorithm to create a model.

2.4.2 Principal Component Analysis (PCA)

Principal Component Analysis a multivariate projection capable of classifying samples through their common traits. It is a technique useful for large complex data sets and it is usually applied at the

beginning of the statistical investigation, since it can give insight on the major information contained in the dataset and on how to reduce dataset dimension by extracting the orthogonal principal components describing data variation [Solanky et al., 2005]. PCA thus generates a summary of data which can graphically analyzed through the score and the loading plots. The first one represents in two dimensions the information underlying in the data set, summarizing the relationships between observation and investigating on the possible patterns, trends or outliers among data. Loading plot, instead, describes the variables mostly responsible for clustering or outliers observed in samples [Zira et al., 2010]. The two plots are thus complementary. PCA can help in the simultaneous analysis of complex datasets, such as NMR spectra, and give insight on the biochemical variation observed in metabolites and caused by physiological factors and changes [Solanky et al., 2005]. The principle behind PCA is a bilinear decomposition, where data will be represented in K dimensions (K equals the number of variable) and reduced to few principal components (PCs), the ones mostly responsible for data variance. This is done without prior knowledge of classes or clusters among samples, thus PCA is considered an “unsupervised” techniques. The PCs will thus be linear combination of the input variables (integrals of bins or spectral points, in the case of spectra) with their respective weights, so that the first principal component will represent maximum data variance and the next PCs will contain decreasingly sized values of variance. By plotting the first two or three PCs it is possible to concentrate original N-dimension data into 2 or 3Ds. These scoreplots are helpful, as stated, to visualized any kind of pattern present in data and due to metabolic response. PCA will thus reduce the X number of variables into Y latent variables by a linear transformation. The first principal component, for example, will be calculated as:

$$Y_1 = a_{1,1}X_1 + a_{1,2}X_2 + ... + a_{1,p}X_p$$

and the weight coefficients, the best correlation coefficients between X and Y, will be found maximizing the variance for the vector in equation a) by equation b).

Eq a $Y_1 = \vec{a}_1 \bullet \vec{X}_1$

Eq b $1 = a_{1,1}{}^2 + a_{1,2}{}^2 + ... + a_{1,p}{}^2$

These coefficients are the LOADINGS. This way, variance of the first PC will be equal to:

$$Var(Y1) = \vec{a}_1{}^T \sum_x \vec{a}_1$$

in which ΣX is X covariance matrix.

All other PCs are found similarly, knowing that in those cases variance should be lower and that all the coefficients for each component will give a sum of their squares equal to 1. Each principal component should be independent from the others. The number of obtainable PCs is equal to the number of variables in the original data matrix. Through software calculation, all Principal Components can be found, though a number of them should be selected to reduce the dataset and remove noise. Generally, the PCs representing 80-90% of variance are kept. Another criterion which could be employed is to keep the principal components with a variance bigger than the average. To interpret the results of the PCA, the three characters of each PC should be considered:

- estimated explained and residual variance
- score values, projects of the observations on the new axes
- loading values, weight of the variables along that PC and correlation between variables

Each variable has in fact a loading value for each PC. These loadings will define how much that variable is responsible for the definition of that PC and thus, of the possible differences among samples along that component. Loadings will be between -1 and +1, being geometrically the cosine of the angle between the variable vector and the principal component. The smaller the angle, the greatest the link between variable and component and thus, the loading. PCA is therefore a fast and powerful technique for data reduction and for initial pattern investigation and outlier removal. It is to be kept in mind, though, that any factors creating data variance will be seen in PCA. Consequently, even parameters of the experimental analysis (such as the solvent-suppression scheme) can influence data variance and distinguish among clusters of data. Hence, all types of variations present in data and not object of the study should be reduced or, at least, considered as possible factors inducing differences in data and excluded when found in a PCA [Potts et al., 2001]. When trends or patterns relative to specific sample classes are found in PCA, supervised MVA is the next step. In supervised analysis information on classes to which samples belong is included in the analysis. The reason for the performance of this type of MVA could be to correctly classify samples or to predict classes of new samples. Supervised data analysis is a two step process: firstly the model is trained on representative data and their classes, then the model is validate and classification success rate is calculated on new data. Examples of this type of technique are Partial Least Squares Regression (PLS), PLS Discriminant Analysis (PLS-DA) or Artificial Probabilistic Neural Networks (PNN).

2.4.3 Partial Least Squares Regression (PLS)

Partial Least Square Regression or PLS is a regression, thus a method of statistical analysis employing an independent variable X and a dependent variable Y to highlight possible correlations. Regressions consist of two steps: calibration and validation. Calibration generates a multivariate model for X and Y through regression. A relation between the two matrix is searched, looking for the equation best describing it. A subset of sample needs to be used in this step. During validation, the previously created model is used to estimate new Y values given new X values. Usually a “test set validation” is employed, considering two sample sets: a training set and a test set for validation. PLS will firstly perform a PCA. In the PCA of the X matrix, a matrix T of scores and a matrix P of loadings is obtained, whilst for the Y matrix, a U matrix of scores and a Q matrix of loadings will be obtained. PLS uses the scores of Y to decompose X and calculate its loadings. It will the substitute Y scores with the X scores in order to decompose Y. In this way decomposition is mutually influenced. The process is iterated until it converges. The algorithm of partial least squares regression revolves around the intercorrelation between X an Y, reducing the variance of X not correlated with Y. PLS Discriminant Analysis (PLS-DA) is a multivariate classification method based on PLS. PLS-DA explains the maximum separation among defined sample classes, thus a previous knowledge of the class which sample belongs to is fundamental. The PLS regression in PLS-DA is done against a dummy matrix Y which accounts for the variations among the classes. The PLS-DA model is calculated and validated and then can be used for prediction of new samples. [Brindle et al., 2002] Classes separation is maximized and thus the research for the metabolites responsible for classification is simplified.

2.4.4 ASCA

Deriving from the univariate statistical technique of ANOVA, ASCA (ANOVA-Simultaneous Component Analysis) is a multivariate statistical tool developed for metabolomics investigations capable of dealing with the typical complex datasets. This method, developed by Smilde et al. [Smilde et al., 2005] merges the estimation of parameters from ANOVA with PCA, though removing the disadvantages of the two techniques and cosidering both the experimental design and the variables covariance [Jensen et al., 2005]. Considering a mean-centred data matrix X, with two main effects A and B – i.e. A is time and B treatment – which could interact and have different levels (A1, A2, A3, etc), then:

$$X = A + B + AB + E$$

in which E considers the information not assignable to A or B. The ASCA algorithm thus calculates

the level estimates for each effects, then performs a Simultaneous Component Analysis (SCA). The results of the SCA are the deviations of each sample for the specified effect (scores) and the weights of the variables for this effect (loadings). Moreover, the algorithm will also calculate the estimates for the effects interaction and its SCA. All the different SCA scores have rotational freedom and can thus be orthogonal, assuming that some rules are followed (i.e. Balanced design). In this way, the ASCA algorithm allows to investigate simultaneously both the single effect contributions and possible interactions. ASCA has been applied successfully in many scientific works, such as in psychology, but its peculiar design appears clearly useful in metabolomic studies, in which multiple effects impact on the data matrix (e.g. Time, treatment, individual response, etc.) and can interact with each other (Fig. 4).

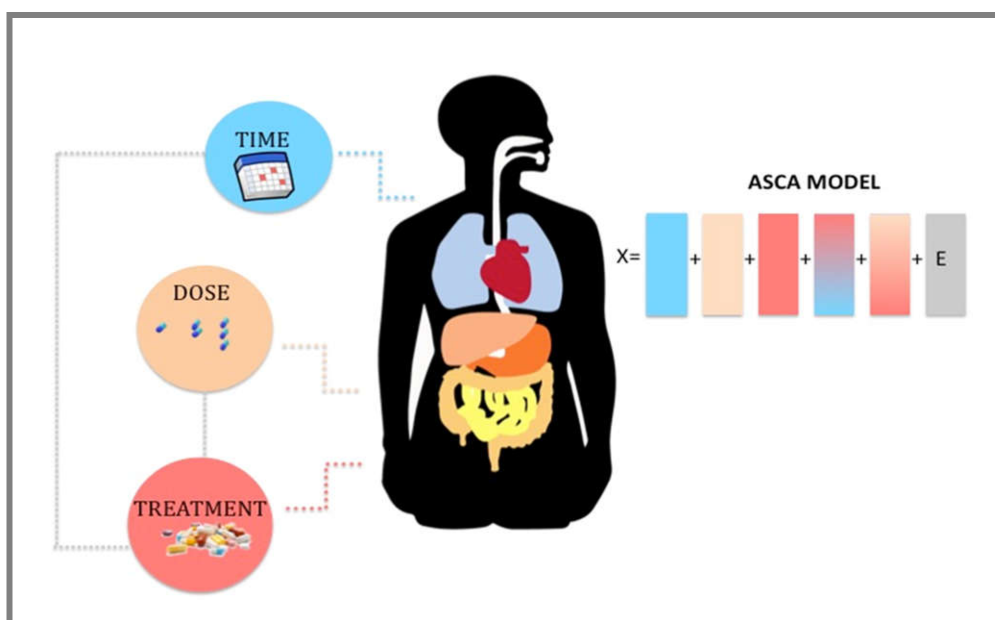


Fig. 4: Schematic representation of ASCA. Three factors impact on the data matrix (time, treatment and dose), together with their interactions (time and treatment, dose and treatment). All of these will be captured and included in the calculation of the ASCA model, together with information that cannot be described by any of these factors (E).

2.4.5 MultiLevel Analysis

In many studies involving both food items and human biofluids, different samples at different time-points can be taken for each object or individual. This could be done, for example, to study storage and matrix stability, ripening of certain foods or human response to a treatment. Thus, in PCA or PLS model, sources of variation coming from the different time-points is confounded with the variation to be investigated. For this reason, multilevel PCA and PLS are considered now a good option when this type of research is carried out. Multilevel models will consist of submodels for all the different levels of variance. This means that intra-sample variation will be separated from inter-sample variation [de Noord & Theobald, 2005].

2.5 Peak Assignment

A necessary step in spectroscopic analysis is the identification of the molecules present in the sample, know as signal/peak assignment. This is fundamental in order to understand the molecular profile of the sample (especially in non target analysis) and to find out which metabolites are responsible for any possible outlier, sample clustering or separation. This chapter focuses on NMR peak assignment, especially through the employment of the Chenomx software, which was greatly used in this research work.

2.5.1 Assignment principles

As described in chapter 2.2.3, signals in NMR spectra are characterized by some particular features:

- Intensity: related to the molecule concentration and the number of protons.
- Chemical Shift: due to the particular functional group and chemical environment.
- J-coupling: correlated to types of bonds and adjacent protons.

These features will characterize specifically the NMR spectrum of each particular molecule, allowing thus the assignment of that metabolite from the investigation of its spectral characteristics. The process of spectrum assignment can therefore be explained as a three-step operation, which can be iterated for every signal:

- signal identification;
- determination of: integration, chemical shift, coupling constant and multiplicity;
- assignment of the signal to the specific atom in the molecule [Banfi & Patiny, 2008].

2.5.2 Assignment through online tools

Nowadays, online through websites such as NMRDB (nmrdb.org) it is possible to draw the structure of a molecule (Fig. 5) or upload its molfile and predict its NMR spectrum, in order to find out whether a specific metabolite has signals in the analysed spectra. After having drawn the molecule or uploaded the molfile, it is thus possible through this applet to gain the information on its NMR spectrum, including chemical shift, J-couping and signal multiplicity (Fig. 6). In this way, it is possible to verify if a molecule we expect to be in our sample is actually present in it. The opposite process, from the NMR spectra to the molecule structure, can be done using our knowledge on the NMR spectrum characteristics or employing online databases (i.e. HMDB www.hmdb.ca, BMRB www.bmrb.wisc.edu, etc.). Through these databases, it is possible, by listing the chemical shifts of our signals (Fig. 7), to get a list of possible molecule and thus compare their spectrum with ours.

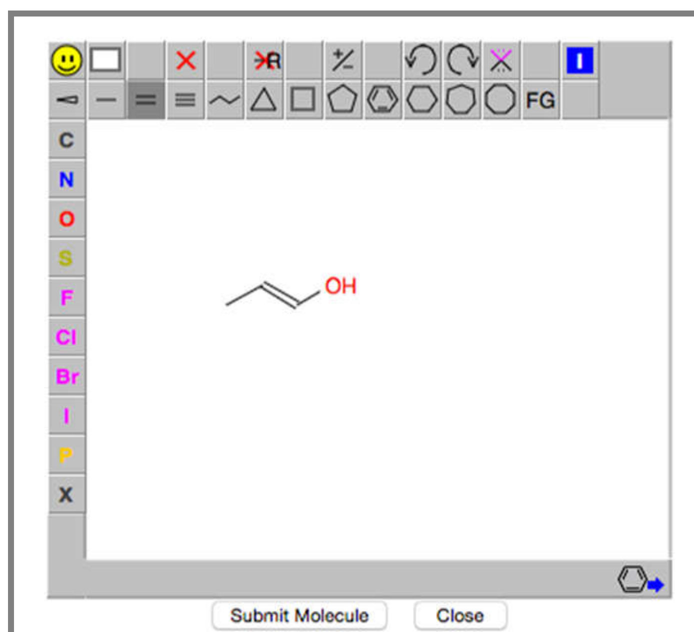


Fig. 5: Molecule drawn on the editor on NMRDB.

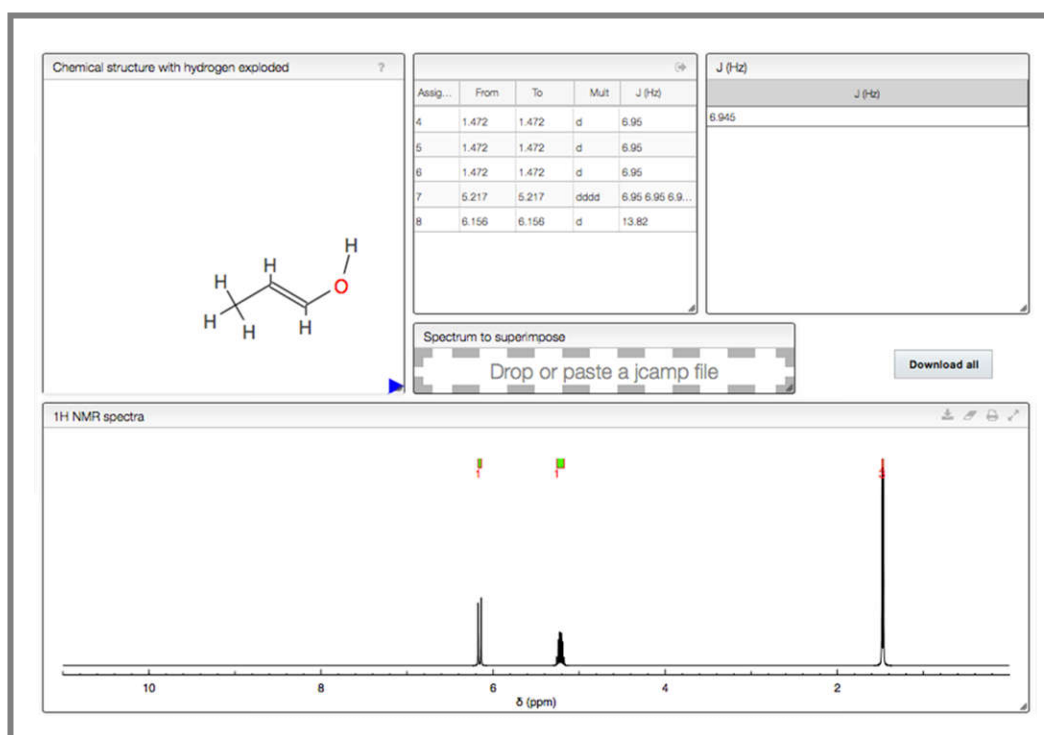


Fig. 6: Calculated 1D ^{1}H-NMR spectrum for the drawn molecule and information on signals (chemical shift, multiplicity and J).

As shown in figure 8, one can list the chemical shifts of the unknown signal and decide the peak tolerance before searching the possible molecular matches. This process can be done for MS spectra and also 2D-NMR spectra. A list of possible molecules is then showed (Fig. 8). In this way, it is possible to visualize their spectrum and match it with the one we are investigating on.

Fig. 7: 1D-NMR peak query on HMDB.

Fig. 8: Results from a spectral list query on HMDB.

2.5.3 Assignment through 2D-NMR spectroscopy

Another way of assigning unknown signals is through the employing of 2D-NMR techniques. 2D-NMR sequences employ multipulse sequences, in fact, to gain information that is not possible or not easy to obtain from 1D-NMR spectra. A simple 2D experiment will consist of a pulse sequence exciting the sample nuclei with two pulses or a group of pulses and then reseiving the FID. Acquisition is carried out for many times, incrementing evolution time (t1) between the two pulse groups. Acquisition time will be labeled as t2. 2D-NMR spectra will be thus composed by a stack of 1D spectra, differing from each other by a slight change in the value of t1. Therefore, in each successive experiment the parameters are kept the same, expect for the pulse phase. Through the

Fourier Transform of the two time domains, it is then possible to draw a map of spin-spin correlations. The axes will be represented by the evolution frequency (f1) and the acquisition frequency (f2). In some cases, signals can appear at different frequency values in f1 and f2 in case their frequencies have changed during t1. For better understanding, usually the 2D spectrum is plotted with the 1D projections on the axes. A 2D-spectrum is characterized by two types of peaks:

- diagonal peaks
- cross-peaks

Diagonal peaks represent the correlation of peaks to themselves and therefore are not very informative. Cross-peaks, instead, give information on the correlation of two different resonances by short inter-atomic distance or by the connection through a bond. In Figure 9 an example of a COSY spectrum for a molecule is shown, and the appearence of the signals for a specific group of protons is highlighed.

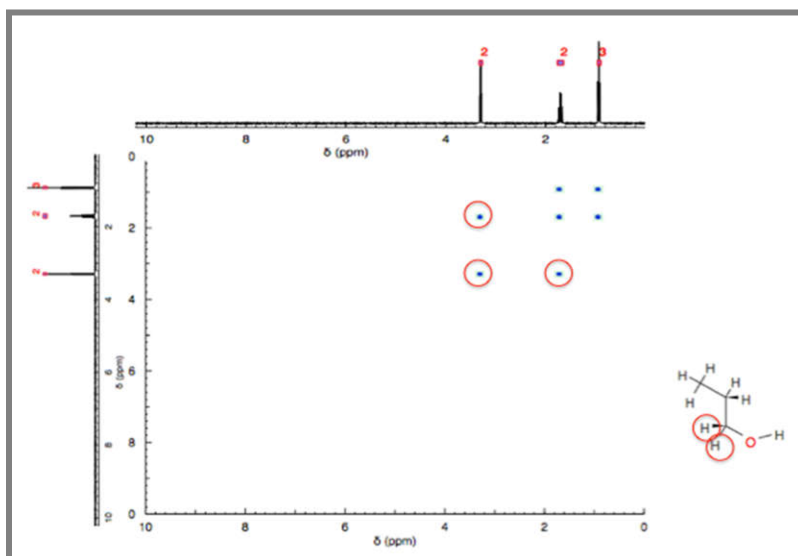


Fig. 9: COSY 2D spectrum for the molecule shown on the right. The hydrogens circled will give the highlighted signals on the 2D plot.

It can be seen that the highlighted protons are correlated to the signal at 1.7 ppm but not to the signal at 0.9 ppm.

2.5.4 Assignment through softwares

To facilitate the complex and time-consuming process of spectral assignment, nowadays specific softwares have been developed for the purpose. These softwares allow the upload of the NMR spectrum file and the matching with spectra of possible molecules, providing the possibility of composing the spectra of different molecules in order to best match the global and complex spectrum from our analysis. One of these softwares is Chenomx, also described in Annex 1. The

Chenomx profiler, in fact, allows both the identification and quantification of many molecules present in the software database in the NMR spectrum uploaded from the user's experiments.
The Chenomx library contains hundreds of molecules together with their information regarding the molecular structure and their biological, clinical and industrial meaning and application (http://www.chenomx.com/). It is possible to automatically fit a compound by searching for the possible compounds on a certain chemical shift by right-clicking on the signal and then fitting automatically the best-fitting molecule by simply hitting the space bar.
An example of a molecule fitting is shown in Figure 10, where dimethylamine was fitted on its singlet at 2.72 ppm.

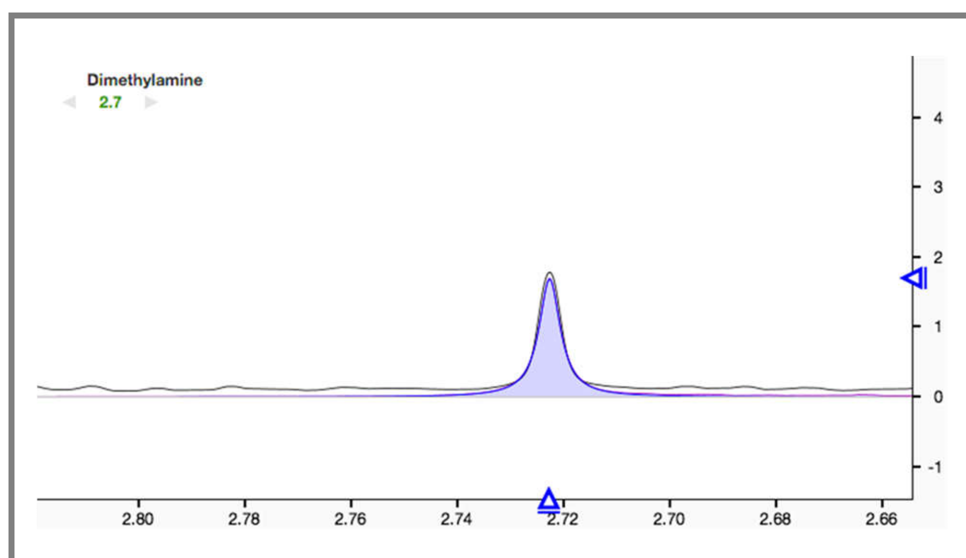


Fig. 10: Signal for dimethylamine fitted on Chenomx.

It is then possible to combine signals from different molecules in order to best match the analysed spectrum, since experimental spectra are usually composed by the combination and superimposition of many different spectra (Fig. 11). This is particularly useful since many times some interesting features for sample discrimination are composed by many different molecular signals and it is usually hard to understand which metabolites are responsible for these signals. It is thus clear of the task of spectral assignment is made a lot easier through the employment of this software (or similar tools), especially since it is possible to adjust the signals to match the sample's pH and other experimental conditions. In this way a table of the assigned metabolites can be easily compiled and this is one of the basic features to show in research paper. Thus this makes this process faster and a lot more user-friendly.

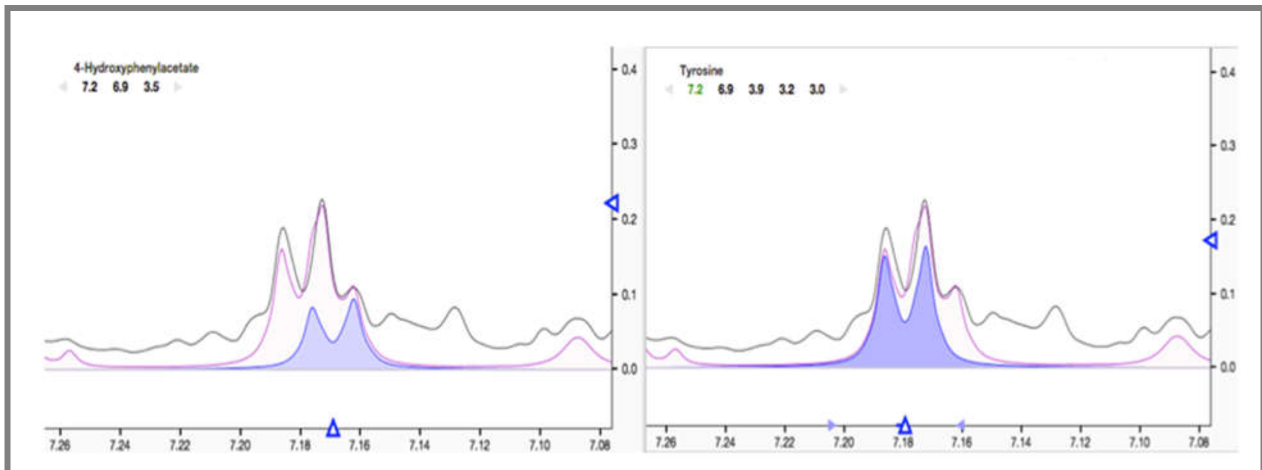


Fig. 11: Composition of a complex signal on a urine spectrum. Through Chenomx it is possible to combine signals from different molecules to understand how the signal is composed.

REFERENCES CHAPTER 2

Banfi, D., & Patiny, L. (2008) "www.nmrdb.org: Resurrecting and processing NMR spectra on-line" Chimia, 62(4), 280-281.

Brindle, J.T., Antti, H., Holmes, E., Tranter, G., Nicholson, J.K., Bethell, H.W.L., Clarke, S., Schofield, P.M., McKilligin, E., Mosedale, D.E., & Grainger, D.J. (2002) "Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics" Nature Medicine, 8(1), 1439-1444.

Cloarec,O., Dumas, M.E., Trygg, J., Craig, A., Barton, R.H., Lindon, J.C., Nicholson, J.K., & Holmes, E. (2005) "Evaluation of the Orthogonal Projection on Latent Structure Model Limitations Caused by Chemical Shift Variability and Improved Visualization of Biomarker Changes in 1H NMR Spectroscopic Metabonomic Studies" Analytical Chemistry, 77(2), 517-526.

Craig, A., Cloarec, O., Holmes, E., Nicholson, J.K., & Lindon, J.C. (2006) "Scaling and Normalization Effects in NMR Spectroscopic Metabonomic Data Set" Analytical Chemistry, 78(7), 2262-2267.

Dai B. & Eads C. D. (2010) "Efficient removal of unwanted signals in NMR spectra using the filter diagonalization method"Magnetic Resonance in Chemistry, 48(3), 230-234.

De Meyer., T., Sinnaeve, D., Van Gasse, B., Rietzschel, E., De Buyzere, M., Langlois, M., Bekaert, S., Matins, J., & Van Criekinge, W. (2010) "Evaluation of standard and advanced preprocessing methods for the univariate analysis of blood serum 1H-NMR spectra" Analytical and Bioanalytical Chemistry, 398(4), 1781-1790.

de Noord, O.E. & Theobald, E. H. (2005) "Multilevel component analysis and multilevel PLS of chemical process data" Journal of Chemometrics, 19(5-7), 301-307.

Dieterle, F., Ross, A., Schlotterbeck, G., & Senn, H. (2006) "Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in 1H NMR Metabonomics" Analytical Chemistry, 78(13), 4281-4290.

Giraudeau, P., Tea, I., Remaud, G.S., & Akoka, S. (2014) "Reference and normalization methods: Essential tools for the intercomparison of NMR spectra" Journal of Pharmaceutical and Biomedical Analysis, 93, 3-16.

Hochrein, J., Zacharias, H.U., Taruttis, F., Samol, C., Engelmann, J.C., Spang, R., Oefner, P.J., & Gronwald, W. (2015) "Data Normalization of 1H-NMR Metabolite Fingerprinting Data Sets in the Presence of Unbalanced Metabolite Regulation", Journal of Proteome Research, 14(8), 3217-3228.

Jansen, J.J., Hoefsloot, H.C., van der Greef, J., Timmerman, M.E., Westerhuis, J.A., & Smilde, A.K. (2005) "ASCA: analysis of multivariate data obtained from an experimental design" Journal of Chemometrics, 19(9), 469–481.

Potts, B. C. M., Deese, A. J., Stevens, G. J., Reily, M. D., Robertson, D. G., & Theiss, J. (2001) "NMR of biofluids and pattern recognition: assessing the impact of NMR parameters on the principal component analysis of urine from rat and mouse." Journal of Pharmaceutical and Biomedical Analysis, 26(3), 463-476.

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.

Smilde, A.K., Jansen, J.J., Hoefsloot, H.C., Lamers, R.J., van der Greef, J., & Timmerman, M.E. (2005) "ANOVA-simultaneous component analysis (ASCA): a new tool for analyzing designed metabolomics data" Bioinformatics, 21(13), 3043-3048.

Solanky, K.S., Bailey, N.J., Beckwith-Hall, B.M., Bingham, S., Davis, A., Holmes, E., Nicholson, J.K., & Cassidy, A. (2005) "Biofluid 1H NMR-based metabonomic techniques in nutrition research metabolic effects of dietary isoflavones in humans" Journal of Nutritional Biochemistry, 16(4), 236 – 244.

Torgrip, R. J. O., Åberg, K. M., Alm, E., Schuppe-Koistinen, I., & Lindberg, J. (2008) "A note on normalization of biofluid

1D 1H-NMR data” Metabolomics, 4(2), 114–121.

Xiao, C., Hao, F., Qin, X., Wang, Y., & ang, H. (2009) “An optimized buffer system for NMR-based urinary metabonomics with effective pH control, chemical shift consistency and dilution minimization” , Analyst, 134(5), 916-925.

Zira, A.N., Theocharis, S.E., Mitropoulos, D., Migdalis, V., & Mikros, E. (2010) “1-H NMR Metabonomic Analysis in Renal Cell Carcinoma: a Possible Diagnostic Tool” Journal of Proteome Research, 9(8), 4038–4044.

CHAPTER 3: DEFINING FOOD QUALITY

Food quality is a term that is defined by various factors: from the basic composition and safety, to technological and microbiological attributes, storage time and behavior, and nutritional value [Capozzi & Trimigno, 2015]. It is thus hard to define the concept of quality, though in recent years food and nutritional science and technology have tried to answer this question appropriately, also facing the ever-changing attributes of food quality. The first issue with food quality has been food safety, defined by the microbiological state of the food item, thanks to the discoveries of Pasteur.

After the rapid growth of Western countries, though, more attributes were linked to food quality and the globalization added the influence of foreign materials, technologies and preparations in the equation [Giusti et al., 2008]. In the last decades, food quality and food safety have been extensively debated by the public, the industry and the scientific community due to many reasons: the occurrence of several food scares that have focused the attention of the public to safety issues, the increased interest of consumers in the specific methods of food production and the stricter requirements and demands of particular consumers' segments. It is evident how these factors influenced food production and processing and prompted the introduction of new food products with added values to the market, just as required by consumers [Grunert, 2005]. The presence of completely innovative food products and the new and greater awareness among consumers, producers and food experts, have caused the change from food quality towards the so-called "Total Food Quality" (TFQ). This new concept can be defined by these factors, that characterize all food products, both by the consumers' and by the scientific point of view:

- Safety: the pre-requisite of food quality, the absence of toxic compounds, contaminants, pathogens or mycotoxins.
- Sensory and organoleptic characters: such as color, texture, taste or aroma.
- Nutritional value: the energy content and the composition in carbohydrates, lipids, proteins, essential aminoacids, vitamins and minerals and bioactive compounds, together with the concepts of digestibility, bioavailability and bioaccessibility of those compounds.
- Functional properties: characteristics of specific ingredients in respect to processing.
- Stability and service: resistance to deterioration during processing, storage, transportation and shelf-life.
- Healthiness: capability of food components to have a beneficial effect on human health.
- Psychological factors: such as price, ease of use or novelty.

Therefore, the TFQ principle is both an objective (scientific) and subjective (consumer-based) concept [Giusti et al., 2008]. Going through these attributes it is clear how they can deeply affect the human body and its wellness state. Scientific advancements and a growing number of epidemiological studies have recently proved the role of diet in disease control and prevention and the role of specific dietary compounds in the occurrence of particular diseases. It was also proved how economic development and globalization shifted lifestyle and diets with a great impact both on the nutritional status and on health of the population, especially in developing countries.

Inappropriate dietary habits such as an high fat or high energy density diets and unhealthy lifestyles such as the decline in physical activity or the increased use of tobacco, have increased the incidence of chronic diseases, such as obesity, diabetes or hypertension and, as a result, an increase in premature death and in the burden on national health budgets. Since nutrition is stepping up as one of the main adjustable determinants of chronic disease and scientific evidence has shown how shifts in diet can affect strongly the health status, both in a positive or negative fashion, it is clear how the scientific community is increasingly interested in the development of specific nutritional patterns and foods that could be capable of preventing and/or helping in the treatment of the aforementioned chronic disease [FAO/WHO, 1990]. In effect, many scientific studies in the last decades have tried to assess the relationship between specific foods or diets and the health of individuals. These studies have also prompted the production of novel foods with added beneficial attributes, such as the presence of bioactive compounds. For the EU legislation, though, it is necessary to prove a specific health claim before having the capability to employ it on the label and advertising of a food product. The EU legislator, in effect, enacted the Regulation 1924/2006, stating the necessary conditions for the use of health claims and the system of scientific evaluation employed on those. This Regulation states that a health claim is any claim saying, suggesting or implying a relationship between a food, a food component or a food category and health. Therefore there needs to be scientific evidence for this relationship demonstrating conclusive evidence of cause and effect [Nocella & Kennedy, 2012]. Only pre-approved claims can be used on food products and in any marketing related material. This is to prevent misleading of consumers; in effect, claims should be worded in a way that is understandable by the average consumer, though there is still no reference on how this understanding should be verified. In addition, the EU regulation, sets boundaries on the type of food items that are allowed to use health claim depending on their nutritional profile (i.e. high saturated fat or sodium content), though even in this case there is no well-defined agreement [EC Regulation No 1924/2006;

Lähteenmäki, 2013].

As stated, there is a lot of interest in the development of bioactive novel foods and in the investigation of their actual benefits on the human body. To do so, though, firstly the actual quality of the novel product has to be assessed in comparison to non-enriched counterparts (i.e. stability, shelf-life, etc.) and secondly the actual effect on humans has to be proved. For this reason, nutritional studies on health claims are long and expensive. New sciences like metabolomics are helping the boost of this kind of assessment thanks to the holistic approach they bring (see Chapter 3.1) and the fact that techniques like NMR are capable of giving insight on many different aspect of a sample like a food product (see Chapter 3.2). An example of the complexity of the definition of food quality and the helpfulness of the metabolomic approach is described in Chapter 3.3.

REFERENCES CHAPTER 3 (INTRODUCTION)

Capozzi, F., & Trimigno, A. (2015) “Using metabolomics to describe food in detail” Metabolomics as a Tool in Nutrition Research, 203.

Giusti A.M., Bignetti, E., & Cannella, C. (2008) “Exploring new frontiers in total food quality definition and assessment: From chemical to neurochemical properties” Food and Bioprocess Technology, 1(2),130-142.

Grunert, K.G. (2005) “Food quality and safety: consumer perception and demand” European Review of Agricultural Economics, 32(3), 369-391.

World Health Organization. (1990). Diet, nutrition and the prevention of chronic diseases: report of a WHO study group [meeting held in Geneva from 6-13 March 1989].

Nocella, G., Kennedy, O. (2012) “Food health claims – What consumers understand” Food Policy, 37(5), 571-580.

Lähteenmäki, L. (2013) “Claiming health in food products” Food Quality and Preference, 27(2), 196-201.

REGULATION (EC) No 1924/2006 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 20 December 2006 on nutrition and health claims made on foods.

Chapter 3: Defining Food Quality

3.1 Describing food quality through metabolomics

F. Capozzi and A. Trimigno (2014): "Using metabolomics to describe food in detail".
Book: "Metabolomics as a Tool in Nutrition Research" by Brennan, L., & Sebedio, J.-L. - Woodhead Publishing, November 2014 – ISBN: 9781782420842.

Role of the PhD candidate: literature research and writing of the review, completed and corrected by the candidate's supervisor

Abstract

Food science has been recently more and more linked to other scientific fields such as medicine, veterinary, agriculture, biology and genetics. Metabolomics can aid in linking those different sciences in the assessment of how food quality might impact on health, while keeping into account that food quality changes as the consequence of the influence of human and environmental perturbations on the food metabolome. This chapter summarizes some metabolomics applications in the evaluation of food quality as affected by genetic selection and modifications, different growing (e.g., organic vs. conventional) and rearing conditions, geographical origin, food manufacturing protocols, up to assess the effect of digestion on the nutrients availability.

Keywords

Foodomics, geographical origin, organic food, digestion, metabolites profile, NMR spectroscopy, Mass spectrometry, multivariate data analysis, nutri-metabolomics

Introduction

Metabolomics is a relatively new 'omic' science aiming at describing in a snapshot the complete set of metabolites constituting a defined biological system. This new omic approach has recently become widespread in many research areas, such as the study of human pathologies and diseases, drug discovery and toxicity assessment, human nutrition and plant analysis. This was possible thanks to technological innovations which nowadays allow the high throughput separation and identification of small molecules even from complex biological matrixes. These techniques are mainly referred to be mass spectroscopy and high-resolution nuclear magnetic resonance spectroscopy, together with some separation methods like capillary electrophoresis and chromatography (gas -GC-, liquid -LC-, and ultra high-pressure or high-performance liquid -UPLC and HPLC chromatography). Together with these scientific advances, the creation of databases,

containing information on spectral data and the corresponding signal assignments, has been a powerful input for the progress of metabolomics in various fields. Metabolomics, together with genomics, transcriptomics and proteomics, provide the holistic definition of food according to the new foodomic approach. Foodomics has been defined as "the discipline that studies the food domain as a whole with the nutrition domain, applying the same advanced omics technologies to different samples, and integrates all results in order to have an overall vision allowing the improvement of health and well-being" (Capozzi et al. 2013). Data obtained through the techniques described above are usually then interpreted through statistical and multivariate data analyses in order to have robust and comparable information, since generally it is necessary to classify samples in separate categories. Thanks to the possibility to identify a great range of metabolites with the aforementioned analytical techniques and to the potential of modern computers to handle a high quantity of data for multivariate analysis, this kind of approach is now widely applied in every life science. Food science has been recently more and more linked to other scientific fields such as medicine, veterinary, agriculture, biology and genetics. Metabolomics can aid in linking those different sciences in the assessment of how food quality might impact on health, while keeping into account that food quality changes as the consequence of the influence of human and environmental perturbations on the food metabolome (Figure 1). For this reason metabolomics applied to the exploration of possible links between food quality and the nutritional value has become one of the main branches in foodomics. More and more studies, in effect, are carried out to investigate on the sources of variance for the quality of food: e.g., how genetic modification of plants can alter the quality of plant food or their derivatives, or how it is possible to assess the origin of a determined food by looking at its metabolic profile. These studies are of increasing importance in our society since consumers are more concerned about food quality, often associated to its origin, and the consequences of food transformation on the nutritional value of the product. Digestibility, as well, is an important attribute of food quality, and its assessment by metabolomics is of paramount importance for the development of foods tailored to fragile categories of population like infants and elderly. Thus, being able to thoroughly prove food quality is a key tool for food producers and factories and for food consumers. This chapter will highlight some of the key studies made in this field and the future possibility for metabolomics in the evaluation of food quality.

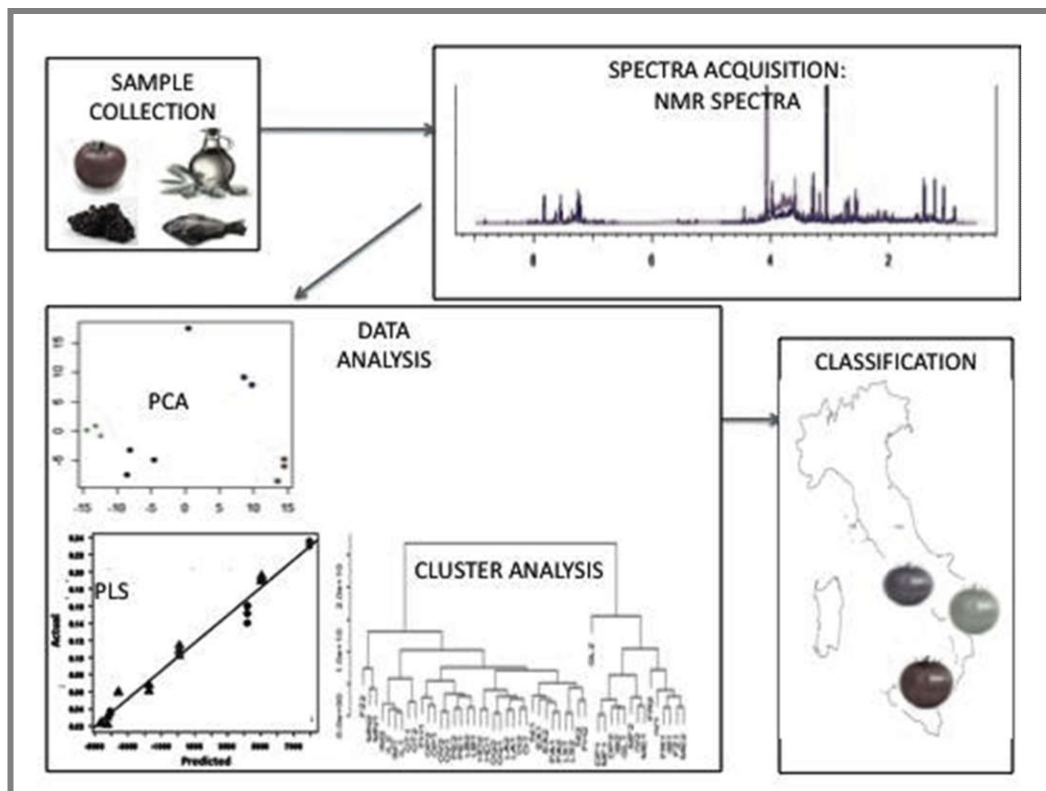


Fig. 1: The metabolomics approach using 1H-NMR spectroscopy. After sample collection, NMR spectra are acquired and subjected to multivariate analysis. As in the illustrated case, usually spectra do not show much difference at first glance. Thus, statistical tools such as multivariate data analysis techniques (PCA, PLS, Cluster Analysis) are employed to emphasize possible differences. Using those methods it will be then possible to classify the samples according to their different origin and characteristics.

Assessing the effect of genetic selection

In recent years a lot of research has focused on the production of genetically modified (GM) or transgenic foods, produced by DNA recombination in order to obtain favourable features. These new products, however, are still seen as potentially harmful by consumers, and further and modification analysis increasing their understanding and investigating their quality and possible unintended effects is necessary. Metabolomics is starting to be employed in this sense, thanks to its great potential of investigation by high throughput techniques, and many studies have now been published. Unintended effects in GMs production can be various: they can have an impact in food quality and safety or not, they can be predictable or not; therefore, research is needed to prove the nature of these possible effects. The "substantial equivalence" (SE) of GMs to their wild type counterpart thus necessitate to be proved by assessing the variation of the overall composition of the modified genotype with respect to the original wild type organism and by comparing such differences with those naturally occurring within the only wild type category. This concept means that two plant variants (GM and not) are similar enough to be considered the

same, and then they can be treated in the same way for what food safety concerns. Therefore, no unintended or different phenotypical variances should be found in GMs crops, especially variations that influence negatively their nutritional profile or can create harmful effects on consumers (Kok et al. 2003). It is clear how metabolomics can help in this assessment, generating full metabolite profiles for GM products and comparing them to their original, unmodified, counterparts. A study from Fraser et al. (2007) examined normal and GM tomatoes with an over expression of phytoene synthase1 through different approaches (protein activity, gene expression, physiological parameters and metabolomics). This is one of the most thorough studies, employing many different techniques: Real-Time PCR for mutant gene expression, enzyme assays, HPLC (to identify carotenoids, isoprenoids, flavonoids and phenylpropanoids), and GC-MS. Metabolic profiles were obtained through HPLC and TOF-MS, showing that transgenic tomatoes were different from the normal ones by various aspects. It was observed that the whole ripening process was different in the transgenic lines. Metabolomics was useful to assess the specific changes occurring at the different ripening stages, therefore linking gene expression and enzyme activity to the actual altered production of metabolites, and allowing to understand even unexpected metabolic deviations caused by the different genetic profiles of those tomatoes. Having data from genomic and enzymatic assays, it was also possible to link the results from those analyses to the metabolomics information, to get a whole panoramic view of what is really happening in the food product after genetic insertion and modification. Another study by Levandi et al. (2008) found differences between transgenic and conventional maize through CE-TOF-MS and subsequent spectral and statistical analysis (PCA and univariate statistical tests). As usual, the best extraction and analytical conditions were assessed at first and, after the optimization, metabolites in the maize samples were identified thanks to literature and metabolite databases. Peaks areas were then compared between GM and wild type samples, to see which molecules were over- or under-expressed in the former type of maize. Twenty seven metabolites were detected and identified in the samples and just two among these were found to have statistical predictivity to differentiate the two types of crops. In effect, L-proline-betaine and, even more, L-carnitine, were typical of GM maize; this was interpreted as a signal of alteration in the fatty acid and glucose metabolism, since L-carnitine is employed in those pathways enhancing the transport of fatty acids in mitochondria for their subsequent oxidation and increasing glycogen storage, respectively. Further multivariate PCA confirmed these compounds as discriminating between the two classes. It is now clear how this kind of approach can provide robust results, since different statistical analysis techniques, both

supervised and unsupervised, are In recent years a lot of research has focused on the production of genetically modified (GM) or transgenic foods, produced by DNA recombination in order to obtain favourable features. These new products, however, are still seen as potentially harmful by consumers, and further analysis increasing their understanding and investigating their quality and possible unintended effects is necessary. Metabolomics is starting to be employed in this sense, thanks to its great potential of investigation by high throughput techniques, and many studies have now been published. Unintended effects in GMs production can be various: they can have an impact in food quality and safety or not, they can be predictable or not; therefore, research is needed to prove the nature of these possible effects. The “substantial equivalence” (SE) of GMs to their wild type counterpart thus necessitate to be proved by assessing the variation of the overall composition of the modified genotype with respect to the original wild type organism and by comparing such differences with those naturally occurring within the only wild type category. This concept means that two plant variants (GM and not) are similar enough to be considered the same, and then they can be treated in the same way for what food safety concerns. Therefore, no unintended or different phenotypical variances should be found in GMs crops, especially variations that influence negatively their nutritional profile or can create harmful effects on consumers (Kok et al. 2003). It is clear how metabolomics can help in this assessment, generating full metabolite profiles for GM products and comparing them to their original, unmodified, counterparts. A study from Fraser et al. (2007) examined normal and GM tomatoes with an over expression of phytoene synthase1 through different approaches (protein activity, gene expression, physiological parameters and metabolomics). This is one of the most thorough studies, employing many different techniques: Real-Time PCR for mutant gene expression, enzyme assays, HPLC (to identify carotenoids, isoprenoids, flavonoids and phenylpropanoids), and GC-MS. Metabolic profiles were obtained through HPLC and TOF-MS, showing that transgenic tomatoes were different from the normal ones by various aspects. It was observed that the whole ripening process was different in the transgenic lines. Metabolomics was useful to assess the specific changes occurring at the different ripening stages, therefore linking gene expression and enzyme activity to the actual altered production of metabolites, and allowing to understand even unexpected metabolic deviations caused by the different genetic profiles of those tomatoes. Having data from genomic and enzymatic assays, it was also possible to link the results from those analyses to the metabolomics information, to get a whole panoramic view of what is really happening in the food product after genetic insertion and modification. Another study by Levandi et al. (2008) found

differences between transgenic and conventional maize through CE-TOF-MS and subsequent spectral and statistical analysis (PCA and univariate statistical tests). As usual, the best extraction and analytical conditions were assessed at first and, after the optimization, metabolites in the maize samples were identified thanks to literature and metabolite databases. Peaks areas were then compared between GM and wild type samples, to see which molecules were over- or under-expressed in the former type of maize. Twenty seven metabolites were detected and identified in the samples and just two among these were found to have statistical predictivity to differentiate the two types of crops. In effect, L-proline-betaine and, even more, L-carnitine, were typical of GM maize; this was interpreted as a signal of alteration in the fatty acid and glucose metabolism, since L-carnitine is employed in those pathways enhancing the transport of fatty acids in mitochondria for their subsequent oxidation and increasing glycogen storage, respectively. Further multivariate PCA confirmed these compounds as discriminating between the two classes. It is now clear how this kind of approach can provide robust results, since different statistical analysis techniques, both supervised and unsupervised, are employed corroborating each other. A later study on GM maize by Leon et al. (2009), employing FT-ICR-MS and CE-TOF-MS followed by PLS-DA, confirmed L-carnitine as a discriminating marker between GM and conventional maize, and highlighted at least other 9 metabolites (mainly from the amino acids pathways) able to distinguish between the two types of maize. At first, again, it was necessary to perform an optimization of the analytical parameters. In this case, after peak identification, a supervised statistical method, PLS-DA, was employed to discriminate among samples, still confirming the results coming out from the previous studies, even by exploiting different statistical techniques to increase the robustness of the results.

Other studies focused on GM wheat vs. conventional wheat (Baker et al. 2006; Baudo et al., 2006) showed no clear difference between the two types of wheat, whilst proved a higher differentiation due to the environmental and farming conditions. The first example (Baker et al. 2006), employed NMR analysis combined with PCA and GC-MS coupled with Analysis of Variance to study the metabolome. Conversely, the study from Baudo et al. (2006) mainly focused on transcriptomics analysis. Combination of different approaches (metabolomics and complementary traditional genomics and transcriptomics techniques) provided more information on how the genetic variation can cause effects on the phenotypical behaviour. Another study by Garcìa-Villalba et al. (2008) was focused on the comparative analysis between transgenic and conventional soybean through CE-ESI-TOF-MS. At first, the optimal conditions for extraction of metabolites from soybean samples were assessed, then the best separation and CE-UV parameters were established.

Afterward, metabolites were identified and quantified, using the best possible combination. Some differences among the samples were found: GM soybean showed greater concentrations of some metabolites like liquiritigenin, p-coumaroyl glucoside, naringenin 7-O-glucoside and 6-methoxytaxifolin; conventional soybean, instead, seemed to contain larger amounts of proline, histidine, asparagine, gluconic acid and trihydroxy-pentanoic acid. The main difference, though, was found to be 4-hydroxy-L-threonine, found only in conventional soybean and only this difference has been ascribed to the genetic modification. Again, this metabolomic study allowed the identification of differences in the metabolic profiles of crops, allowing the drawing of the biochemical modifications and the identification of the possible control mechanisms of the metabolic pathways. In this specific case, the amino acids and derivatives showing different concentrations between GM maize and the corresponding wild type counterpart were sharing a common precursor. Therefore, it was possible to verify that the 5- enolpyruvylshikimate-3-phosphate synthase, differentiating the transgenic line, affected the synthesis of aromatic amino acids. Picone et al. (2011 a) studied the possible substantial equivalence of GM grapes through a metabolomic approach employing NMR spectroscopy. The spectral matrix was subjected to multivariate statistical analysis (PCA) in order to project the overall variance of all wild type and transgenic grapes on the orthogonal components of a metabolomics space. Subsequent ANOVA was applied to the first 20 PC scores, and it was found that only the first two principal components were responsible for the discrimination between wild type and GM grapes. Studying the respective PC loadings, it was clear that, rather than a few metabolites, the whole metabolic profile was responsible for a clear differentiation between GM and conventional grapes. Among the spectral regions, however, the most affected were those referring to aromatic molecules (tryptophan and indole derivatives), with pathways expected to be affected by the genetic modification. Surprisingly, the region with organic acids also showed significant differences, thus pointing out that the ripening process could also be affected by the genetic insertion. Interestingly, only one of the two studied cultivars showed a discriminant metabolic profile between GM and wild-type, being larger the inter- than the intra-category variance. In this case, it was demonstrated that the effect of the genetic modification was modulated by the host genotype. This metabolomics study proved to be an effective tool in investigating the substantial equivalence, since the whole plant metabolic profile was being analyzed and quantified and not just some pre-selected molecules. NMR spectroscopy and multivariate data analysis (unsupervised in this case), can give robust results in this sense. An untargeted approach in metabolomics is fundamental as a preliminary

analysis for the assessment of pros and cons of new products. The studies mentioned above demonstrate how metabolomics reveals even the unexpected effects caused by a genetic alteration. This is not always predictable by taking into account only the specific modified pathway, because cascade effects might occur, generated by feedback regulations. Thanks to the metabolomic approach, which allows to analyse the whole metabolic profile, it is possible to detect all variations, including those which may not be captured by target analysis.

Assessing the effect of different growing and rearing conditions

a. Organic vs. conventional

Many studies have been performed trying to assess differences in the metabolome of organic or conventional products. In effect, in recent years there has been an increasing request for organic foods, mainly due to consumer concerns about food quality and safety and the resulting perception that organic products are safer and healthier than the conventional counterparts. Nowadays, many unprocessed and processed organic food products are present on the market, and numerous studies have been so far addressed to compare the actual composition and the nutritional value of these food matrices. Organic foods are considered of higher quality than the one conventionally produced, since consumers believe that they contain lower levels of pesticide and chemical fertilizers, whilst they possess higher level of nutrients and beneficial phytochemicals, together with better sensory characteristics (Williams et al., 2001; Harper et al., 2002; Makatouni et al., 2002). On the contrary, however, organic products may be more contaminated by microorganism and, as a consequence, by microbial products such as biogenic amines, as a result of the application of manure and the reduced utilization of antibiotics and fungicides.. Organic products are usually sold at higher prices in comparison to their conventional counterparts and this generates the risk of fraud. Therefore, again, proper analytical techniques are required to assess the true organic origin of these foods. Usually, traditional analyses were carried out in order to compare organic and conventional food products, these techniques including target analyses of specific macro (sugars, lipids, proteins) or micronutrients (minerals, vitamins, bioactive compounds, etc.). Other studies have been focused on the analysis of stable isotopes of light elements (H, C, N, O, S) in order to verify, for example, the soil of cultivation and fertilizers used: synthetic nitrogen fertilizers, banned in organic farming, have, in effect, lower δ 15N values, whilst manure, used in organically grown crops, has higher content of this isotope. However, these techniques resulted mainly in unreliable data for the classification of farming

procedures, unless they are combined with other multivariate approaches (Capuano et al., 2012). Untargeted screening is therefore promising for the achievement of robust and unbiased results. Metabolomics is clearly one of the best applicable strategies in order to identify and quantify molecules for the comparison of different products. Since the different farming procedures and practices might influence cellular processes and plant response, a metabolomics approach should be able to detect a change in the resulting plant metabolites (Wishart, 2008). Mass spectrometry (MS) coupled with other separation techniques has been widely used in recent studies with the aim of discriminating between organically and conventionally grown food crops. An untargeted study carried out by Chen et al. (2010) proved discrimination of organic or conventional grapefruit samples (Citrus paradisi cv. Rio Red) analyzed by mass spectrometry with two different separation techniques: FI-ESI-TOF-MS (Flow Injection ElectroSpray Ionisation Time-Of-Flight Mass Spectroscopy) and FI-ESI-IT-MS (Flow Injection ElectroSpray Ionisation Ion Trap Mass Spectroscopy). Data obtained from mass spectrometry were then analyzed by ANOVA and PCA. Year of harvest was the first source of variance among spectra, whilst farming mode was the second one. This study, though, was not capable of identifying metabolites giving rise to those differences, but the metabolomics approach pointed out that a metabolic modification exists between the grapefruit cultivated with different protocols. Zorb et al. (2006) found out that it was possible, through GC-MS analysis, to discriminate between dynamic, bio-organic and conventional wheat grains looking at only 8 out of 52 different detected metabolites (amino acids, sugars, alcohols, phosphates, nucleotides and organic acids). Wheat grains extracts were analyzed by GC-MS. Tukey's test was then applied to the concentrations of 52 metabolites. Seven replicates were made and the values were accepted only when the difference between those replicates was below 5%. Since only 8 metabolites were found to be differently contained in grains according to their cultivation practices, it was concluded that the farming does not have much impact on the nutritional value of the products, since sugar and sugar alcohols, were not different in their content. In addition, there was no sign of limited photosynthesis in organic plants. In 2010, Röhlig et al. investigated the influence of the farming system on the metabolome of maize kernels, showing that the main differences were due to cultivars and environmental factors, whilst only minor variation was caused by the farming system (organic vs. conventional). The molecular GC-MS profiles of organic and conventional maize were compared by ANOVA and PCA. Before PCA, data were scaled by the standard deviation of each variable in order to reduce the influence of abundant metabolites. A clear separation based on the metabolic profiles among the three

analysed cultivars resulted by PCA on the first two components, accounting for a total 39-59% of expressed variance. Within each cluster, organic productions were slightly separated from conventional ones. ANOVA, , applied to the relative amount of each analyte, revealed that 29% of 125 compounds so far considered in such univariate analysis were capable of discriminating between locations/cultivation systems. The main differences were found by ANOVA in the concentrations of myo-inositol, malic acid and phosphate, which showed higher levels in the organic crops. Myo-inositol, for instance, is an important metabolite in many biochemical pathways, intervening in stress response and osmoregulation. Its increased expression in organic kernels, though, has not yet been explained. Phosphate, instead, is very important for plant growth and has been proved to change in various organically grown plants, depending on their farming procedures and protocols (i.e. organic fertilization) (Steiner et al., 2006). A more recent study carried out by Bonte et al. (2013) by GC-MS profiling, performed on 11 wheat cultivars grown in controlled conditions, directed to compare conventional vs. organic system, found significant differences in the concentration of specific metabolites in winter cultivars. Only a few metabolites were identified: myo-inositol, in particular, and malic acid, like in the study from Röhlig et al., were particularly high in grains from the organic crops whilst some free amino acids (such as aspartate, asparagine, alanine) resulted having a higher concentration in the conventional farming system.

As stated, mass spectrometry is usually employed in metabolomics studies. Ambient mass spectrometry was recently found to be a very useful tool, capable to generate a specific fingerprint for low molecular weight metabolites (<1kDa). This type of MS requires very little sample preparation and does not necessitate any kind of separation prior to spectral acquisition. Newly designed ambient MS techniques are DESI (desorption electrospray ionization), ASAP (atmospheric solids analysis probe) and DART (direct analysis in real time). A pilot study tried comparing tomatoes and sweet bell peppers, either organically or conventionally grown, through the use of DART and TOFMS (Time Of Flight Mass Spectrometry) analysis (Novotnà. et al., 2012). This research demonstrated that this new technique might be applicable for a rapid assessment of plant samples and showed a reasonable differentiation of samples due to the type of farming system, though it assigned a greater impact on the metabolic fingerprint given by the production year. Models generated through PCA and subsequent LDA on those data, however, were not completely reliable since a higher number of samples would be needed to generate a more robust classification (Novotnà et al., 2012). Very few researchers have been carried out studies on food processed from organic versus conventional farming. One of these studies concerned ketchup

(Vallverdù-Queralt et al., 2011) and demonstrated, through LC-ESI-QqQ (Liquid Chromatography coupled to MS in tandem mode) and statistical analysis (ANOVA), some differences in the composition of the organic or conventional products. The former showed significantly higher amounts of caffeoylquinic and dicaffeoylquinic acids, caffeic and caffeic acid hexosides, kaemp-ferol-3- O-rutinoside, ferulic-O-hexoside, naringenin-7-O-glucoside, naringenin, rutin and quercetin. The latter products, instead, contained typical compounds (glutamylphenylalanine and Nmalonyltryptophan) that were not found in the organic ones. The authors concluded that the metabolites found in organic ketchups are due to secondary metabolism of plant, related to its self-defence mechanisms. A further step in the application of metabolomics concerns the study from Laghi et al. (2014) employing NMR spectroscopy to assess the differences between red wines obtained from either organic or biodynamic grapes. Principal component analysis was applied to NMR spectra and showed a great variation due to the vinification protocol (biodynamic vs. organic). Tyrosine-related metabolites seemed to be the major source of distinction, exhibiting a higher concentration in organic wines. The second PC (principal component) showed the effect of vineyard management causing a greater concentration of resveratrol and a lower amount of trans-caffeic acid in organic grapes. Further cluster analysis showed again a marked difference between samples due to the vinification procedure and, to less extent, to the vineyard management. Also ANOVA applied to NMR spectra pointed out that the concentration of trans-caffeic acid was higher in biodynamic wines, whilst containing a lower concentration of glutamine. This inverse correlation was described by Megarejo et al. (2010), to wit that some polyphenols might target biogenic amines producing enzymes. This was further investigated by studying the spectral differences in wines from successive years, proving that switching organic to biodynamic farming modifies the phenylpropanoid pathways, thus corroborating the hypothesis that the decrease of glutamine concentration is related to an antagonistic effect of polyphenols biosynthesis. The application of the metabolomic approach has been thus proved to be helpful in the investigation of the possible differences between organic and conventional food products. Mass spectrometry, coupled to newly advanced techniques, has been particularly exploited, but also NMR spectroscopy was found to be useful in those studies. In this way it is feasible to highlight the patterns of metabolites responsible for the discrimination between the different cultivation protocols, and give proof of possible distinctive characteristics in organically produced foods that could be investigated by nutritionists and food scientists when looking for additional positive properties. Whilst univariate statistical analysis is not able to provide a clear set of diagnostic metabolites differentiating organic

from conventional products, multivariate data analyses, such as PCA, provides combinations of compounds useful for classification, suggesting that patterns of metabolites rather than single molecules may be used as biomarkers of quality.

b. Geographical origin

Another important parameter for food quality is the geographical origin, which becomes fundamental for some products. In effect, a particular geographical origin might give an added value to a product in comparison to others, thus generating the necessity of protecting the denomination of origin of some particular foods. The added value of such protected products has created the increased production of fraudulent fake food products, sold as food with protected denomination of origin. It has thus become essential to be able to assess the authenticity of these counterfeit products to confirm this attribute of food quality. Metabolomics has been proposed as a new technique to aid the certification of geographical origin, to select some suitable metabolic markers for the substantiation of the claim. Many recent studies have employed techniques based on mass spectrometry (MS) coupled with statistical and multivariate data analysis in order to be able to assess metabolic differences in food products related to their geographical origin. Most of the researches is made by using an approach based on Induced Couple Plasma coupled to mass spectrometry (ICP-MS), combined with multivariate data analysis, such a PCA, canonical or linear discriminant analysis, in order to classify samples according to their origin. This approach was applied, for example, to honey (Chudizinska et al., 2010), olive oil (Benincasa et al., 2007), paprika (Brunner et al., 2010), tomatoes (Savorani et al., 2009; Mallamace et al., 2014) and tomato products (Lo Feudo et al., 2010). Meat quality is very important both for health and for nutritional reasons and metabolomics studies demonstrate how a great variation can be seen in samples from different geographical origin. This can also be exploited for certifications of origin that are now requested in many countries, either for security and for quality assessment aims. Jung et al. (2010) investigated the effects of geographical origin on the metabolome of beef by applying 1H-NMR spectroscopy. PCA and OPLS-DA were employed to establish whether some metabolites were able to discriminate between beef from different geographical regions (Australia, Korea, New Zealand, USA). At first, PCA was carried out in order to test, in an unsupervised manner, if there were differences among the different types of samples showing, indeed, a separation among the four groups. Thus, the supervised OPLS-DA approach was used to maximize these differences, the model so far obtained further tested with validation procedures, and the results were found accurate. A few orthogonal components were sufficient to distinguish between beef origins, and

many metabolites were important for that classification: acetate, anserine, betaine, carnitine, carnosine, choline, creatine, creatinine, fumarate, glycerol, hypoxanthine, lactate, niacinamide, succinate and all amino acids were involved in discriminating patterns. Most of the metabolites are clearly higher in the New Zealand samples, especially amino acids, whilst Australian beef exhibited lower levels of that molecules; US samples were the highest in succinate concentration. It is clear that many conditions can alter beef composition, like breed, feed, rearing system, preand post-slaughter parameters and environmental conditions. Breed can cause differences in amino acids, both essential and non-essential; the amino acids also increase thanks to postmortem aging, due to the hydrolysis of peptides and proteins. Diet can also affect the level of amino acids found in beef: if fed with silage, meat will show higher concentration of many amino acids in comparison with concentrate-fed cattle. Further studies have been addressed to analyze differences in the fatty acid composition, metabolites than can again vary due to diet, environmental effects and type of production system. Also for vegetable products, metabolomics has been exploited as a tool to prevent frauds, where geographical origin certification is a key element for certification of food quality. Two case studies were carried out (Savorani et al., 2009; Mallamace et al., 2014) to establish differences among cherry tomatoes with respect to their origin (Pachino province in Sicily vs. other Italian productions). They both employed NMR techniques and multivariate statistical analysis to assess the metabolic profile of those products and identify possible differences from cherry tomatoes of other geographical origin. The study showed how the metabolomics approach could be useful for tracking food origin, Pachino tomatoes proven to contain higher concentration of metabolites like fructose and glucose, glutamine, glutamate, aspartate and 4- aminobutyric acid (GABA) and lower levels of fatty acids, alanine, methanol and acetylglutamic acid in comparison to products cultivated in other, even neighbouring, areas. Another example of application of metabolomics to the geographical origin is described by Cajka et al. (2010) who studied the possibility of recognize beer origin through metabolomics, employing DART-TOFMS and statistical analysis (PCA, PLS-DA, Linear Discriminant Analysis and Artificial Neural Network with Multilayer Perceptrons). Markers were chosen among the DART-TOFMS profiles after careful inspection, in order to select the ones which could perform better upon chemometric analysis. At first, the prediction ability of both positive and negative ions was tested through leave-one-out cross validation of a preliminary LDA model. PCA was then calculated in order to investigate possible clustering among samples and it showed how beers from Czech Republic were different from Belgian and Dutch beers. PLS-DA was then performed to assess the source of variance between

one type of beer over the rest of the productions and between Trappist and non-Trappist beers: models with high recognition performances were obtained, though their predictive power was lower. LDA was used as an alternative supervised technique to find out the metabolite patterns giving maximum separation among sample classes. In this case, a better model from PLS-DA was generated for the classification of one beer vs. all the others, rather than that obtained by comparing Trappist and non-Trappist beers. Finally ANN-MLP was tested as further approach. The latter technique applies for situations in which there is a complex relationship between input classifiers and predicted variables. Even in this case, the model built up by comparing one type over the other beers was again better than the model obtained when testing if beers were Trappist or not. Moreover, models had better performance in the latter approach than in PLS-DA and LDA, although the results were found similar. These models, all together, can establish robust statistical differences between beer productions, by finding the features of the metabolite fingerprint responsible for discrimination. Finally, the same study also compared another different approach by applying, on the same samples, SPME-GC-TOFMS instead of DART-TOFMS. The different SPME-GC-TOFMS trial had proven to be more effective in the classification, while requiring more (although simple) preparation steps and thus more analytical time. Although metabolomics proves that the geographic origin of a food product may be reflected in a unique molecular fingerprint, the actual source of differentiation is still questionable, since multiple factors contribute to define a “local production”, including the restriction to use only autochthon genotypes, the practice of using traditional productive protocols, and the existence of confined pedoclimatic conditions. For this reason, the definition of a geographic origin must rely on updated databases, having the purpose to monitor the changes in one or more of the multiple factors affecting the metabolite profile of the local production. In this way, robust models developed to predict the geographic origin of a product will include correction functions based on the complete set of cofactors contributing to define its molecular composition. Only then, robust models are entitled to be adopted to assess fraudulent declarations of a certified origin, when the latter is an added value.

c. Rearing conditions

Many elements contribute to the establishment of the quality of food products of animal origin, like the feeding, the use of antibiotics and medicines or the choice of rearing conditions. Just recently metabolomics studies have been challenged to show, in the metabolic profiles of animal products, the differences attributable to the rearing protocols. Most studies, though, have been focused on the evaluation of differences in food products caused by the use of hormones or other

chemicals (i.e. antibiotics) to improve cattle growth. A study from Regal et al. (2011) showed statistical differences between LC-HRMS spectra of sera sampled from control cattle and the one treated with estradiol and progesterone. This was performed by multivariate statistical analysis based on PCA, OPLS and OPLS-DA. First unsupervised PCA was performed, in order to find out possible sources of variations among samples. Then, OPLS and OPLS-DA were used to built more robust model for samples classification. Eight molecular markers found by OPLS were finally selected for discrimination between treated and untreated animals and were therefore associated to perturbed metabolic pathways of hormones. Another study by Graham et al. (2012) was focused on bovine plasma samples in order to highlight the presence of illicit growth-promoting agents. These products are illegally used to improve feed efficiency, increasing animal weight gain. Growth-promoters cause alteration in blood and tissues composition, and many studies were made to detect those biological changes. Since metabolomics could be of great help, being able to give a rapid and highthroughput screening of metabolic profile, this research employed two methods of NMR spectroscopy (classical 1-D and CPMG pulse sequence), followed by multivariate statistical analysis (OPLS-DA) to differentiate treated and non-treated cattle. The two NMR techniques were used in order to develop a high-throughput screening technique, since they can show different classes of metabolites. NMR peaks were identified through libraries and literature data and OPLS-DA was then performed on spectral data in order to visualize differences and clusters among samples. Treated animals showed altered metabolic profiles, especially in markers of metabolic balance and nitrogen flux, proving how growth-promoting agents can alter different biochemical processes like protein metabolism, gluconeogenesis and glycogen deposition. The impact of different rearing conditions on the quality of bovine milk was investigated by Boudonck et al., (2009), through the aid of GC-MS and LC-MS/MS metabolomics, followed by unsupervised data analysis (PCA). Ten different types of bovine milk were analyzed. Samples were different in brand, percentage of fat, expiring date, rearing method and type of package. Metabolites from GC and LC mass spectra were identified by using specific libraries using three criteria: retention time, experimental precursor mass match to library standard and MS/MS scores. Internal standards were used in both analytical methods in order to calibrate the experiment and control it. Data were then normalized dividing the raw metabolite area by its median value in each run-day, in order to correct for inter-day possible variation. After this pre-statistical transformation, data were analyzed first with the unsupervised PCA technique to explore possible separation and clustering, and then the expected and interesting classes were further tested with analysis of variance. Principal

component analysis showed that the first PC, accounting for the maximum variance, was related to difference in rearing practices (organic vs. conventional), whilst the second orthogonal PC was influenced by the fat percentage of milk. One of the most interesting findings was the relation between farming (organic vs. conventional) and the consequent metabolic differences. A one-way ANOVA was performed on milks and 25 metabolites were found to be significantly different in their concentration between the two types of milk (whilst 11 metabolites would be expected different in case of random classification). Organic milk showed higher concentrations of tyrosine, isoleucine, mannose, glycerate, ribose, carnitine, hyppurate and butyrylcarnitine when compared to conventional milk, whilst proline, trans-4-hydroxyproline, glucose-1-phosphate, ribose-5-phosphate, glycerol-3- phosphate and glycerol-2-phosphate were found in lower concentration. Hyppurate was one of the main marker for organic milk and it has been positively linked to i) fiber intake in diets (Holmes et al., 2008), ii) consumption of products rich in polyphenols (Walsh et al., 2007) or iii) exposure to determined environmental conditions (Walsh et al., 2007; Holmes et al., 2008). Thus, the observed discrimination is consistent with the different diets fed to animals in the two types of rearing, since organically raised cattle is mainly fed with forage (grass and clover), whilst conventional cows consume grain-based feed with cereals, maize and protein supplements.

Two studies from Van Ruth et al. (2011, 2013) demonstrated differences in the carotenoid profile between organic, free-range and barn eggs through high-performance liquid chromatography (HPLC) and multivariate statistical analysis. A model was built with knearest-neighbour clustering analysis allowing to correctly predict organic (100% accuracy) and non-organic (free-range and barn) samples (92% accuracy). The model calibrated on Dutch eggs was then further tested with eggs from different countries and it proved successful again in discriminating samples by their production system. This metabolomics approach was again proved to be useful, verifying the hypothesis that the different rearing conditions and, in particular, the access to different sources of feed for organic hens created differences in carotenoid profiles in eggs. Other studies have focused on fish quality in relation to rearing practices. A first study from Savorani et al. (2010) tried to assess the differences in the metabolic profiles from gilthead sea bream in relation to aquaculture methods through 1H-NMR metabonomics. NMR spectra were acquired on percloric acid extracts. After the normal processing steps (Fourier transform, baseline and phase correction), spectra were normalized to total unit area in order to correct for possible vertical scale errors due to dilution biases. After that, spectra were aligned using the iCoshift algorithm to allow the highest stability in the signal position among the different spectra. Finally, multivariate data analyses (PCA, PLS, ECVA

and iECVA) were performed on the spectral matrix. Only few out of the many signals present in the spectra could be assigned by using literature data; therefore, the nature of the other unknown signals was revealed by addition of pure standard compounds, and finally 38 additional peaks were assigned. Spectral analysis was useful to follow the evolution of the main metabolites content during fish storage, and to capture the most important features to be recorded to describe the decay of fish quality. The main metabolites are amino acids and nucleotides. PCA was calculated by selecting only the aromatic region of the spectra: the results pointed out that inosine and inosine 5'-monophosphate were mainly responsible of differentiation due to storage time. However, specimens extracted after 16 days of storage under ice, evidenced differences mostly caused by the farming system. Supervised ECVA confirmed that aquaculture practices give a higher variance of metabolite composition in stored fish rather than fresh fish. By using iECVA, and omitting the effect of storage, it was found that the compounds responsible for discrimination were glycine, histidine, alanine and glycogen. It was therefore shown that post-mortem glycolysis is faster for fish farmed in 'pseudo-natural' environments, like lagoons and off-shore cages, with respect to those reared in tanks, therefore suggesting that the former aquaculture systems can induce higher levels of stress in animals or that a greater glucose metabolism rate is caused by a higher exercise. A later study by Picone et al. (2011 b) was carried out on another fish species, namely Spaurus aurata, by applying a different chemometric approach. The same metabolites as those found in the previous study were again identified, with a paired t-test, by comparing the metabolic profile evolution during the storage: obviously, glycogen decreased significantly as a consequence of glycolysis; nucleotides were degraded into hypoxanthine as the consequence of post mortem processes, and TMAO was converted into TMA by microbial attack; the amount of lactate, conversely, did not change because already depleted at the moment of the starting observation. Some differences were captured by PCA depending on the farming system, where PC1 mainly collected spectral changes in the hydroxylic region (aminoacids, sugars, hydrophilic molecules), and PC2 being dominated by aliphatic signals. The effects of storage temperature in the amino acid profile of Bogue fish was studied by Ciampa et al. (2012) by NMR metabolomics on spectra acquired at 600 MHz. In this case, the novelty of the approach concerned the normalization algorithm which was based on the total nucleotides amount, constituting a pool of inter-converting metabolites. In this case, the normalization to the total unitary area was impracticable, since the solubilisation of compounds from protein hydrolysis during storage causes a clear increment of the total area in the spectra. The alternative normalization step, based on the internal standard

addition, was neither a good choice, because of the unknown dilution of extracts. For these reasons, the entire pool of metabolites originating from the ATP degradation pathway was used for normalization. After this pre-statistical step, the individual amino acid content was studied, pointing out different evolutions: the concentration of taurine, not involved in any lysis, did not change much during storage; the amount of alanine, phenylalanine, glycine, tryptophan, methionine, isoleucine, leucine, valine, glutamate and tyrosine increased during storage whilst a simultaneous decrease of histidine and serine was observed. These results confirmed similar changes in the amino acids content observed by other techniques. The advantage of using NMR spectroscopy is associated to its ability to quantify several classes of compound in one shot. By analyzing the same spectra, it was observed that, whilst the amount of amino acids changed at a constant rate during storage, some other molecules slowly changed up to day 4, then started varying faster, due to bacterial development. To prove this, typical metabolites of bacterial development were quantified (i.e. lactate, acetate, glucose, succinate and ethanol) and the hypothesis was then demonstrated. This study emphasizes how a holistic view, obtained by applying the metabolomics principles to the whole molecular profile, can be helpful to understand the processes going on during food storage and this approach can be even more robust when a freshness index is built by taking into account all metabolites undergoing transformation. Basing their selection on a non-targeted approach some unexpected alteration can be found and be key for the product quality assessment.

Assessing the effect of food processing

Another key aspect affecting food quality is represented by the production technology and the processes which transform the raw material in food (fermentation, pasteurization, cooking, etc.). Nowadays, more and more food products are subject to many transformation steps and processes, which might alter their nutritional, sensory and biochemical properties. It is therefore necessary to assess any possible alteration caused by these procedures, in order to investigate the nature of transformed foods and identify any possible differences caused by varying the manufacturing protocols or the presence/absence of treatments. For example, milk is usually manipulated to lower its fat content, which can change in a range of 3 units of percentage. Only recently, though, studies have been carried out to assess the secondary changes occurring in the milk matrix regarding other nutrients. The aforementioned study from Boudonck et al. (2009) found further differences in milk due to the reduction or not of fat content. GC-MS and LC-MS/MS spectra were

subjected to a two sample t-test, in order to compare fat free with whole milk and reduced fat with fat free milk. It was found that 78 metabolites differed significantly in at least one of the four categories characterized by different fat percentages. Obviously, whole milk compared to reduced or fat free milk contained higher amounts of the major free fatty acids (i.e. palmitate, oleate, stearate and myristate), more cholesterol, 1,2-dipalmitoylglycerol and alpha-tocopherol. The nutritional value related to these nutrients is thus altered in the process of fat extraction, without any relation to milk brands. Other metabolites (amino acids, sugars, purines and pyrimidines, succinate, fumarate and casein peptides) showed altered concentrations between the groups, though these were also related to brands. Another field of application is the production of vegetable conserves. The effect of processing in the industrial production of tomato paste was studied by Capanoglu et al. (2008). Samples taken from different processing steps and from different tomato batches were analyzed by LC-QTOF-MS and untargeted metabolomics. The authors concluded that this approach is more high-throughput than a conventional technique based on the HPLC simply coupled to an online antioxidants detector. LC-MS, in effect, provided a better insight in the actual alteration caused by the processing of tomato for paste production. Not just the amount of caroteinoids and vitamin C, indeed, varied, but 40% of the 3177 mapped metabolites showed altered signals during the process. It seemed clear that two processing steps were accountable for the main differences: the removal of seed and skin and the transition from fruit to breaker. The first causes a great reduction in the concentration of flavonoids and alkaloids, since seeds and skin are richer in those compounds and they do not get completely extracted in the process. The other mentioned step showed, instead, an increase in flavonoids and alkaloids, maybe due to a wound response in the fruit tissues, though the total antioxidant activity decreases because of loss in vitamin C during tomato breaking. Again, metabolomics proved to be a rapid and useful tool to investigate critical processes in the production of foods and in the future it might aid factories to assess the best ways to reduce nutrient loss and alteration. In this light, knowledge-driven industrial production systems would take advantage from metabolomics also in the improvement of traditional food. Beleggia et al. (2011), investigated the effect of pasta making, by following the process from semolina to the final product. Different techniques were employed: GC-MS, LC-MS, HPLC for carotenoid profiling, mineral profiling, and successive statistical analysis through ANOVA and PCA-FA (factor analysis). Results showed that metabolites varied because of the processing steps and also of the processing conditions. PCA scores resulted in the samples discrimination according to types of pasta (PC1) and processing phases (PC2). ANOVA confirmed

these results. Investigation on the quantities of minerals and metabolites was then performed and the differences among pasta samples prepared with varying processing conditions were identified. Phytosterols degrade because of oxidation and this is influenced by factors such as temperature and reaction time. It was found accordingly that the drying step caused a reduction of this metabolites together with hydroxy fatty acids, tocopherols and carotenoids, suggesting also that an accurate temperature adjustment (lower drying temperature) might help in the maintenance of the levels of these molecules and, thus, of a greater nutritional value, in pasta. Variations were also observed in carotenoid degradation probably due to the kneading process, which favours lipoxygenase in its oxidation of PUFA and, as a consequence, of carotenoids. Different types of semolina showed different behaviours in this step, due to their characteristics and enzyme activities. Another change was observed in the content of total sugars, which showed a first increase during the mixing step, due to optimal conditions for alpha-amylase, and a later reduction after the drying phase, probably due to Maillard reaction and the inactivation of amylase. Again, it is clear how the metabolomics approach gives insight in many of the different biochemical changes occurring in food matrices during transformation steps, pointing out that the quality of the raw material is not the only parameter which influences the final product, but also the optimal processing conditions, fundamental to the maintenance of high nutritional values in conventional as well as in the functional foods. It is thus clear how an analytical approach employing different techniques and ensuing multivariate and statistical data analysis is able to extract the relevant information out from the many different operating conditions affecting the behaviour of food components and metabolites through the processing procedures. This is necessary for food industry in order to assess and then control the processing steps and conditions for the production of the best food possible.

Assessing the effect of digestion

Having assessed the quality of a food product is not enough to prove the real benefits it might have on human health. The food matrix will undergo the whole digestive process and its beneficial components will need to be released in a correct form, in order to diffuse, through the digestion fluid, towards the gut membrane , where they will be absorbed and transferred to the blood stream, thus becoming bioavailable, and ready to exert their bioactivity on the human body.

Just in recent years, research has started focusing on the effect of digestion on the composition and structure of food products and a very few studies were addressed to distinguish foods on the

basis of their digestibility. This, though, is fundamental for the understanding of true nutritional and health properties of food products and for their promotion. A study from Toydemir et al. (2013) investigated the effect of simulated gastrointestinal (GI) digestion for sour cherry fruit and nectar samples. They proved that there was a greater anthocyanin recovery after the digestion of nectar rather than that of fruits. This might be due to a greater stability of these metabolites in the nectar, maybe because of the matrix structure, the pH, the temperature and other components present in the nectar. In effect, the nectar had more than 50% sucrose added, and it was already shown how this ingredient might preserve and stabilise anthocyanins (Wrolstad et al., 1990). This analysis employed spectrophotometric assays and HPLC, in order to get a full phenolic profiling. Again, the use of compatible techniques is necessary to test the results obtained with each method. Another research by Bordoni et al. (2011) compared in vitro digestion of Parmigiano Reggiano cheese, characterized by two different aging times (15 and 30 months), by analysing NMR spectra of digestates at different digestive steps. It was assessed that spectra of undigested aqueous extracts showed a greater presence of sharp signals (amino acids) in the 30-months-aged Parmigiano due to microbial fermentation during aging. The main changes during the digestive steps were found to be caused by the increase of broad signals from casein digestion. The study then focused on the analysis of particular spectral regions (amide and aromatic protons). It was assessed that higher amide areas were found in the 15- month-aged Parmigiano, associated to large-sized protein fragments, which were released during digestion of this cheese. The aromatic area, instead, showed the same trend in both types of samples, indicating that this kind of amino acids gets released majorly as bound to small peptides. This research exploited a metabolomics approach, employing two different NMR techniques (1D-NOESY and DOSY), while comparing different food products digested with the same protocol, thus highlighting the changes of their digestibility, as the consequence of the food production protocol. In this way it is possible to evaluate the actual digestibility of foodstuff, and the multivariate data analysis would separate the digestion process in different components, each characterized by a different molecular profile, observable during the sequential digestion phases (i.e., protein hydrolysis and peptide release in the oral, gastric and duodenal tracts). Three studies published in 2014 (Pan et al., Bordoni et al., Ferranti et al.) shed light on meat digestion by in vitro simulation and metabolomics. The research from Bordoni et al. (2014) focused on the evaluation of protein bio-accessibility of Bresaola product (cured beef). Samples were digested and collected at 5 different check points, in order to assess the protein hydrolysis pathway during the various digestive steps. Protein digestion was

evaluated with a holistic approach, by employing a combination of Bradford assays, SDS-PAGE, 1H-NMR and TD-NMR. 1H-NMR spectra were normalized with reference to the L-carnitine signal, a metabolite which keeps an unaltered concentration during digestion. In this case other recurrent normalisation procedures could not be used, because the total spectral area changes during the digestive phases, and the internal standard TSP, added for frequency calibration, disappears from solution during digestion due to its absorption on the matrix molecules. The obtained spectra were used to corroborate the results of the other techniques, by following the appearance of small and medium-sized peptides not detectable by either the Bradford assay or the SDS-PAGE, the latter techniques only witnessing the disappearance of large proteins. TD-NMR gives further information on the modification of the matrix structure, indicating that mastication provokes, in the oral phase, the exit of 50% of water from myofibrils and the entrance in the fibre bundles of 15% of digestion juices; moreover, the same technique shows unbundling of meat fibres during the gastric phase, favouring the action of pepsin. The study from Picone et al. (2014) also focused on the Breasola digestibility in order to assess the possible release of bioactive peptides from the food matrix. In effect, protein-rich matrices release, during digestion, small peptides with potential benefits on human health, since some of them have been proved to have antihypertensive or antioxidant effects. This research employed in-vitro digestion and sampling at 5 different check points along the digestion process. Samples were then analysed through SDS-PAGE, 2-dimensional electrophoresis (2-DE), subsequent protein spot hydrolysis, MALDI-TOF-MS analysis and nano-HPLC-ESI-Q-TOF-MS/MS analysis of peptides. Through these various techniques the profiles of peptide and protein mixtures have been established. 1D-SDS-PAGE and the coupled MS analysis showed that some proteins were already hydrolyzed in the Bresaola samples before gastric digestion since, in the post-mortem phase, proteins from the sarcoplasmic and myofibrillar tissues undergo hydrolysis by calpains and other endogenous proteases. The undigested compounds, studied by 2-DE and MALDI-TOF-MS, are found to be serum albumin, actin, tropomyosins, myosin light chain 1 and 3, and some of their fragments. During gastric and duodenal digestion, Bresaola proteins are subject to the attack of different proteases. The gastric phase showed the degradation of some sarcoplasmic proteins and the appearance of polypeptides, the latter undetectable by SDS-PAGE. RPHPLC and MALDI-TOF-MS were mainly employed to monitor the evolution of the peptide profiles during the digestive steps. RP-HPLC showed that myofibrillar proteins are demolished in the duodenal digestion after being released during the gastric phase, whereas MALDI-TOF-MS spectra disclosed the size of the newly formed smaller peptides. Nano- HPLC-ESI-Q-

TOF-MS/MS analysis showed the presence of bioactive, ACE-inhibitors, peptides or their precursors after digestion, though it still needs validation of their bioavailability in in-vivo systems. This study showed how omics based on complementary analytical techniques can explain many underlying processes happening during food digestion. Pan et al. (2014), instead, studied the digestion of ham to investigate its nutrients' bioaccessibility using SDS-PAGE, microscopy and 1H-NMR spectroscopy. SDS-PAGE and NMR spectra showed the kinetics of protein degradation into smaller peptides. Lactate, because unaffected by digestion, was used as a normalisation reference for NMR spectra in order to quantify the nutrients in the different digestive steps and understand their fate. It appears how lipids and macromolecules increase greatly during gastric and duodenal digestion, though they stop growing in quantity after 60 minutes of digestion. NMR spectra also showed the presence in samples, after digestion, of potential bioactive or positive compounds, like carnosine or choline, and this can be used as an index of product digestibility or food quality, associated to the accessibility of nutrients released upon digestion.

Conclusions

Metabolomics is nowadays a widespread omic approach fruitfully exploited in many applied sciences. The key aspects of this new science, like the use of rapid and effective techniques, made it even fundamental in many fields. Food science has not been an exception and it is now employing this approach in many researches. The increased consumers’ demand for higher food quality and its relative substantiation has made necessary to develop new techniques able to screen as much as possible the metabolic and nutritional profile of food products. It is thus clear how metabolomics fitted perfectly in this hole and was therefore used as a new explorative and classifying tool in this sense. Many studies are now been made in various assessments of food quality employing analytical techniques suitable for metabolomics, such as mass spectrometry or NMR spectroscopy and subsequent multivariate data analysis (Table 1). This type of research methodology is today the basis in many food researches. This chapter highlighted just some of the many uses of metabolomics in food science, like in assessing differences between GM and non-GM products or in identifying particular metabolites typical of food from a certified geographical origin. It is thus evident how metabolomics will increasingly be an aid in this research field and needs to develop new high-technology methods and protocols in order to be able to assess completely and automatically food quality.

REFERENCES CHAPTER 3.1

Baker, J. M., Hawkins, N. D., Ward, J. L., Lovegrove, A., Napier, J. A., Shewry, P. R., Beale, M. H. "A metabolomic study of substantial equivalence of field-grown genetically modified wheat." Plant Biotechnology Journal 4 (2006): 381-392.

Baudo, M. M., Lyons, R., Powers, S., Pastori, G. M., Edwards, K. J., Holdsworth, M. J., Shewry, P. R. "Transgenesis has less impact on the transcriptome of wheat grain than conventional breeding." Plant Biotechnology Journal 4 (2006): 369-380.

Beleggia, R., Platani, C., Papa, R., Di Chio, A., Barros, E., Mashaba, C., Wirth, J., Fammartino, A., Sautter, C., Conner, S., Rauscher, J., Stewart, D., Cattivelli, L. "Metabolomics and food processing: from semolina to pasta." Journal of agricultural and food chemistry 59 (2011): 9366-9377.

Benincasa, C., Lewis, J., Perri, E., Sindona, G., Tagarelli, A. "Determination of trace element in Italian virgin olive oils and their characterization according to geographical origin by statistical analysis." Analytica chimica acta 585 (2007): 366-370.

Bonte, Anja, Neuweger, H., Goesmann, A., Thonar, C., M.der, P., Langenk.mper, G., Niehaus, K. "Metabolite profiling on wheat grain to enable a distinction of samples from organic and conventional farming systems." Journal of the science of food and agriculture (2014). doi: 10.1002/jsfa.6566.

Bordoni, A., Picone, G., Babini, E., Vignali, M., Danesi, F., Valli, V., Di Nunzio, M., Laghi, L., Capozzi, F. "NMR comparison of in vitro digestion of Parmigiano Reggiano cheese aged 15 and 30 months." Magnetic Resonance in Chemistry 49 (2011): S61-S70.

Bordoni, A., Laghi, L., Babini, E., Di Nunzio, M., Picone, G., Ciampa, A., Valli, V., Danesi, F., Capozzi, F. "The foodomics approach for the evaluation of protein bio-accessibility in processed meat upon in vitro digestion. " Electrophoresis (2014). doi:10.1002/elps.201300579.

Boudonck, K. J., Mitchell, M. W., Wulff, J., Ryals, J. A. "Characterization of the biochemical variability of bovine milk using metabolomics." Metabolomics 5 (2009): 375-386.

Brunner, M., Katona, R., Stef.nka, Z., Prohaska, T. "Determination of the geographical origin of processed spice using multielement and isotopic pattern on the example of Szegedi paprika." European Food Research and Technology 231 (2010): 623-634.

Cajka, T., Riddellova, K., Tomaniova, M., Hajslova, J. "Ambient mass spectrometry employing a DART ion source for metabolomic fingerprinting/profiling: a powerful tool for beer origin recognition." Metabolomics 7 (2011): 500-508.

Capanoglu, E., Beekwilder, J., Boyacioglu, D., De Vos, R. C., Hall, R. D. "Changes in antioxidant and metabolite profiles during production of tomato paste." Journal of agricultural and food chemistry 56 (2008): 964-973.

Chen, P., Harnly, J. M., Lester, G. M. "Flow injection mass spectral fingerprints demonstrate chemical differences in rio red grapefruit with respect to year, harvest time, and conventional versus organic farming." Journal of agricultural and food chemistry 58 (2010): 4545-4553.

Chudzinska, M., Baralkiewicz, D. "Estimation of honey authenticity by multielements characteristics using inductively coupled plasma-mass spectrometry (ICP-MS) combined with chemometrics." Food and chemical toxicology 48 (2010): 284-290.

Ciampa, Alessandra, Picone, G., Laghi, L., Nikzad, H., Capozzi, F. "Changes in the Amino Acid Composition of Bogue (Boops boops) Fish during Storage at Different Temperatures by 1H-NMR Spectroscopy." Nutrients 4 (2012): 542-553.

Ferranti , P., Nitride, C., Nicolai, M. A., Mamone, G., Picariello, G., Bordoni, A., Valli, V., Di Nunzio, M., Babini, E., Marcolini, E., Capozzi, F. "In vitro digestion of Bresaola proteins and release of potential bioactive peptides." Food Research International (2014). doi: 10.1016/j.foodres.2014.02.008.

Fraser, P. D., Enfissi, E. M., Halket, J. M., Truesdale, M. R., Yu, D., Gerrish, C., Bramley, P. M. "Manipulation of phytoene levels in tomato fruit: effects on isoprenoids, plastids, and intermediary metabolism." The Plant Cell Online 19 (2007):

3194-3211.

Garcìa-Villalba, R., Le.n, C., Dinelli, G., Segura-Carretero, A., Fernàndez-Gutiérrez, A., Garcia-Cañas, V., Cifuentes, A. "Comparative metabolomic study of transgenic versus conventional soybean using capillary electrophoresis–time-of-flight mass spectrometry." Journal of Chromatography A 1195 (2008): 164-173.

Graham, S. F., Ruiz-Aracama, A., Lommen, A., Cannizzo, F. T., Biolatti, B., Elliott, C. T., Mooney, M. H. "Use of NMR metabolomic plasma profiling methodologies to identify illicit growth-promoting administrations." Analytical and bioanalytical chemistry 403 (2012): 573-582.

Holmes, E., Loo, R. L., Stamler, J., Bictash, M., Yap, I. K. S., Chan, Q., Ebbels, T., De Iorio, M., Brown, I. J., Veselkov, K. A., Daviglus, M. L., Kesteloot, H., Ueshima, H., Zhao, L., Nicholson, J. K., Elliott, P. "Human metabolic phenotype diversity and its association with diet and blood pressure." Nature 453 (2008): 396-400.

Hong, V., Wrolstad, R. E. "Characterization of anthocyanin-containing colorants and fruit juices by HPLC/photodiode array detection." Journal of Agricultural and Food Chemistry 38 (1990): 698-708. Jung, Y., Lee, J., Kwon, J., Lee, K. S., Ryu, D. H., Hwang, G. S. "Discrimination of the geographical origin of beef by 1H NMR-based metabolomics." Journal of agricultural and food chemistry 58 (2010): 10458-10466.

Kok E. J., Kuiper H. A. "Comparative safety assessment for biotech crops." Trends Biotechnol. 21 (2003) 439–444.

Laghi, L., Versari, A., Marcolini, E., Parpinello, G. P. "Metabonomic Investigation by 1HNMR to Discriminate between Red Wines from Organic and Biodynamic Grapes." Food and Nutrition Sciences 5 (2014): 52.

Leon, C., Rodriguez-Meizoso, I., Lucio, M., Garcia-Ca.as, V., Iba.ez, E., Schmitt-Kopplin, P., Cifuentes, A. "Metabolomics of transgenic maize combining Fourier transform-ion cyclotron resonance-mass spectrometry, capillary electrophoresis-mass spectrometry and pressurized liquid extraction." Journal of Chromatography A 1216 (2009): 7314-7323.

Levandi, T., Leon, C., Kaljurand, M., Garcia-Ca.as, V., Cifuentes, A. "Capillary electrophoresis time-of-flight mass spectrometry for comparative metabolomics of transgenic versus conventional maize." Analytical chemistry 80 (2008): 6329-6335.

Lo Feudo, Gabriella, Naccarato, A., Sindona, G., Tagarelli, A. "Investigating the origin of tomatoes and triple concentrated tomato pastes through multielement determination by inductively coupled plasma mass spectrometry and statistical analysis." Journal of agricultural and food chemistry 58 (2010): 3801-3807.

Mallamace, D., Corsaro, C., Salvo, A., Cicero, N., Macaluso, A., Giangrosso, G., Ferrantelli, V., Dugo, G. "A multivariate statistical analysis coming from the NMR metabolic profile of cherry tomatoes.(The Sicilian Pachino case)." Physica A: Statistical Mechanics and its Applications (2014). doi: 10.1016/j.physa.2013.12.054

Melgarejo, E., Urdiales, J. L., S.nchez-Jim.nez, F., Medina, M. "Targeting polyamines and biogenic amines by green tea epigallocatechin-3-gallate." Amino acids 38 (2010): 519-523.

Novotna, H., Kmiecik, O., Gałązka, M., Krtkov., M.,Hurajov., A., Schulzov., V., Hallmann, E., Rembiałkowska, E., Hajšlov., J. "Metabolomic fingerprinting employing DARTTOFMS for authentication of tomatoes and peppers from organic and conventional farming." Food Additives & Contaminants: Part A 29 (2012): 1335-1346.

Pan, X., Smith, F., Cliff, M. T., Capozzi, F., Mills, E. C."The Application of Nutrimetabolomics to Investigating the Bioaccessibility of Nutrients in Ham Using a Batch in Vitro Digestion Model." Food and Nutrition Sciences 5 (2014): 17.

Picone, G., Mezzetti, B., Babini, E., Capocasa, F., Placucci, G., Capozzi, F "Unsupervised principal component analysis of NMR metabolic profiles for the assessment of substantial equivalence of transgenic grapes (Vitis vinifera)." Journal of agricultural and food chemistry 59 (2011 a): 9271-9279.

Picone, G., Balling Engelsen, S., Savorani, F., Testi, S., Badiani, A., & Capozzi, F. "Metabolomics as a powerful tool for molecular quality assessment of the fish Sparus aurata." Nutrients 3 (2011 b): 212-227.

Regal, P., Anizan, S., Antignac, J. P., Le Bizec, B., Cepeda, A., Fente, C. "Metabolomic approach based on liquid chromatography coupled to high resolution mass spectrometry to screen for the illegal use of estradiol and progesterone in cattle." Analytica chimica acta 700 (2011): 16-25.

Röhlig, R. M., Engel, K.-H. "Influence of the input system (conventional versus organic farming) on metabolite profiles of maize (Zea mays) kernels." Journal of agricultural and food chemistry 58 (2010): 3022-3030.

Savorani, F., Picone, G., Badiani, A., Fagioli, P., Capozzi, F., & Engelsen, S. B. "Metabolic profiling and aquaculture differentiation of gilthead sea bream by 1H NMR metabonomics." Food chemistry 120 (2010): 907-914.

Savorani, F., Capozzi, F., Engelsen, S.B., Dell'Abate, M. T., Sequi, P. "Pomodoro di Pachino: An Authentication Study Using 1H-NMR and Chemometrics-Protecting Its P.G.I. European Certification" Magnetic Resonance in Food Science: Challenges in A Changing World (2009):158-166.

Steiner, C., Teixeira, W. G., Lehmann, J., Nehls, T., de Mac.do, J. L. V., Blum, W. E., Zech, W. "Long term effects of manure, charcoal and mineral fertilization on crop production and fertility on a highly weathered Central Amazonian upland soil." Plant and soil 291 (2007): 275-290.

Toydemir, G., Capanoglu, E., Kamiloglu, S., Boyacioglu, D., de Vos, R. C., Hall, R. D., Beekwilder, J. "Changes in sour cherry (Prunus cerasus L.) antioxidants during nectar processing and in vitro gastrointestinal digestion." Journal of Functional Foods 5 (2013): 1402-1413.

Vallverdù.-Queralt, A., Medina-Remòn, A., Casals-Ribes, I., Amat, M., & Lamuela- Raventòs, R. M."A metabolomic approach differentiates between conventional and organic ketchups." Journal of agricultural and food chemistry 59 (2011): 11703-11710.

Van Ruth, S. M., Alewijn, M., Rogers, K., Newton-Smith, E., Tena, N., Bollen, M., Koot, A. "Authentication of organic and conventional eggs by carotenoid profiling." Food chemistry 126 (2011): 1299-1305.

Van Ruth, S. M., Koot, A. H., Brouwer, S. E., Boivin, N., Carcea, M., Zerva, C. N., Haugen, J.-E., Höhl, A., Köroglu, D., Mafra, I., Rom, S. "Eggspectation: organic egg authentication method challenged with produce from ten different countries." Quality Assurance and Safety of crops & foods 5 (2013): 7-14.

Walsh, Marianne C., Brennan, L., Pujos-Guillot, E., Sébédio, J.-L., Scalbert, A., Fagan, A., Higgins, D. H., Gibney, M. J. , Influence of acute phytochemical intake on human urinary metabolomic profiles." The American journal of clinical nutrition 86 (2007): 1687-1693.

Wishart, D. S. "Metabolomics: applications to food science and nutrition research." Trends in Food Science & Technology 19 (2008): 482-493.

Woese, K., Lange, D., Boess, C., Bögl, K. W. "A comparison of organically and conventionally grown foods—results of a review of the relevant literature." Journal of the Science of Food and Agriculture 74 (1997): 281-293.

Zörb, C., Langenkämper, G., Betsche, T., Niehaus, K., Barsch, A."Metabolite profiling of wheat grains (Triticum aestivum L.) from organic and conventional agriculture." Journal of agricultural and food chemistry 54 (2006): 8301-8306.

Table 1. Metabolomics studies applied to food science

Type of study	Techniques	Statistical Tools	Food Analysed	Discriminant Metabolites	Reference Article
Genetic modification	Real Time PCR, enzyme assays, HPLC, TOF-MS	-	Tomatoes	-	Fraser et. al (2007)
Genetic modification	CE-TOF-MS	PCA, Student's t-test	Maize	L-proline-betaine, L-carnitine (typical of GM maize)	Levandi et al. (2008)
Genetic modification	FT-ICR-MS, CE-TOF—MS	PCA, PLS-DA	Maize	L-carnitine	Leon et al. (2009)
Genetic modification	NMR, GC-MS	PCA, ANOVA	Wheat	-	Baker et al. (2006)
Genetic modification	Transcriptomic	-	Wheat	-	Baudo et al. (2006)
Genetic modification	CE-ESI-TOF-MS		Soybean	4-hydroxy-L-threonine (only in conventional soybean)	García-Villalba et al. (2008)
Genetic modification	NMR	PCA, ANOVA, Student's t-test	Grapes	tryptophan and indole derivatives (typical of GM grapes)	Picone et al. (2011 a)
Organic vs. conventional	FI-ESI-TOF-MS, FI-ESI-IT-MS	ANOVA, PCA	Grapefruit	-	Chen et al. (2010)
Organic vs. conventional	GC-MS	Tukey's test	Wheat	-	Zörb et al. (2006)
Organic vs. conventional	GC-MS	ANOVA, PCA	Maize	Myo-inositol, malic acid and phosphate (in organic samples)	Röhlig et al (2010)
Organic vs. conventional	GC-MS	Student's t-test, PCA	Wheat	Myo-inositol and malic acid (in organic crops)	Bonte et al. (2013)
Organic vs. conventional	DART, TOFMS	PCA, LDA	Sweet bell peppers	-	Novotnà et al. (2012)
Organic vs. conventional	LC-ESI-QqQ	ANOVA	Ketchup	caffeoylquinic and dicaffeoylquinic acids, caffeic and caffeic acid hexosides, kaemp-ferol-3-O-rutinoside, ferulic-O-hexoside, naringenin-7-O-glucoside, naringenin, rutin and quercetin (in organic ketchup); glutamylphenylalanine and N-malonyltryptophan (in conventional ketchup)	Vallverdù-Queralt et al. (2011)
Organic vs. conventional	NMR	ANOVA, PCA	Red wines	Tyrosine, trans-caffeic acid, glutamine	Laghi et al (2014)
Geographical origin	NMR	PCA, OPLS-DA	Beef	Different aminoacids depending on the origin	Jung et al.

				(Korea, Australia, US, New Zealand)	(2010)
Geographical origin	NMR	PCA, LDA	Cherry tomatoes	-	Savorani et al. (2009)
Geographical origin	HR-MAS NMR	PCA, Student's t-test	Cherry tomatoes	fructose and glucose, glutamine, glutamate, aspartate and GABA (higher in Pachino tomatoes), fatty acids, alanine, methanol and acetylglutamic acid (lower in Pachino tomatoes)	Mallamace et al. (2014)
Geographical origin	DART-TOFMS, SPME-GC-TOFMS	PCA, PLS-DA, LDA, ANN-MLP	Beer	-	Cajka et al. (2010)
Rearing conditions	GC-MS, LC-MS/MS	PCA, ANOVA	Bovine milk	Hippuric acid (for organic milk)	Boudonck et al. (2009)
Rearing conditions	HPLC	k-nearest-neighbour	Eggs	-	Van Ruth et al. (2011)
Rearing conditions	HPLS	k-nearest-neighbour	Eggs	-	Van Ruth et al. (2013)
Rearing conditions	NMR	PCA, PLS, ECVA, iECVA	Gilthead sea bream	Molecules from the glucose metabolism (higher in farmed fish)	Savorani et al. (2010)
Rearing conditions	NMR	Paired t-test, PCA	Sparus aurata	Hydroxylic and aliphatic regions	Picone et al (2011 b)
Rearing conditions	NMR	-	Bogue fish	Amino acids	Ciampa et al. (2012)
Processing	LC-QTOF-MS, HPLC		Tomato paste	Flavonoids, alkaloids, vitamin C	Capanoglu et al. (2008)
Processing	GC-MS, LC-MS, HPLC	ANOVA, PCA-FA	Pasta	Minerals, phytosterols, fatty acids, carotenoids, tocopherols, sugars	Beleggia et al. (2011)
Digestion	Spectrophotomer, HPLC	-	Cherry fruit	-	Toydemir et al. (2013)
Digestion	NMR		Parmigiano Reggiano	Amino acids	Bordoni et al. (2011)
Digestion	SDS-PAGE, microscopy, NMR	PCA	Ham	-	Pan et al. (2014)
Digestion	Bradford assay, SDS-PAGE, NMR, TD-NMR	-	Breasola	-	Bordoni et al. (2014)
Digestion	SDS-PAGE, 2-DE, MALDI-TOF-MS, nano-HPLC-ESI-Q-TOF-MS/MS	-	Bresaola	-	Ferranti et al. (2014)

3.2 Definition of food quality by nmr-based foodomics

Alessia Trimigno[a], Flaminia Cesare Marincola[b], Nicolò Dellarosa[a], Gianfranco Picone[a], Luca Laghi[a]*

[a]Department of Agricultural and Food Sciences, University of Bologna, Piazza Goidanich 60, 47521 Cesena, Forlì Cesena, Italy

[b]Department of Chemical and Geological Sciences, University of Cagliari, Cittadella Universitaria, SS 554, km 4.5, Monserrato, 09042 Cagliari, Italy

Journal: Current Opinion in Food Science – Volume 4, August 2015, Pages 99-105

Corresponding author

*Luca Laghi (l.laghi@unibo.it - Department of Agricultural and Food Sciences, University of Bologna, Piazza Goidanich 60, 47521 Cesena, Forlì-Cesena, Italy - Tel +39/0547338106)

Role of the PhD candidate: help in the definition of each contributor's role in the writing of the review, organization and correction of the written parts, together with the writing of the main section of the review.

Abstract

Quality definition of food includes several complex factors like physical, compositional and microbial features, modifications induced by technological processes or storage, nutritional value and safety. Foodomics is a holistic approach applying omics technologies to observe food along the entire production/consumption chain. In the present review, we present key applications of nuclear magnetic resonance in foodomics described in the 2012-2015 period, in the quest for robust and thorough information required by the scientific community. In doing so, we summarize the issues connected to food traceability and authenticity, composition and physical characteristics, processing and storage and health, that mostly impact food quality.

Highlights

- Foodomics has been proven a promising approach for food quality investigation.
- NMR has boosted an omic approach to food quality, due to its high reproducibility.
- A common trait of key foodomics papers is transparency of each investigation step.
- Advancements in foodomics are foreseen through interdisciplinary networking activity.

Introduction

Soon after the advent of genomics, transcriptomics, proteomics and metabolomics, aimed to a holistic understanding of the complex human biology and physiology, it has been natural to apply the same approaches to food. The information collected in this way at each step of the production/consumption chain has been enclosed into the single definition of foodomics, which now is defined as “ the discipline that studies the food and nutrition domains through the application and integration of advanced omics technologies to improve consumer's well-being, health, and confidence” [1]. Among the techniques used for this discipline, Nuclear Magnetic Resonance (NMR) has given a great boost to the new approach, thanks, in particular, to the high reproducibility of its observations [2]*. Goodacre [3], in a recent divertissement, has noted that the growth of the metabolomics literature reminded that of microorganisms. The lag phase is represented by a handful of papers, listed by Goodacre [3], which created the conceptual framework. A key element of the initial phase can be identified also in the development of software and algorithms (i.e. Projection on Latent Structures [4]) specifically tailored to highlight the useful features in the overwhelming information represented by large experimental datasets.
The phase of rapid growth can be felt in the works focusing on the most diverse biofluids, foods and raw materials, aiming at verifying the applicability of the metabolomics approach. A common trait of these papers is some degree of failure in making each step of the investigation (i.e. experimental design, data generated and analytical tools) totally transparent. The stationary, desirable, phase is represented by works where experimental design, data generated and the means of analysis are made publicly available, and the level of metabolite identification is properly assigned [5], as well as the level of confidence of each key statement. Papers actively contribute to a generalized increase of quality, when they fulfill such requirements better than any other of their own field, by forcing the scientific community to adapt. In the present review, we mainly focus on the literature of the 2012-2015 period and present key applications of NMR in foodomics, in the quest for robust and thorough information required by the scientific community. In doing that, we will run through the steps of the food production chain, from the origin of the raw material to the transformation that food undergoes during storage, that mostly affect food quality, as summarized in figure 1.

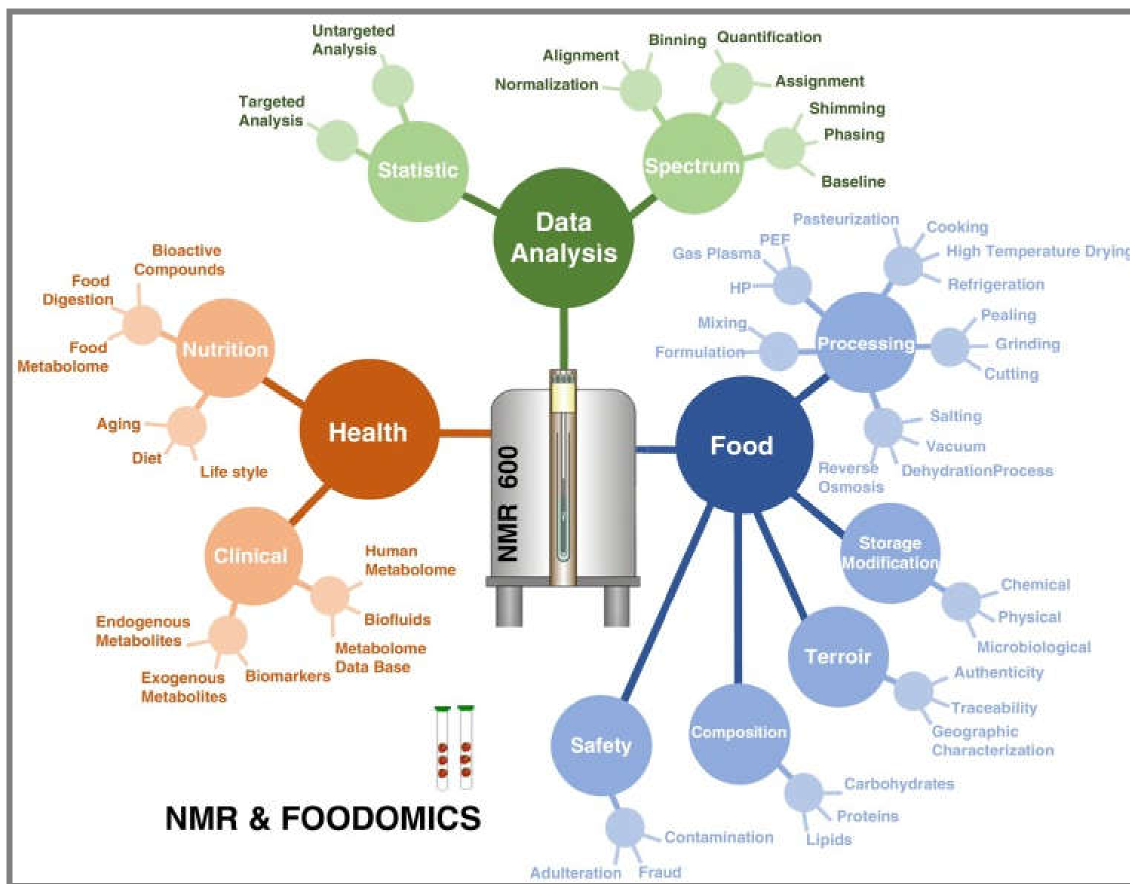


Figure 1. Infografic of the information that can be obtained along the food production/consumption chain by NMR, through a foodomic approach. For the spectra, processing steps see reference [2].

Traceability, authenticity and safety of food

A relevant percentage of the commercial value of several foods and beverages relies on the ensemble of climate, land, cultural practices and history of the raw material, collectively enclosed in the term “terroir”. NMR has been used from the eighties with the purpose of geographic characterization, by studying the distribution of stable isotopes of the bio-molecules [6]. A natural consequence of the advent of the “omic” analytical techniques has been to consider the entire metabolome in the perspective of terroir characterization. This is generally observed in a non-targeted fashion [7], that is without focusing on specific molecules but letting features of the NMR spectra emerge from the entire spectrum profile, through the appropriate mathematical treatments. Examples of this approach are the work by Gallo et *al.* [8] on table grape, and the work by Hohman et *al.* [9] about tomato. Unfortunately, each aspect of a food terroir potentially affects the metabolome, so that general applicability represents the weak point of any mathematical model trying to relate features of an NMR spectrum with a single aspect of the origin of a food. This is probably why no method based on non-targeted fingerprinting has been accepted for food official controls [7]. The recently published work that can be considered a to-date benchmark in

this respect is the one by Godelmann et *al.* [10]** testing wine screening method WineScreener™. The analysis of 600 German wine samples, produced during 2 years in 5 areas from 10 grape varieties, led to the maximum correct prediction of geographical origin, year of vintage and grape variety (89%, 97% and 95% respectively). It has been suggested that the studies on food safety, similarly to those on terroir, would bring information to a higher level if including an omic-untargeted approach, because many features that raise concerns about the healthiness of food, as genetic modification [11] or microorganisms development [12,13], are likely to influence large portions of food or raw material molecular profile [14]. The main challenge for scientists facing non-targeted analysis is to correctly define appropriate biomarkers from raw NMR spectra containing hundreds of metabolites. The rationale of this practice is that the inclusion of a feature of a spectrum unrelated to a characteristic of interest leads to lower correct prediction rates or less parsimonious models. Consequentially, an increasing number of papers has been expressly focused on metabolites selection as the key part of mathematical data modeling [15,16]*. As a general trend, in the last decade data treatment has taken advantages of computer-aided multivariate analysis tools that allow the simultaneous model building and variable selection by associating to each metabolite a proper weight in the model. Examples are interval partial least square (*i*PLS), interval extended canonical variable analysis (*i*ECVA) [17] and sparse PLS [18]. Those models have largely replaced statistical methods based on Normal distributions, for example t-test or ANOVA, because less prone to false discoveries, i.e. false positives or negatives, which typically affect univariate analysis [19]. Even though computer-aided methods have considerably improved the data analysis performance, their incorrect use, typically the lack of a proper validation step, can lead to false correlation between metabolites and the characteristic of interest. There is consensus that a key step towards reliable non-targeted fingerprinting methods is the exchange and comparison of data between the stakeholders involved in foodomics observations, through databases dedicated to food, where standard format NMR spectra are enriched by metadata and powerful data mining engines [3]. The last few years have seen for the purpose the launch of MetaboLights by the European Bioinformatic Institute [20], the launch of NIH Metabolomics Workbench (http://www.metabolomicsworkbench.org), establishing similar data storage infrastructures, and the collection of huge amount of data on single food products [21,22].

Food composition and physical characteristics

The extensive implementation of automatic spectrometer setup procedures achieved in the last

decades has made quantitative investigations on modern NMR instruments limitedly user dependent. In addition, the effects of suboptimal instrument setup can be accurately mapped [23]. The number of quantitative NMR (qNMR) applications targeted towards specific molecules is therefore increasing [24,25], and this is particularly useful when all the legal requirements of a food can be entirely assessed by means of NMR spectroscopy. This is the case of egg yolk based liqueurs [26], where total sugar and alcohol have been directly quantified, while the egg yolk content has been successfully estimated. Targeted NMR based applications are particularly suitable for (semi-) automatized signals deconvolution procedures, fundamental in sight of the standardized/harmonized operating procedures [7]. Significant steps forward in this respect are represented by the software products like Chenomx, (Chenomx inc., Edmonton, Ca) merging automatic procedures and users guidance though a game-like interaction with the software, Batman [27]** R (www.R-project.org) package, with many pros among which being licenced under the GNU general public license, and MVPACK [28], promising to follow the entire pipeline of NMR spectra processing and data mining. The quantitative applications targeted towards *a priori* selected molecules is stimulating (and is stimulated by) the marketing of cost-effective bench-top, air-cooled, medium field spectrometers. In these instruments, the problems of signal resolution, caused by the magnetic field lower than the one of the cryogenic-cooled counterparts, are partially solved by working on field homogeneity, with modifications of the permanent magnets arrangement originally described by Halbach [29]*. Mathematical relationships granting good rates of correct predictions have been established between NMR spectra and even physical characteristics of food, of key value for transformation. This is the case of meat, where tenderness [30] and water-holding capacity [31], the most important characteristics of meat, together with appearance [32], have been successfully modelled through NMR. This is the case also for milk, the coagulating properties of which have been successfully related to the metabolites profile, observed by ^{1}H-NMR and Principal Component Analysis (PCA) [33].

Food processing and storage

The ideal food combines nutritional and sensorial quality, but the design of the appropriate manufacturing processes is still a considerable challenge [34]. Food processing-technologies and modification phenomena occurring during storage have a general impact on the metabolic pathways of food cells and microorganisms and, in turn, on the food metabolome, making holistic analytical techniques invaluable for the characterization of food quality. Despite the great

potential, their employment in investigations on the consequences of technological processes to the quality of food is erratic. Great attention has been paid by the NMR community to chemical and microbiological evolution phenomena occurring during food storage, with particular emphasis on degradation processes. Observations are accumulating at a fast pace for fish [35,36], meat [12], vegetables [37] and spices [38], taking into consideration not just storage time, but also storage conditions, such as composition of modified atmospheres and temperature [35]. Mechanical treatments, such as peeling, chopping, shredding, and heat transfer treatments still appear to be investigated below their potentialities. Lopez-Sanchez [34] was able to follow the effects of different combinations of heating and blending on the phytochemical composition of tomato, broccoli and carrots purees. In the review by Erikson et *al.* [39], effects of frying and boiling of different species of fish are described. Roasting effects have been observed for coffee [40] and laver product [41]. Freezing causes massive water migration among cell compartments and causes cell membranes breakage, due to the formation of ice crystals. This may allow freeze stored food to be distinguished from the refrigerated one. An example is represented by fish freezing, leading to the formation of dimethylamine [42], that instead can be found only in traces in refrigerated fish. Effects on fish molecular profile due to mass transfer have also already been noticed because of salting [35]. The request for minimally processed food is stimulating the research on non-thermal technological treatments for bacteria reduction or degradative enzymes inactivation, such as irradiation, use of high pressures or application of gas plasma [43]. NMR-based foodomics applications are emerging also in this field, as in the case of ground beef [44]. The high demand for functional food is increasing the research on mass transfer processings alternative to osmosis, such as vacuum impregnation [45]. We found no examples of foodomics investigation in literature, but such gap is likely to be soon filled. The relationships between specific processes applied to food and the features of the NMR spectra profile must be considered with caution, due to the presence of confounding factors. During fish storage, for example, trimethylamine is produced by bacterial spoilage of proteins [39], but its concentration cannot be reliably used as an universal index for the correctness of fish conservation, because deeply influenced by fish breeding too.

Food and health

Today, we are witnessing an increasingly growing opinion that a proper nutrition, along with an adequate lifestyle, plays a key role in the prevention, onset and control of many diseases, among which metabolic syndrome, diabetes and cancer [46]. In order to gain insight into this issue, the

knowledge of which food components influence the health status and their mechanisms of action is crucial. In this context, research in food science and nutrition has started to move in the direction of integrative analysis, finding in foodomics an holistic way to study the complex hierarchical structure (from genes to proteins to metabolites) linking food and health [47-49]. Indeed, the foodomics approach can help the investigation of both the food metabolome and the correlated human metabolome: the characterization of the whole metabolic fingerprint of food might greatly help the analysis of the mechanisms of nutrition at the molecular level, while, on the other side, the investigation of human metabolome in response to a specific diet can be useful to identify novel biomarkers of food intake not necessarily predictable by the sole food composition. This new methodology, giving insights on the effects of nutritional exposure, thus also on the nutritional status and nutritional impact on diseases, is proven to be particularly promising in the prospect of the development of tailored dietary and health recommendations. The foodomics studies that have been setup to investigate the issues connected to nutrition, often collectively considered in the definition of nutri-metabonomics, can be grouped into three main categories. The first and more traditional is the nutritional intervention study, where a group of subjects is given a particular diet and the consequent metabolome changes are investigated. In this case the metabolic profile presents two kinds of metabolites, the exogenous, that can be considered as a marker of the specific food intake, and the endogenous, metabolites generated by our bodies as a consequence of the consumption of that food product. One example is the NMR study from Heinzmann [50] where subjects were followed after the acute ingestion of specific food products (fish, fruits, wine and grapes) and both the direct effects (presence of biomarkers) and indirect effects (metabolic pathway alterations) were observed on urine samples. Rasmussen et *al.*, [51] investigated both the effects of the consumption of a high/low protein diet and of fiber and dietary glycemic index. The second type of nutritional assessment is the one that analyses both dietary data and metabolome from biofluids of a selected population. In this case, dietary patterns and trends are observed in the population by the statistical analysis of the dietary information collected. Biomarkers are looked for in the human metabolome as a hint to the real consumption of the declared items. Savorani et *al.*, [52] performed this kind of observational study and assessed the presence of three distinct dietary patterns both with the employment of the food diaries and the urine and plasma metabolome. Similar work has been carried out by De Filippis et *al.*, [53] where NMR-based metabolomics was employed to extract the relevant dietary patterns of vegans, vegetarians and omnivores and then to find their impacts on the saliva metabolome. The last type

of nutri-metabonomic research is the one studying population with different health conditions. In this case, a nutritional intervention is usually employed, and dietary data might be collected in order to assess the compliance to the diet. Following the intervention, the metabolome is analyzed, in order to assess the specific metabolic profile and thus, identify the biomarkers, of the effect of food ingestion on health and metabolic pathways. Lehtonen [54] observed the modifications of post prandial fingerpirnts of human urine after consumption of lingonberries as a supplement to an oil-rich meal. Moazzami [55] investigated the effect of rye bread on postmenopausal women, showing that the dietary intervention leads to shifts in metabolic pathways that can have beneficial effects on the selected population.

Conclusions

In conclusion, foodomics has been proven a powerful tool for many different aspects of food quality definition. Its high-throughput approach can give insights on the whole metabolic profile of food products, helping the characterization and the definition of specific quality features that make certain foods unique. Contributions toward this direction have been provided in studies of food authentication or in investigations concerning processing and storage procedures. Furthermore, foodomics has been giving boost to new kind of nutritional studies, aimed at understanding how metabolites contained in food can influence human metabolism and health. The Foodomics community foresees advancements in this new omic field through an intense networking activity. Indeed, there is an evident necessity to increase the level of collaboration within experts of different disciplines, such as bioinformatics, chemometrics, analytical chemistry, biochemistry or statistics [56]. This networking approach will help creating more accessible and reliable information through the employment of specific compounds databases and the definition of good operating procedures and standard protocols in order to generate a more common perspective and more robust data.

Conflict of interest statement

The authors declare no conflict of interest

REFERENCES CHAPTER 3.2

[1] Cifuentes A: Special issue: advanced separation methods in food analysis. J Chromatogr A 2009, 1216:7109-7358.

[2] Laghi L, Picone G, Capozzi F: Nuclear magnetic resonance for foodomics beyond food analysis. TrAC-Trend Anal Chem 2014, 59:93-102.

[3] Goodacre R: I spy with my little eye something beginning with … 'H'. Metabolomics 2015, 11:6-8.

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

[5] Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, Fan TM, Fiehn O, Goodacre R, Griffin JL, et al.: Proposed minimum reporting standards for chemical analysis. Metabolomics 2007, 3:211-221.

[6] Versari A, Laurie VF, Ricci A, Laghi L, Parpinello GP: Progress in authentication, typification and traceability of grapes and wines by chemometric approaches. Food Res Int 2014, 60:2-18.

[7] Esslinger S, Riedl J, Fauhl-Hassek C: Potential and limitations of non-targeted fingerprinting for authentication of food in official control. Food Res Int 2014, 60:189-204.

[8] Gallo V, Mastrorilli P, Cafagna I, Nitti GI, Latronico M, Longobardi F, Minoja AP, Napoli C, Romito VA, Schäfer H, et al.: Effects of agronomical practices on chemical composition of table grapes evaluated by NMR spectroscopy. J Food Compos Anal 2014, 35:44-52.

[9] Hohmann M, Christoph N, Wachter H, Holzgrabe U: 1H NMR Profiling as an Approach To Differentiate Conventionally and Organically Grown Tomatoes. J Agr Food Chem 2014, 62:8530-8540.

[10] Godelmann R, Fang F, Humpfer E, Schütz B, Bansbach M, Schäfer H, Spraul M: 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 Agr Food Chem 2013, 61:5610-5619.

[11] Picone G, Mezzetti B, Babini E, Capocasa F, Placucci G, Capozzi F: Unsupervised Principal Component Analysis of NMR Metabolic Profiles for the Assessment of Substantial Equivalence of Transgenic Grapes (Vitis vinifera). J Agr Food Chem 2011, 59:9271-9279.

[12] Ercolini D, Ferrocino I, Nasi A, Ndagijimana M, Vernocchi P, La Storia A, Laghi L, Mauriello G, Guerzoni ME, Villani F: Monitoring of Microbial Metabolites and Bacterial Diversity in Beef Stored under Different Packaging Conditions. Appl Environ Microbiology 2011, 77:7372-7381.

[13] Picone G, Laghi L, Gardini F, Lanciotti R, Siroli L, Capozzi F: Evaluation of the effect of carvacrol on the Escherichia coli 555 metabolome by using 1H-NMR spectroscopy. Food Chem 2013, 141:4367-4374.

[14] Davies H: A role for "omics" technologies in food safety assessment. Food Control 2010, 21:1601-1610.

[15] Mehmood T, Liland KH, Snipen L, Sæbø S: A review of variable selection methods in Partial Least Squares Regression. Chemometr Intell Lab 2012, 118:62-69.

[16] Savorani F, Rasmussen MA, Rinnan Å, Engelsen SB: Interval based chemometric methods in NMR-foodomics.. In Chemometrics in Food Chemistry. Edited by Marini F. Elsevier Science; 2013: 449-486.

[17] Savorani F, Picone G, Badiani A, Fagioli P, Capozzi F, Engelsen SB: Metabolic profiling and aquaculture differentiation of gilthead sea bream by 1H NMR metabonomics. Food Chem 2010, 120:907-914.

[18] Jiang M, Wang C, Zhang Y, Feng Y, Wang Y, Zhu Y: Sparse Partial-least-squares Discriminant Analysis for Different Geographical Origins of Salvia miltiorrhiza by 1H-NMR-based Metabolomics. Phytochem Analysis 2014, 25:50-58.

[19] Broadhurst D,Kell D: Statistical strategies for avoiding false discoveries in metabolomics and related experiments. Metabolomics 2006, 2:171-196.

[20] Haug K, Salek RM, Conesa P, Hastings J, de Matos P, Rijnbeek M, Mahendraker T, Williams M, Neumann S, Rocca-Serra P, et al.: MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data. Nucleic Acids Res 2012.

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

[22] Cagliani LR, Culeddu N, Chessa M, Consonni R: NMR investigations for a quality assessment of Italian PDO saffron (Crocus sativus L.). Food Control 2015, 50:342-348.

[23] Giraudeau P, Silvestre V, Akoka S: Optimizing water suppression for quantitative NMR-based metabolomics: a tutorial review. Metabolomics 2015:1-15.

[24] Li ZY, Welbeck E, Wang RF, Liu Q, Yang YB, Chou GX, Bi KS, Wang ZT: A Universal Quantitative 1H Nuclear Magnetic Resonance (qNMR) Method for Assessing the Purity of Dammarane-type Ginsenosides. Phytochem Analysis 2015, 26:8-14.

[25] Cao R, Nonaka A, Komura F, Matsui T: Application of diffusion ordered-1H-nuclear magnetic resonance spectroscopy to quantify sucrose in beverages. Food Chem 2015, 171:8-12.

[26] Hohmann M, Koospal V, Bauer-Christoph C, Christoph N, Wachter H, Diehl B, Holzgrabe U: Quantitative 1H NMR Analysis of Egg Yolk, Alcohol, and Total Sugar Content in Egg Liqueurs. J Agr Food Chem 2015, 63:4112-4119.

[27] Hao J, Astle W, De Iorio M, Ebbels TMD: BATMAN—an R package for the automated quantification of metabolites from nuclear magnetic resonance spectra using a Bayesian model. Bioinformatics 2012, 28:2088-2090.

[28] Worley B,Powers R: MVAPACK: A Complete Data Handling Package for NMR Metabolomics. ACS Chemical Biology 2014, 9:1138-1144.

[29] Armstrong-Smith I, A Briefcase full of NMR, in The Analytical Scientist. 2015. p. 24-29.

[30] D'Alessandro A,Zolla L: Foodomics to investigate meat tenderness. Trac-Trend Anal Chem 2013, 52:47-53.

[31] Bowker BC, Grant AL, Forrest JC, Gerrard DE: Muscle metabolism and PSE pork. J Anim Sci 2000, 79:1-8.

[32] Petracci M, Bianchi M, Mudalal S, Cavani C: Functional ingredients for poultry meat products. Trends Food Sci Tech 2013, 33:27-39.

[33] Sundekilde UK, Frederiksen PD, Clausen MR, Larsen LB, Bertram HC: Relationship between the Metabolite Profile and Technological Properties of Bovine Milk from Two Dairy Breeds Elucidated by NMR-Based Metabolomics. J Agr Food Chem 2011, 59:7360-7367.

[34] Lopez-Sanchez P, de Vos RCH, Jonker HH, Mumm R, Hall RD, Bialek L, Leenman R, Strassburg K, Vreeken R, Hankemeier T, et al.: Comprehensive metabolomics to evaluate the impact of industrial processing on the phytochemical composition of vegetable purees. Food Chem 2015, 168:348-355.

[35] Piras C, Scano P, Locci E, Sanna R, Cesare Marincola F: Analysing the effects of frozen storage and processing on the metabolite profile of raw mullet roes using 1H NMR spectroscopy. Food Chem 2014, 159:71-79.

[36] Shumilina E, Ciampa A, Capozzi F, Rustad T, Dikiy A: NMR approach for monitoring post-mortem changes in Atlantic salmon fillets stored at 0 and 4° C. Food Chem 2015, 184:12-22.

[37] Kortesniemi M, Vuorinen AL, Sinkkonen J, Yang B, Rajala A, Kallio H: NMR metabolomics of ripened and developing

oilseed rape (Brassica napus) and turnip rape (Brassica rapa). Food Chem 2015, 172:63-70.

[38] Ordoudi SA, Cagliani LR, Lalou S, Naziri E, Tsimidou MZ, Consonni R: 1H NMR-based metabolomics of saffron reveals markers for its quality deterioration. Food Res Int 2015, 70:1-6.

[39] Erikson U, Standal IB, Aursand IG, Veliyulin E, Aursand M: Use of NMR in fish processing optimization: a review of recent progress. Magn Reson Chem 2012, 50:471-480.

[40] Wei F, Furihata K, Miyakawa T, Tanokura M: A pilot study of NMR-based sensory prediction of roasted coffee bean extracts. Food Chem 2014, 152:363-369.

[41] Ye Y, Yang R, Lou Y, Chen J, Yan X, Tang H: Effects of Food Processing on the Nutrient Composition of Pyropia yezoensis Products Revealed by NMR-based Metabolomic Analysis. J Food Nutr Res 2014, 2:749-756.

[42] Martinez I, Bathen T, Standal IB, Halvorsen J, Aursand M, Gribbestad IS, Axelson DE: Bioactive Compounds in Cod (Gadus morhua) Products and Suitability of 1H NMR Metabolite Profiling for Classification of the Products Using Multivariate Data Analyses. J Agr Food Chem 2005, 53:6889-6895.

[43] Tappi S, Berardinelli A, Ragni L, Dalla Rosa M, Guarnieri A, Rocculi P: Atmospheric gas plasma treatment of fresh-cut apples. Innov Food Sci Emerg 2014, 21:114-122.

[44] Zanardi E, Caligiani A, Palla L, Mariani M, Ghidini S, Di Ciccio PA, Palla G, Ianieri A: Metabolic profiling by 1H NMR of ground beef irradiated at different irradiation doses. Meat Sci 2015, 103:83-89.

[45] Panarese V, Dejmek P, Rocculi P, Gómez Galindo F: Microscopic studies providing insight into the mechanisms of mass transfer in vacuum impregnation. Innov Food Sci Emerg 2013, 18:169-176.

[46] Kumar PA, Chitra PS, Reddy GB: Metabolic syndrome and associated chronic kidney diseases: Nutritional interventions. Rev Endocr Metab Disord 2013, 14:273-286.

[47] Ibáñez C, Valdés A, García-Cañas V, Simó C, Celebier M, Rocamora-Reverte L, Gómez-Martínez Á, Herrero M, Castro-Puyana M, Segura-Carretero A, et al.: Global Foodomics strategy to investigate the health benefits of dietary constituents. J Chromatogr A 2012, 1248:139-153.

[48] Valdés A, García-Cañas V, Simó C, Ibáñez C, Micol V, Ferragut JA, Cifuentes A: Comprehensive Foodomics Study on the Mechanisms Operating at Various Molecular Levels in Cancer Cells in Response to Individual Rosemary Polyphenols. Anal Chem 2014, 86:9807-9815.

[49] Valdés A, Simó C, Ibáñez C, Rocamora-Reverte L, Ferragut JA, García-Cañas V, Cifuentes A: Effect of dietary polyphenols on K562 leukemia cells: A Foodomics approach. ELECTROPHORESIS 2012, 33:2314-2327.

[50] Heinzmann SS, Merrifield CA, Rezzi S, Kochhar S, Lindon JC, Holmes E, Nicholson JK: Stability and Robustness of Human Metabolic Phenotypes in Response to Sequential Food Challenges. J Proteome Res 2012, 11:643-655.

[51] Rasmussen L, Winning H, Savorani F, Ritz C, Engelsen S, Astrup A, Larsen T, Dragsted L: Assessment of dietary exposure related to dietary GI and fibre intake in a nutritional metabolomic study of human urine. Genes Nutr 2012, 7:281-293.

[52] Savorani F, Rasmussen MA, Mikkelsen MS, Engelsen SB: A primer to nutritional metabolomics by NMR spectroscopy and chemometrics. Food Research International 2013, 54:1131-1145.

[53] De Filippis F, Vannini L, La Storia A, Laghi L, Piombino P, Stellato G, Serrazanetti DI, Gozzi G, Turroni S, Ferrocino I, et al.: The Same Microbiota and a Potentially Discriminant Metabolome in the Saliva of Omnivore, Ovo-Lacto-Vegetarian and Vegan Individuals. PlosOne 2014, 9:1-9.

[54] Lehtonen H-M, Lindstedt A, Järvinen R, Sinkkonen J, Graça G, Viitanen M, Kallio H, Gil AM: 1H NMR-based metabolic fingerprinting of urine metabolites after consumption of lingonberries (Vaccinium vitis-idaea) with a high-fat meal. Food Chem 2013, 138:982-990.

[55] Moazzami AA, Bondia-Pons I, Hanhineva K, Juntunen K, Antl N, Poutanen K, Mykkanen H: Metabolomics reveals the metabolic shifts following an intervention with rye bread in postmenopausal women- a randomized control trial. Nutr J 2012, 11:88.

[56] García-Cañas V, Simó C, Herrero M, Ibáñez E, Cifuentes A: Present and Future Challenges in Food Analysis: Foodomics. Anal Chem 2012, 84:10150-10159.

3.3 Definition of meat quality through metabolomics

Capozzi, F., Ferranti, P., & Trimigno, A.

Chapter 11- Proteomics, transcriptomics, and metabolomics in relation to poultry meat quality in “Poultry Quality Evaluation 1st Edition – Quality Attributes and Consumer Values”, Editors: Petracci M. & Berri, C., Woodhead Publishing ISBN: 9780081007631 – Published Date: 1st June 2017

Role of the PhD candidate: structuring of the chapter, writing of the NMR-related sections, assembling of the NMR and MS parts and production of the final draft, then corrected by the candidate’s supervisor.

Introduction to proteomics and metabolomics

Definition of proteomics and metabolomics

During the last two decades, the so-called “omics” sciences were emerging and have greatly developed in a broad number of research fields. Genomics and transcriptomics originated from the discoveries in genetics, focusing on the analysis of DNA and gene expression modulation, constituting also a basic approach in food research [Capozzi et al., 2013]. Following this holistic approach, proteomics developed as the science capable of analysing the whole set of proteins that is expressed in an organism or in its structures (organs, tissues, cells, enzymes) as well as of the related multiple isoforms and fragments thereof, also including the different post-translational modifications [Paredi et al., 2012; Ferranti et al., 2015]. It is clear how this science can be very useful to understand biological processes and the influence of different genotypes and external factors on protein expression and how it is relevant for the definition of protein-rich food products [D'Alessandro et al., 2011]. To further analyse biological system from its genome to its biological functions, metabolomics arose as the research field that studies all the cellular molecules taking part into living cells reactions (metabolites) [Ferranti et al., 2015]. Departing from the genome, modulated by external modifications and epigenetic alterations, it is necessary, in fact, to study the expression of the proteome, which constitutes, together with the metabolome, the whole living system. In effect, metabolomics is considered the thorough and concurrent determination of endogenous metabolites at the molecular level, and their global and dynamic changes over time, in complex multi-cellular systems as a consequence of biological stimuli or genetic manipulation or both [Nicholson et al., 2008]. These metabolites represent the mark of specific metabolic pathways and biochemical activities, giving insights into how a system's biochemistry responds to specific

factors (diseases, drugs, diets) [Hu et al., 2013]. The comprehensive knowledge of the three "omes" (genome, proteome and metabolome) will result in a great scientific progress, allowing to learn how the whole biological system is regulated. By focusing all omics in the holistic definition of the food matter and its link to healthy nutrition, a new discipline has been coined in 2009 taking the name of Foodomics [Bordoni et al. 2013]. In the development of the advanced omics platforms, because of their potential to profile complex mixtures of biomolecules, mass spectrometry techniques have assumed an unquestionable role. The analytical capability of MS became highly descriptive, but still merely dependent on the resolution power of the coupled separation devices (HPLC, GC, CE, SFC, PAGE): synergistically together, spectrometry and chromatography are able to characterize, at the molecular level, the entire panel of the components of a complex system [Ferranti et al., 2015]. NMR spectroscopy is the other elective platform for metabolomics and, although less sensitive than MS, provides further robust information about the whole system, as it does not necessarily require a separation step preliminary to analyses. For these reasons MS, together with NMR and other spectroscopic techniques, are the core essence of the omics technologies [Ferranti et al., 2015]. Comprehensive analysis of omics data can give scientists the power to study not only individual biomarkers, but also the occurrence of patterns of proteins, lipids, and metabolites, useful to trace a food from raw materials to the end products, also establishing structure/function links for the design of novel functional foods and to monitor food chain contamination by microbial and chemical agents [Ferranti et al., 2015].

Main platforms employed

Proteomics techniques

"In the past ten years, scientifical and technological advancements have been made in order to find and perfect new methodologies capable of separating, purifying and characterizing different food samples. [Ferranti et al., 2015]. In spite of this, food characterization still remains a challenge, due to the great variability of food matrices and related samples (i.e. source, type of food, storage parameters, processing, contaminants, sampling procedures). [Ferranti et al., 2015]

In biomedical samples, the most employed chemical and biochemical techniques for sample purification and analysis are chromatography (GC), high performance liquid chromatography (HPLC), capillary electrophoresis (CE), mono-dimensional (1D) or two-dimensional (2D) polyacrylamide gel electrophoresis (PAGE), and ELISA (Enzyme-Linked Immuno-Sorbent Assay).

[Ferranti et al., 2015]

Even though these methods can be extremely helpful for routine analyses, they can merely be descriptive. In fact, they can simply compare a reference value or profile with the tested samples, analysed employing the same conditions. In order to overcome these limitations, modern omic approaches were developed and are now available. Spectroscopic methods such as, most often, MS and NMR, usually combined with separation techniques and chemometric and bioinformatic tools, are the basis of omic approaches such as proteomics or metabolomics. [Ferranti et al., 2015].

In the last decades, mass spectrometry has been more and more employed due to its high versatility, sensitivity and specificity, great informative level and capability of high-throughput screening. [Hood et al., 2012; Ferranti et al., 2015; Porcari et al., 2016].

The output of MS technology is the molecular mass of an analyte, which is a key feature for any molecule. Moreover, with the information gained from the fragmentation of the molecules of interest, it is possible to identify the separated components of a mixture and even structurally characterize unknown compounds. [Ferranti et al., 2015]. In MS-based proteomics, for instance, it is possible to explore the whole range of proteins and peptides in a sample employing MS technologies coupled with a prior separation step (with high resolution electrophoresis or chromatography).

There are two main ways to perform an MS-based proteomic workflows. These strategies are called the bottom-up and the top-down approach, [Angel et al., 2012; Ferranti et al., 2015] and differ for the type of MS instruments used and the separation required.

In the first one (bottom-up), generally a separation by two-dimensional gel electrophoresis (2-DE) is followed by the analysis of the peptides produced by in-gel proteolytic digestion (e.g. with trypsin) by MALDI-TOF MS or LC-MS/MS and subsequent database search (Peptide Mass Fingerprinting, PMF). The employment of 2-DE allows the resolution of complex proteomes, but it has the disadvantages of being very laborious and having a limited dynamic range. [Ferranti et al., 2015]

Compared to the use of electrophoretic techniques, the coupling of LC with MS seems easier and more straightforward. In the past, MS has been generally coupled with GC to be used in both environmental and biological analysis. The determination of the most suitable form of LC to be coupled to MS or tandem MS depends largely on the application. For example, in the clinical screening of bile acids in urine, the best option could be to inject a crude extract. Most generally, in LC-MS applications, the ultimate stationary phase employed is a reverse phase (RP), whilst the

mobile phase is usually acetonitrile or aqueous alcohol [Ferranti et al., 2015].

Metabolomics techniques

Metabolomics studies the subset of metabolites (metabolome) with specific biochemistry and functions in the biological sample of interest. The metabolome, in fact, can be considered the entire metabolite cohort of an organism, but also the metabolic profile of its specific parts, such as organs, tissues or cells. Blood and urine can be studied in their specific metabolome since they also are affected by both the genome and external perturbations. Due to the disparateness of characteristics and concentrations of those metabolites, suited analytical techniques are necessary for qualitative and quantitative metabolomics investigation depending on the studied metabolome [Hu et al., 2013]. These technologies require being "high-throughput" and fast in the acquisition of a large number of measurements [Capozzi et al., 2013].

One of the main technology employed is mass spectrometry (MS), which is usually chosen for its great sensitivity and high specificity, capable of the detection and quantification of even small-molecular and low-concentrated metabolites.

Mass spectrometry is usually coupled with techniques capable of separating metabolites from the sample, in order to be able to discriminate them. Gas chromatography-mass spectrometry (GC-MS) is generally employed to analyse volatile components, whilst liquid chromatography-mass spectrometry (LC-MS) is chosen for the metabolites that are not volatile at conditions compatible with their stability. Hydrophilic Interaction Chromatography (HILIC) is instead used for very polar or ionic molecules, whilst Reverse Phase High Performance Liquid Chromatography (RP-HPLC) is able to investigate on both polar and non-polar metabolites, such as lipids. New technologies in this field are represented by Ambient Desorption Ionization (DESI) techniques since they allow the direct examination of the sample in open atmosphere with very little sample preparation, reducing therefore the time of analysis per sample (<1 min per sample instead of 15-60 min with GC- or LC-MS). Similar to DESI, there is Direct Analysis in Real Time Ionization (DART), which represents one of the Atmospheric Pressure Chemical Ionization (APCI) methods and it is greatly employed for the soft ionization of both polar and non-polar molecules [Cajka et al., 2013].

Nuclear magnetic resonance (NMR) spectroscopy is another high-throughput technique employed for metabolomics. NMR is in effect direct and fast and it is able to analyse a wide range of metabolites without much sample preparation [Hu et al., 2013]. This technique is indeed able to provide the whole molecular profile of the sample and thus, the entire metabolome and can be applied to a great range of both liquid and solid matrices without sample alteration nor the

production of dangerous waste [Capozzi et al., 2013].

This technique started to be employed in organic chemistry in the late 40s, in order to investigate molecular structure. In the 80s, after the development of new instruments and the development of further knowledge, its application was expanded also to food science [Marcone et al., 2013].

Table 1: Advantages and disadvantages of the two main technological platforms employed in metabolomics: mass spectrometry and nuclear magnetic resonance.

	Advantages	Disadvantages
MASS SPECTROMETRY	•Higher sensitivity •Requires lower sample size •Machines cost less than NMR. •More databases of spectral information.	•High dependence on the experimental conditions (lab, technicians, etc.) •Quantitative analysis can only be comparative between samples. •Need of separation techniques and sample preparation (usually lengthy).
NUCLEAR MAGNETIC RESONANCE	•Fast •Non-invasive •Low cost per experiment •Minimal sample preparation needed •Sample can be recovered •Information on the molecular structure (i.e. position of functional groups) can be gained. •Quantitative	•Lower sensitivity •Expensive instrument •Greater sample size required.

Proteomics and metabolomics in the investigation of meat quality

Meat quality has been defined in literature by four types of product characteristics: (i) nutritional value (i.e. content of fat, protein, carbohydrates, minerals, digestibility), (ii) processing quality (pH, fatness, water holding capacity, shear-force, length of sarcomeres), (iii) safety (contaminants, residues, additives, microbiological status, etc.) and (iv) sensorial quality (texture, colour, flavour) [Becker, 2000].

These characters can be influenced by various intrinsic and extrinsic factors such as genetics (including gender, accounting generally for 30% of variation), breed, feeding, rearing conditions (space allowance, activity, climate, etc.), possible diseases, transport and pre-slaughtering conditions, type of slaughtering, post-slaughtering handling (e.g. chilling and storage), meat processing (e.g. heat treatments), product formulation and additives [Olsson et al., 2005]. The pro-capita consumption of poultry meat from 1990 to 2009 has increased of almost 80% and in future projections it is still expected to grow. The global trends in meat consumption show that income and price will be factors losing importance, whilst quality will become a greater influencing factor [Henchion et al., 2014]. The quality of meat, in effect, has been recently increasingly recognized as

a very important character and a great number of studies has been carried out to evaluate the link between the influencing factors and the main features of this quality such as texture, aroma, colour, oxidation and drip-loss.

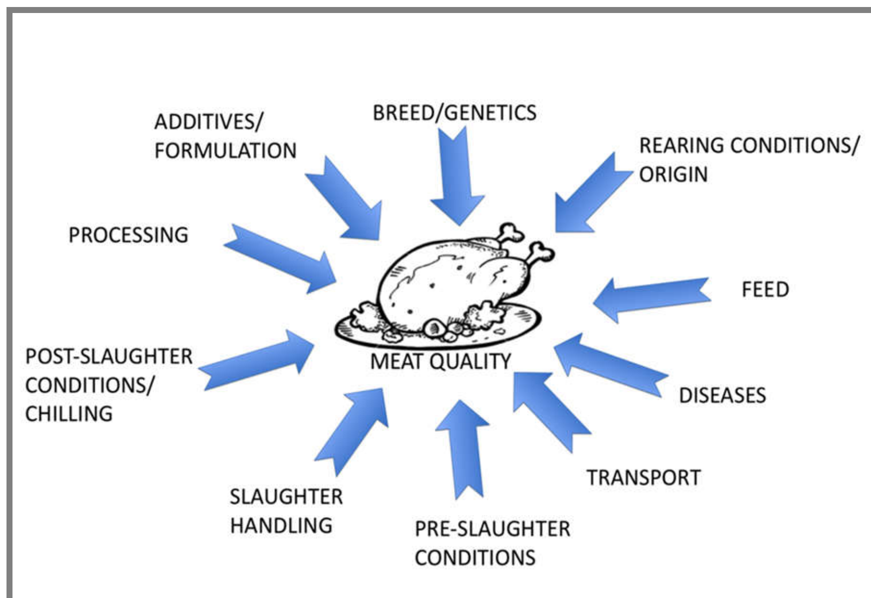


Figure 1: Factors influencing meat and meat-products quality.

a) Genetics and breed

As stated, genetics and breeds have a great impact on the final characteristics of meat and meat products. Proteomic technologies in food science are capable of defining the entire and detailed biochemical composition of a food product and its modifications caused by specific factors [Mamone et al., 2013].

One of the tools emerging for the discovery of biomarkers and for food authentication is MALDI-TOF-MS, capable of performing a molecular profiling of a sample. Through this approach, it is thus possible to assess species and specific genetic lines. Traditional strategies, such as PCR, can therefore be complemented or substituted by MS-based technologies in studies regarding, for example, the identification of food-borne bacteria or the authentication of meat and fish. [Mamone et al., 2013]. Currently, there is a very poor presence in literature of applications of such approaches for studies focused on poultry meat. However, it is not speculative to affirm that the same protocols and methods so far employed for different meat (e.g., porcine and bovine) and seafood could be easily adapted to assess genetic origin and assist breeding schemes for the poultry sector.

Mazzeo et al., applied a newly developed MALDI-TOF-MS approach for the authentication of fish. [Mazzeo et al., 2008]. In this case, 25 different fish species were analysed and their highly specific

MS profiles were obtained. The biomarkers found were signals from protein with an approximate molecular weight of 11 kDa. The method thus developed can also be employed for the assessment of the authenticity of commercial products and to prevent fraudulent substitutions in species which are commonly subjected to this [Mamone et al., 2013].

Pineiro and colleagues [Pineiro et al., 2003] employed proteomics to spot new protein markers for the authentication of seafood. Gel-based methods were used and a polypeptide was chosen to differentiate between hake species of high value and the ones considered less valuable. Other marker proteins were also suggested for the assessment of seafood freshness and shelf-life and to differentiate frozen fish from fresh one [Pineiro et al., 2003; Mamone at al., 2013].

Breed is an important factor in meat quality also because particular genetic lines are considered endowed with higher quality and therefore will be sold at greater prices to consumers. For this reason, fraudulent labelling occurs and tests to reduce this crime are recently developed and investigated. Many metabolomics studies have been applied to investigate meat in this sense. Straadt et al. employed HR-NMR spectroscopy to investigate the potential of metabolomics in the assessment of the meat quality as revealed by the metabolic profile of new pig-breeds (uncommon and novel pig crossing between Iberian and Mangalitza vs. the typical Duroc and Landrace pigs) [Straadt et al., 2011]. The profile of meat, especially affecting the drip, revealed that the post-mortem formation of lactate, measured through 1H-NMR techniques, is correlated to breed and this would explain also the differences in WHC. Ritota et al. employed 1H-NMR-MAS in order to acquire the metabolic profiles of two beef muscles from four different breeds (Buffalo, Chianina, Holstein Friesian and Maremmana) and tried to discriminate between two muscles in the different breed groups through the use of chemometric tools such as PCA, PLS-DA and OPLS-DA [Ritota et al., 2012]. It was possible to identify two muscle origins, namely Buffalo and Chianina, and the use of VIP (Variable Importance in Projection) values allowed to determine the most relevant metabolites for these classifications. Genetics plays also an important role in the selection of specific breeds with improved growth rates and, in the case of poultry, breast yield, since these factors will also increase commercial meat production. These advances, though, can be associated with various alterations in meat quality. For chickens, the genetic improvements related to the rapid and large growth of birds, result also in histological and biochemical modifications of muscle tissues that lead to different types of myopathies [Mudalal et al., 2015]. The increased rate of muscle growth, at a rate which is greater than the physiologically sustainable one, leads to muscle damage and myodegeneration. Breast muscle fibres of the fast-growing strains are characterised

by a shift towards glycolytic metabolism. In addition, compared to the unselected breeds, they have an increased fibre diameter and intercapillary distance, a decreased capillary-to-fibre ratio and rate of protein degradation [Mudalal et al., 2015]. Due to these changes, some abnormalities related to structure and metabolism have appeared and are usually coupled to muscle myodegeneration and regeneration [Petracci and Cavani, 2012]. The main abnormalities observed are the so-called “white striping” [Kuttappan et al., 2009] and “wooden breast” [Sihvo et al., 2014]. White-striping of breast fillets is an abnormalities characterised by the presence of white striations parallel to muscle fibres on the surface of the Pectoralis major muscle and affects severely approximately 3% of Italian medium-sized birds (average live weight 2.75 kg) from both standard and high-growth genotypes raised under commercial condition [Petracci et al., 2013]. This abnormality can also be accompanied by the “wooden breast” defect, which affects the fillet with a visible hard, bulging and pale area in the caudal part [Sihvo et al., 2014]. These histological abnormalities affect the aesthetic characteristics of meat which are strongly impaired for consumer acceptance and the product is consequently downgraded by the poultry industry. Besides the detrimental consequence on the sensorial quality, some studies have shown that meat affected with white-striping has also inferior technological properties, like a poor texture and a decrease in the binding and water holding capacities and poor texture and then used for the production of processed meat products with blending with other amending ingredients [Petracci et al., 2013]. It is thus important to investigate on the quality of these meat types and classify the production according to their composition, in order to decide the best processing chain better valorising the product. A study by Bertram et al. has recently investigated the molecular profile of “wooden breast” fillets employing metabolomics based on HR-MAS proton NMR spectroscopy [Sundekilde et al., 2017]. Different experimental conditions for spectra acquisition have been adopted, namely one-dimensional NOESY and CPMG pulse sequences with two different echo times (50 and 400 ms), in order to enable the best characterization of the molecular profile of the chicken muscle tissue. Multivariate data analysis proved that the muscle affected by the abnormality had lower concentrations of anserine, carnosine and creatine compared with normal muscles. Since creatine is an important energy reservoir in muscle tissue, while anserine and carnosine play a role in homeostasis of muscles, these variations imply alterations in the buffering and anti-oxidative capacities of the muscle. In addition, lipid fraction was higher in both NOESY and CPMG spectra of wooden breast muscles. Because the same result emerges within both experimental conditions, it is proven that the occurred modification is consistent with an increased amount of intracellular

lipids. All this information is clearly very useful for the understanding of the underlying molecular mechanisms that are involved in this abnormality and prove that these metabolomics techniques can provide early biomarkers for the definition of chicken quality, especially as a tool to drive the newest challenges in the selection of poultry destined to production of meat with quality attributes more satisfactory for consumers and industries.

b) Rearing conditions/Origin of meat

Geographical origin of meat has been increasingly considered by consumers and producers both for safety reasons (i.e. BSE issues) and for quality standards (Certified Origin Products and specific rearing conditions such as organic). Usually, this determination was done through the analysis of stable isotopes ratios, though NMR has recently started to be employed on this assessment, due to its ability to determine rapidly the whole molecular profile of a food product. Again, there are no current application reported in literature for poultry meat but the knowledge emerging from metabolomics studies applied on meat of different origin could be easily foreseen as applicable also to poultry. Jung et al. employed 1H-NMR spectroscopy and multivariate data analysis on extracts of beef originated from four different countries (Australia, Korea, New Zealand and USA) in order to highlight the differences in the metabolite profile of raw beef depending on its origin [Jung et al., 2010]. PCA and OPLS-DA showed significant differences among the four countries; in particular OPLS-DA showed that the metabolites mostly responsible for this separation were succinate and amino acids such as isoleucine, leucine, methionine, tyrosine and valine. A one-way ANOVA was also employed to further validate these results. Thus, metabolomics based on NMR provided links between composition and the geographical origin of a meat product, although other factors must be kept in mind in the evaluation of the results: in this case, for example, succinate is widely used as a substitute for salt in feedings and therefore the differences in this metabolite might be due to different dietary regimens. Another aspect that might influence meat quality is the animal welfare standards followed within the production systems. The dense rearing conditions might in effect cause various level of stress in the animals, influencing the final characteristics of meat. The presence of biomarkers in meat related to animal stress may be considered as a further indicator of origin due to specific production standards defined by regulations adopted by producers consortia, based on specific territories, which could differentiate some farms from others. The commercial value of best practise for animal welfare is not negligible, as modern consumers tend to prefer products with a certificated production respecting animal welfare. It is thus important to have scientific and objective tools to assess indicators of animal welfare to be

put in relation to the farm originating the meat. Marco-Ramell et al. investigated serum proteome from three groups of cows subjected to three ranges of welfare: good, semiferal conditions and hardest conditions (in the mountains) [Marco-Ramell et al., 2012]. Difference gel electrophoresis (DIGE) labelling of serum, followed by two-dimensional electrophoresis and MALDI-MS or ion trap MS was carried out to demonstrate that the proteins responsible for discrimination among the three groups belonged to three main pathways: the oxidative stress (glutathione peroxidase and paraoxonase), the acute phase protein family (Heremans Schmid glycoprotein alpha2) and the complement system. Other biomarkers of the redox status such as superoxide dismutase and glutathione reductase were altered depending on the rearing conditions of the animals. These results show that the hardest living conditions, with food restriction and less human care and contact, will induce increased levels of carbonyl content in plasma proteins and a greater activity of superoxide dismutase and glutathione peroxidase, meaning animals are at a higher level of oxidative stress. This study also proved that biomarkers for animal welfare and stress can be selected for classification, such as glutathione peroxidase, Heremans Schmid glycoprotein alpha2 and cholesterol and faecal corticosterone.

c) Feed

As already stated, diet can have impact on the metabolic profile of an organism in two ways: feed molecules become host metabolites once digested and adsorbed by the animal, and they modulate gene transcription by epigenetics modifications. In fact, feed components might directly result as meat metabolites and thus influence the characteristics of meat such as flavour or colour, oxidizability or its nutritional value. In other cases, the specific nutrition could give rise to modifications in the biological performance through epigenetic modulation of the genetic traits.

In 2015, Watanabe et al. investigated the relationship between feeds and glutamate, the most taste-active component of meat [Watanabe et al., 2015]. Broiler chicks of 14 days were fed two different diets for 10 days: one containing the 100% of the recommended lysine content (precursor for glutamate) and one containing 150% of the recommended lysine intake. The concentration of free amino acids were measured in muscle tissue through the employment of CE-TOFMS. It was found that, of the 127 metabolites identified, 68 were up-regulated and 22 down-regulated in the group on the diet with a higher content of lysine: more specifically, the products from the degradation of lysine, saccharopine, alfa-aminoadipic and pipecolid acid, were greatly increased in this group. In addition, the muscular content of free glutamate in the 150% lysine diet group was increased by 44% in comparison to the control group. The epigenetic mechanism of such effect

was further validated by the measurement of the mRNA level of the relative enzymes. These results shows that all the products from lysine degradation are increased when lysine content is increased in the diet and the content of free glutamate is also regulated by this degradation, showing that a higher lysine diet can improve the taste of meat. Another key characteristic in modern meat quality is leanness and this trait could be easily linked to the animal diet. Health awareness among the consumers has been increased, consequently directing their preference towards lean meat. This latter may be produced by abusing feed additives like L-carnitine, ractopamine, nucleotides, due to their lipolytic activity and better effect on growth performance of birds [Zuo et al., 2010; Yasmeen et al., 2013]. Costa-Lima et al. investigated on the effect of ractopamine on the sarcoplasmic proteome of Longissimus thoracis pork muscle [Costa-Lima et al., 2015]. The relative proteome was analysed through two-dimensional electrophoresis and mass-spectrometry and nine spots were found different between the treated and control group and these were correlated to glycolytic enzymes. In this way they proved how ractopamine had an effect, not only by perturbing lipolysis, but also affecting the proteome, thus influencing the conversion of muscle to meat. Another important aspect of feed is related to the consumers' perception that domestic meat is believed to be of higher quality owing to the feeds employed for its production [Henchion et al., 2014]. It is therefore important to develop specific methods to determine the link between product quality and feeding, in order to scientifically substantiate promotion of local meat consumption only when appropriate. Local productions, besides, usually focus on specific characteristics for the meat product, which are recognized and appreciated by the consumers, and studies have been focused on the development of specific feeds for selected animals enhancing the esteemed sensorial attributes, therefore generating added value products. Sánchez del Pulgar et al. identified, through PTR-ToF-MS, the metabolite profile of lean and subcutaneus fat from dry-cured Iberian ham in order to identify possible differences due to different diets [Sánchez del Pulgar et al., 2013]. Pigs fattened outdoors with pasture and acorn or pigs fed with a high-oleic concentrated feed were compared by application of univariate (ANOVA) and multivariate data analysis (PCA and PDA, Penalized Discriminant Analysis) to mass spectra. Samples were successfully discriminated in relation to the feed employed: the first group, fed on acorn and pasture, showed higher concentration of ketones and aldehydes, probably for the greater amount of fatty acids in the feed, and lower contents of sulfur-containing compounds. Again, the metabolomics approach proved useful for this kind of quality assessment: the method employed was very rapid and might be employed in the future for non-destructive on-line

monitoring of quality. Jurado et al. also studied Iberian pigs after two different diets: in this case "extensive" and "intensive" fattening systems, and focused on the use of 4-methylsterols and 4,4-dimethylsterols as biomarkers [Jurado et al., 2013]. GC-MS and GC-FID were employed to identify and quantify the compounds in the samples and, in this case, six methylsterols were determined and described for the first time in these samples. These metabolites were analysed with pattern recognition techniques, such as PCA and LDA, in order to discriminate between the two diets. In the end, a multilayer perceptron artificial neural network was able to differentiate the two systems with a 91.7% classification performance. The importance of feeding as determinant for the consumers' choices is also attributable to the request of locality, i.e. very-short supply chains, for animal production because the far industrially produced fodders are negatively perceived in comparison to local crops. Zancanaro et al. employed NMR spectroscopy to discriminate muscles of pigs fed with either an industrial feed or a 0-miles feed with a similar composition [Zancanaro et al., 2011]. Lipidomics showed that the molar percentage of 18C fatty acids and the mean fatty acid chain length were different between groups. Principal component analysis was also applied on the NMR spectra and showed some discriminating potential markers between the two groups due to phosphorylated compounds and some amino acids. Again, NMR metabolomics has not yet employed in the poultry sector, but the experimental design adopted in the above studies can be usefully employed in the assessment of the effect of external factors such as feed on poultry meat quality. Investigation on meat quality in relation to feed could be also correlated to the importance of tackling against food frauds. This is of particular importance when frauds have also implications on food safety. After food scares such as bovine spongiform encephalopathy (BSE), a 2002 EU regulation prohibited the use of animal by-products and processed products of a certain animal species from the purpose of feeding the same species (Regulation (EC) No 1774 /2002). Nonetheless, feeds with processed animal protein from parts and bodies of animal of the same species might be still employed fraudulently, due to their extreme cheapness. Therefore, such a fraud might also represent a safety issue. Most recently, therefore, many studies investigating this issue were carried out. Unfortunately, only a few of them were addressed to poultry, although animal by-products, e.g. protein hydrolysates, are explored as ingredients for their feeds, in the effort to promote sustainable systems which more and more ask to exploit and valorise wastes. For instance, a purified protein hydrolysate with zero chromium from chrome-tanned leather waste has been tested as poultry feed [Chaudhary et al., 2016]. The study produced evidence that protein hydrolysate can replace up to 75 % of soybean meal in broiler diets without affecting either growth

performance or meat characteristics. However, the traditional analyses employed to evaluate the meat quality, e.g. ultimate muscle pH, colour, protein, fat and mineral content, are far away from the holistic view provided by the omics approach and a definitive response for the equivalence of such feeding is not reached yet. In another study, Cajka et al. investigated on the possible differences in chicken meat in relation to a feed with (5-8% w/w) and without the addition of chicken bone meal (banned component) employing DART-MS and multivariate data analysis both on the feed and on chicken muscle [Cajka et al., 2013]. The metabolite fingerprint was assessed for both feed and muscle extracts and these were evaluated through PCA and OPLS-DA. In this way it was possible to discriminate between the two diets on polar and non-polar extracts, and this was assessed positively also after 6 months. One thing that was questioned, though, was that triglycerides (in non-polar extract) showed pattern variation due also to season and other raw materials used in feed preparation, thus complicating the assessment of the effect of the chicken bone meal.

d) Pre- and post-slaughter conditions

Some of the main defects in poultry meat are the so-called pale-soft-exudative (PSE) and dark-firm-dry (DFD) meat. These defects interest meat colour, mainly, and the texture, with detrimental effects provoking consumers discard. PSE and DFD are due to unusual post-mortem pH evolution in meat: very low pH in PSE meat and high pH in DFD meat are reached ultimately, giving higher chances of microbial growth. These defects are caused by stress-related shocks in the animals before death, causing alteration in their metabolism and thus post-mortem variation in meat. For PSE meat, genetic links were found in pigs, whilst in chicken this defect is still mainly related to pre-slaughter stress and inadequate meat chilling. In any case, these factors will cause either a fast post-mortem glycolysis and thus a fast pH descent or just abnormally low values of pH for the meat. In DFD meats, instead, prolonged stress will cause the use of glycogen supplies and thus, post-mortem, the absence of lactate will cause higher pH values. Many test are now available to assess these two types of defects; in PSE also genetic tests can be possible for swine, though most recently metabolomics was proposed as a fast alternative. In poultry, metabolomics would be essential since genetic links are still not evident. However, publications concerning studies for PSE and DFD signatures based on metabolomics are still absent in literature for birds and, again, the potentiality of this approach will be described below for pork. Nuclear magnetic resonance was used to study the possible consequences of pre-slaughter exercise on the metabolic profile of plasma, employing 40 pigs divided in different groups: control group and exercise on a treadmill

followed by 0, 1 or 3 hours of rest pre-slaughter [Bertram et al., 2010]. It was seen that plasma lactate greatly increased in the group with no rest between exercise and slaughter, whilst when rest was present, the stress-related variation in the metabolic profile were attenuated. It was found that plasma lactate was greatly correlated with WHC, muscle temperature 1 minute post-mortem and meat pH. In this way, it was possible to see that even without directly measuring the whole metabolic profile of meat it is possible to evaluate meat quality through a specific metabolite content in pre-slaughtering animals and adopt it as the marker for predicting the impact of stress. The same research group employed 31P-NMR spectroscopy to investigate the phosphorylated metabolome in an attempt to determine whether two different post-mortem cooling profiles, commercial batch (soft) and tunnel chilling (fast), would generate differences in pork meat quality [Bertram et al., 2001]. Phosphocreatine was degraded less rapidly post-mortem in tunnel chilling, whilst ATP showed no degradation differences in the two cooling methods. pH values also differed, with a maximum variation between groups of 0.25 units after 150 minutes post-mortem. It was also calculated that the fraction of inactivated myosin was 2.5 times greater using the soft method, indicating that the measured differences in pH and temperature among the two regimes could influence protein degradation and WHC of meat. These studies show the great potential of metabolomics as a fast tool to assess the presence of meat defects and it could be easily employed on chicken meat, too, in order to spot the defects in time and avoid investing PSE/DFD meats for the formulation of products which are not suitable for. In effect, PSE meat could be used in special products formulated in order to restore a proper colour and protein functionality, so to improve the texture and contain yield loss, whilst DFD meat cuts, due to their microbiological instability, could be used in heat-processed products [Lesiów et al., 2003]. It is clear that metabolome perturbations are mirrors of a more general alteration in the muscle physiology, reflected also at the level of the proteome characterising the muscular apparatus, as affected by different enzymatic activities, e.g. by hydrolases. Zanetti et al. investigated the proteome of chicken Pectoralis major muscle in order to evaluate possible changes due to in transit time before slaughtering (90 or 220 minutes) [Zanetti et al., 2013]. Two-dimensional electrophoresis, followed by image analysis and LC-MS/MS of the spots of interest, chosen by means of statistical analysis (Wilcoxon's test and PCA) were applied as blood tests. They showed that the animals with longer transit times had higher stress levels as evidenced by the presence of 2 spots related to time in transit, associated to proteins linked to cellular housekeeping functions (metabolism, cell division, apoptosis-control). Principal component analysis did not show a clear separation in the proteome

of chicken with different in-transit times, though some small differences were identified. For future studies, thus, employing more different times and other tissues could be more helpful to gain insight on the stress-related proteome changes. It was clear, though, how this pre-slaughtering process can hinder good meat quality by causing stress to the birds. Another important factor affecting meat quality is the slaughtering procedure itself. Salwani et al. investigated the differences in the proteome of the skeletal muscle of broiler chickens with or without gas stunning prior to slaughtering [Salwani et al., 2015]. Meat quality parameters (colour, pH, shear-force and drip/cooking-loss) were also measured and differences were found between the two treatments. Gas-stunned chicken had lower redness, pH and shear-force, whilst they showed greater drip and cooking loss both at 4 and 24h post-mortem. In addition, proteomics, through two-dimensional polyacrylamide gel electrophoresis on samples at 7 minutes post-mortem, showed that gas-stunning up-regulated the expression of creatine and pyruvate kinase and beta-enolase. These results show that pre-slaughtering treatments having negative effect on meat quality (WHC and colour, especially) provoke increased speed of the post-mortem energy metabolism. Most recently, Beauclercq and colleagues [Beauclercq et al., 2016], employed high-resolution NMR (1H and 31P) in order to investigate on the actual variation in chicken Pectoralis major muscle and serum due to the different values of meat ultimate pH, a characteristic that, as was mentioned above, is fundamental for the quality of poultry meat and related to DFD/PSE meats. Two genetic lines of chickens with either high or low ultimate pH (pHu+ or pHu-) were selected: 20 male broilers were chosen from the first one and 19 from the second one. The pHu+ line showed a significantly lower glycolytic potential and was darker and less exudative. OPLS-DA was carried out on the concentration of spectral metabolites and a total of 26 molecules (from 1H and 1P spectra) were found discriminative in muscles and 20 (just from 1H spectra) in serum. The pHu- line was found to have higher carbohydrate content in both muscle and serum, together with other metabolites linked to production of energy. On the other hand, the pHu+ line, showed a metabolism more switched towards ketogenic amino acid degradation and lipid beta-oxidation, since less glucose and glycogen are present and thus amino acids such as alanine are employed for energy production. This study thus paved the way into the clearing of the mechanisms behind the different glycogen levels at slaughter (and thus different quality traits) and suggested a fast screening technique to exclude meat cuts with poor quality, by employing a defined set of spectral metabolites to exclude pHu+ cuts.

e) Meat processing

Modern consumers require meat products that are both convenient and of high quality, possibly appearing as close as possible to minimally processed foods. Thus, it is important for the meat producers to develop new technologies capable of providing these products, while maintaining good preservation of the quality attributes. Whatever is the processing technology applied, food composition and structure evolve during production and storage, and the main purpose of the industry is to take control over meat transformation processes, by monitoring opportune markers of quality. During a processing step such as ripening, for example, the peptidic profile of a food product changes greatly and can give hints on the origin and the evolution of the specific protein fraction. In effect, physical, chemical and enzymatic processes occur during ripening and modify the food matrix. Reactions such as proteolysis, lipolysis and fermentation (i.e. lactic and propionic), change the chemical components of the food product, impacting on its final sensory properties.

Flavour, for example, is greatly influenced by proteolysis, which could contribute both positively (with the release of specific amino acids and peptides), and negatively, with the production of off-flavours (such as from bitter, hydrophobic peptides). Moreover, it could liberate substrates used in other reactions. Therefore, it is fundamental to gain the desired, balanced, break-down of proteins into specific amino acids and peptides, in order to obtain an acceptable flavor for that food product [Mamone et al., 2013]. Poultry, especially turkey, is more and more sold as a cured product to extend shelf-life and increase consumers' acceptance. Thus, the experience gained so far in other more consolidated meat-curing industry, including adoption of omics approaches for quality control, is of inspiration for developing analogous systems also in less traditional sectors such as that of cured poultry meat.

Proteomics has investigated the biochemical changes occurring in dry-cured ham during its processing: during the post-mortem phase, endogenous proteolytic enzymes are activated and this will induce the degradation of muscle proteins [Mamone et al., 2013]. In fact, muscle sarcoplasmic proteins have been found to be relevant substrate for proteolysis by the identification of many small peptides, which are released by enzymes from the glycolytic pathway. [Mora et al., 2011a].

Due to the fact that there is a lack of cleavage specificity, it is usually hard to identify the peptides released in food products. In spite of this, information on the endogenous proteases both from early stages (calpains) and medium- and long-term (cathepsins), and their action, can be gained through the employment of multistage MS-based peptide sequencing [Mamone et al., 2013].

Mora et al. [Mora et al., 2009] have demonstrated in dry-curing hams that there is an additional intense secondary proteolysis from amino- and carboxy-peptidases. This was assessed through the

identification of oligopeptides with consecutive loss of terminal amino-acids.

Moreover, both small and large fragments of heavy and light chains of myosin and actin were identified through MS methods in dry-cured ham, proving that also muscle myofibrillar proteins experience extensive proteolysis [Di Luccia et al., 2005; Sentandreu et al., 2007; Mora et al., 2011b]. In fermented sausages, it was studied how lactic acid bacteria cause an even greater proteolysis [Picariello et al., 2006]. Nonetheless, only a few studies regarding these products were published in the field of peptidomics, and it could be interesting to employ this approach to define the specific proteolytic patterns in dry fermented sausages in order to assess, or even predict, the stability of determined strains of starter cultures of lactic acid bacteria. [Mamone et al., 2013].

Recently, Sentandreu et al. [Sentandreu et al., 2011] have proposed MS for the identification of species-specific biomarkers in order to determine the authenticity of meat-derived products

Moreover, MS analysis allowed the characterization of the domains of food proteins stable to digestion. In cured beef Bresaola, which is extensively hydrolysed by endogenous proteases, a large variety of peptides are released since the early post-mortem phase. Many of these peptides can exert several biological activities in human body [Ferranti et al. 2014]. After ingestion, the proteins, polypeptides and peptides from Bresaola are further degraded by the proteases found in the gastrointestinal (GI) tract. Ferranti and colleagues [Ferranti et al. 2014] have employed a static in vitro digestion model, including oral, gastric and duodenal phases, to evaluate the effect of different curing protocols on the digestibility of Bresaola meat. The obtained digestates were profiled by mass spectrometry-based proteomic and peptidomics strategies [Ferranti et al. 2014].

It was seen that, at the end of digestion, all samples were almost identical for the majority of the released peptides. Only some slight differences, due, possibly, to differing technological procedures or nature of the raw material, were visible. The only thing that could be associated with the characteristics of the food was the kinetics of the peptide release, thus the evolution in time of the quantitative composition of the profile of the digestates (digestome).

In this case, sarcoplasmic proteins were degraded relatively fast, whilst the myofibrillar chains would not be completely hydrolysed by the action of gastro-duodenal proteases, even after a prolonged hydrolysis. Moreover, these chains require a previous step of proteolytic release. [Ferranti et al. 2014]. This study identified more than 170 peptides, which were liberated from both sarcoplasmic and structural (actin and myosin) muscle proteins. Of note, many of the identified peptides are considered precursors of sequences which are potentially antioxidant and antihypertensive [Ferranti et al. 2014].

Bordoni et al. also investigated on Bresaola samples through the metabolomics approach [Bordoni et al., 2014]. Samples were again digested in vitro and digestates were investigated through 1H-NMR to study the influence of the matrix in the bioavailability of nutrients. In effect, the meat matrix has been shown to modulate the free diffusion of specific nutrients, e.g. carnosine, in the digestion media towards the absorption sites, thus affecting the nutritional quality of the product. The developed protocol is applicable for screening of the nutritional value of protein-rich foods, being able to assess the bio-accessibility of proteins end-products.

f) Product formulation

The specific formulation of a meat product is important both for sensory and nutritional quality and for its technological characteristics, though in some cases it is important to determine it to avoid fraudulent practices. A study from Lassoued et al. proved, for chicken meat, that there is a link between risk aversion and ethically motivated behaviour on one side, and the trust the consumers have in a brand and in the declared quality attributes on the other side [Lassoued et al., 2015]. It is therefore important for food industries to build trusting relationship with consumers by providing scientific evidence, i.e. objective descriptors, to discriminate different products in terms of their quality attributes. On the other hand, the promotion of frauds is stimulated because producers are facing a very competitive global market, thus they are in need to find the cheapest way to produce their meat bypassing the quality check indicators. It is therefore necessary to find new scientific methodologies to reinforce the quality check system and to fight food frauds, as also mentioned in the case of feeds. One of the main ingredients employed in the formulation of meat products is mechanically recovered meat (MRM). This ingredient is industrially obtained by recovering residual meat from animal bones or poultry carcasses from which the bulk of the meat has been already removed. MRM has the appearance of finely comminuted meat and is incorporated in a wide range of meat products, as a cheap source of meat. Similarly to MRM, desinewed meat (DSM) is produced using a low pressure technique to remove meat from animal bones. MRM and DSM meat are similar to hand de-boned meat (HDM) in their chemical composition, though they have a lower commercial value. [Mamone et al. 2013]. Moreover, they have been excluded from the EU definition of meat with the Directive of the European Parliament N. 101/2001. This was also caused by the fact that consumers were concerned about their safety [Mamone et al. 2013]. It is thus necessary to discriminate between MRM and DSM, and HDM with reliable analytical methods. Recently, Surowiec and colleagues [Surowiec et al., 2011a] have carried out a proteomic study aiming at identifying potential markers capable of detecting chicken

from MRM. In this research, proteins were extracted intact from raw meat and analysed by the employment of OFF-GEL electrophoresis, SDS-PAGE and, finally, nano-LC-MS/MS to identify the markers. Researchers were thus able to define the key proteins from the processed meat samples and identify the specific markers for MRM chicken [Surowiec et al., 2011a; Mamone et al. 2013]. Surowiec et al. also employed metabolomics techniques for the same kind of assessment on HDM pork and chicken meat, MRM meat and DSM meat samples of different origin [Surowiec et al., 2011b]. In this case, GC-MS was used to determine the molecular profiles of samples and chemometric tools (PCA and OPLS-DA) were employed for the identification of specific differences between types of meat. It was found that the supervised statistical technique (OPLS-DA) was capable to separate all classes of meat-products (HDM, MRM, DSM) for both pork and chicken. The metabolites that generated this separation were then confirmed by Student's t-test and were tentatively identified by comparison with specific databases. However, these metabolites appeared to belong to numerous chemical classes and were mostly fat-related, thus, currently, none could serve as a specific biomarker, but future research could help in this assessment, in order to develop a fast methodology for the identification of HDM, MRM and DSM meats. The blending of meat from different species in a food product might be considered fraud when not declared, and can be a relevant character of quality control for determined consumers, such as the ones concerned about ethical aspects (i.e. against the use of horse meat) or religious requirements (e.g. no pork consumption). Recently, also, some food scandals occurred in Europe with horse-meat contamination of “beef” lasagna and traces of pork in Muslims school meals of chicken in the UK [D'Alessandro et al., 2013]. It is necessary, therefore, to be able to identify these frauds in order to protect consumers. Von Bargen et al. employed mass spectrometry in order to detect trace contamination of pork and horse meat and determine the specific biomarkers of these two types of meat against beef, chicken and lamb [von Bargen et al., 2013]. Through a proteomic approach, the typical biomarker peptides of the 5 different species were identified and these were taken into consideration in the study. Researchers were able to detect 0.55% horse contamination and 0.13% of pork contamination in beef matrices, showing that this proteomics technique can be of great help in the assessment of frauds. A fast and sensitive method, based on multiple reaction monitoring was developed and this can be applied in a routine manner in quality control laboratories employing tripe-quadrupole MS instruments, representing an alternative to genomics and classical proteomics techniques that are more expensive (i.e. ELISA) or still problematic for this type of assessment (i.e. PCR). Sentandreu et al. [Sentandreu et al., 2010] employed a proteomic

approach capable of detecting chicken meat in mixed meat preparations. In this study, myofibrillar proteins are extracted, the OFFGEL isoelectric focusing is used to enrich target proteins, then myosin light chain 3 is digested through in-solution trypsin digestion. Finally, LC-MS/MS was employed to analysed the peptides thus generated. Through this method, researchers were capable of detecting, with a high confidence, 0.5% of contaminating chicken in pork meat. AQUA stable isotope peptides, synthesized using sequences of species-specific peptide biomarkers, which were selected previously, allowed the quantitative detection of chicken meat. There was a linearity between the amount of chicken in the meat mixture and the amount of peptide biomarkers found. The method now needs to be validated by other independent replications, though it proved to be very simple and effective. This method could be used with both cooked and raw meat. It could also be adapted and employed for highly processed food, where immunoassays are precluded due to the modifications occurring to the tertiary protein structure during processing. Moreover, the proteomic analysis employed in this study was capable in identifying a definitive discriminatory sequence of peptides. The traditional DNA and PCR based methods can not allow this, instead. The researchers also proposed and expect future applications of this method as a regular procedure, by translating this technology into routine MS instruments. This will be very useful in laboratories used for public inspections [Sentandreu et al., 2010; Mamone et al., 2013]. Another important application in the determination of specific food ingredients in a formulation is in the case of food allergens, such as soy. Soy proteins can be added as emulsifiers in meat-derived proteins, to work during heat-treatment. Their use, though, could be limited or forbidden by specific regulatory authorities in some countries. Suitable analytical methods, capable of detecting and/or quantifying soybean proteins (Soy Protein Isolate, SPI) in meat products are thus needed, so that it will be possible to assess whether the threshold levels for their contents are respected [Mamone et al., 2013]. At the moment, only a few methods are available for their semi-quantitative determination. Authorities such as the FDA (USA Food and Drug Administration) have criticized the Official Method (AOAC Official Method of Analysis 988.10), which is based on an ELISA assay. This method is, in fact, not considered too reliable or accurate. [Mamone et al., 2013]. Moreover, methods capable of detecting even low ppm concentrations of soybean proteins are needed, since, as stated, these molecules have an allergenic potential [Koppelman et al., 2004; Mamone et al., 2013]. The alternative to the traditional method, are LC-based assays, but these could only work with raw meats. Recently, a proteomic study [Leitner et al., 2006] aimed at identifying SPI adulteration in heat-processed meat products from differen origins (chicken, turkey, pork, beef),

employed multi-dimensional HPLC-MS/MS. In this research, in order to isolate the interesting peaks originating from SPI-containing meat products, it was necessary to pre-fractionate samples through the employment of perfusion chromatography. The fractions were then digested using trypsin and finally anlysed by nanoflow LC-MS/MS. Glycinin and conglycinin (major seed proteins), in many variants and subunits, were spotted in SPI samples, together with other two proteins. Meat samples containing SPI were characterized by peak from different glycinin A subunits. The subunit A4 from glycinin G4 was in fact found in all SPI samples and could be used for new analytical methods for the identification of SPI in processed meat [Leitner et al., 2006; Mamone et al., 2013].

Conclusions and future applications

Proteomics and metabolomics techniques and their associated methodologies have helped in many research field to find interactions between genome and external factors able to modify the characteristics of a biological system. The three approaches have been increasingly employed also in food science and in the assessment of meat quality. As stated, the quality of meat products depends on many different characteristics that, in turn, are influenced by genetics, but also by various external factors, such as the rearing methods and feeds employed, the handling of slaughtering and the processing and formulation of meat-based products.

Table 2: Reported proteomics and metabolomics studies on the factors that impact on meat quality.

	PROTEOMICS	METABOLOMICS
BREED/GENETICS	Mazzeo et al., 2008 Pineiro et al., 2003	Straadt et al., 2011 Ritota et al., 2012 Sundekilde et al., 2017
REARING CONDITIONS/ORIGIN	Marco-Ramell et al., 2012	Jung et al., 2010
FEED	Costa-Lima et al., 2015	Watanabe et al., 2015 Sanchez del Pulgar et al., 2013 Jurado et al., 2013 Zancanaro et el., 2011 Cajka et al., 2013
TRANSPORT/ PRE-SLAUGHTER CONDITIONS	Zanetti et al., 2013 Salwani et al., 2015	Bertram et al., 2010 Beauclercq et al., 2016
POST-SLAUGHTER HANDLING/CHILLING		Bertram et al., 2001
PROCESSING	Mora et al., 2011a Mora et al., 2009 Di Luccia et al., 2005 Sentandreu et al., 2007 Mora et al., 2011b Picariello et al., 2006 Sentandreu et al., 2011 Ferranti et al., 2014	Bordoni et al., 2014
ADDITIVES/FORMULATION	Surowiec et al., 2011a Von Bargen et al., 2013 Sentandreu et al., 2010 Leitner et al., 2006	Surowiec et al., 2011b

A lot of reported studies were based on cattle or swine, though the same principles stands and could be applied to chicken meat. This is especially the case for the investigation of the occurrence of PSE and DFD meat. In this sense, the employment of conjoint genomics, proteomics and metabolomics studies could help to have a complete insight on these defects that majorly impact chicken meat. In addition, these techniques could be helpful for the investigation of other defects such as white striping or wooden breast. The development of fast methodologies that could be applied on-line would be of great help in the determination of the destination of specific meat cuts (i.e. meat-based products or full cuts). Moreover, in the modern market it is necessary to give consumers as much objective and biomarker-assisted information as possible and it is thus of great help to have techniques such as proteomics or metabolomics to be able to confirm the quality of a product and the absence of frauds. This is important, for example, in the determination of the nutritional quality of chicken meat, considered a lean food, with many important bioactives, or for the certification of specific types of rearing (i.e. free-range) considering the increasing interest in animal welfare. To conclude, both proteomics and metabolomics will have a great impact on the evaluation of chicken meat quality and could help breeders and industries to evaluate genetic selection on the quality of the food products in their future.

REFERENCES CHAPTER 3.3

Angel, T. E., Aryal, U. K., Hengel, S. H., Baker, E. S., Kelly, R. T., Robinson E. W., & Smith, R. D. (2012) Mass spectrometry-based proteomics: existing capabilities and future directions. Chemical Society Reviews 41(10), 3912-3928.

Beauclercq, S., Nadal-Desbarats, L., Hennequet-Antier, C., Collin, A., Tesseraud, S., Bourin, M., Le Bihan-Duval, E., & Berri, C. (2016) Serum and Muscle Metabolomics for the Prediction of Ultimate pH, a Key Factor for Chicken-Meat Quality. Journal of Proteome Research, 15(4), 1168-1178.

Becker, T. (2000). Consumer perception of fresh meat quality: a framework for analysis. British Food Journal, 102(3), 158-176.

Bertram, H. C., Dønstrup, S., Karlsson, A. H., Andersen, H. J., & Stødkilde-Jørgensen, H. (2001). Post mortem energy metabolism and pH development in porcine M. longissimus dorsi as affected by two different cooling regimes. A 31 P-NMR spectroscopic study. Magnetic resonance imaging, 19(7), 993-1000.

Bertram, H. C., Oksbjerg, N., & Young, J. F. (2010). NMR-based metabonomics reveals relationship between pre-slaughter exercise stress, the plasma metabolite profile at time of slaughter, and water-holding capacity in pigs. Meat science, 84(1), 108-113.

Bordoni, A., Laghi, L., Babini, E., Di Nunzio, M., Picone, G., Ciampa, A., Valli, V., Danesi, F., & Capozzi, F. (2014). The foodomics approach for the evaluation of protein bioaccessibility in processed meat upon in vitro digestion. Electrophoresis, 35(11), 1607-1614.

Cajka, T., Danhelova, H., Zachariasova, M., Riddellova, K., & Hajslova, J. (2013). Application of direct analysis in real time ionization–mass spectrometry (DART–MS) in chicken meat metabolomics aiming at the retrospective control of feed fraud. Metabolomics, 9(3), 545-557.

Capozzi, F., & Bordoni, A. (2013). Foodomics: a new comprehensive approach to food and nutrition. Genes & nutrition, 8(1),1-4.

Chaudhary, R. & Pati, A. (2016).Poultry feed based on protein hydrolysate derived from chrome-tanned leather solid waste: creating value from waste. Environmental Science and Pollution Research, 23(8), 8120-8124.

Costa-Lima B., Suman S., Li S., Beach C., Silva T., Silveira E., Bohrer B. & Boler D. (2015). Dietary ractopamine influences sarcoplasmic proteome profile of pork Longissimus thoracis. Meat Science, 103, 7-12

D'Alessandro, A., & Zolla, L. (2013). Foodomics to investigate meat tenderness. TrAC: Trends in Analytical Chemistry, 52, 47-53.

D'Alessandro, A., Marrocco, C., Zolla, V., & Zolla, L. (2011). Meat quality of the longissimus lumborum muscle of Casertana and Large White pigs: Metabolomics and proteomics intertwined. Journal of proteomics, 75(2), 610-627.

Damez, J.L., & Clerjon, S. (2016). Recent Advances in Meat Quality Assessment. Handbook of MEAT AND MEAT PROCESSING, Edition: 2nd, Chapter: 8, Publisher: CRC Press, Editors: Y.H. Hui, pp.161-175

Di Luccia, A., Picariello, G., Cacace, G., Scaloni, A., Faccia, M., Liuzzi, V., Alviti, A., & Musso, S. S. (2005). Proteomic analysis of water soluble and myofibrillar protein changes occurring in dry-cured hams. Meat Science, 69(3), 479-491.

Ferranti, P., Nitride, C., Nicolai, M. A., Mamone, G., Picariello, G., Bordoni, A., Valli, V., Di Nunzio, M., Babini, E., Marcolini, E., & Capozzi, F. (2014). In vitro digestion of Bresaola proteins and release of potential bioactive peptides. Food Research International, 63, 157-169.

Ferranti, P., Nitride, C. & Gallo, M. (2015) "Omics" and Biomedical Applications. Analytical Separation Science.

Henchion, M., McCarthy, M., Resconi, V. C., & Troy, D. (2014). Meat consumption: Trends and quality matters. Meat

science, 98(3), 561-568.

Hood, L. E., Omenn, G. S., Moritz, R. L., Aebersold, R., Yamamoto K. R., Amos, M., Hunter-Cevera, J., Locascio, L. & Workshop Participants (2012) New and improved proteomics technologies for understanding complex biological systems: addressing a grand challenge in the life sciences. Proteomics, 12(18), 2773-2783.

Hu, C., & Xu, G. (2013). Mass-spectrometry-based metabolomics analysis for foodomics. TrAC Trends in Analytical Chemistry, 52, 36-46.

Jung, Y., Lee, J., Kwon, J., Lee, K. S., Ryu, D. H., & Hwang, G. S. (2010). Discrimination of the geographical origin of beef by 1H NMR-based metabolomics. Journal of agricultural and food chemistry, 58(19), 10458-10466.

Jurado, J. M., Jiménez-Lirola, A., Narváez-Rivas, M., Gallardo, E., Pablos, F., & León-Camacho, M. (2013). Characterization and quantification of 4-methylsterols and 4, 4-dimethylsterols from Iberian pig subcutaneous fat by gas chromatography–mass spectrometry and gas chromatography–flame ionization detector and their use to authenticate the fattening systems. Talanta, 106, 14-19.

Koppelman, S. J., Lakemond, C. M., Vlooswijk, R., & Hefle, S. L. (2004). Detection of soy proteins in processed foods: literature overview and new experimental work. Journal of AOAC International, 87(6), 1398-1407.

Kuttappan, V.A., Brewer, V.B., Clark, F.D., McKee, S.R., Meullenet, J.F., Emmert, J.L. & Owens, C.M. (2009). Effect of white striping on the histological and meat quality characteristics of broiler fillets. Poultry Science, 88 (E-Suppl. 1), 136–137.

Lassoued, R., Hobbs, J. E., Micheels, E. T., & Zhang, D. D. (2015). Consumer Trust in Chicken Brands: A Structural Equation Model. Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie, 63(4), 621-647.

Leitner, A., Castro-Rubio, F., Marina, M. L., & Lindner, W. (2006). Identification of marker proteins for the adulteration of meat products with soybean proteins by multidimensional liquid chromatography-tandem mass spectrometry. Journal of proteome research, 5(9), 2424-2430.

Lesiów, T., & Kijowski, J. (2003). Impact of PSE and DFD meat on poultry processing-a review. Polish journal of food and nutrition sciences, 12(2), 3-8.

Mamone, G., Picariello, G., Nitride, C., Addeo, F., & Ferranti, P. (2013). The role of proteomics in the discovery of marker proteins of food adulteration. In *Proteomics in Foods* (pp. 465-501). Springer US.

Marcone, M. F., Wang, S., Albabish, W., Nie, S., Somnarain, D., & Hill, A. (2013). Diverse food-based applications of nuclear magnetic resonance (NMR) technology. Food Research International, 51(2), 729-747.

Marco-Ramell, A., Arroyo, L., Saco, Y., García-Heredia, A., Camps, J., Fina, M., Piedrafita, J., & Bassols, A. (2012). Proteomic analysis reveals oxidative stress response as the main adaptative physiological mechanism in cows under different production systems. Journal of roteomics, 75(14), 4399-4411.

Mazzeo, M. F., Giulio, B. D., Guerriero, G., Ciarcia, G., Malorni, A., Russo, G. L., & Siciliano, R. A. (2008). Fish authentication by MALDI-TOF mass spectrometry. Journal of agricultural and food chemistry, 56(23), 11071-11076.

Mora, L., Sentandreu, M. A., Fraser, P. D., Toldrá, F., & Bramley, P. M. (2009). Oligopeptides arising from the degradation of creatine kinase in Spanish dry-cured ham. Journal of agricultural and food chemistry, 57(19), 8982-8988.

Mora, L., Valero, M. L., Del Pino, M. S., Sentandreu, M. A., & Toldrá, F. (2011). Small peptides released from muscle glycolytic enzymes during dry-cured ham processing. Journal of proteomics, 74(4), 442-450. a

Mora, L., Sentandreu, M. A., & Toldrá, F. (2011). Intense degradation of myosin light chain isoforms in Spanish dry-cured ham. Journal of agricultural and food chemistry, 59(8), 3884-3892. b

Mudalal, S., Lorenzi, M., Soglia, F., Cavani, C. & Petracci, M. (2015) Implications of white striping and wooden breast

abnormalities on quality traits of raw and marinated chicken meat. Animal, 9(4), 728–734.

Nicholson, J. K., and Lindon, J. C. (2008). Systems biology: metabonomics. Nature, 455(7216), 1054-1056.
Olsson, V., & Pickova, J. (2005). The influence of production systems on meat quality, with emphasis on pork. AMBIO: A Journal of the Human Environment,34(4), 338-343.

Paredi, G., Raboni, S., Bendixen, E., de Almeida, A. M., & Mozzarelli, A. (2012). “Muscle to meat” molecular events and technological transformations: the proteomics insight. Journal of proteomics, 75(14), 4275-4289.

Petracci, M. & Cavani, C. (2012). Muscle growth and poultry meat quality issues. Nutrients, 4, 1–12.

Petracci, M., Mudalal, S., Bonfiglio, A. & Cavani, C. (2013). Occurrence of white striping and its impact on breast meat quality in broiler chickens. Poultry Science, 92(6), 1670–1675.

Picariello, G., De Martino, A., Mamone, G., Ferranti, P., Addeo, F., Faccia, M., SoagnaMusso, S., & Di Luccia, A. (2006). Proteomic study of muscle sarcoplasmic proteins using AUT-PAGE/SDS-PAGE as two-dimensional gel electrophoresis. Journal of Chromatography B, 833(1), 101-108.

Piñeiro, C., Barros-Velázquez, J., Vázquez, J., Figueras, A., & Gallardo, J. M. (2003). Proteomics as a tool for the investigation of seafood and other marine products. Journal of proteome research, 2(2), 127-135.

Porcari, A. M., Fernandes, G. D., Barrera-Arellano, D., Eberlin, M. N., & Alberici R. M. (2016) Food quality and authenticity screening via easy ambient sonic-spray ionization mass spectrometry. Analyst, 141(4), 1172-1184.

Ritota, M., Casciani, L., Failla, S., & Valentini, M. (2012). HRMAS-NMR spectroscopy and multivariate analysis meat characterisation. Meat science, 92(4), 754-761.

Salwani, M. S., Adeyemi, K. D., Sarah, S. A., Vejayan, J., Zulkifli, I., & Sazili, A. Q. (2015). Skeletal muscle proteome and meat quality of broiler chickens subjected to gas stunning prior slaughter or slaughtered without stunning. CyTA-Journal of Food, 1-7.

Sánchez del Pulgar, J., Soukoulis, C., Carrapiso, A. I., Cappellin, L., Granitto, P., Aprea, E., Romano, A., Gasperi, F., & Biasioli, F. (2013). Effect of the pig rearing system on the final volatile profile of Iberian dry-cured ham as detected by PTR-ToF-MS. Meat science, 93(3), 420-428.

Sentandreu, M. A., Armenteros, M., Calvete, J. J., Ouali, A., Aristoy, M. C., & Toldrá, F. (2007). Proteomic identification of actin-derived oligopeptides in dry-cured ham. Journal of agricultural and food chemistry, 55(9), 3613-3619.

Sentandreu, M. A., Fraser, P. D., Halket, J., Patel, R., & Bramley, P. M. (2010). A proteomic-based approach for detection of chicken in meat mixes. Journal of proteome research, 9(7), 3374-3383.

Sentandreu, M. A., & Sentandreu, E. (2011). Peptide biomarkers as a way to determine meat authenticity. Meat science, 89(3), 280-285.

Sihvo, H.K., Immonen, K. & Puolanne, E. (2014). Myodegeneration with fibrosis and regeneration in the pectoralis major muscle of broilers. Veterinary Pathology, 51(3), 619–623.

Straadt, I. K., Aaslyng, M. D., & Bertram, H. C. (2011). Assessment of meat quality by NMR—an investigation of pork products originating from different breeds. Magnetic Resonance in Chemistry, 49(S1), S71-S78.

Sundekilde, U. K., Rasmussen, M. K., Young, J. F., & Bertram , H. C. (2017). High resolution magic angle spinning NMR spectroscopy reveals that pectoralis muscle dystrophy in chicken is associated with reduced muscle content of anserine and carnosine. Food Chemistry, 217, 151–154.

Surowiec, I., Koistinen, K. M., Fraser, P. D., & Bramley, P. M. (2011). Proteomic approach for the detection of chicken mechanically recovered meat. Meat science, 89(2), 233-237. a

Surowiec, I., Fraser, P. D., Patel, R., Halket, J., & Bramley, P. M. (2011). Metabolomic approach for the detection of mechanically recovered meat in food products. Food Chemistry, 125(4), 1468-1475. b

von Bargen, C., Dojahn, J., Waidelich, D., Humpf, H. U., & Brockmeyer, J. (2013). New sensitive high-performance liquid chromatography–tandem mass spectrometry method for the detection of horse and pork in halal beef. Journal of agricultural and food chemistry, 61(49), 11986-11994.

Watanabe, G., Kobayashi, H., Shibata, M., Kubota, M., Kadowaki, M., & Fujimura, S. (2015). Regulation of free glutamate content in meat by dietary lysine in broilers. Animal Science Journal, 86(4), 435-442.

Yasmeen, A., F.M. Anjum, Z. Nasreen, R. Parveen, S. Ali and M.S. Arshad, (2013). Enhancement of broiler growth performance and meat quality with β-adrenergic agonist (salbutamol) and α-tocopherol acetate supplementation in feed. Journal of Public Health and Biological Science, 2, 253-264.

Zancanaro, C., Fontanella, M., Nicolato, E., Mosconi, E., Pellicciari, C., & Malatesta, M. (2011). Ex-vivo nuclear magnetic resonance characterization of pig muscles. International Journal of Meat Science, 1-15.

Zanetti, E., Masi, A., Pivato, M., Tolin, S., Trentin, A. R., Guler, C., Yalçin, S., & Cassandro, M. (2013). A note on protein expression changes in chicken breast muscle in response to time in transit before slaughtering. Proteome science, 11(1), 1.

Zuo, P., Y. Zhang, J. Liu and B.C. Ye (2010). Determination of β-adrenergic agonists by hapten microarray. Talanta, 82(1), 61-66.

CHAPTER 4: FOODOMICS FOR THE INVESTIGATION OF FOOD QUALITY IN PLANT-DERIVED PRODUCTS

4.1 Investigation on the impact of farming system

Recently, due to the increasing demand for organic foods, many studies have been carried out trying to investigate the possible metabolite difference between conventional and organic foods.
In the last decades, consumers have become more and more concerned about the quality and safety of food products and the general perception is that organic food items are healthier and safer than the conventional ones due to the fact that they are thought to have lower levels of chemical fertilizers and pesticides and higher nutritional and sensory quality. This is why organic food products now represent a big subset in the food market and why it is necessary to assess the real benefits and nutritional qualities of this type of products. [Williams & Hammit, 2001; Harper & Makatouni, 2002; Makatouni, 2002; Capozzi & Trimigno, 2015]. On the other hand, though, organic foods can actually experience more microbiological contamination and, as a consequence, higher concentration of molecules such as biogenic amines, since less antibiotics and fungicides are used and manure is employed as a fertilizer. Organic foods are sold on the market at higher prices in opposition to conventional products and this is why, sometimes, fraud occurs. For all these reasons, it is thus necessary to establish correct analytical methods to assess the organic quality of food products and their possible higher quality traits. Traditionally, target analysis on macro and micro nutrients - from sugar to vitamin contents – were employed. In other cases, stable isotopes of elements such as hydrogen, carbon or nitrogen were analysed to assess the soil were products were cultivated on and the type of fertilizer employed, since organic farming mostly employes manure, which causes a higher content of the 15N isotope. Traditional approaches, though, lack the proper reliability unless they were actually combined with multivariate methods. For this reason, untarget analysis employing metabolomic techniques can be a better choice for the attainment of reliable and robust results, also due to the fact that the different methods of farming and cultivation can influence metabolic processes and responses. [Wishart, 2008; Capozzi & Trimigno, 2015]. In addition, new farming approaches, derived from the organic one, have been lately developed and still need to be assessed in their outcomes. In effect, biodynamic agriculture was developed from the organic farming system. Biodynamic farming employes organic fertilizers, but specifically particular fermented herbal preparations added to compost and field sprays. These preparations are composed by minerals or fermented plant with animal organs, water and/or soil

and are used to improve the quality of soil and crops. Biodynamic farming thus aims to reproduce the natural cycle of energy and nutrients. By doing this, biodynamic methods should affect the microorganisms found in soil and, as a consequence, their products. It is thus interesting also to find out how this farming management affects crops. In order to investigate the effects or organic and biodynamic farming, a metabolomic approach was applied on grapes and reported in this thesis.

Journal: Food Chemistry, Volume 213, 15 December 2016, Pages 187-195

Title

^{1}H-NMR foodomics reveals that the biodynamic and the organic cultivation managements produce different grape berries (Vitis *vinifera* L. cv. Sangiovese)

Authors

Gianfranco Picone[a], Alessia Trimigno[a], Paola Tessarin[b], Silvia Donnini[b], Adamo Domenico Rombolà[b*] and Francesco Capozzi[a*]

[a]Department of Agricultural and Food Sciences, Alma Mater Studiorum - University of Bologna, Piazza Goidanich 60, 47521 Cesena (FC), Italy (gianfranco.picone@unibo.it; alessia.trimigno2@unibo.it)
[b]Department of Agricultural Sciences, Alma Mater Studiorum - University of Bologna, Viale G. Fanin, 44, 40127 Bologna, Italy (paola.tessarin2@unibo.it; silvia.donnini@unibo.it; adamo.rombola@unibo.it)
*Corresponding authors: Francesco Capozzi (francesco.capozzi@unibo.it); Adamo Domenico Rombolà (adamo.rombola@unibo.it)

Highlights

- The agronomical effects on the composition of grape are studied.
- Results are consistent with a shift towards the anaerobic metabolism in biodynamic berries.
- Proline, valine and isoleucine concentrations are higher in biodynamic than organic grapes.

Role of the PhD candidate: statistical analysis of the NMR spectra, together with the writing of the related Results and Discussion, in collaboration with Gianfranco Picone.

Abstract

The increasing demand for natural foods and beverages has boosted the adoption of organic and biodynamic cultivation methods which are based on protocols avoiding use of synthetic pesticides. This trend is striking in viticulture, since wine production is largely shaped by the varying drinking attitudes of environment-friendly consumers. Using ^{1}H-NMR, the compositions of cv. Sangiovese (*Vitis vinifera* L.) grape berries, collected at harvest in 2009 and 2011, in experimental plots

cultivated either with biodynamic or organic methods, were compared. Although the analysis provides a comprehensive metabolic profile of berries, the resulting distinctive pattern between the two cultivation methods consists of a few molecules. Lower content of sugars, coumaric and caffeic acids, as well as higher amount of γ-aminobutyric acid (GABA) were observed in biodynamic grapes with respect to the organic ones. The ^{1}H-NMR foodomics approach, applied for the first time to assess the effect of the biodynamic practice on the composition of grape berries, evidenced a diverse fruit metabolome that could be associated to a different physiological response of plants to the agronomic environment.

Keywords

Foodomics; Grape; Organic; Biodynamic, HR-NMR; Metabolomics; GABA, Polyphenols.

Introduction

The increasing demand for safe food products have boosted the diffusion of alternative agronomical strategies, such as organic and biodynamic farming which can provide numerous benefits in terms of sustainability and soil quality. This trend is remarkable in viticulture , mainly as a consequence of the cultural and social prominent role of wine and of the special attention historically devoted to viticulture and winemaking. Over 280,000 ha of organic grapes were grown in 2014, according to FIBL and IFOAM surveys. Worldwide approximately 147,000 ha are managed according to Demeter biodynamic standards. There are 520 Demeter wineries in the world, with 8000 ha of vineyards. Organic and conventional farming adopt distinct management practices and processing methods , which in many countries are regulated by specific laws. For example, the European Community recently enacted a regulation which states that “organic wines” should only originate from organic grapes with limited use of sulphur dioxide during the vinification process and storage, whereas other oenological practices are either restricted or prohibited. By contrast, there is still a lack of official European regulation for biodynamic viticulture and winemaking. Regardless of certification procedures, biodynamic agriculture differs from organic management in the use of specific fermented preparations, applied on crops or soil in very small amounts, which are claimed to stimulate soil nutrient cycle, photosynthesis and optimal evolution of compost, enhancing both soil and crop quality. Few scientific studies have been devoted to understand the effects of biodynamic preparations/method on plant physiology and yield, edible parts composition and nutritional value. Recently it has been reported that some biodynamic preparations (500, 500K, 501 and Fladen) led to an increase in leaf enzymatic activities of chitinase

and β-1,3-glucanase, which are typically correlated with plant biotic and abiotic stresses and associated with induced plant resistance. Regarding to yields, both Reeve, et al. (2005) and Botelho, et al. (2015) detected similar values when the organic cultivated vines where compared to the biodynamic ones. In addition, Döring, et al. (2015) observed similar yields in vines under biodynamic and organic farming; however, these cultivation methods showed lower values when compared to the integrated management. Scant information is available on the characteristics of biodynamic grapes and wine. With regard to berry chemistry, a pioneering study from Reeve, et al. (2005) showed that biodynamically cultivated wine grapes had significantly higher soluble solids, total phenols and anthocyanins in the last harvest year compared to those from organically managed plants. In the white cv. Riesling no differences on must soluble solids, pH and total acidity were detected among organic, biodynamic and integrated treatments, whereas α-amino-acids changed according to the cultivation method (Döring et al., 2015). A study by Tassoni, et al. (2013) compared the biogenic amine content and the polyphenol profiles of berries (cvs. Pignoletto and Sangiovese) grown with conventional, organic and biodynamic practices, highlighting no significant differences among samples. In a recent work with other cultivars (Albana and Lambrusco), Tassoni, et al. (2014) recorded changes in the concentration of catechins and stilbens. However, the analysis did not generally evidence significant differences among wine samples coming from different agricultural and winemaking practices. A recent study demonstrated that the quality of cv. Sangiovese red wines was largely influenced by the application of biodynamic preparations. Moreover, by using ^{1}H-NMR, it was possible to discriminate between red wines from organic and biodynamic grapes. The present paper, based on the foodomics approach , proposes a holistic approach combining high resolution nuclear magnetic resonance spectroscopy (HR-NMR) with multivariate data analysis, to reveal differences between the metabolomes of organic and biodynamic grape berry. To date, there are no studies using NMR and foodomics to evaluate the effect of the biodynamic agronomical practices on the grape berry metabolic composition. The NMR spectroscopy is a method of choice for foodomics due to ease of sample preparation, reducing the risks of manipulation artefacts, and shortening the overall analysis time, largely used in high-throughput applications. Despite the lower sensitivity, with respect to some other techniques commonly applied for metabolite profiling, NMR spectroscopy has the advantage to acquire spectra that directly represent the molecular composition of the extracts. NMR spectra, indeed, may be subjected to immediate data analysis, taking advantage of their extreme reproducibility, high dynamic concentration range, and the universal correlation between the

concentration and the area of the corresponding signal, as the response factor is the same for every hydrogen atom irrespective of the molecule it belongs to. Thus, the NMR spectrum of a biological extract reflects, as a whole, its metabolic status, while the loss of important information is reduced as much as possible. Most of the redundant information associated to the NMR spectra is condensed in a few orthogonal patterns by application of multivariate data analysis, Such a simplification allows the recognition of meaningful information about the response of an organism to external perturbations. The main aim of the current study is to measure the overall molecular profile of biodynamic grapes and compare it to the organic ones. The proposed approach applies metabolomics to evaluate the metabolic profile of grapes, as determined on their perchloric acid extracts. The changes induced by agronomic techniques on the composition of the fruit are combined to other meta-information related to the sample nature in order to interpret the chemical perturbations. To the best of the authors' knowledge, this is the first study which, starting from the analysis of the overall NMR spectra of samples produced with different agronomic practices, identifies a pool of compounds able to discriminate between grape berries produced by biodynamic and organic cultivation methods. We discuss here possible interpretation of mechanisms of action underlying the observed modifications.

Material and Methods

Site description, experimental design and vineyard management

The experiment was performed in a mature vineyard of cv. Sangiovese (clone FEDIT 30 ESAVE), *Vitis vinifera* L., grafted on Kober 5BB, trained to Cordon du Royat. The vineyard was located in Tebano, Italy (44°17′7″ N, 11°52′59″E, 117 m a.s.l.), in a medium hill slope, with South-East/North-West and downhill oriented rows. Vines were spaced in 2.8 m x 1.0 m (3,571 plants ha^{-1}). Starting in 2007, the vineyard was managed as organic (ORG) in accordance with Regulation EC 834/2007.

Then, in 2008, the total surface (20,000 m^2) was divided in two large uniform areas with similar soil physic-chemical characteristics, each of them submitted to a specific cultivation method: organic (ORG) farming, managed according to the same Regulation CE 834/2007 ; biodynamic (BD) farming, based on organic management, except for the applications of biodynamic preparations. Each treatment, ORG and BD, was applied to 7 replications, for a total of 14 experimental plots, each including 12 sample plants (total 168 sample plants). The biodynamic preparations 500, fladen, 500 K, 501 were used from 2008 until 2011 and trunk paste in 2010, as described in Botelho, et al. (2015). To prepare fladen, a hole was dug, covered with sticks of birch, filled with fresh and

compact cow manure, without straw and subsequently inoculated with compost preparations. Eggshells and basalt were mixed together for 1 h with a spade, then placed in the ground, inside a box with its base open to the soil, for 8 weeks. The fladen preparation was used when dark brown, crumbly and free of manure smell. In 2008, fladen was applied 3 times, twice in spring and once in autumn; 501 4 times during summer; 500 once in autumn. In 2009, fladen was sprayed 3 times in spring; 500 4 times in spring; 501, twice in summer; 500 K twice in autumn. In 2010, fladen was applied twice in spring; 500 twice in spring, 501 three times in summer, 500 K once in autumn. In 2011 500 was applied twice in spring; fladen twice in spring; 501 three times in summer; of 500 K once in autumn. The application time was modulated according to environmental conditions and observations on plants. In 2008–2011, the number of buds (12–14) was adjusted by winter pruning, whereas the number of bunches (11–15) was adjusted by cluster thinning only in 2011. At the end of each vegetative season herbaceous species were sown in alternate planting rows, such as fava bean (*Vicia faba*), barley (*Hordeum vulgare*), subterranean clover (*Trifolium subterraneum*) and mustard green (*Brassica juncea*) in both ORG and BD plots. Soil was managed by mowing the vegetation during late spring, which maintained biomass on the soil surface. Both, ORG and BD experimental plots were treated against diseases and pests in the same manner, using the biological products that are allowed by Regulations EC 2092/91 Treatments consisted mainly of copper (an average of 6 kg/ha/yr) and sulfur (an average of 70 kg/ha/yr), enabling control of fungal pathogens. In 2009, the average T recorded from the end of April to harvest (September 22nd) was 22.0 °C; maximum T of 38.7 °C occurred on July 22nd, total rainfall was 190 mm and predominantly took place in April and during the first week of July. Overall, the 2011 vegetative season was marked by average temperatures well above the seasonal normal, with temperature peaks of 30°C in August. From bud burst to harvest, the average relative humidity (RH) varied from 40 to 70%; highest values were observed during spring (92%) and the lowest (38%) during the latter part of August. The total rainfall (204 mm) from bud burst to harvest was sporadic during spring and almost absent during ripening.

Analysis of berries

At harvest, on 50 berries randomly collected in each experimental plot, the following parameters were evaluated: berry weight; expressed as g per berry (Gibertini Elettronica S.r.l., Milan, Italy); total soluble solids (TSS; Brix; Electronic Refractometer Maselli Misure S.P.A., Parma, Italy), titratable acidity (TA; expressed as g/L) and pH (Crison Compact Titrator, Crison Instrument SA, Barcelona, Spain). At harvest, two sub-samples of 50 berries per plot were randomly collected from

the same side, North-West oriented, of the vegetative wall. The samples were brought on ice to the laboratory and stored at -80° C until extraction.

HR ^{1}H-NMR Sample Preparation

Polar metabolites were extracted from a single berry per time (2009 and 2011) using a 3M perchloric acid (PCA, 65%, Sigma-Aldrich Inc., St. Louis, MO, USA) to stop enzymatic activity and to remove proteins. Subsequently, the acid mixtures were subjected to a first centrifugation for 5 min at 14,000 rpm (Microfuge® 18 centrifuge, Beckman Coulter, Brea, California) in order to remove insoluble tissues. Each resulting supernatant was dispensed in 1.00 mL aliquots and neutralized using 9 M KOH to pH 7.5 and then centrifuged at 8000 rpm for 5 min at 4 °C in order to remove potassium perchlorate precipitate. NMR samples were prepared by adding 100 µL of deuterium oxide (D_2O) and 10 µL of 3-TrimethylSilyl-Propanoic-2,2,3,3-d4 acid sodium salt (TSP) at a final concentration of 1.4 mM to 600 µL of supernatant.

^{1}H-NMR spectra acquisition and processing

All NMR spectra were recorded at 300 K on a Bruker US+ Avance III spectrometer operating at 600 MHz, equipped with a BBI-z probe and a B-ACS 60 sampler for automation (Bruker BioSpin, Karlsruhe, Germany) according to a protocol already published. Each spectrum was acquired using 32 K data points over a 7211.54 Hz spectral width and adding 32 transients and an acquisition time of 2.27 s. A recycle delay of 5 s and a 90° pulse of 11.4 µs were set up. HOD residual signal was suppressed by applying the ZGPR sequence (a standard pulse sequence included in the Bruker library) by irradiating it during the recycle delay at δ equal to 4.703 ppm (parts per million by frequency). Each spectrum was automatically processed adjusting phase and baseline with the command apk0.noe using TopSpin version 3.0 (Bruker BioSpin, Karlsruhe, Germany) and applying a line broadening factor of 0.3 Hz. The chemical shifts were internally referenced to the Trimethylsilyl propanoic acid (TSP, Cambridge Isotope Laboratories, USA) at 0.00 ppm. After the Fourier Transformation (FT) and prior to multivariate analysis, data underwent to a pre statistical improvement, such as normalization by Probabilistic Quotient Normalization (PQN) and peaks' alignment by spectral reduction into 192 bins of 0.04 ppm of length. The first step is aimed at removing possible dilution effects , while the second step avoids the effect of peaks misalignments among different spectra due to variations in chemical shift of signals belonging to some titratable acids. Both these treatments, normalization and binning, were performed by using homemade algorithms written in the R 3.2.2 program language (http://www.r-project.org/). The spectral peaks were assigned by comparing their chemical shift and multiplicity with the literature and by using

Chenomx software (Chenomx Inc., Edmonton, Canada).

Statistical analysis of berries data

Comparison of means and analysis of variance between treatments were done by using SAS 6.04 software (SAS INSTITUTE, CARY, NC, USA). Means were compared by the Student-Newman-Keuls test ($P = 0.05$).

Multivariate statistical analysis of NMR data

Multivariate data analysis of the NMR spectra was carried out by using homemade algorithms written in the R 3.2.2 program language. A mean-centered principal component analysis (PCA) was applied on the PQN and binned data set as an explorative analysis to examine the intrinsic variation in the data set.

Results and Discussion

Analysis of berries

The cultivation method did not modify berry growth and technological parameters, regardless of the year. In fact, in 2009, BD and ORG vines displayed similar values of berry weight (ORG: 2.15±0.26 g; BD 1.93±0.29 g), TSS (ORG: 22.5±1.6 Brix; BD 20.9±2.6 Brix), pH (ORG: 3.31±0.13; BD 3.3±0.071) and TA (ORG: 4.03±0,46 g L^{-1}; BD: 4.26±0,77 g L^{-1}). Also in 2011 BD and ORG plants showed similar berry weight (ORG: 2.91±0.24 g; BD 1.65±0.281 g), TSS (ORG: 25.5±0.7 Brix; BD 25.6±1.0 Brix), pH (ORG: 3.38±0.049; BD 3.41±0.0) and TA (ORG: 7.58±0.51 g L^{-1}; BD: 7.64±0.61 g L^{-1}).

NMR analysis.

A representative one dimensional ^{1}H-NMR spectrum of ORG grape berry, from 2009 harvest, acquired from pulp and skin extracts is represented in Fig. 1. Twenty-eight metabolites were identified in the spectrum and their assignments are reported in the picture with their relative chemical shifts. The grape berry extract is characterized by the presence of high concentration of sugar compounds which resonances in the ^{1}H-NMR spectrum are dominant. This is reflected by their weight as source of variance among all samples, as shown in Fig. 2B which represents the loadings' plot resulting from PCA applied to the entire data array, according to standard protocols for metabolomics. The PC scores plot in Figure 2A shows a tendency of samples to differentiate along PC1according to the cultivation method (BD and ORG), and along PC2 according the year (2009 white and black circles and 2011 white and black squares). As sugar compounds are abundant in grape berry, PC1 loadings' plot is mainly characterized by sugar resonances. Positive

loadings, in the sugar region, mainly involve glucose and fructose signals. Thus, it appears that sugars are more abundant in ORG than in BD, since the latter are in the negative right part of the plot. This trend is confirmed for both 2009 and 2011 years (Table 1 and 2).

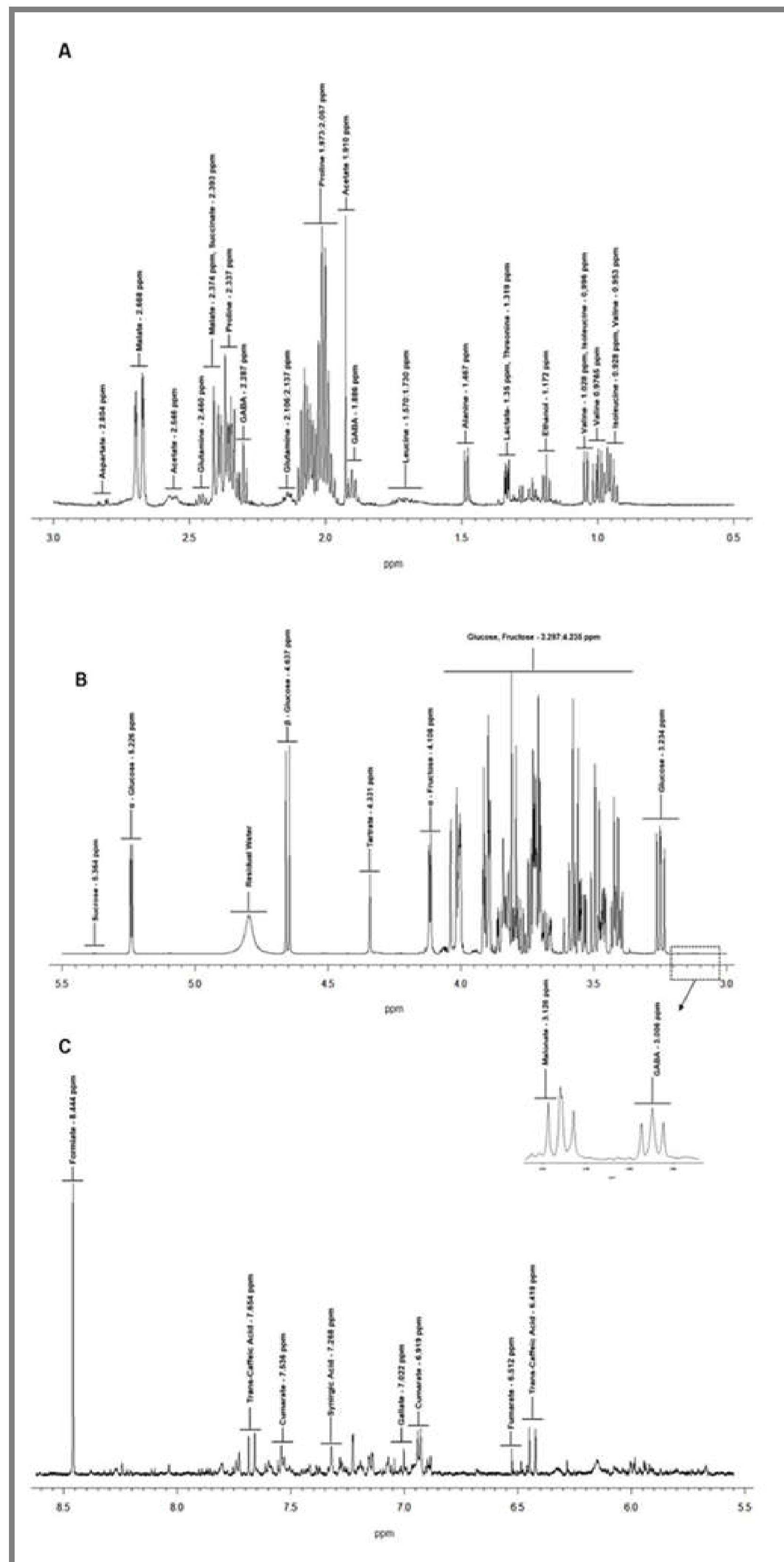


Fig. 1. ^{1}H-NMR spectrum registered on a grape berry. The spectrum has been split into three parts for the sake of clarity. Some resonances have been assigned and labelled accordingly in the following regions: A) Aliphatic region, characterized by the presence of signals belonging to amino acids and organic acids; B) Carbohydrate regions, characterized by the presence of signals belonging to sugars and C) Aromatic region, characterized by the presence of signals belonging to phenolic and other aromatic compounds.

In order to have information from the other parts of the spectrum, a PCA was applied separately on both aliphatic (Fig. 3) and aromatic (Fig. 4) regions.

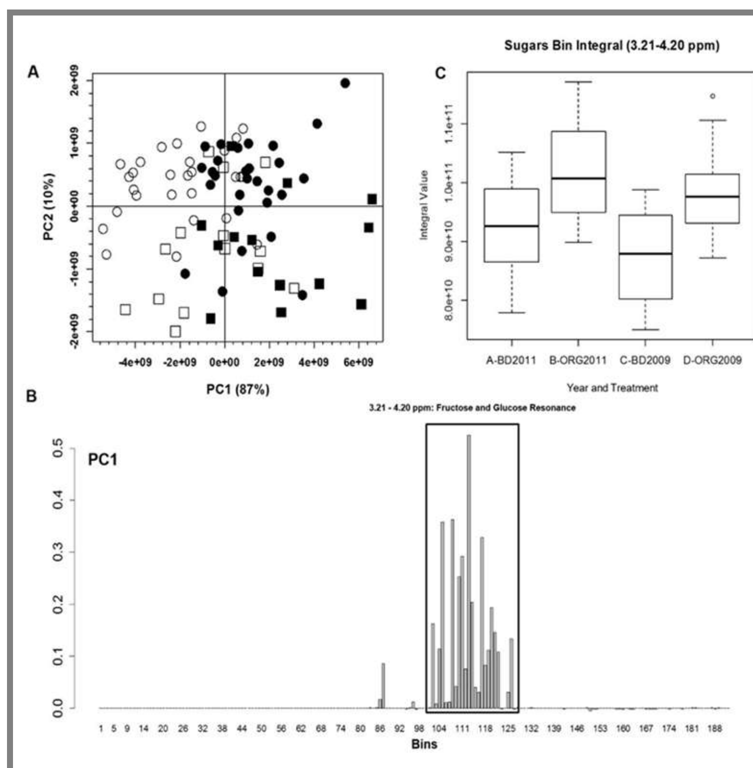


Fig. 2. Biodynamic and organic grape berry sample multivariate analysis. A) PC score plot from 84 NMR spectra of skins and pulp extract of BD (open symbols) and ORG (filled symbols) berries harvested in 2009 (circle) and 2011 (square). The first two PCs explain 87% (PC1) and 10% (PC2), respectively, of the total variance. B) Loading bar-plot for spectral bins along PC1. C) Boxplot of sugar regions from 3.21 to 4.20 ppm, characterized by the presence of fructose and glucose signals.

The loadings' bar-plot along PC1 highlights the metabolites of the aliphatic spectral region contributing to the differentiation between BD and ORG grape berries. Most of these, especially organic acids malate and lactate, are more abundant in the BD samples. Although the differences between refractometric °Brix index of ORG and BD grapes are not statistically significant, it has to be point out that the water content may be the main source of variance, thus enlarging the standard deviations of each group masking possible differences among groups. The NMR data, operated in a way that spectra are normalized on the organic compounds content, indeed point out that a different amount of water among berries (dilution factors) are responsible for a large intra-group variance, preventing the differences between groups to be statistically significant. The

PQN algorithm, so far adopted to normalize spectra, emancipates from dilutions. Together with organic acids, also some amino acids such as proline, valine and isoleucine are much higher in biodynamic samples together with γ-aminobutyric acid (GABA) (Fig. 3B). As far as the aromatic compounds are concerned, it is worth noting here that only molecules soluble in the water phase are extracted and visible in the NMR spectra.

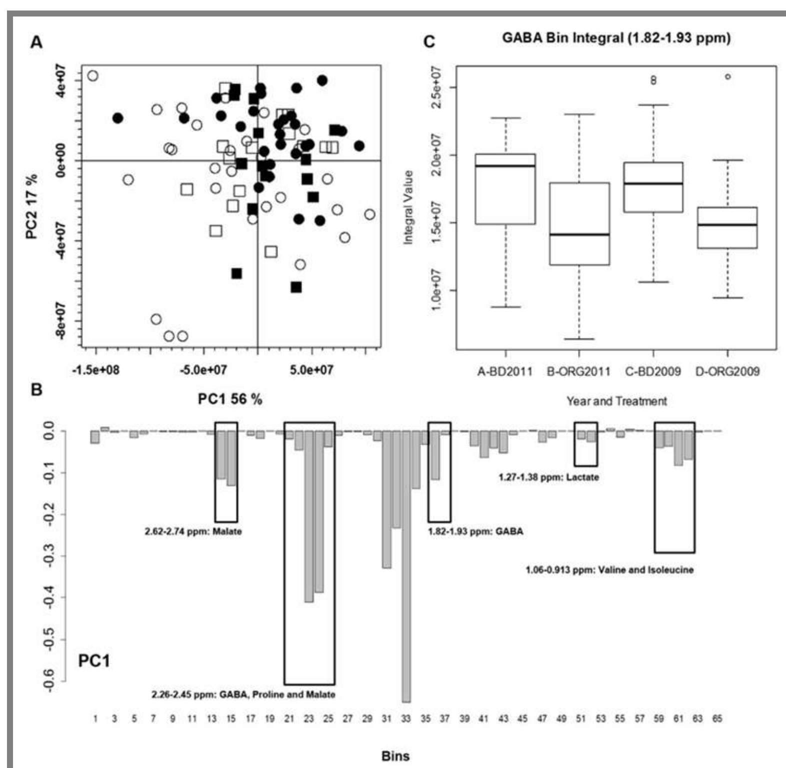


Fig. 3. Biodynamic and organic grape berry samples multivariate analysis. A) PC score plot of aliphatic region (from 0.51 to 3.21 ppm) from 84 spectra of extracts from skins and pulp of BD (open symbols) and ORG (filled symbols) berries harvested in 2009 (circle) and 2011 (square). The first two PCs explain 56% (PC1) and 17% (PC2), respectively, of the total variance. B) Loadings bar-plot for spectral bins along PC1. C) Boxplot of GABA interval from 1.82 to 1.93 ppm.

Although the intensity of signals reflects a poor abundance with respect to organic acids and sugars, a multitude of peaks is present revealing a complex mixture of monomeric phenolic compounds as evidenced by the PC analysis confined on the aromatic part of the NMR spectrum with exclusion of the formic acid singlet (Fig. 4). Along the main component (PC1), which explains the 45% of the total variance, there is a clear separation of the two groups, BD and ORG. The analysis of the loadings (Figure 4) confirms that ORG samples are characterized by higher levels of coumaric acid (Fig. 4C) and trans caffeic acid (Fig. 4D). The experimental conditions (experimental design, number of replicates, sampling methods, etc.), adopted in the present study, allowed us to minimize possible interference exerted by environmental and agronomical factors in the comparison of the cultivation method. The vineyard hosting the trial was accurately controlled by

monitoring physiological parameters (e.g. leaf chlorophyll and nutrient concentration, photosynthetic activity, leaf water potential), yield and fruit quality. Plants subjected to both BD and ORG methods displayed similar vegetative (e.g. pruning weight), productive (e.g. yield) (data not shown) and technological parameters (e.g. berry soluble solids, pH and titratable acidity).

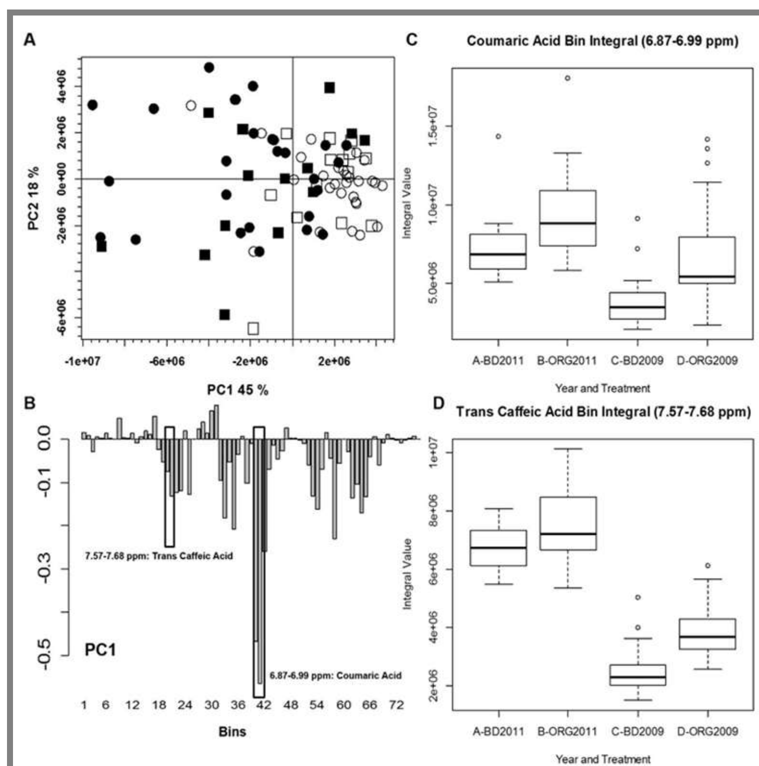


Fig. 4. Biodynamic and organic berry samples multivariate analysis. A) PC score plot of aromatic region (from 5.55 to 8.4 ppm) from 84 spectra from extracts of skins and pulp of BD (open symbols) and ORG (filled symbols) berries harvested in 2009 (circle) and 2011 (square). The first two PCs explain 45% (PC1) and 18% (PC2), respectively, of the total variance. B) Loadings bar-plot for spectral bins along PC1. C) Boxplot of Coumaric Acid interval from 6.87 to 6.99 ppm. D) Boxplot of Trans Caffeic Acid interval from 7.57 to 7.68 ppm.

Therefore, the differences detected by ^{1}H-NMR on berries from the two cultivation methods should be ascribed to the use of biodynamic preparations. Based on the results of this work it can be referred that berries from BD agriculture, compared to those from ORG cultivation, display an activation of glycolysis toward fermentative pathway. The fermentative pathway, occurring under the absence or low oxygen concentration, implies that pyruvate is converted into lactate or ethanol. Under non limiting oxygen levels, pyruvate is oxidized to yield acetyl CoA that enters in the Krebs cycle. Indeed, the lower concentration of sugars (mainly glucose and fructose) along with the increase concentration of organic acids such as lactate and malate and ethanol might suggest the activation of the anaerobic metabolism in the BD berries respect to ORG berries. Grapevines of cv. Sangiovese, under BD management exhibited lower leaf stomatal conductance as compared to organically cultivated vines. In a study conducted on cv. Riesling this parameter did not differ

between ORG and BD vines (Döring et al., 2015). The different effects of the cultivation method could be related to plant water status and genotypic responsiveness toward it (Chaves et al., 2010). Furthermore, reduction in stomatal conductance has been associated with enhanced plant tolerance toward biotic and abiotic stresses , which is considered to be one of the key aspects of BD viticulture. A lower stomatal conductance is associated with an increase in the concentration of internal CO_2. These findings might support the possible predominance of anaerobic metabolism in BD berries. Accordingly, the observed accumulation of formic acid in BD berries would result from the activation of formate dehydrogenase that under specific conditions might consume CO_2 and NADH to produce formate. Considering that this reaction might regenerate NAD^+, the accumulation of formate in BD berries might suggests the occurrence of an enhanced turnover of reducing power presumably associated to compensate the deficit of aerobic metabolism (mitochondrial respiration and Krebs cycle). In fact, together with organic acids, also some amino acids such as proline, valine and isoleucine are much higher in biodynamic samples together with γ-aminobutyric acid (GABA) (Fig. 3). In particular, isoleucine and valine are formed from pyruvate: the increase in their concentration in the BD berries suggests that a reduced amount of pyruvate enters in the Krebs cycle and could be directed to other routes. Since the vines submitted to BD cultivation method exhibited lower leaf stomatal conductance , it would be interesting to further explore the possible relationship between this physiological parameter and the fruit anaerobic metabolism. The ^{1}H-NMR technique, by emancipating from possible detrimental artefacts introduced by more manipulating methods, provided complex information on the chemical composition of berries, embedded in hundreds signals generated by as much compounds. However, the complexity of information is simplified by multivariate data analysis, which mines the relevant spectral features responsible for variance, and this approach highlighted the prevalent role of few molecules that clearly discriminated between the two cultivation methods. Among these substances there are some polyphenols, such as coumaric and caffeic acids that represent a class of phenolic acids largely important for their influence on the wine sensory value. In fact, they contributes to the colour, astringency, bitterness and the aroma of wine, either directly or as a result of a whole interactions with other molecules such as proteins, polysaccharides or other polyphenols. In our experiment, coumaric and caffeic acids, together with other polyphenolic compounds, were lower in BD grapes with respect to ORG ones. Thus, between BD and ORG berries there are differences in the polyphenolic profile that may reflect in the deriving wine. Wines derived from BD grapes harvest in 2009 were characterized by a lower concentration of

total polyphenols and anthocyanins than ORG wines of the same year , and the observed trend is confirmed in the corresponding berries in the present work. In spite of their importance in grape and wine quality, polyphenols are hard to compare among different cultivation methods, because their content is heavily affected by environmental and other agronomical practices (e.g. canopy and soil management) as well as vegetal material (e.g. variety, rootstock). Therefore, limited published literature is present on this subject. The conclusions drawn by Tassoni, et al. (2013) and are based on a single investigated vineyard per each cultivation method. Moreover, the concentration of specific polyphenols was not reported by Reeve, et al. (2005). It is noteworthy that the concentration of GABA, significantly higher in berries obtained by the BD management in the 2009 harvest, effectively discriminated between the two cultivation methods. Although the difference in not significant in 2011, the same trend is observed confirming the tendency to be more accumulated in BD than in ORG. However, it has to be argued that the environmental conditions, fluctuating year by year, may have influence in attenuating such a difference. This molecule appears ubiquitously in living organisms and is considered an indicator of the capability of plants to react to environmental stimuli, including attacks by insects and pathogens. GABA has different functions within the plant: defense against insects , protection from oxidative stress and transmission of signals in the plant. In particular, in response to a signal mediated by cytosolic Ca^{2+} ions, GABA is rapidly synthesized at high amounts. In addition, GABA represents a substrate for yeasts. *Saccharomyces cerevisiae* can in fact efficiently use it as a source of assimilable nitrogen, especially in conditions of poor availability of YAN. The biodynamic cultivation method induced in wines effects on GABA concentration similar to those found in the present paper. The new information may suggest that the source of such a difference could be assigned prevalently to the cultivation practice rather than to yeasts co-metabolism. Consequently, GABA represents an interesting molecule as putative marker for sustainable agricultural systems. Furthermore, in humans GABA works an inhibitory neurotransmitter with beneficial health effects; hence special attention is devoted to improve its concentration in food and beverages. As a component of food imparting health benefits (relief from stress, better immune system), biodynamic applications could be object of future research in the. As a component of food impaeastudy that highlights changes of GABA in grape berries associated to biodynamic BD method.

Conclusions

The comparison of grape composition between two different cultivation methods by ^{1}H-NMR-

based foodomics demonstrated changes in molecular composition, related to quality attributes, here assigned to specific compounds testifying the extraordinary capability of plants to respond to challenging surrounding environment. The foodomics approach, previously applied to berries for the first time to evaluate a possible substantial equivalence between conventional and transgenic genotypes , stimulated us to further explore plant responses to the biodynamic cultivation method with special attention to edible parts. The comparison of grape composition between two different cultivation methods by ^{1}H-NMR-based foodomics demonstrated changes in molecular composition, related to quality attributes, here assigned to specific compounds, which are relevant for their food and enological implications.

Acknowledgements

This work was supported by CL.A.N (Cluster Agrifood Nazionale D.D. MIUR 30.5.2012 prot.N.257/RIC) through the PROSIT Project. The Consortium CIRMMP is acknowledged as well for providing support and facilities.

Table 1. Sugars, Coumaric acid, Caffeic acid and GABA Area arbitrary unitsin cv. Sangiovese berries obtained in 2009 from experimental plots under organic (ORG) and biodynamic (BD) management.

2009				
Treatment	**Sugars**	**Coumaric**	**Caffeic**	**GABA**
ORG	103501,8638	12,696081	5,079097	14,477993
BD	92253,38848	7,471296	3,205697	18,007534
Significance	**	**	***	**

****Significant at *P ≤ 0.01*; ** significant at *P ≤ 0.001*.**

Table 2. Sugars, Coumaric acid, Caffeic acid and GABA Area arbitrary unitsin cv. Sangiovese berries obtained in 2011 from experimental plots under organic (ORG) and biodynamic (BD) management.

2011				
Treatment	**Sugars**	**Coumaric**	**Caffeic**	**GABA**
ORG	106975,6726	17,656448	9,096597	14,536449
BD	96410,60039	15,207398	8,162619	16,682954
Significance	**	**	**	*n.s.*

**** significant at *P ≤ 0.01*; n.s., not significant.**

REFERENCES CHAPTER 4.1

Abdou, A. M., Higashiguchi, S., Horie, K., Kim, M., Hatta, H., & Yokogoshi, H. (2006). Relaxation and immunity enhancement effects of γ-aminobutyric acid (GABA) administration in humans. Biofactors, 26(3), 201-208.

Bach, B., Sauvage, F.-X., Dequin, S., & Camarasa, C. (2009). Role of γ-aminobutyric acid as a source of nitrogen and succinate in wine. American Journal of Enology and Viticulture, 60(4), 508-516.

Botelho, R. V., Roberti, R., Tessarin, P., Garcia-Mina, J. M., & Rombolà, A. D. (2015). Physiological responses of grapevines to biodynamic management. Renewable Agriculture and Food Systems, 1-12.

Bouché, N., Lacombe, B. t., & Fromm, H. (2003). GABA signaling: a conserved and ubiquitous mechanism. Trends in cell biology, 13(12), 607-610.

Bown, A. W., Hall, D. E., & MacGregor, K. B. (2002). Insect footsteps on leaves stimulate the accumulation of 4-aminobutyrate and can be visualized through increased chlorophyll fluorescence and superoxide production. Plant Physiology, 129(4), 1430-1434.

Capozzi, F., & Trimigno, A. (2014). “Using metabolomics to describe food in detail” Metabolomics as a Tool in Nutrition Research, 203.

Collins, C., Penfold , C., Johnston, L., Marschner, P., & Bastian, S. (2015). The relative sustainability of organic, biodynamic and conventional viticulture In 19th International meeting of Viticulture GiESCO, (pp. 53-55). Pech Rouge-Montpellier.

Demeter, I. (2012) International Viticulture Congress). Colmar, France.

Dieterle, F., Ross, A., Schlotterbeck, G., & Senn, H. (2006). Probabilistic Quotient Normalization as Robust Method to Account for Dilution of Complex Biological Mixtures. Application in 1H NMR Metabonomics. Analytical Chemistry, 78(13), 4281-4290.

Döring, J., Frisch, M., Tittmann, S., Stoll, M., & Kauer, R. (2015). Growth, Yield and Fruit Quality of Grapevines under Organic and Biodynamic Management. PLoS ONE, 10(10), 1-28.

Downey, M. O., Dokoozlian, N. K., & Krstic, M. P. (2006). Cultural practice and environmental impacts on the flavonoid composition of grapes and wine: a review of recent research. American Journal of Enology and Viticulture, 57(3), 257-268.

EC. (1991). Council Regulation (EEC) No 2092/91 of 24 June 1991 on organic production of agricultural products and indications referring thereto on agricultural products and foodstuffs. Official Journal of the European Communities, L198, 1-15.

EC. (2007). Council Regulation (EC) No 834/2007 of 28 June 2007 on organic production and labelling of organic products and repealing Regulation (EEC) No 2092/91. Official Journal of the European Union, L189, 1-23.

EC. (2012). Specific rules for the making of wine. In E. Commission (Ed.), vol. No 203 (pp. 71/42-71/47).

Gattinger, A., Muller, A., Haeni, M., Skinner, C., Fliessbach, A., Buchmann, N., Mäder, P., Stolze, M., Smith, P., & Scialabba, N. E.-H. (2012). Enhanced top soil carbon stocks under organic farming. Proceedings of the National Academy of Sciences, 109(44), 18226-18231.

Harper, G. C., & Makatouni, A. (2002). “Consumer perception of organic food production and farm animal welfare.” British Food Journal, 104(3/4/5), 287-299.

Kinnersley, A. M., & Turano, F. J. (2000). Gamma aminobutyric acid (GABA) and plant responses to stress. Critical Reviews in Plant Sciences, 19(6), 479-509.

Koepf, H., Schaumann, H., & Haccins, M. (2001). Agricoltura Biodinamica (Vol. 366).

Koepf, H. H., Pettersson, B., & Schaumann, W. (1976). Biodynamic agriculture. Spring Valley, NY: The Anthroposophic Press.

Kruger, N. J., Troncoso-Ponce, M. A., & Ratcliffe, R. G. (2008). 1H NMR metabolite fingerprinting and metabolomic analysis of perchloric acid extracts from plant tissues. Nat. Protocols, 3(6), 1001-1012.

Laghi, L., Picone, G., & Capozzi, F. (2014). Nuclear magnetic resonance for foodomics beyond food analysis. TrAC Trends in Analytical Chemistry, 59, 93-102.

Laghi, L., Versari, A., Marcolini, E., & Parpinello, G. P. (2014). Metabonomic investigation by 1H-NMR to discriminate between red wines from organic and biodynamic grapes. Food and Nutrition Sciences, 2014, 52-59.

Loveless, K., Mueller, S., Lockshin, L., & Corsi, A. (2010). The relative importance of sustainability, quality control standards and traceability for wine consumers: a cross-national segmentation. In Australian and New Zealand Marketing Academy Conference, University of Canterbury, New Zealand, vol. 29).

Mäder, P., Fliessbach, A., Dubois, D., Gunst, L., Fried, P., & Niggli, U. (2002). Soil Fertility and Biodiversity in Organic Farming. Science, 296(5573), 1694-1697.

Makatouni, A. (2002). “What motivates consumers to buy organic food in the UK? Results from a qualitative study.” British Food Journal, 104(3/4/5), 345-352.

Mann, S., Ferjani, A., & Reissig, L. (2012). What matters to consumers of organic wine? British Food Journal, 114(2), 272-284.

Mulas, G., Galaffu, M. G., Pretti, L., Nieddu, G., Mercenaro, L., Tonelli, R., & Anedda, R. (2011). NMR Analysis of Seven Selections of Vermentino Grape Berry: Metabolites Composition and Development. Journal of Agricultural and Food Chemistry, 59(3), 793-802.

Murdoch, J., & Miele, M. (1999). ‘Back to nature’: changing ‘worlds of production’in the food sector. Sociologia ruralis, 39(4), 465-483.

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

Parpinello, G. P., Rombolà, A. D., Simoni, M., & Versari, A. (2015). Chemical and sensory characterisation of Sangiovese red wines: Comparison between biodynamic and organic management. Food Chemistry, 167, 145-152.

Picone, G., Laghi, L., Gardini, F., Lanciotti, R., Siroli, L., & Capozzi, F. (2013). Evaluation of the effect of carvacrol on the Escherichia coli 555 metabolome by using 1 H-NMR spectroscopy. Food Chemistry, 141(4), 4367-4374.

Picone, G., Mezzetti, B., Babini, E., Capocasa, F., Placucci, G., & Capozzi, F. (2011). Unsupervised principal component analysis of NMR metabolic profiles for the assessment of substantial equivalence of transgenic grapes (Vitis vinifera). Journal of Agricultural and Food Chemistry, 59(17), 9271-9279.

Plahuta, P., & Raspor, P. (2007). Comparison of hazards: Current vs. GMO wine. Food control, 18(5), 492-502.

Ponzio, C., Gangatharan, R., & Neri, D. (2013). Organic and Biodynamic Agriculture: A Review in Relation to Sustainability. Int. J. Plant Soil Sci, 2(1), 95-110.

Preys, S., Mazerolles, G., Courcoux, P., Samson, A., Fischer, U., Hanafi, M., Bertrand, D., & Cheynier, V. (2006). Relationship between polyphenolic composition and some sensory properties in red wines using multiway analyses. Analytica Chimica Acta, 563(1–2), 126-136.

Ramputh, A.-I., & Bown, A. W. (1996). Rapid [gamma]-aminobutyric acid synthesis and the inhibition of the growth and

development of oblique-banded leaf-roller larvae. Plant Physiology, 111(4), 1349-1352.

Reeve, J. R., Carpenter-Boggs, L., Reganold, J. P., York, A. L., McGourty, G., & McCloskey, L. P. (2005). Soil and winegrape quality in biodynamically and organically managed vineyards. American Journal of Enology and Viticulture, 56(4), 367-376.

Reganold, J. P., Palmer, A. S., Lockhart, J. C., & Macgregor, A. N. (1993). Soil Quality and Financial Performance of Biodynamic and Conventional Farms in New Zealand. Science, 260(5106), 344-349.

Reganold, J. P., Palmer, A. S., Lockhart, J. C., & Macgregor, A. N. (1993). Soil quality and financial performance of biodynamic and conventional farms in New Zealand. Science, 260, 344-344.

Rombolà, A. D., Tassarin, P., Tumbarello, G., Parpinello, G. P., & Versari, A. (2015). Evaluation of Soil and Canpoy Management Strategy in High Sustainable Viticultural Systems. In 19th International meeting of Viticulture GiESCO, (pp. 95-98). Pech Rouge-Montpellier.

Ross, C. F., Weller, K. M., Blue, R. B., & Reganold, J. P. (2009). Difference testing of Merlot produced from biodynamically and organically grown wine grapes. Journal of Wine Research, 20(2), 85-94.

Salazar-Parra, C., Aguirreolea, J., Sánchez-Díaz, M., Irigoyen, J. J., & Morales, F. (2012). Climate change (elevated CO2, elevated temperature and moderate drought) triggers the antioxidant enzymes' response of grapevine cv. Tempranillo, avoiding oxidative damage. Physiologia plantarum, 144(2), 99-110.

Savorani, F., Picone, G., Badiani, A., Fagioli, P., Capozzi, F., & Engelsen, S. B. (2010). Metabolic profiling and aquaculture differentiation of gilthead sea bream by 1 H NMR metabonomics. Food Chemistry, 120(3), 907-914.

Shelp, B. J., Bown, A. W., & McLean, M. D. (1999). Metabolism and functions of gamma-aminobutyric acid. Trends in plant science, 4(11), 446-452.

Tassoni, A., Tango, N., & Ferri, M. (2013). Comparison of biogenic amine and polyphenol profiles of grape berries and wines obtained following conventional, organic and biodynamic agricultural and oenological practices. Food Chemistry, 139(1), 405-413.

Tassoni, A., Tango, N., & Ferri, M. (2014). Polyphenol and biogenic amine profiles of Albana and Lambrusco grape berries and wines obtained following different agricultural and oenological practices. Food and Nutrition Sciences, 2014.

Turinek, M., Grobelnik-Mlakar, S., Bavec, M., & Bavec, F. (2009). Biodynamic agriculture research progress and priorities. Renewable Agriculture and Food Systems, 24(02), 146-154.

Vigani, G., Chittó, A., De Nisi, P., & Zocchi, G. (2012). cDNA-AFLP analysis reveals a set of new genes differentially expressed in cucumber root apexes in response to iron deficiency. Biologia plantarum, 56(3), 502-508.

Vigani, G., & Zocchi, G. (2010). Effect of Fe deficiency on mitochondrial alternative NAD (P) H dehydrogenases in cucumber roots. Journal of plant physiology, 167(8), 666-669.

Wallace, W., Secor, J., & Schrader, L. E. (1984). Rapid accumulation of γ-aminobutyric acid and alanine in soybean leaves in response to an abrupt transfer to lower temperature, darkness, or mechanical manipulation. Plant Physiology, 75(1), 170-175.

Willer, H., Lernoud, J., & Schlatter, B. (2014). Current statistics on organic agriculture worldwide: Organic area, producers and market (Vol. 46).

Williams, P. R., & Hammitt, J. K. (2001). “Perceived risks of conventional and organic produce: pesticides, pathogens, and natural toxins.” Risk analysis, 21(2), 319-330
Wishart, D. S. (2008). “Metabolomics: applications to food science and nutrition research.” Trends in Food Science & Technology,19(9), 482-493.

Zeng, W., Melotto, M., & He, S. Y. (2010). Plant stomata: a checkpoint of host immunity and pathogen virulence. Current opinion in biotechnology, 21(5), 599-603.

Zucca, G., Smith, D. E., & Mitry, D. J. (2009). The Sustainable Viticulture and Winery Practices in California: What is it, and do Customers Care? International Journal of Wine Research, 2(1), 189-194.

4.2 Foodomics to investigate genetic modification

Another recent trend in food production is the employment of genetically modified (GM) crops and transgenic foods. These are obtained by the recombination of the DNA in order to achieve specific and favourable characters (higher yield, better nutritional quality, etc.). Research is though needed to understand if these genetic modifications can induce changes that could be harmful for the consumers and metabolomics could give answers in this topic, thanks to its great analytical power. Genetic modification can in fact impact on food safety and quality, thus the so-called “substantial equivalence” of GMO in relation to the wild type (WT) counterpart needs to be assessed. The substantial equivalence means that the two variants are similar enough that can be considered the same and treated equally in regards of food safety. In addition to the necessary assessment of the substantial equivalence, the beneficial effects of the GM need to be proved and this is way modern research and, more specifically omic sciences, have been more and more employed to investigate on GMOs. In 2011, Picone et al., tried to assess the substantial equivalence of GM table grapes employing 1H-NMR spectroscopy and metabolomics. This study showed that the whole grape metabolome, but more specifically organic acids and aromatic compounds, could discriminate between GM and wild-type and this difference was more significant for one GM cultivar, thus the genetic modification was modulated by the host genotype. This study also proved that the metabolomic approach employing NMR spectroscopy and multivariate statistical analysis could help in the investigation of GM crops and products. Starting from this study, then, the characteristics of the GM lines compared to the WT one of the investigated cultivars were investigated in the following study. This was done in order to better understand what happens when the transgene is inserted.

Journal: Metabolomics, Volume 12, October 2016, Page 150

Title

Metabolic changes of genetically engineered grapes (*Vitis vinifera* L.) studied by ^{1}H-NMR, metabolite heatmaps and *i*PLS

Authors

Gianfranco Picone[1,¤,], Francesco Savorani[2,3¤], Alessia Trimigno[1], Bruno Mezzetti[4], Francesco Capozzi[1] and Søren Balling Engelsen[2*]

[1]Department of Agricultural and Food Sciences, Alma Mater Studiorum - University of Bologna, Piazza Goidanich 60,

47521 Cesena (FC), Italy; E-mails: gianfranco.picone@unibo.it (G.P.); alessia.trimigno2@unibo.it (A.T.); francesco.capozzi@unibo.it (F.C.)
[2]Quality & Technology, Department of Food Science, Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 30, 1958 Frederiksberg C, Denmark; E-mails: se@food.ku.dk (S.B.E.); frsa@food.ku.dk (F.S.)
[3]Department of Applied Science and Technology (DISAT), Polytechnic University of Turin – Corso Duca degli Abruzzi 24, 10129 Torino (TO), Italy; E-mail: francesco.savorani@polito.it (F.S.)
[4]Department of Agricultural, Food and Environmental Sciences, Marche Polytechnic University, Via Brecce Bianche 60100 Ancona (AN), Italy; E-mail: b.mezzetti@univpm.it (B.M.)

¤ Picone and Savorani contributed equally to the research.

* To whom correspondence should be addressed: E-mail: se@food.ku.dk

Tel: +45 35 33 32 05 – Fax: +45 20 20 00 64

Role of the PhD candidate: statistical analysis of the NMR spectra and, more specifically the production of the metabolite heatmaps, the writing of the Materials and Methods and Results and Discussion sections together with Gianfranco Picone.

Abstract

Introduction: The *Deficiens* Homologue 9-iaaM (*DefH9-iaaM*) gene is an ovule-specific auxin-synthesizing gene which is expressed specifically in placenta/ovules and promotes auxin-synthesis. It was introduced into the genome of two grape cultivars Thompson Seedless and Silcora (SIL) and both transgenic cultivars had an increased number of berries per bunch.

Objectives: This study investigates the down-stream metabolic changes of Silcora and Thompson seedless grape cultivars when genetically modified through the insertion of the *DefH9-iaaM* gene into their genome.

Methods: The effects of the genetic modification upon the grape metabolome were evaluated through ^{1}H-NMR and exploratory data analysis. Chemometric tools such as Interval Partial Least Squares regression and metabolite heatmaps were employed for scrutinizing the changes in the transgenic metabolome as compared to the wild type one.

Results: The results show that the pleiotropic effect on the grape metabolome as a function of the gene modifications is relatively low, although the insertion of the transgene caused a decrement in malic acid and proline and an increment in p-coumaric acid content. In addition, the concentration of malic acid was successfully correlated with the number of inserted copies of transgene in the Silcora cultivar, proving that the increased production of berries, promoted by the inserted gene, is achieved at the expense of a decrement in malic acid concentration.

Conclusion: NMR together with chemometrics is able to identify specific metabolites that were up-

or down regulated in the genetically engineered plants allowing highlighting alterations in the down-stream metabolic pathways due to the up-stream genetic modifications.

Keywords: *Vitis vinifera* L.; GMO grape; ^{1}H-NMR, metabolomics; chemometrics; *i*PLS

Introduction

Grape berries (Vitis vinifera L.) present a double sigmoid developing growth pattern that can be divided into three phases. The first phase (fruit set, Supplementary Fig. S1) represents the initial rapid size increment along with the formation of the berry skin and flesh (berry pericarp). During the second phase (ripening), grape berries reach full ripeness with a rapid accumulation of sugars as well as a decrement in acidity. The two phases are separated by a transition phase characterized by the so-called véraison or onset of grape ripening. In the first phase, tartaric acid together with hydroxycinnamic acids, tannins and monomeric catechins are accumulated in the berries. Hydroxycinnamic acids play an important role as precursors of volatile phenols like p-coumaric acid, caffeic acid, and ferulic acid , as well as for the non-volatile tannins that are responsible for the bitter and astringent sensory properties of red wine. At the later stages of grape berry development, accumulation of other compounds occurs, such as sugars (sucrose, glucose and fructose), amino acids such as proline and arginine that represent most of the nitrogen content in grape berries , minerals and micronutrients related to the plant survival as well as some aroma compounds such as methoxypyrazines. The second phase, representing the fruit ripening, starts with the véraison that marks the end of the lag phase in which berries have remained steady and is characterized by drastic morphological and chemical changes. In this phase the berries become larger, softer, sweeter, less acidic and strongly colored and flavored. The enhancement in flavor is a consequence of a double ripening effect: firstly, the production of flavor and aromatic compounds such as norisoprenoids, terpenes and thiols in the form of sugar or amino acid conjugates and, secondly, the modification of the balance between organic acids and sugars. Just after the véraison there is a sudden and massive increment in the monosaccharides glucose and fructose due to the breakdown of sucrose that has been transported into the berries from the leaves. Also organic acids are synthesized at this stage, the largest part is represented by tartaric and malic acids, which typically account for 90% of acids in grapes. The remaining part of organic acids is characterized by minor amounts of citric, succinic, lactic and acetic acid. After véraison, malic acid decreases rapidly because its production ceases while the respiration increases. In contrast, tartaric acid only decreases slightly, mainly because of dilution effects and a partial conversion to its salt forms. The

Deficiens Homologue 9-iaaM (DefH9-iaaM) gene, an ovule-specific auxin-synthesizing gene, was introduced into the genome of two grape cultivars with different flowering habit: Thompson Seedless (THS) long cane flowering habit and Silcora (SIL) short cane flowering habit. Both transgenic cultivars had an increased number of berries per bunch, but this effect was stronger for THS than for SIL. In a preliminary study, Picone et al. (2011) showed first that the effect of the insertion of different copies of the DefH9-iaaM construct on the entire metabolome of SIL and THS could be detected by the application of unsupervised multivariate analysis of 1H-NMR spectra data of grape hydro-alcoholic extracts. In addition, it was assessed whether the genetic modification shifts the metabolome towards the same direction by investigating both cultivars. This study instead, is aimed at characterizing and understanding the changes in the downstream grape metabolome induced by the up-stream genetic modifications. For this purpose an advanced chemometric tool, the interval Partial Least Squares regression (iPLS) has been employed in order to characterize the minute metabolic changes resulting from the genetic modification in comparison with the no GM corresponding control. 1H-NMR in combination with chemometrics have proven great potential for the generation of unambiguous biological information on how an organism responds to an external perturbation such as genetic factors or environmental factors. For this reason, the NMR based metabolomics approach has been largely employed in the enology field for the exploration of the changes in the non-volatile metabolome of wine and grape berries due to these effects. Recently the NMR techniques started to be considered as a tool for the detection of both intended and unintended effects due to the genetic modification of organisms. The main advantage of NMR spectroscopy undoubtedly relies in its very large dynamic range so that major compounds, like sugars, can be evaluated together with minor components, like phenols and other aromatic compounds, in a single analysis. The total amount of sugars, acids, aromatics, and all other compounds is directly proportional to the area in the corresponding spectral regions. The same NMR spectrum simultaneously provides information on the presence and amount of acids, aromatics and other chemicals, provided they are soluble and above the detection limit, whilst with other analytical techniques separate runs are often necessary. This advantage, along with its high reproducibility, non-destructivity and easy sample preparation, make NMR spectroscopy one of the most exploited techniques for metabolomics, complementary to GC- and HPLC-MS platforms which excel for their much higher sensitivity. Since the metabolomic study of genetically modified food is commonly based on MS-based approaches , to the authors' knowledge this is the first study in which the metabolic effects of genetically engineered grapes are

investigated using NMR spectroscopy.

Materials and methods

Plant Material

The experiment was performed on both control, wild type (WT) samples, and transgenic lines of THS and SIL. For the SIL grape cultivar, two different genetic lines were considered, distinguishable by the number of copies of the transgene introduced into the host genome: SIL line with a single copy of the new inserted gene (SIL-GM1) and SIL line with three copies of the transgene (SIL-GM2). For what concerns the THS cultivar, only the genetic line with one copy of transgene inserted was considered (THS-GM1). The transgene consists of the *iaaM* gene from *Pseudomonas syringae* pv. *Savastanoi*, under the control of the placenta and ovule-specific promoter *DefH9* gene isolated from *Antirrhinum majus*. The *iaaM* gene encodes the *Trp-2-monooxygenase* enzyme that catalyzes the conversion of tryptophan (Trp) into indole-3-acetamide (IAM) which is then hydrolyzed into indole-3-acetic acid, (IAA) , as shown in Supplementary Fig. S2.

Experimental trial

Transgenic *DefH9-iaaM* lines of both cultivars were cultivated (1999-2006) under open field conditions to compare their plant development, yield and fruit quality with the corresponding non-transgenic lines of THS and SIL. The open field experimental trial with transgenic and control clones was established at the Experimental Farm of the Marche Polytechnic University in March 2001, following the EC (CE 2001/18) rules for transgenic plants field evaluation. All plants, both transgenic and wild type, were cultivated in the same vineyards according to identical agronomic practices. Field trial was set including 4 plots of 4 plants for each clone (control and transgenic lines).

Experimental design and sample preparation

Fruit sampling: NMR samples were prepared from the extracts of berries of THS-WT and of the *DefH9-iaaM* line (THS-GM1), and berries of SIL-WT and of the two *DefH9-iaaM* modified lines, SIL-GM1 and SIL-GM2. All berries were harvested in the same period (year 2007). At harvest, samples were freeze dried at -80° and then shipped in dry ice at the laboratory of analyses.

Samples preparation: For each genotype (control and transgenic lines of both cultivars), a bulk of 100 ± 1 grams of berries was treated; both skin and pulps of each sample were homogenized under ice chilling and three aliquots of about 10.0 g were vortexed separately with 10.0 ml of a mixture of methanol and 50 mM CH_3COOH/Na^+ buffer pH 5.0 (2:1) in order to perform the solvent extraction. The suspensions were centrifuged at 10,000 rpm for 20 minutes at 4° C. The resulting hydro-

alcoholic solution was dispensed in different Eppendorf tubes, as 1 ml aliquots, and stored at -80° C. Before the NMR acquisition, 10% (v/v) D_2O was added to each 1 ml extract, thus centrifuged at 14,000 rpm for 5 minutes at room temperature. pH was not further adjusted, but it was measured for each sample yielding an average of 4.29 with a standard deviation of 0.08, confirming the buffer efficacy. A volume of 800 µl was transferred to a 5-mm NMR tube in order to acquire a single NMR FID (Free Induction Decay). Three extracts (E1-3) for each homogenate (H1-3) obtained from all genotypes (G1-5) were subjected to NMR analysis, by alternating samples so that the five extracts E1:H1:G1-5 were firstly analyzed, then followed by the series E2:H2:G1-5. When the last series E3:H3:G1-5 was analyzed, a total of 45 spectra was acquired. With this experimental design, represented in Supplementary Fig. S3, it was also possible to asses both accuracy and precision of the instrumental analysis.

Chemicals

All chemical reagents were of analytical grade and were purchased from SIGMA-ALDRICH, Inc., St. Louis, MO.

^{1}H-NMR Data acquisition and processing

The ^{1}H-NMR spectra were recorded at T= 300K using a Varian Mercury AS/400 NMR spectrometer operating at 9.4 T, corresponding to 400.098 MHz ^{1}H Larmor frequency. For each spectrum a WET (Water suppression Enhanced through T1 effects) sequence was used. The WET sequence was employed in order to remove water and methanol singlets from the spectra and this is carried out through the generation of a shaped pulse which solely excites these solvent's resonances and, after that, these excited resonances are suppressed. The advantage of the WET method is that it can also facilitate the decoupling and removal of ^{13}C satellites of organic solvents as methanol. The sequence has been created using the following command: wet1D_setpulses('30','1H water frequency', '30', '1H methanol frequency'). A total of 2048 scans were accumulated, with data collected into 16k data points with a spectral width of 16 ppm, a pulse angle of 60°, a recycle delay of 1.0 s and acquisition time of 2.561 s. Prior to Fourier transformation, each FID was apodized by Lorentzian line broadening of 1 Hz and the corresponding spectra were phase and baseline corrected manually using MestRe-C 4.9.8.0 (www.mestrec.com, Mestreab Research SL, Santiago de Compostela, Spain). The same software was used to export data in ASCII file format. Chemical shift referencing was performed by setting the glucose's β-anomeric signal to 4.60 ppm. The following spectral regions were removed prior to data analysis: the regions including only noise (the spectrum edges between 14.05 and 8.7 ppm and between 0.4 and -1.93 ppm) and the NMR signals

which are strongly affected by the residual solvent peak (water, between 5.15 and 4.65 ppm, methanol, between 3.28 and 3.41 ppm and acetic acid, between 1.97 and 2.18 ppm). As a result, the total number of spectral data points was reduced to 7585 points. A spectral dataset matrix with the dimensions 45 × 7585 was then constructed for subsequent multivariate data analysis.

NMR data preprocessing and analysis

Prior to multivariate analysis, the NMR data was preprocessed in order to minimize unwanted sources of variation due to slightly different instrumental conditions, imperfect baseline and phase corrections, and sample preparation artifacts. Firstly, as no internal standard was used all ^{1}H-NMR spectra were normalized to unit area ; secondly, the signals were aligned in order to correct for small differences (<0.05 units) in the pH of the extract. The alignment of the spectra was performed using the *i*coshift tool available at http://www.models.life.ku.dk/algorithms/. This algorithm is able to perform a tailored alignment in specific regions of the spectrum, allowing the correction of single shift differences instead of the entire spectrum. Spectral alignment is required in order to conserve the bilinearity of the data and to avoid using binning which is a crude method to "align by peak-broadening" which may cause the loss of valuable spectral information.

Chemometric tools

In order to explore the data, unsupervised Principal Component Analysis (PCA) was initially carried out on the mean-centered data matrix, divided into three different ^{1}H-NMR spectral regions: phenolic region, carbohydrate region and organic acid region (see Fig. 1). Using this segmentation, it was possible to investigate which spectral region contains most information about the grape metabolome perturbations induced by the genotype changes. The three different ^{1}H-NMR spectrum regions were analyzed using the chemometric software LatentiX™ 2.0 (www.latentix.com, Latent5, Copenhagen, Denmark). In order to maximize the separation between the genotypes and thus enable the identification of important and descriptive regions in NMR spectroscopic data, *i*PLS regression has been performed for both grape cultivars separately: this method has been carried out in Matlab version R2014a (MathWorks Inc., Natick, MA, USA) using the *i*Toolbox freely available for download from http://www.models.life.ku.dk/algorithms/. Finally, a metabolic correlation analysis was conducted in order to highlight the interactions amongst metabolites, and the relationship between metabolites and the gene copies.

The Pearson correlation analysis generates a coefficient called the Pearson correlation coefficient, denoted as *r*. Its value can range from -1 for a perfect negative linear relationship to +1 for a perfect positive linear relationship. A value of 0 (zero) indicates no relationship between two

variables.

Results

As previously shown by Costantini , the genetic modification by the *DefH9-iaaM* gene induces in THS a larger increment in number of inflorescences per shoot, in comparison to the control (THS-WT). Gene expression studies showed that this effect corresponded to a high expression of the *iaaM* gene which increased the production of IAA at the fruit set (first phase in Supplementary Fig. 1) with a consequent increment of auxins accumulation in the first stage of fruit development. For the SIL cultivar, the SIL-GM1 line showed a higher effect on increasing fruit set and total production. For all lines very small differences were detected in terms of fruit quality in comparison to the corresponding controls (THS-WT and SIL-WT) as reported in a previous study .

3.1 ^{1}H-NMR profiles

In order to further investigate the effect of these genetic modifications on the metabolome, NMR spectroscopy was employed on THS and SIL samples, both WT and GM and the results are presented and discussed in the following. Fig. 1 shows a representative ^{1}H-NMR spectrum of a methanol-water extract of THS-WT including the metabolite assignment numbers used in the assignment table (Table 1).

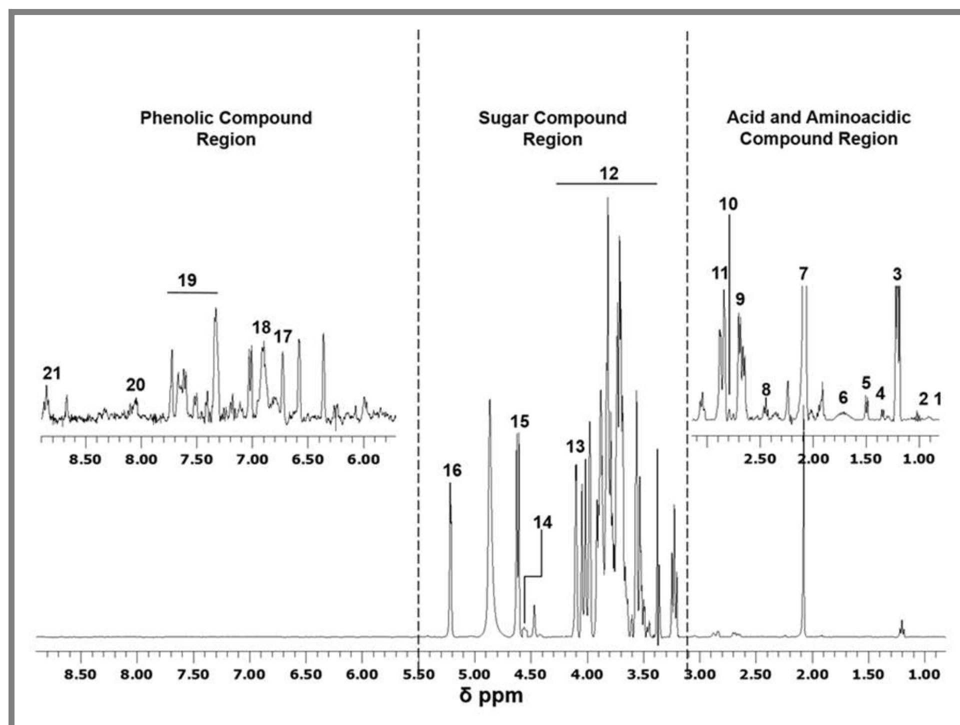


Fig. 1 ^{1}H-NMR spectrum of hydro-alcoholic extract of THS-WT grape (*Vitis vinifera* L) recorded at 400 MHz. Downfield and upfield regions are magnified on the vertical scale, respectively 400 and 20 times, in order to better represent the presence of low-intensity signals.

Table 1 lists the molecular assignments obtained through a combination of literature assignments , standard addition (spiking) and by the use of a multimedia library of pure compounds believed to be present in the extracts (http://www.bmrb.wisc.edu/).

Table 1 Assignment table of the identified metabolites present in the 1H-NMR spectra of hydro-alcoholic extracts of grape cultivars at pH 4.29. Chemical shift values are referenced to the glucose's β-anomeric proton signal at 4.60 ppm*

Number	Compound	Functional group	Multiplicity[a]	δ (ppm)
1	Valine	γ-CH_3	D	0.96
2	Valine	$γ^1$-CH_3	D	1.02
3	Ethanol	CH_2	T	1.17
4	Lactate	β-CH_3	D	1.33
5	Alanine	β-CH_3	D	1.47
6	Arginine	γ-CH_2	M	1.69
7	GABA	β-CH_2	M	1.90
8	Acetate	CH_3	S	2.07
9	Proline	β-CH	M	2.35
7a	GABA	α-CH_2	T	2.430
10	Malate	β-CH	Dd	2.63
11	Citrate	α-CH + γ-CH	D	2.77
12	Malate	$β^1$-CH	Dd	2.81
7b	GABA	γ-CH_2	T	3.02
13	Glucose-Fructose-Sucrose	CH	M	3.49-4.08
14	Fructose	CH	D	4.08
15	Ascorbate	CH	D	4.55
16	β-D-Glucopyranose	CH	D	4.60
17	α-D-Glucopyranose	CH	D	5.21
18	Fumarate	α, β-CH=CH	S	6.57
19	*p*-Coumarate	CH	D	6.88
20	Cinnamic acids	Ring Protons	M	7.30-7.80
21	Trigonelline	CH	M	8.05
22	Trigonelline	CH	M	8.85

[a]d, doublet; dd, doublet of doublets; m, multiplet; s, singlet; t, triplet
*All the identified compounds correspond to the second level of metabolite identification (putatively annotated compounds) according to MSI

The ^{1}H-NMR spectrum is mainly dominated by the overlapping of carbohydrate peaks belonging to sucrose, glucose and fructose (between 3.49 and 4.49 ppm) which characterized the midfield region of the spectrum. In this region there are also peaks belonging to the α- and β-anomeric of glucose whose protons can be found at 5.25 (α) and 4.60 (β) ppm. In the upfield region (between 0.99 and 3.49 ppm), the most abundant peaks correspond to amino acids: valine (0.96 and 1.02 ppm), alanine (1.47 ppm), arginine (1.69 ppm) and proline (2.34 ppm), and to the organic acids malate and citrate (from 2.87 to 2.64 ppm). Also signals from lactic acid (1.33 ppm), 4-aminobutanoic acid (GABA - 1.907, 2.430 and 3.025 ppm) and acetic acid (2.07 ppm) are present, the latter used in the extraction buffer. In the downfield region some signals of different phenolic compounds can be observed, but only the signals of p-coumaric acid from 6.84 to 6.94 ppm has

been assigned.

Multivariate data analysis

In order to explore differences in the metabolome of the five different genotypes, individual PCA models were calculated on the three main spectral regions: aromatic, carbohydrate and organic acid regions, respectively. The results are shown in Fig. 2 and the loadings plots for the aromatic and organic acid regions are presented in Supplementary Fig. 5a and 5b respectively. In order to further identify specific metabolites which are correlated to the genetic modification and to the number of transgene copies, different *i*PLS models were calculated on the spectra from each cultivar. For this purpose, the spectral data sets were subdivided into 37 meaningful equidistant intervals by the *i*PLS routine. Fig. 3 and 4 show the *i*PLS summarizing plots for the prediction of the number of transgene's copies for the THS and SIL grapes, respectively. The two figures show, superimposed onto the corresponding average spectra, interval bars representing the PLS model Root-Mean-Square Error of Cross-Validation (RMSECV) for each spectral interval. The interval PLS performance is conveniently compared to the PLS performance obtained using the full spectral area (global RMSECV) indicated by the stipulated horizontal line. The digits in the bars indicate the number of components, i.e. the latent variables (LVs) used for calculating each local PLS model. Models were calculated both with a fixed number of LVs, equal for each interval and with the optimal number of LV's obtained through Cross Validation for each interval (data not shown). The first approach resulted in the most meaningful outcome by describing the relevance of each interval in characterizing the metabolic changes due to the genetic modification. Because of their higher complexity, only the models using the full spectral area were left free to use their optimal number of LVs. Fig. 3 shows the global PLS model for THS samples which has a RMSECV of 0.25 (in the range from 0 (WT) to 1 for the (THS-GM1)) and requires 6 LVs (black dotted line).

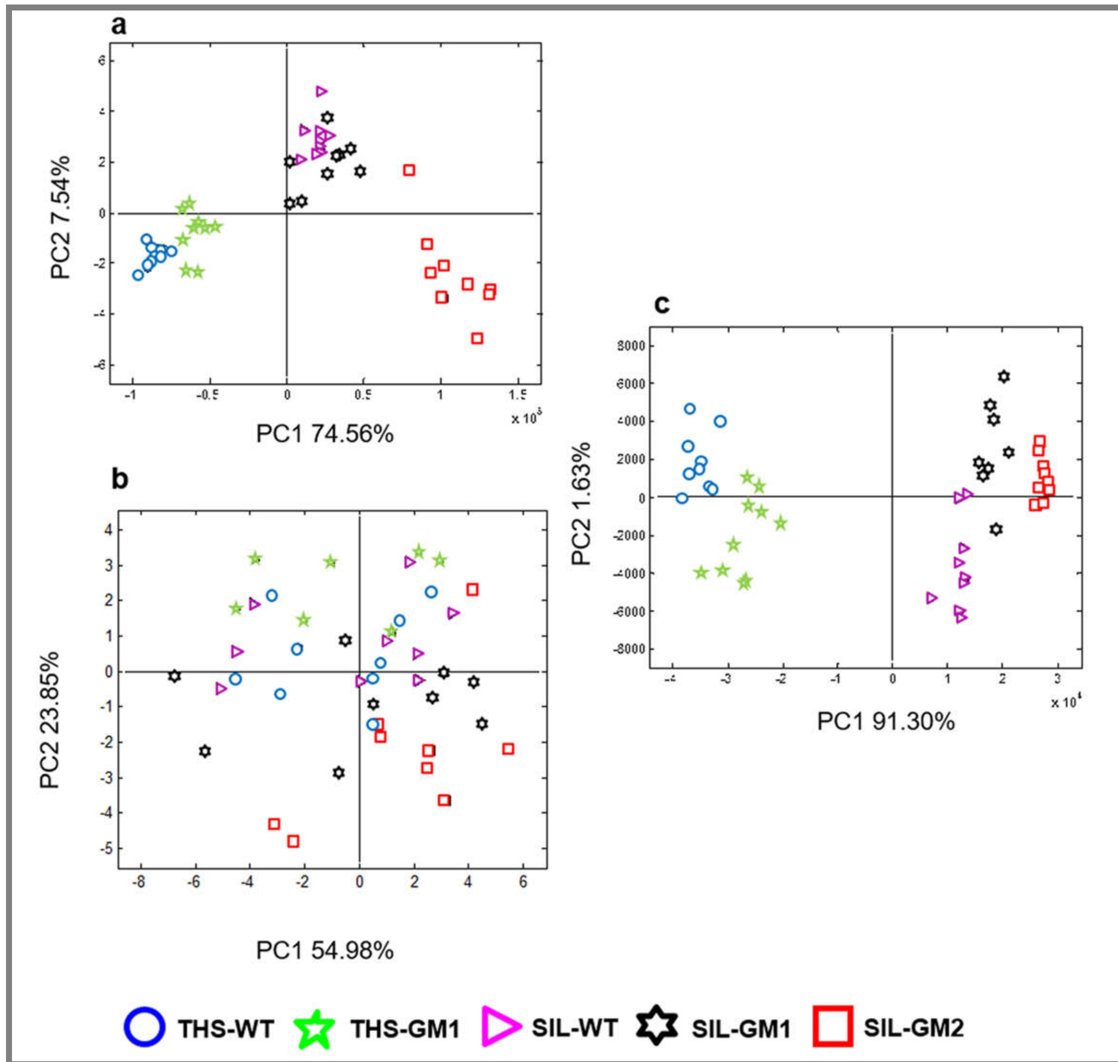


Fig. 2 Scores plot of the first two principal components of the PCAs for the three regions of the ^{1}H-NMR spectra; a) Downfield region (phenolic compounds region); b) Midfield region (carbohydrate compounds region) and c) Upfield region (Acid and aminoacidic compounds region).

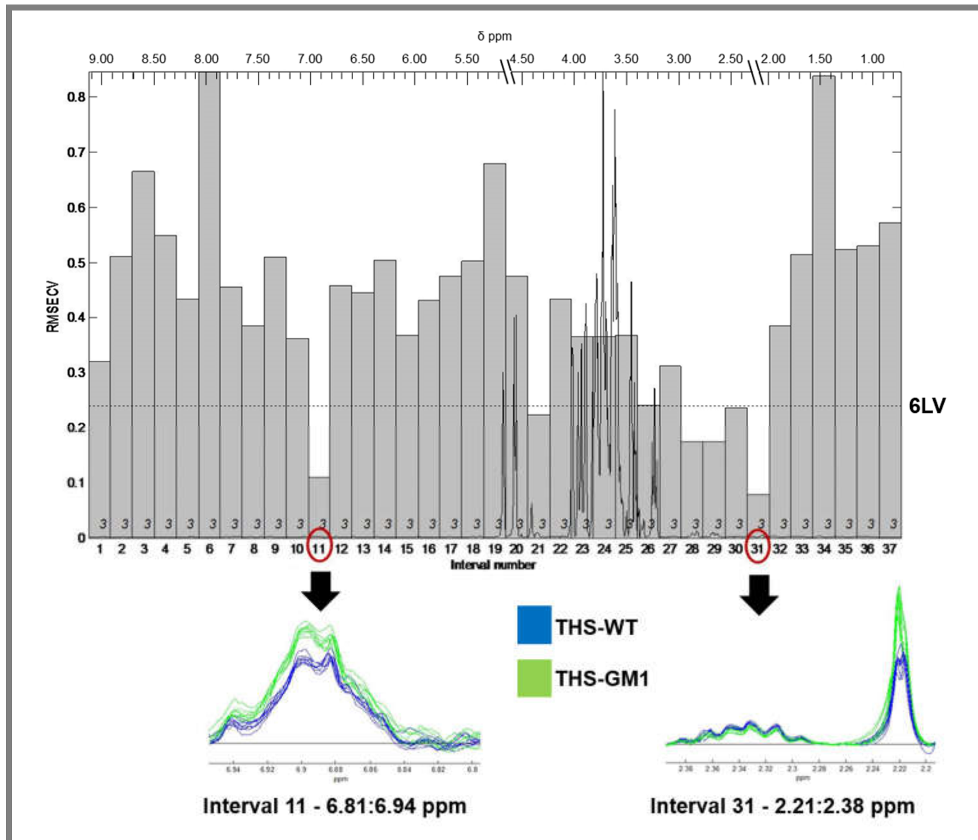


Fig. 3 *i*PLS plot with interval number versus RMSECV for the prediction of the number of transgenic copies in THS (zero for WT and 1 for the THS-GM1). The figure shows the bars representing the RMSECV for each spectral interval compared to the global PLS performance (the stipulated horizontal line) and superimposed onto the average spectrum. Digits in the columns indicate the number of latent variables (LV) used in each local PLS model. The selected intervals (11 and 31) provide a performance superior to the overall calibration obtained using the full ^{1}H-NMR spectrum

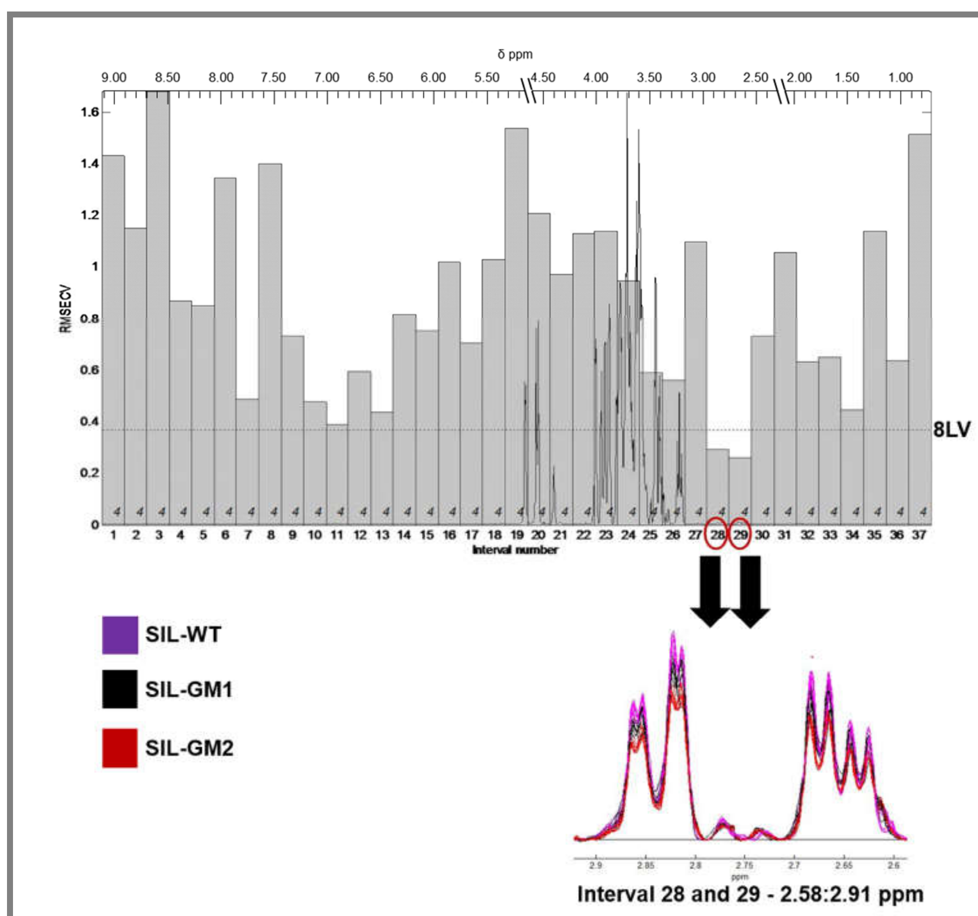


Fig. 4 *i*PLS plot with interval number versus RMSECV for the prediction of the number of transgenic copies in SIL (zero for WT, 1 and 3 for the transgenic types SIL-GM1 and SIL-GM2 respectively). The figure shows the bars representing the RMSECV for each spectral interval compared to the global PLS performance (the stipulated horizontal line) and superimposed onto the average spectrum. Digits in the columns indicate the number of latent variables (LV) used in each local PLS model. The selected intervals (28 and 29) provide a performance superior to the overall calibration obtained using the full ^{1}H-NMR spectrum.

However, by only considering interval 11 (6.80-6.95 ppm) or interval 31 (2.20-2.40 ppm) the PLS prediction error (RMSECV) drastically decreases to lower than 0.1 in the range from 0 for WT to 1 for the THS-GM1. Moreover, these intervals provide a more parsimonious PLS models as they require only three latent variables. Accordingly, these two intervals contain the primary or best information for distinguishing the metabolome of THS with one transgenic copy and the metabolome of the wild type. The same data mining approach was used for the SIL samples, this time with zero, one or three copies of the gene, and the results are summarized in Fig. 4. The global model obtained using the SIL samples has RMSECV= 0.39 [0; 2] and requires 8 LV. The most informative intervals are the neighboring no. 28 and no. 29 (2.90-2.60 ppm) requiring only four latent variables to develop a regression model with a RMSECV lower than 0.3 [0; 2]. In addition, in order to investigate the perturbation on the grape metabolomes of the two grape varieties, metabolite-metabolite correlations were established and represented using the so-called heat

maps (Fig. 5).

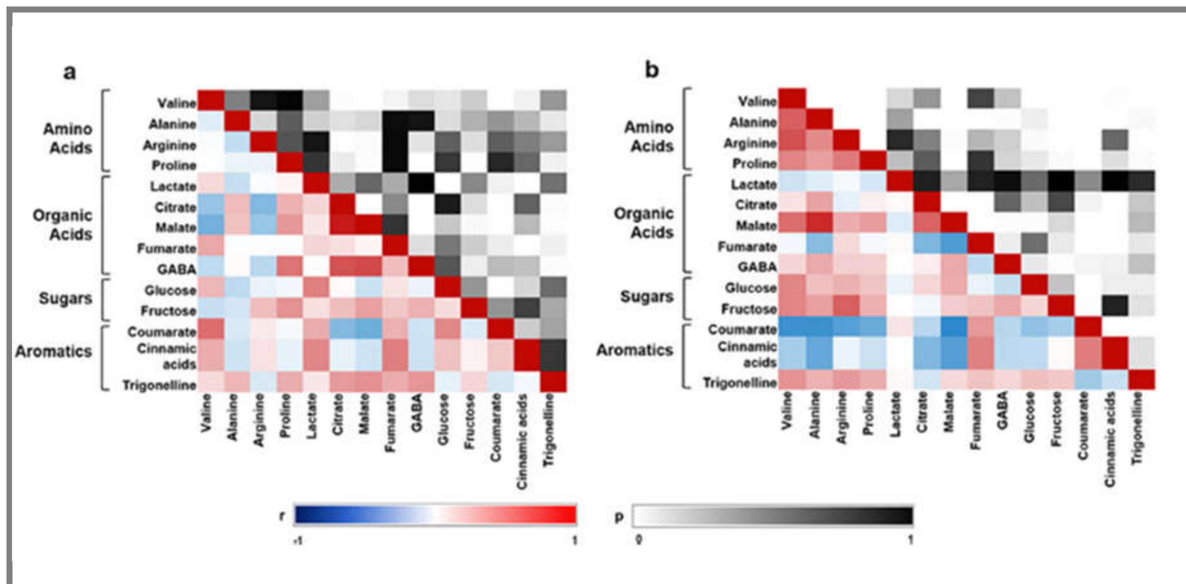


Fig. 5 Heat maps (Pearson Correlation) representing the r coefficients (left side) and the corresponding p values (right side) for the correlation of selected and identified metabolites listed in Table 1: A) using all THS spectra (WT and GM1) and B) using all SIL spectra (WT, GM1 and GM2).

Discussion

Exploratory data analysis

As it can be observed from Fig. 2, the phenolic (Fig. 2a) and the organic acid (Fig. 2c) regions both show some clustering of the genotypes. Curiously, the carbohydrate region (Fig. 2b), which in intensity dominates the spectra, contains no information that can discriminate between the diverse genotypes. This behavior demonstrates that the genetic modifications do not significantly and consistently alter the carbohydrate composition and content of the berries. As the carbohydrate content largely dominates the berries metabolome, a PCA performed on the whole spectral region will yield practically identical result as the carbohydrates region alone (Supplementary Fig. S4). Thus, in order not to overlook minor metabolic differences among the different grape lines, it is of paramount importance that the three spectral regions are considered separately. For both the phenolic and the organic acid regions, the difference between the grape cultivars, THS and SIL, is dominant (along the first principal component), whereas the variation between the transgenic grapes and the WT only accounts for a smaller variation, still mainly along PC1. Inspection of the variance described shows that the PCA of the organic acid region explains the highest percentage of the total systematic variance (92.7%) (Fig. 2c). Scrutinizing this PCA scores plot reveals that the genotypes are clustered according to the number of transgene's copies: THS-WT; THS-GM1 (one copy), SIL-WT; SIL-GM1 (one copy) and SIL-GM2 (three copies). A similar

trend is observed also in the aromatic region but to a lesser extent, as previously demonstrated by Picone et al. (2011). This clustering is evident in the *i*PLS modelling performed separately for each cultivar (Fig. 3 and 4), in which the number of copies of inserted transgene was used as *y* regression vector.

Thompson seedless data mining

In Fig. 3, interval 31 includes both signals belonging to proline and another signal which we were not able to assign, also considering that there is no such reference in literature. In order to verify the predictive power of proline, a PLS-DA was calculated and validated by Cross Validation (Supplementary Fig. S6). Using 3 LVs, a RMSECV of 0.175 and R^2 of 0.8778 were obtained for the obtained model. The concentration of proline decreases from the THS-WT to the THS-GM1. Proline is commonly the predominant free amino acid in grape berry, and its accumulation depends on several factors such as grape berry maturity and cultivar. Moreover, this amino acid appears to accumulate in the berry when this is grown under high salinity and/or water deficit conditions.

Interval 11 (6.80-6.95 ppm) contains signals belonging to *p*-coumaric acid, a hydroxycinnamic acid, which is biosynthesized from cinnamic acid by a hydroxylation controlled by cinnamate-4-hydroxylase. Variations in the concentration of phenolic components occur only around and after *véraison*. As demonstrated by Fernández de Simón et al. (1992), the phenolic composition changes from one rich in hydroxycinnamic acids (HCA), such as caftaric (caffeic acid conjugated with tartaric acid), *p*-coutaric (*p*-coumaric acid conjugated with tartaric acid) and fertaric (ferulic acid conjugated with tartaric acid) acids into another rich in a variety of benzoic and cinnamic acids which are derivatives of HCA. Ali et al. (2010) reported that the HCA concentration depends on different factors including grape variety and growing conditions; it seems likely that also genetic modifications can induce changes in the concentration of these metabolites by up-regulating or down-regulating the production of HCA. The increasing area of the *p*-coumaric acid signals in interval 11, from THS-WT to THS-GM1, appears to substantiate this hypothesis. Also intervals 28 and 29 performed significantly better than the global spectrum but this was more evident for the *i*PLS models obtained for the SIL cultivar and it is discussed in the following.

Silcora data mining

According to the *i*PLS on SIL cultivar (Fig. 4), the two intervals contain the same chemical information since they split the complex NMR signal from the malic acid protons in two. It is therefore not surprising to see that they are performing nearly identically, quantitatively speaking. Indeed, malic acid concentration is found to be significantly higher in wild type samples. As

mentioned in the introduction, malic acid, together with tartaric acid, represents the most abundant organic acid in grape berries and, together, they account for 69–92% of all organic acids. After *véraison* malic acid decreases drastically since its biosynthesis is stopped and the respiration continues to increase. As malic acid is identified as the most discriminating compound, an attempt to model its concentration by means of PLS regression was conducted using the average concentrations of this organic acid as measured by Costantini et al. (2007) in both cultivars and genotypes (except for SIL-GM2 in which it was not measured). The aim was to test the correlation between the concentration of malic acid and the number of copies of the inserted transgene and to evaluate the quantitative effect on the metabolome when multiples copies of the same transgene are inserted (as it was done for SIL). As it can be observed from Fig. 4, the concentration of the malic acid decreases when the number of transgene copies increases; the more the gene copies the lower the concentration of malic acid. This is also demonstrated by a PLS regression performed on the intervals containing the malic acid's signals (Supplementary Fig. S7). The obtained model shows that the difference in concentration (g/l) from THS-WT to THS-GM1 is estimated to be smaller than what reported by Costantini et al. (2007). In particular, the predicted values for the samples THS-WT are dispersed in a range between 3.02 and 3.29 g/l (consistently lower than the previously reported 3.40 average value), which is in perfect agreement with the concentrations showed by the THS-GM1 predicted samples ranging between 2.63 and 2.83 g/l (consistently higher than the previously reported 2.40 average value). For what it concerns the SIL samples, it is possible to test also the hypothesized correlation between malic acid content and number of transgene copies, but it must be remembered that only the average values relative to malic acid concentrations in SIL-WT (1.5 g/l) and GM1 (1.3 g/l) were available and used in the PLS model as actual values. The obtained regression model not only confirms that these values are a good estimate of the average predicted values, but also predicts the amounts to be ascribed to SIL-GM2 samples, containing 3 copies of the transgene, to be, in average, 0.92 g/l. This result is in nearly perfect agreement with a Δ-concentration between SIL-WT and SIL-GM2 being three times lower than that obtained between SIL-WT and SIL-GM1 in which only one copy of the transgene is involved, and thus chemically substantiates the validity of the obtained models.

Metabolite-Metabolite correlation analysis

The two heat maps presented in Fig. 5a and 5b, related to THS and SIL respectively, appear to exhibit a similar type of perturbation. However, the SIL heat map presents both more positive and negative strong correlations. For example, the amino acids are highly positively inter-correlated

and alanine strongly positively correlated to malic acid. In contrast the amino acids are strongly negatively correlated to p-coumaric acid (and partially to cinnamic acid). p-coumaric acid further exhibit a strong negative correlation to malic acid. The THS heat map appear flatter, presumably due to the lower correlation dimension (only WT and GM1) but similar trends as observed for the SIL heat map may be observed. The most noticeable difference is a strong correlation between malic acid and GABA. These results largely confirm the findings from the multivariate statistical analysis (*i*PLS and *i*PLS-DA), showing that the insertion of the transgene caused a decrement in the intensities of malate and proline and an increment in p-coumarate concentration.

Conclusions

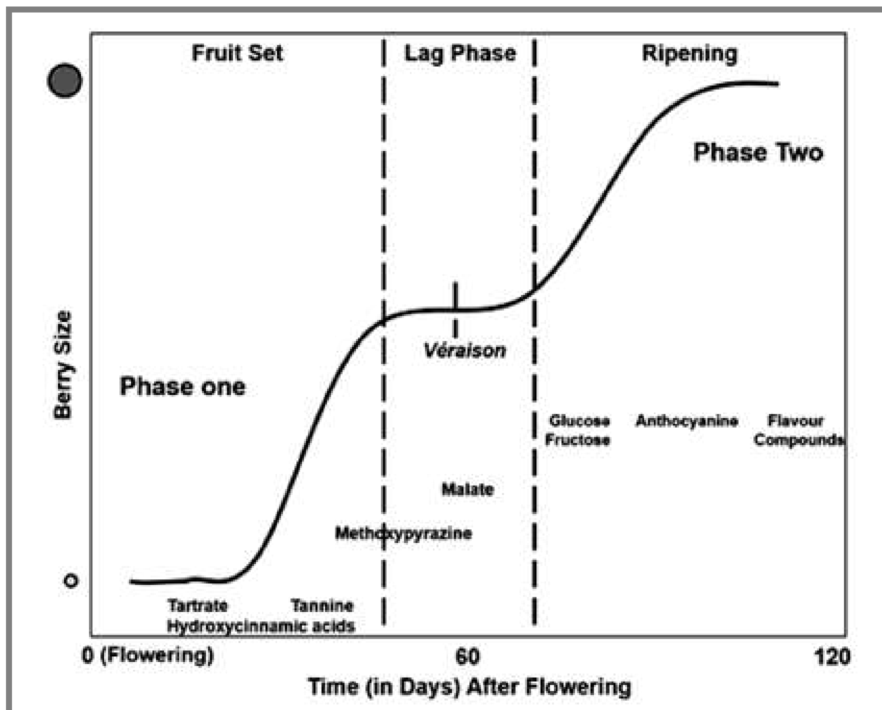
The combination of NMR spectroscopy and chemometrics was able to identify specific metabolites that were up- or down regulated in the genetically engineered plants, and thus able to highlight alterations in the down-stream metabolic pathways of grape berries due to the up-stream genetic modifications. This study shows that plant metabolomics can play an important role as the missing link in functional genomics. The performed study also proves that ^{1}H-NMR spectroscopy is able to measure metabolic differences both among agronomic varieties of grapes as well as between various levels of genetic modification. The results show that the concentration of malic acid is correlated with the gene copy number in each genetically modified grape line, proving that the increased production of berries, promoted by the inserted gene, is achieved at the expense of a greater and greater decrement in malic acid concentration in the berries. Overall, these findings are confirming that the *DefH9-iaaM* gene is inducing positive benefits in table grape production with reduced changes in berry chemical composition and with no negative effects on its sensorial quality, which is in fact associated to compounds influencing the taste, e.g. sugars, acids and tannins.

Compliance with ethical standards

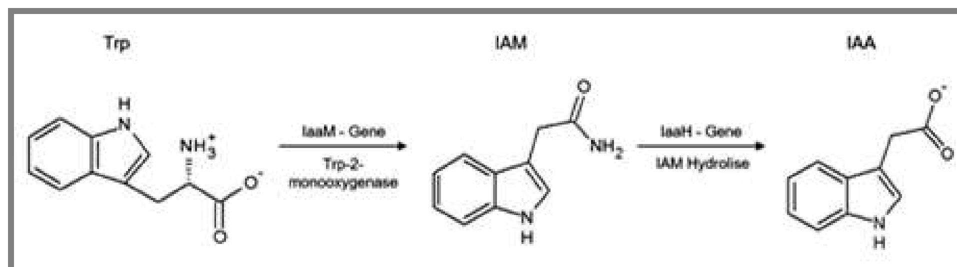
1. *Research involving human participants and/or animals*: This article does not contain any studies with human participants or animals performed by any of the authors.
2. *Informed Consent*: No informed consent was required for this study since no human participant was involved.
3. *Conflict of Interests:* Picone G., Savorani F., Trimigno A., Mezzetti B., Capozzi F. and Engelsen S.B. declare that they have no conflict of interest.

SUPPLEMENTARY MATERIAL

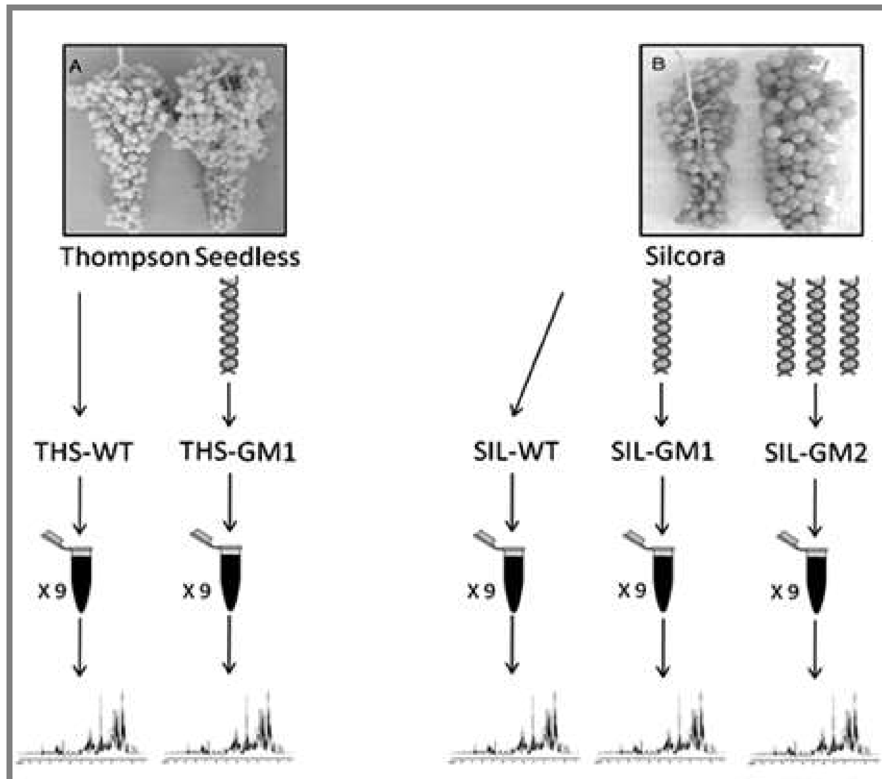
Supplementary Figures



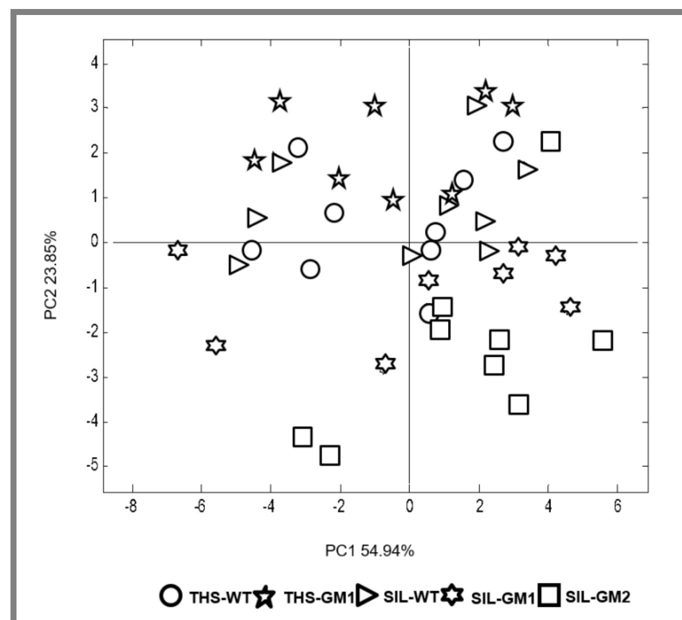
Supplementary Figure S1. Schematic representation of grape berry development. The curve indicates changes in berry size (y-axis) as a function of days after flowering (x-axis). Accumulated compounds during growing and ripening days after flowering are also reported (modified from Ali et al. (2011))



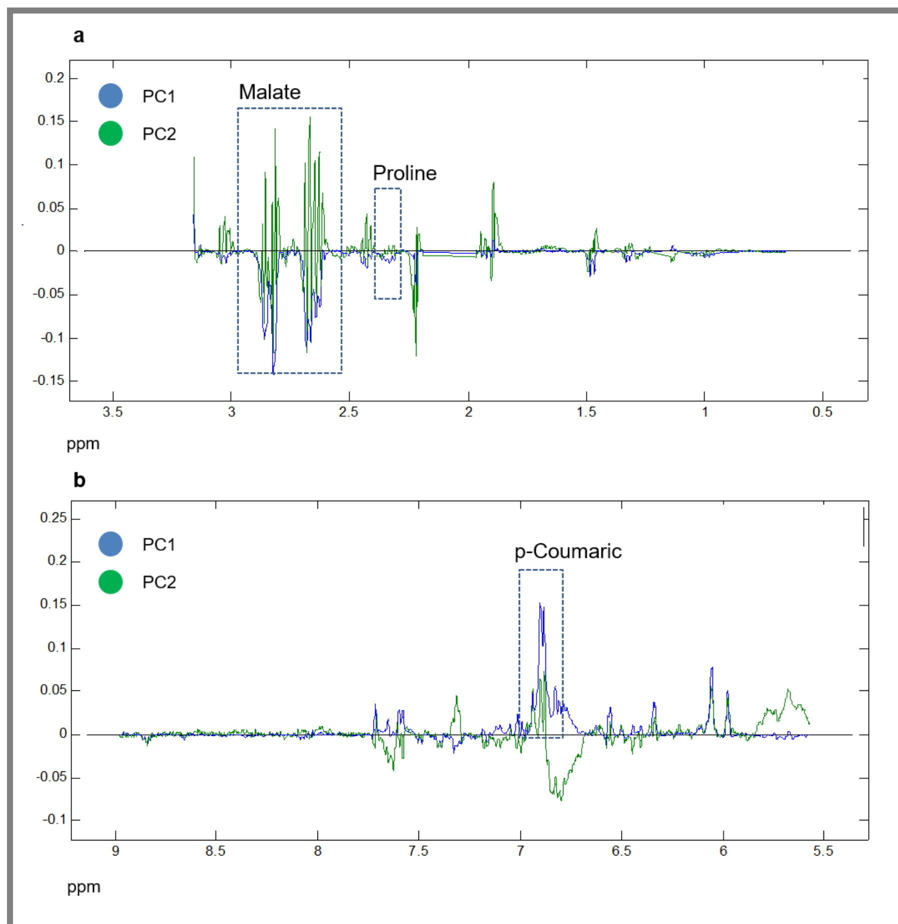
Supplementary Figure S2. IAA synthetic pathway starting from Trp (molecules have been drawn by using ACD/ChemSketch (freeware) 2015.2.5.)



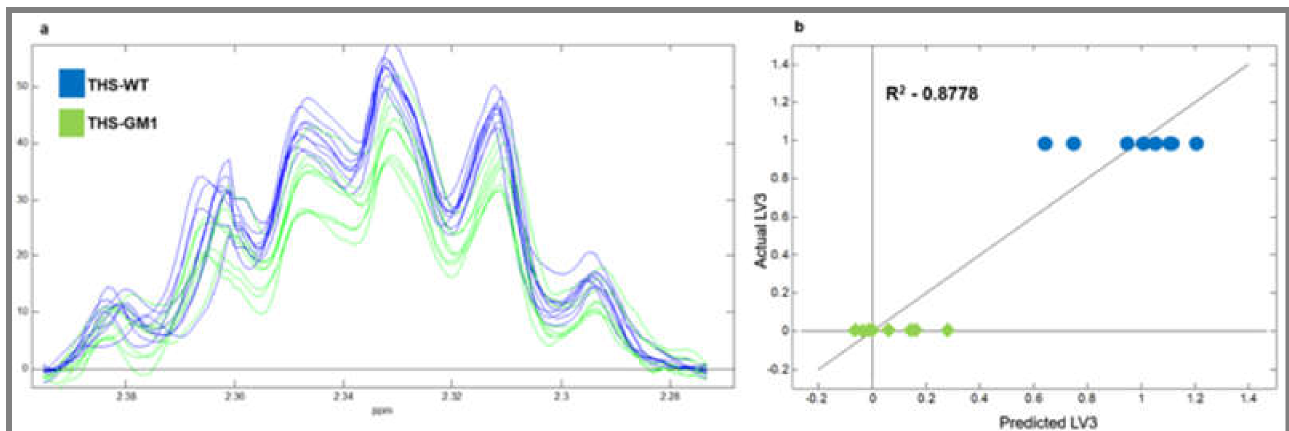
Supplementary Figure S3. Experimental setup. For each genotype, 9 ^{1}H-NMR spectra (3 extractions × 3 replicates) were acquired, adding up to a total of 45 samples



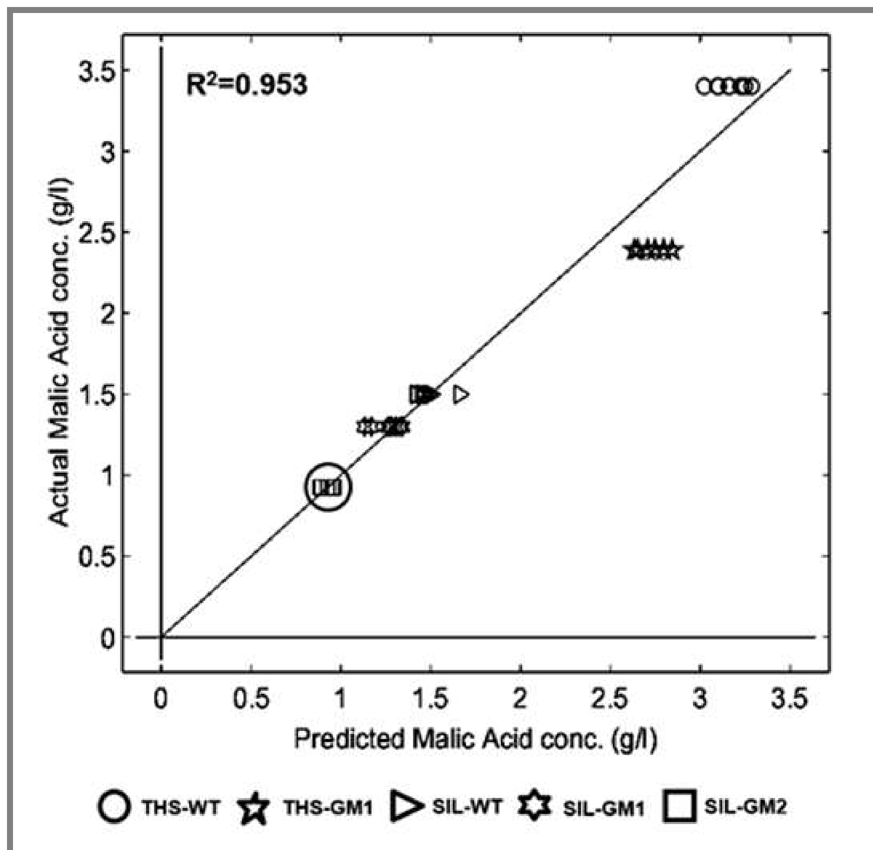
Supplementary Figure S4. PCA scores plot of the entire spectral region



Supplementary Figure S5. Loadings plot for the PCA models shown in Fig. 2; a) Loadings #1 and #2 of the organic acids region; b) Loadings #1 and #2 of the aromatic acids region. The most discriminant metabolites are highlighted



Supplementary Figure S6. a) Superimposition of the THS spectra showing the proline signal NMR region; **b)** PLS-DA actual vs. predicted plot of the model calculated on the proline interval predicting THS-WT and -GM1



Supplementary Figure S7. PLS actual vs predicted cross validated plot calculated using only the NMR spectral intervals containing the malic acid signal (ints. 28-29). A nearly perfect correlation between the NMR estimated values and the chemical analysis is obtained (R^2=0.953). The values for the circled samples, referring to SIL-GM2, were predicted through the PLS model

REFERENCE CHAPTER 4.2

Picone, G., Mezzetti, B., Babini, E., Capocasa, F., Placucci, G., & Capozzi, F. (2011). "Unsupervised principal component analysis of NMR metabolic profiles for the assessment of substantial equivalence of transgenic grapes (Vitis vinifera)." Journal of agricultural and food chemistry, 59(17), 9271-9279.

Adamski, J., & Suhre, K. (2013). Metabolomics platforms for genome wide association studies—linking the genome to the metabolome. *Current opinion in biotechnology, 24*(1), 39-47, doi:10.1016/j.copbio.2012.10.003.

Ali, K., Maltese, F., Choi, Y. H., & Verpoorte, R. (2010). Metabolic constituents of grapevine and grape-derived products. *Phytochemistry Reviews, 9*(3), 357-378, doi:10.1007/s11101-009-9158-0.

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

Anastasiadi, M., Zira, A., Magiatis, P., Haroutounian, S. A., Skaltsounis, A. L., & Mikros, E. (2009). 1H NMR-Based Metabonomics for the Classification of Greek Wines According to Variety, Region, and Vintage. Comparison with HPLC Data. *Journal of Agricultural and Food Chemistry, 57*(23), 11067-11074, doi:10.1021/jf902137e.

Balmer, D., Flors, V., Glauser, G., & Mauch-Mani, B. (2014). Metabolomics of cereals under biotic stress: current knowledge and techniques. *Induced plant responses to microbes and insects*, 146, doi:10.3389/fpls.2013.00082.

Cheynier, V., Duenas-Paton, M., Salas, E., Maury, C., Souquet, J.-M., Sarni-Manchado, P., et al. (2006). Structure and Properties of Wine Pigments and Tannins. *American Journal of Enology and Viticulture, 57*(3), 298-305.

Conde, C., Silva, P., Fontes, N., Dias, A. C. P., Tavares, R. M., Sousa, M. J., et al. (2007). Biochemical changes throughout grape berry development and fruit and wine quality. Global Science Books.

Costantini, E., Landi, L., Silvestroni, O., Pandolfini, T., Spena, A., & Mezzetti, B. (2007). Auxin Synthesis-Encoding Transgene Enhances Grape Fecundity. *Plant Physiology, 143*(4), 1689-1694, doi:10.1104/pp.106.095232.

Craig, A., Cloarec, O., Holmes, E., Nicholson, J. K., & Lindon, J. C. (2006). Scaling and Normalization Effects in NMR Spectroscopic Metabonomic Data Sets. *Analytical Chemistry, 78*(7), 2262-2267, doi:10.1021/ac0519312.

Dai, Z. W., Léon, C., Feil, R., Lunn, J. E., Delrot, S., & Gomès, E. (2013). Metabolic profiling reveals coordinated switches in primary carbohydrate metabolism in grape berry (Vitis vinifera L.), a non-climacteric fleshy fruit. *Journal of Experimental Botany, 64*(5), 1345-1355, doi:10.1093/jxb/ers396.

Davis, C., & Robinson, S. (1996). Sugar accumulation in grape berries. *Plant Physiology, 111*, 275-283, doi:http://dx.doi.org/10.1104/pp.111.1.275.

EFSA (2006). Guidance document of the Scientific Panel on Genetically Modified Organisms for the risk assessment of genetically modified plants and derived food and feed. *EFSA Journal, 99*, 1-100.

Fernández de Simón, B., Hernández, T., Estrella, I., & Gómez-Cordovés, C. (1992). Variation in phenol content in grapes during ripening: low-molecular-weight phenols. *European Food Research and Technology, 194*(4), 351-354.

Gallo, V., Mastrorilli, P., Cafagna, I., Nitti, G. I., Latronico, M., Longobardi, F., et al. (2014). Effects of agronomical practices on chemical composition of table grapes evaluated by NMR spectroscopy. *Journal of Food Composition and Analysis, 35*(1), 44-52, doi:10.1016/j.jfca.2014.04.004.

Garrido, J., & Borges, F. (2013). Wine and grape polyphenols—A chemical perspective. *Food research international, 54*(2), 1844-1858, doi:10.1016/j.foodres.2013.08.002.

Gerós, H., Chaves, M., & Delrot, S. (2012). *The Biochemistry of the Grape Berry*: Bentham Science Publishers.

Griffin, J. L. (2006). The Cinderella story of metabolic profiling: does metabolomics get to go to the functional genomics ball? *Philosophical Transactions of the Royal Society B: Biological Sciences, 361*(1465), 147-161, doi:10.1098/rstb.2005.1734.

Hale, C. R. (1962). Synthesis of Organic Acids in the Fruit of the Grape. *Nature, 195*(4844), 917-918, doi:10.1038/195917a0.

Hall, R., Beale, M., Fiehn, O., Hardy, N., Sumner, L., & Bino, R. (2002). Plant metabolomics the missing link in functional genomics strategies. *The Plant Cell, 14*(7), 1437-1440, doi: http://dx.doi.org/10.1105/tpc.140720.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology, 24*(7), 417-441, doi:http://dx.doi.org/10.1037/h0071325.

Inouye, M., Kettunen, J., Soininen, P., Silander, K., Ripatti, S., Kumpula, L. S., et al. (2010). Metabonomic, transcriptomic, and genomic variation of a population cohort. *Molecular System Biology, 6*(1), 1-10, doi:10.1038/msb.2010.93.

James A, K., Gordon J, T., John R, P., Donald R, H., David, H., Charles R, H., et al. (2000). Development of seed polyphenols in berries from *Vitis vinifera* L. cv. Shiraz. *Australian Journal of Grape and Wine Research, 6*(3), 244-254, doi:10.1111/j.1755-0238.2000.tb00185.x.

Kliewer, W. M. (1966). Sugars and Organic Acids of Vitis vinifera *Plant Physiology, 41*(6), 923–931, doi:http://dx.doi.org/10.1104/pp.41.6.923.

Laghi, L., Versari, A., Marcolini, E., & Parpinello, G. P. (2014). Metabonomic investigation by 1H-NMR to discriminate between red wines from organic and biodynamic grapes. *Food and Nutrition Sciences, 5*, 52-59, doi:10.4236/fns.2014.51007.

López-Rituerto, E., Savorani, F., Avenoza, A., Busto, J. H., Peregrina, J. M., & Engelsen, S. B. (2012). Investigations of La Rioja terroir for wine production using 1H NMR metabolomics. *Journal of Agicultural and Food Chemistry, 60*(13), 3452−3461, doi:http://dx.doi.org/10.1021/jf204361d.

Mezzetti, B., Pandolfini, T., Navacchi, O., & Landi, L. (2002). Genetic transformation of *Vitis vinifera* via organogenesis. *BMC Biotechnology, 2*(18), 1-10, doi:10.1186/1472-6750-2-18.

Mohanta, T. K. (2013). Plant metabolomics: missing link in next generation functional genomics era. *Journal of Applied Biology & Biotechnology, 1*, 001-010, doi:10.7324/JABB.2013.1301.

Mulas, G., Galaffu, M. G., Pretti, L., Nieddu, G., Mercenaro, L., Tonelli, R., et al. (2011). NMR Analysis of Seven Selections of Vermentino Grape Berry: Metabolites Composition and Development. *Journal of Agricultural and Food Chemistry, 59*(3), 793-802, doi:10.1021/jf103285f.

Nørgaard, L., Saudland, A., Wagner, J., Nielsen, J. P., Munck, L., & Engelsen, S. B. (2000). Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. *Applied Spectroscopy, 54*(3), 413-419.

Oliveira, J., de Freitas, V., Silva, A. M. S., & Mateus, N. (2007). Reaction between Hydroxycinnamic Acids and Anthocyanin - Pyruvic Acid Adducts Yielding New Portisins. *Journal of Agricultural and Food Chemistry, 55*(15), 6349-6356, doi:10.1021/jf070968f.

Pereira, G. E., Gaudillere, J.-P., Leeuwen, C. v., Hilbert, G., Maucourt, M., Deborde, C., et al. (2006). 1H NMR metabolite fingerprints of grape berry: Comparison of vintage and soil effects in Bordeaux grapevine growing areas. *Analytica Chimica Acta, 563*(1-2), 346-352, doi:10.1016/j.aca.2005.11.007.

Picone, G., Laghi, L., Gardini, F., Lanciotti, R., Siroli, L., & Capozzi, F. (2013). Evaluation of the effect of carvacrol on the Escherichia coli 555 metabolome by using 1 H-NMR spectroscopy. *Food Chemistry, 141*(4), 4367-4374,

doi:10.1016/j.foodchem.2013.07.004.

Picone, G., Mezzetti, B., Babini, E., Capocasa, F., Placucci, G., & Capozzi, F. (2011). Unsupervised Principal Component Analysis of NMR metabolic profiles for the assessment of substantial equivalence of transgenic grapes (Vitis vinifera). *Journal of Agricultural and Food Chemistry, 59*(17), 9271-9279, doi:10.1021/jf2020717.

Romeyer, F. M., Macheix, J. J., Goiffon, J. J., Reminiac, C. C., & Sapis, J. C. (2002). Browning capacity of grapes. 3. Changes and importance of hydroxycinnamic acid-tartaric acid esters during development and maturation of the fruit. *Journal of Agricultural and Food Chemistry, 31*(2), 346-349, doi:10.1021/jf00116a040.

Rotino, G. L., Perri, E., Zottini, M., Sommer, H., & Spena, A. (1997). Genetic engineering of parhenocarpic plants. *Nature Biotechnology, 15*, 1398-1320, doi:10.1038/nbt1297-1398.

Saito, K., & Kasai, Z. (1968). Accumulation of tartaric acid in the ripening process of grapes. *Plant Cell Physiology, 9*(3), 529-537.

Savage, A. K., Tucker, G., van Duynhoven, J. P. M., Wulfert, F., & Daykin, C. A. (2009). Nutrimetabolomics: development of a bio-identification toolbox to determine the bioactive compounds in grape juice. *Bioanalysis, 1*(9), 1537-1549, doi:10.4155/bio.09.147.

Savorani, F., Picone, G., Badiani, A., Fagioli, P., Capozzi, F., & Engelsen, S. B. (2010). Metabolic profiling and aquaculture differentiation of gilthead sea bream by 1H NMR metabonomics. *Food Chemistry, 120*(3), 907-914 doi:10.1016/j.foodchem.2009.10.071.

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, doi:10.1016/j.jmr.2009.11.012.

Schlosser, J., Olsson, N., Weis, M., Reid, K., Peng, F., Lund, S., et al. (2008). Cellular expansion and gene expression in the developing grape (*Vitis vinifera* L.). *Protoplasma, 232*(3), 255-265, doi:10.1007/s00709-008-0280-9.

Scorza, R., Cordts, J., Gray, D., Gonsalves, D., Emershad, R., & Ramming, D. (1996). Producing TransgenicThompson Seedless' Grape (Vitis vinifera L.) Plants. *Journal of the American Society for Horticultural Science, 121*(4), 616-619.

Simó, C., Ibáez, C., Valdés, A., Cifuentes, A., & García-Cañas, V. (2014). Metabolomics of genetically modified crops. *International journal of molecular sciences, 15*(10), 18941-18966, doi:10.3390/ijms151018941.

Son, H.-S., Hwang, G.-S., Kim, K. M., Ahn, H.-J., Park, W.-M., Van Den Berg, F., et al. (2009). Metabolomic Studies on Geographical Grapes and Their Wines Using 1H NMR Analysis Coupled with Multivariate Statistics. *Journal of Agricultural and Food Chemistry, 57*(4), 1481-1490, doi:10.1021/jf803388w.

Stines, A. P., Grubb, J., Gockowiak, H., Henschke, P. A., Høj, P. B., & Heeswijck, R. (2000). Proline and arginine accumulation in developing berries of *Vitis vinifera* L. in Australian vineyards: Influence of vine cultivar, berry maturity and tissue type. *Australian Journal of Grape and Wine Research, 6*(2), 150-158, doi:10.1111/j.1755-0238.2000.tb00174.x.

Zhao, Y. (2010). Auxin Biosynthesis and Its Role in Plant Development. *Annual Review of Plant Biology, 61*, 49-64, doi: 10.1146/annurev-arplant-042809-112308.

Zhao, Y. (2014). Auxin Biosynthesis. *The Arabidopsis Book / American Society of Plant Biologists, 12*, e0173, doi:10.1199/tab.0173.

CHAPTER 5: FOODOMICS TO INVESTIGATE THE "MATRIX EFFECT"

Food industries are nowadays facing increasing demands from customers and consumers, who are requesting more and more beneficial effects associated to food products. [Kilcast & Subramaniam, 2000]. One of the greatest challenge and trend in the last decades is, in effect, the production of food products enriched with bioactive compounds exerting positive actions on health. This was driven by recent discoveries on the preventive and beneficial potential of foods and, in some cases, individual molecules isolated from them [Sinn et al., 2012; Samuelsen et al., 2014; Yousouf et al., 2015]. After the assessment of the biological function of specific bioactive compounds, the latter are included in foods with a widespread consumption, in order to favour their regular intake [Ganesh et al., 2015; Augustin & Sanguansri, 2015; Li-Chan, 2015]. The main issue with these new enriched products is the investigation on their residual effect when embedded in complex food matrices [Yang et al., 2011; Moro et al., 2015], which modulate the bio-accessibility and consequently the bio-availability of the beneficial molecules.

In addition, these molecules are generally added not in their pure form, but employed as food-grade ingredients (i.e. plant extracts, egg products enriched with PUFA, from hens fed with enriched feeds, etc.). These ingredients are commonly added in high percentages and, consequently, may impact the stability of the foodstuff and the release of bio-active molecules during the shelf-life of the product. For these reasons, the assessment of the matrix stability should be a mandatory task during the formulation of enriched food products. In order to investigate the characteristic of the matrix before and after the enrichment with bioactive compounds, several techniques were traditionally employed. The first aspect that must be taken into consideration is the structure of the matrix itself. The knowledge about the specific interactions between the ingredients added and the basic food matrix can drive the formulation of specific products, starting from the selection of the best matrices for each type of enrichment. Traditional methods usually involves image techniques like light or electron microscopy. Other techniques focused on the assessment of functional properties due to the matrix' structure, like rheological studies. [Heertje, 2014]. The other important thing to know is the total content of bioactive compounds contained in the food matrix. These studies are in effect very useful also to understand whether the structure and the characteristics of the matrix protect the bioactive compounds added during storage. The employed methods are usually destructive, with different level of aggressiveness towards the food matrix. In order to obtain a complete recover of the bioactive molecule, in effect, the matrix needs

to be modified and, in most cases, destroyed. For these reasons, no real information on the matrix and its effect on bio-accessibility is obtained, apart from its protective effect in case the quantity of bioactive recovered is the same as its initial quantity. There are no direct molecular methods for this type of assessment, while the indirect techniques concentrate on the evaluation of different characteristics, such as the (1) water status, (2) release of bioactive compound upon in-vitro simulated digestion of the matrix, (3) release of soluble components by dispersing the matrix in aqueous media.

Water is usually one of the main components of foods and it will thus greatly influence structure and physico-chemical properties of the food matrix, therefore, the knowledge of its normal behaviour and its variation is of great importance when evaluating food stability. The main techniques employed are: the evaluation of the water activity (a_w), the assessment of the glass transition temperature (T_g) and most recently the measurement of water proton relaxation time through NMR spectroscopy [Mc Clements, 2015]. Other methods increasingly carried out, employ *in vitro* and *in vivo* digestion techniques and are very useful for the determination of how the molecules are released and, eventually, absorbed during the digestive processes and how the complex structure of the food product and the processes it has been through influence this behaviour [Turgeon et al., 2011; Norton et al., 2014; Fundo et al., 2015]. Finally, other destructive techniques focus on the study of the soluble molecules that can be released by the matrix. The molecular method employed to assess the matrix stability should not introduce artefacts due to the solvent employed for extraction: for this reason the analysis should be conducted on media, which simulate as much as possible the physiological environment of the fluid that makes the nutrients and the other bioactive molecules accessible. The composition of such fluid must be determined in order to find variation in the concentration of the soluble fraction of the matrix as affected by the enrichment. Then, the formation of new solutes or the consumption of some other expected molecules during the food storage, due, respectively, to hydrolysis of or to condensation with the solid phase, could be used as a mirror of changes in the matrix itself. The analytical technique, must be able to detect all soluble products in an extent directly proportional to the mass of product formed or consumed, and not dependent on the response factors of the detection.

The traditional techniques employed generally focused on the identification and quantification of specific components of the matrix [López-Córdoba et al., 2014; Salcedo-Sandoval et al., 2015], or on the determination of specifically related properties [Dave et al., 2014; Edel et al., 2015], but not

on the whole picture of the food matrix. Nowadays, though, new molecular methods capable of giving information on the total soluble profile are available. One of these techniques is ^{1}H-NMR spectroscopy, capable of giving an insight on how the matrix impacts on the bioaccessibility of the bioactive molecules. This is made possible by the employment of an extracting fluid simulating the physiological compartment of our body in which the bioactive compound becomes available. This can be considered a different form of in-vitro digestion since here the solid matrix is simply extracted by dispersion in a fluid, in the easiest way. The extract can then be measured by NMR spectroscopy many times. When samples are already in a solution, there is no need for the extraction step, thus in this case samples are not destroyed. The NMR technique has the advantage of supplying detailed information on the molecular structure of the biological material observed. It can register the complexity of a mixture since every component contains at least one visible hydrogen atom and thus can generate signals on the NMR spectrum that form the unique pattern specific for that molecule. For this reason, in a typical sample, almost all metabolites will give rise to signals in an ^{1}H NMR spectrum, as long as they are present in concentrations above the detection limit. In this way, the NMR spectrum results to be the superposition, also commonly called "fingerprint", of the spectra of all soluble metabolites in the sample. One of the advantages of NMR is that the sample requires minimal treatment prior to the analysis [Laghi et al., 2014], making it easier and quicker, and this is what is required by food industries. Even in the case where the ingredients of the matrix are oxidised during storage and the bioactive compounds might be polymerised and condensed, the NMR analysis will detect a change between spectra at different time-points of storage [Ciampa et al., 2012], even though the specific molecules will become undetectable in the latter time-point due to their new insolubility. Since the NMR technique is sensitive to all these small changes in the matrix, it is clear how it can provide a useful insight into the investigation of the stability of the product during storage. The matrices investigated in this study were thus analysed by means of NMR spectroscopy, with the purpose of achieving a more global overview during shelf-life. In this study, at first, three enriched products were overwied and investigated. The molecular profile of three foods (Pancake, Biscuit and Milkshake powder), enriched with two combination of three different bioactives (docosahexaenoic acid and anthocyanins, DHA+AC, and docosahexaenoic acid and beta-glucans, DHA+BG), was in fact evaluated by high-resolution proton nuclear magnetic resonance (HR-^{1}HNMR). After this, a more specific analysis of the enriched pancakes and their placebo counterparts was carried out. The products were selected on the basis of their frequency of consumption in Europe, and as

representative of a food category with peculiar matrix effects. Pancakes have, in effect, a high protein content, giving origin to an entrapping wet network during cooking, that must be destroyed during digestion in order to release the bioactive compounds.

The NMR spectra of the selected products were acquired and compared at different time-points during storage. Since no previous methodology was defined, this work tried to write up a protocol for the definition of the parameters that could be employed to measure the similarities and differences among spectral profiles of the matrix during shelf-life.

5.1 Investigation on the stability of three different food matrices

Samples and storage designs

Milkshakes samples were enriched with BG-DHA, whilst cookies and pancakes were enriched with AC-DHA. Bioactive enriched foods (BEF) were compared to placebo (PLA) counterparts. Four sampling times were selected: at the beginning of storage (T0) and after 20, 40 and 60 days of storage (T20, T40, T60). The first sampling point was the day when samples arrived in the laboratory. Then, samples were stored at -20°C and sampled after 20, 40 and 60 days from the day of their arrival.

Sample preparation and NMR analysis

For pancakes and milkshakes, at each sampling time a trichloroacetic acid extraction (TCA) was performed on each sample: for this purpose, 5 g of product were added to 15 mL of 7% (w/w) TCA and minced by means of a vertical homogenizer (Ultra-Turrax, Ika®). The resulting product was filtered with filter paper (No. 4) from Whatman (Little Chalfont, Buckinghamshire, HP7 9NA, UK). The pH of a 850 µL aliquot was adjusted to 7.5 using 9 M KOH in an Eppendorf microfuge tube and centrifuged at 14k rpm for 5 min in order to remove the potassium trichloroacetate precipitate. For biscuits samples, a different protocol was adopted, since samples showed an acid hydrolysis during time. The TCA solution was substituted by a phosphate buffer 50 mM, pH 7.4. The so obtained supernatant was stored at −80 °C until ^{1}H-NMR measurements were performed. The samples were prepared for NMR analysis by adding 150 µL of a D_2O and 10 µL solution of 3-(trimethylsilyl)-propionic-2,2,3,3-d4 acid sodium salt (TSP) 100 mM to the thawed samples. ^{1}H- Samples were analysed by applying the NOESY 1D pulse sequence. NMR spectra were recorded at 298 K with a Bruker (Milano, Italy) AVANCE spectrometer, operating at a frequency of 600.13 MHz, equipped with an autosampler with 60 holders. Each spectrum was acquired using 32K data points over a 7211.54 Hz spectral width and adding 256 transients. A recycle delay of 5 s and a 90° pulse of 11.4

µs were set up. Acquisition time (2.27 s) and recycle delay were adjusted to be 5 times longer than the T1 of the protons under investigation, which was considered not to be longer than 1.4 s. Saturation of residual water signal was achieved by irradiating it during the recycle delay at δ equal to 4.703 ppm. Each spectrum was processed with Top Spin 3.0 (Bruker) by using an automatic command apk0.noe, which performs in one shot the baseline and phase correction and by applying a line-broadening factor of 0.3 Hz. The peaks were assigned by comparing their chemical shift and multiplicity with the literature and by using Chenomx NMR suit 8.1 software.

Pre-processing

After Fourier Transformation, phase and baseline correction, spectra were calibrated with reference to the chemical shift of 0.0 ppm assigned to the internal standard TSP, spectral peripheral regions and the water signal were removed. After this, spectra were normalized referring to the anomeric signals of sugars: sucrose in the case of biscuits, lactose in the case of milkshakes. After normalization and prior to any possible statistical analysis, spectra are binned into intervals of 100 data-points. As a result, the new spectral profile consisted of 655 data-points instead of 65535.

Results and discussion

Milkshake

Enriched samples show different signals in milkshake spectra, as shown in the example in Figure 1.

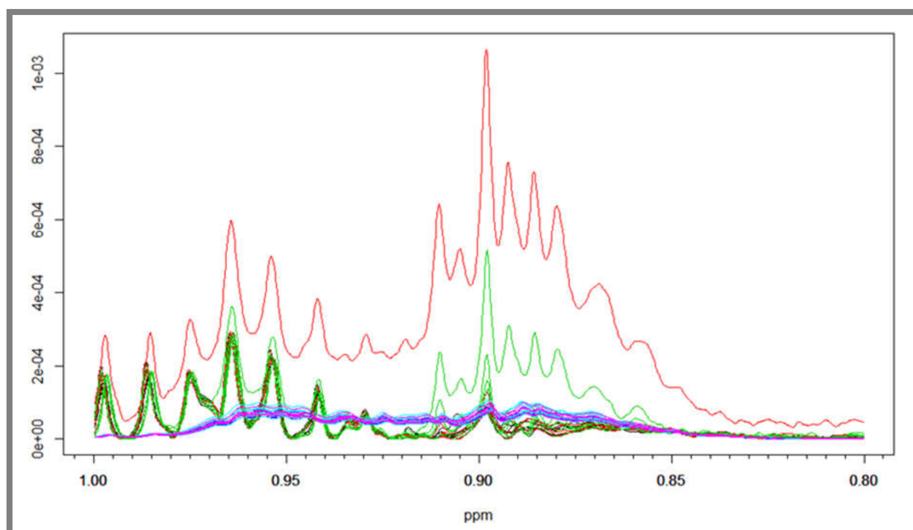


Fig. 1: T0 NOESY1D NMR spectra of enriched (black, red and green) and placebo (blue, light-blue and purple) samples from the first batch. As visible, enriched spectra show many different signals in comparison to placebo samples and also seem to be less reproducible.

Apart from the sugar region, enriched samples all show many different signals not visible in

placebo spectra. These signals are mainly due to organic acids, amino acids (as in Fig. 1) and to aromatic molecules.

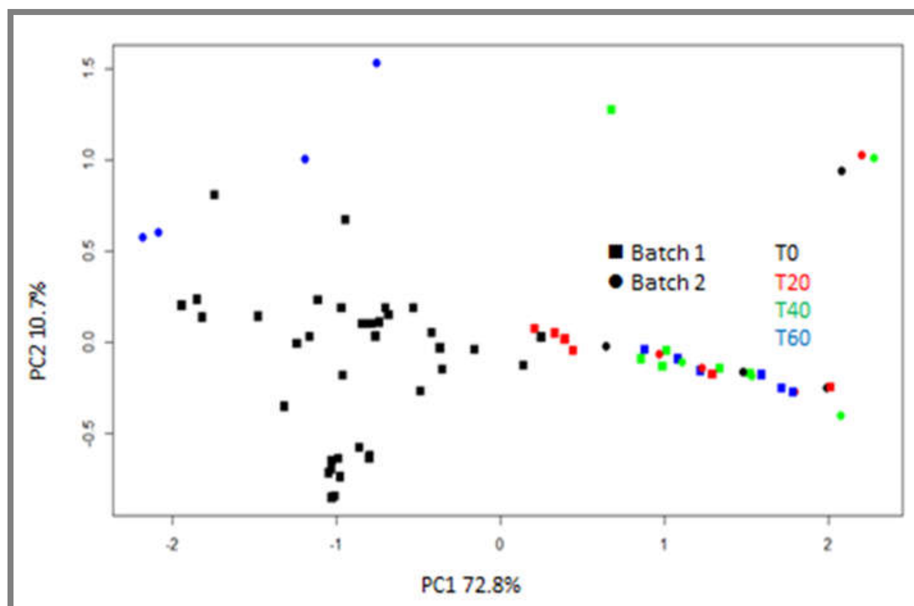


Fig. 2: PCA scoreplot calculated for NOESY1D NMR spectra of enriched and placebo milkshake samples from the first two batches during the 4 different time points.

When samples from different time-points and different batches are evaluated through PCA, it is clear how there are many differences among samples and not mainly caused by the shelf-life itself. As visible in Figure 2, samples from different time-points later than T0 tend to cluster, though with no specific storage-related behavior. Samples from the first batch at T0 seem to differ from any other sample and samples from the second batch at T60 move far from any other type of sample at later storage points and closer to samples at T0 from the first batch.

This behavior might just prove that a lot of difference is there already in the powders at the beginning of storage and these differences can be attenuate by storage time, though this is not always the case. No real differences between placebo and enriched samples is visible by this analysis. In fact, samples seem to differ from each other independently of the enrichment as visible in Figure 3. This shows that the enrichment has no real effect on the matrix behavior during storage.

Biscuits

From the first visual inspection of biscuits spectra, BEF samples had distinctive signals, though just in the aromatic region, whilst other signals just seemed to have a higher intensity for BEF samples, such as in the aliphatic region. Samples from different batches were investigated during the chosen shelf-life to study their behavior.

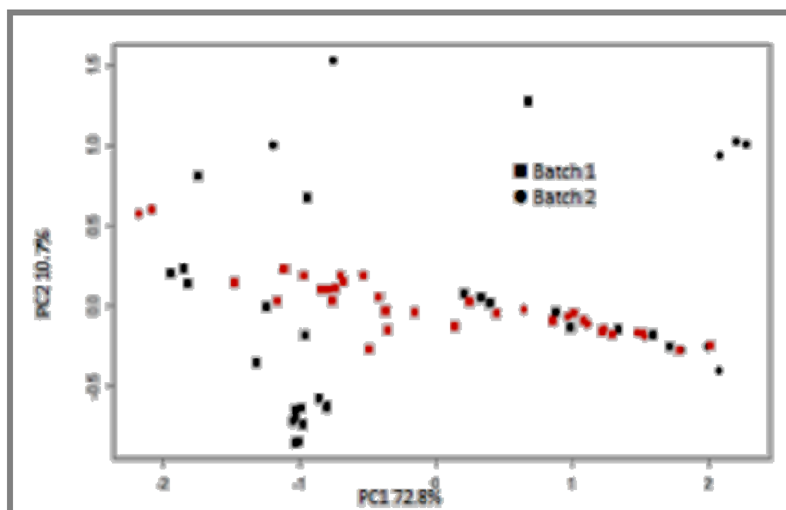


Fig. 3: PCA scoreplot calculated on NOESY1D NMR spectra of enriched and placebo milkshake samples from the first two batches (first represented with squares and second with dots) during the 4 different time points colored according to the matrix composition, PLA in red and BEF in black.

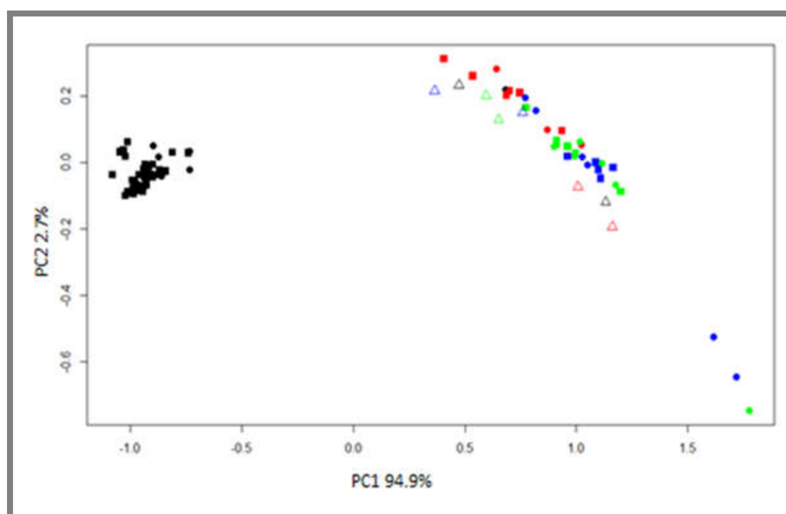


Fig. 4: PCA scoreplot calculated on NOESY1D NMR spectra of BEF and PLA biscuit samples from the first three batches during the 4 different time points. Batches are represented as follows: 1) squares, 2) dots, 3) triangles; time-points are coloured in black fro T0, red for T20, green for T40 and blue for T60. As visible, some samples from the second batch appear to be very different from any other samples and T0 samples from the third batches seemed more similar to the other samples at later shelf-life stages.

PCA was calculated on the spectra after normalization and binning at it showed some interesting results (Fig. 4). Difference between batches and due to the storage behavior generated more differences among samples than BEF vs PLA, as also visible in the zoomed region shown in Figure 5. The main differences seen in spectra causing the groupings in the PCA scoreplot are investigated through the PCA loadings. These show that some differences exist in the sugar region, where signals change in shape and position during storage time. As it can be seen from Figure 6 for the samples of the second batch, the signal around 3.83 ppm changes greatly from T0 (black) to the

later shelf-life stages. The two T60 samples and one T40 sample showing as outliers in the PCA scoreplot are the ones here showing other differences in the signal (lower intensity at 3.83 ppm and different chemical shift also at 3.76-3.8 ppm). One T0 sample, instead, seems more similar to T20 and T40 spectra and this, again, can be seen in the PCA scoreplot shown before.

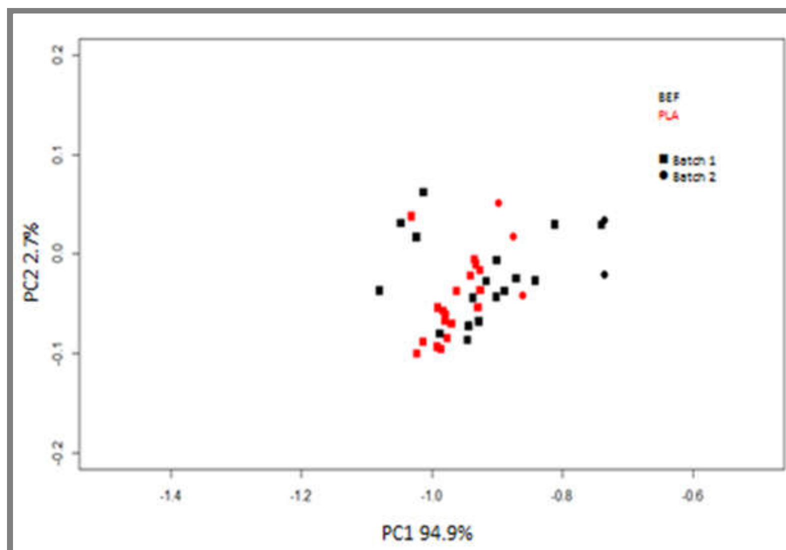


Fig. 5: Zoomed PCA scoreplot to show that there is a very little difference between BEF (black) and PLA (red) biscuit samples. Two batches are shown: first one as squares and second one as dots.

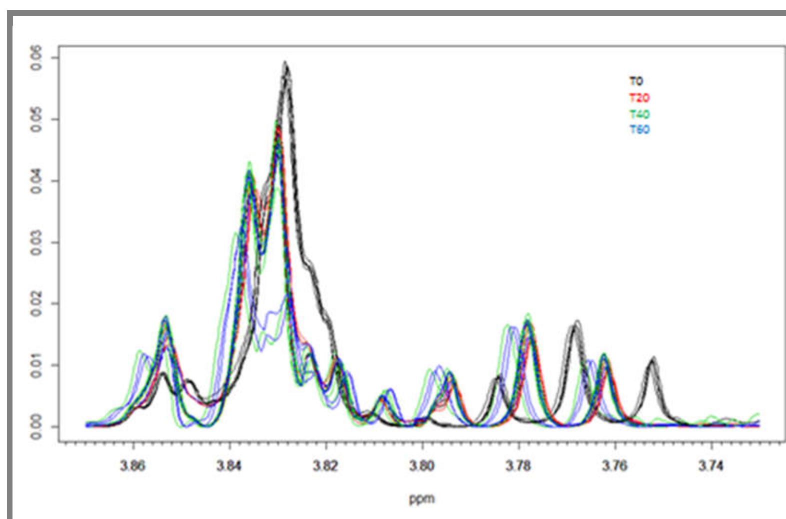


Fig. 6: Part of the sugar region of the NOESY1D NMR spectra for samples from the second batch. Time-points are colored as follows: T0 in black, T20 in red, T40 in green and T60 in blue.

All other samples in later shelf-life stages from batch 1 and batch 3 show a spectra very similar to that of batch 2 samples at T20, and T0 samples of batch 3 also have a this kind of signal shape, showing how they are more similar to samples at a higher shelf-life level. These alterations in the sugar signals might be due to structural changes in the molecules that occur during storage (conformational or configurational changes of monosaccharides, different linkages, etc.).

Pancakes

Pancake samples were analyzed during the whole storage time for one particular batch. Other samples were available from other batches, but it was immediately clear from PCA that the samples were different and could not be considered together for the analysis of storage behavior. For this reason, we focused on the assessment of the matrix behavior during storage only on the aforementioned batch.

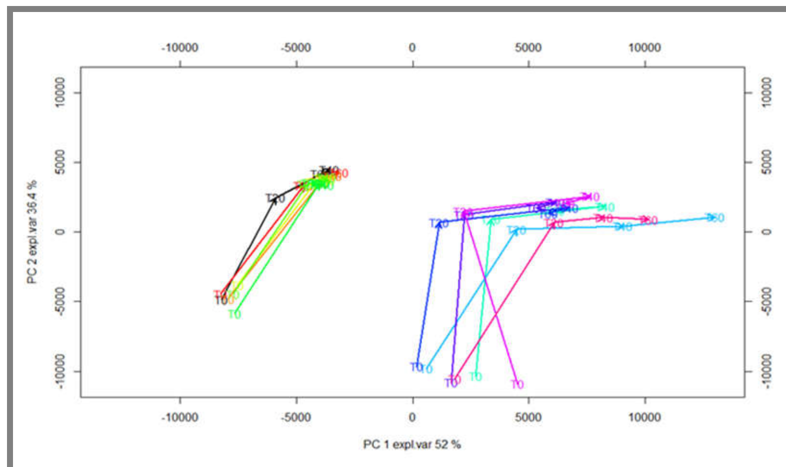


Fig. 7: PCA of pancakes, BEF on the left side and PLA on the right side, showing the behavior during storage.

As visible from the PCA scoreplot shown in Figure 7, BEF samples seemed to be more reproducible and more stable during storage. This image shows how the PLA counterparts are more spread out in the plot and move more along both PC axes during storage. This means that PLA samples show a lot of difference among them from the beginning and then are affected by greater changes in the matrix composition during shelf-life.

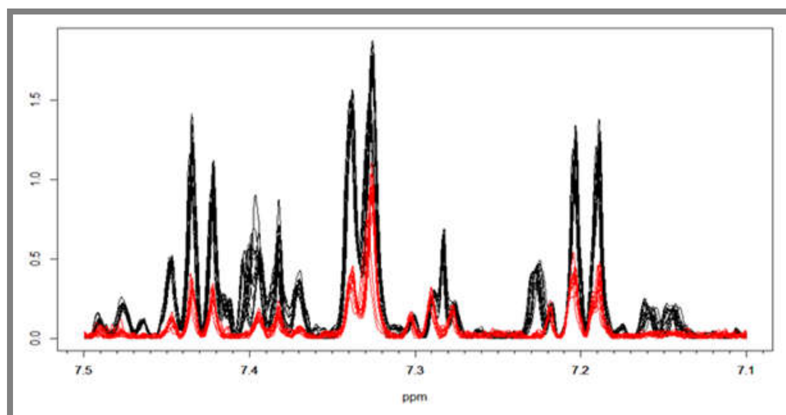


Fig. 8: Some differences in the aromatic region observed for BEF (black) and PLA (red) NOESY1D NMR spectra.

As visible in Figure 8, BEF samples showed a higher concentration of metabolites in the aromatic region at the end of the storage period. This was observed in all time points, though the profile of

both types of samples slightly changed during shelf-life.

Conclusion

These analyses of different batches of pancakes, milkshakes and biscuits have given us a greater insight on the variability and stability of the products. Through the analysis of a batch of pancakes during storage it was proved that BEF samples were more stable and more standardized than their PLA counterparts. In this case, the separation caused by the additional ingredients was greater than the one caused by storage, though batches seemed to be a lot more different between them.
The research on biscuits, instead, has shown how there are more differences caused by storage and, in some part, by the batch, than from the presence or absence of bioactive ingredients.
This shows how the matrix itself has a greater effect in determining the shelf-life behavior for biscuits and the addition of the bioactive compounds does not affect it. For what concerns milkshakes, samples seem to have more differences intrinsically than for any other possible effect (enrichment, batch or storage) and that the enrichment with bioactive compounds does not affect the behavior of the matrix along the shelf-life, being the matrix itself very stable.

5.2 Investigation on the stability of enriched pancakes

Materials and Methods

Spectral acquisition

Five different batches of pancakes were employed for this study in order to test the product reproducibility. Four sampling times were selected: at the beginning of storage (T0) and after 20, 40 and 60 days of storage (T20, T40, T60). At each sampling point, a trichloroacetic acid extraction (TCA) was performed on each sample: for this purpose, 5 g of product were added to 15 mL of 7% (w/w) TCA and minced by means of a vertical homogenizer (Ultra-Turrax, Ika®). The resulting product was filtered with filter paper (No. 4) from Whatman (Little Chalfont, Buckinghamshire, HP7 9NA, UK). The pH of a 850 µL aliquot was adjusted to 7.5 using 9 M KOH in an Eppendorf microfuge tube and centrifuged at 14 K rpm for 5 min in order to remove the potassium trichloroacetate precipitate. The so obtained supernatant was stored at −80 °C until ^{1}H-NMR measurements were performed [Picone et al., 2011]. The samples were prepared for NMR analysis by adding 150 µL of a D_2O and 10 µL solution of 3-(trimethylsilyl)-propionic-2,2,3,3-d4 acid sodium salt (TSP) 100 mM to the thawed samples. ^{1}H- Samples were analysed by applying the NOESY 1D pulse sequence. NMR spectra were recorded at 298 K with a Bruker (Milano, Italy) AVANCE spectrometer, operating

at a frequency of 600.13 MHz, equipped with an autosampler with 60 holders. Each spectrum was acquired using 32K data points over a 7211.54 Hz spectral width and adding 256 transients. A recycle delay of 5 s and a 90° pulse of 11.4 µs were set up. Acquisition time (2.27 s) and recycle delay were adjusted to be 5 times longer than the T1 of the protons under investigation, which was considered not to be longer than 1.4 s. Saturation of residual water signal was achieved by irradiating it during the recycle delay at δ equal to 4.703 ppm. Each spectrum was processed with Top Spin 3.0 (Bruker) by using an automatic command apk0.noe, which performs in one shot the baseline and phase correction and by applying a line-broadening factor of 0.3 Hz. The peaks were assigned by comparing their chemical shift and multiplicity with the literature and by using Chenomx NMR suite 8.2 software.

Pre-processing

After Fourier Transformation, phase and baseline correction, spectra were calibrated with reference to the chemical shift of 0.0 ppm assigned to the internal standard TSP and binned into intervals of 100 data-points. As a result, the new spectral profile consisted of 655 data-points instead of 65535. Samples at T0 were selected to be normalized by the Probabilistic Quotient Normalization algorithm [Dieterle et al., 2006]. The median of these samples was calculated and employed then as the reference spectra for the PQN among every separate batch (without T0). After normalization, all the spectra were put together again and investigated through multivariate statistical analysis.

Multivariate statistical analysis

Principal component analysis (PCA) was calculated to assess if the two types of pancakes (PLA and BEF) showed differences. Thus, after assessing that the main difference among samples was due to the enrichment, the samples were treated in order to remove the variables (bins) responsible for this bias. This was done through the employment of Principal Least Squares Discriminant Analysis (PLS-DA), setting the classes to be BEF and PLA and removing the bins mostly correlated to this separation (cut-off: 0.5). Spectra formed by the remaining bins were then investigated through correlation analysis to find the spectral clusters of correlated bins (corr > 0.7). These clusters were selected and their area was calculated and compared in its average between the different batches.

Results and discussion

1H-NMR spectrum

The NMR spectra recorded for each sample are a measure of the molecular profile of the food

matrix and will contain the signals of the chemical and functional groups of the soluble molecules composing the product analysed.

Most of the soluble fraction is composed of sugars (central part of the spectra), though other molecules are clearly visible through visual inspection of the spectrum, both in the upfield and downfield regions of the spectrum. The upfield region contains signals from aliphatic molecules, whilst the downfield region is where aromatic molecules, such as polyphenols are found. The assigned metabolites are listed in Table 1.

Table 1: List of assigned metabolites, with the main peaks found and the relative chemical shift and multiplicity.

Metabolite	Ppm
Acetate	1.92 (s)
Alanine	1.49 (d)
Citrate	2.53 (d); 2.67 (d)
Formate	8.46 (s)
Fumarate	6.52 (s)
Glucose	3.2-4; 4.68 (d)
Isoleucine	0.99 (d)
Lactate	1.33 (d)
Lactose	3.2-4; 4.5 (d)
Leucine	0.96 (d); 0.97 (d)
Maltose	3.2-4; 5.42 (d)
Phenylalanine	7.34 (d); 7.38 (t); 7.43 (t)
Succinate	2.41 (s)
Tryptophan	7.55 (d); 7.74 (d)
Tyrosine	6.91 (d); 7.20 (d)
Valine	1.045 (d)

Stability analysis

Samples from T0 of the first batch were immediately removed from this analysis, since the appeared as clear outliers, differing from all other T0 samples (data not shown).

PCA (Fig. 9) showed a clear separation along the first principal component (PC1), explaining most of the variance (76.3%), between BEF samples (on the left) and PLA samples (on the right). BEF samples also seemed more reproducible, being less scattered in the scoreplot. A trend seem to be visible, where samples at T20 (in red) seem to differ from other time-points as seen from PC2. Investigating the spectral loadings, it seemed that T20 samples had a lower sugar concentration (Fig. 10), so this had to be further investigated.

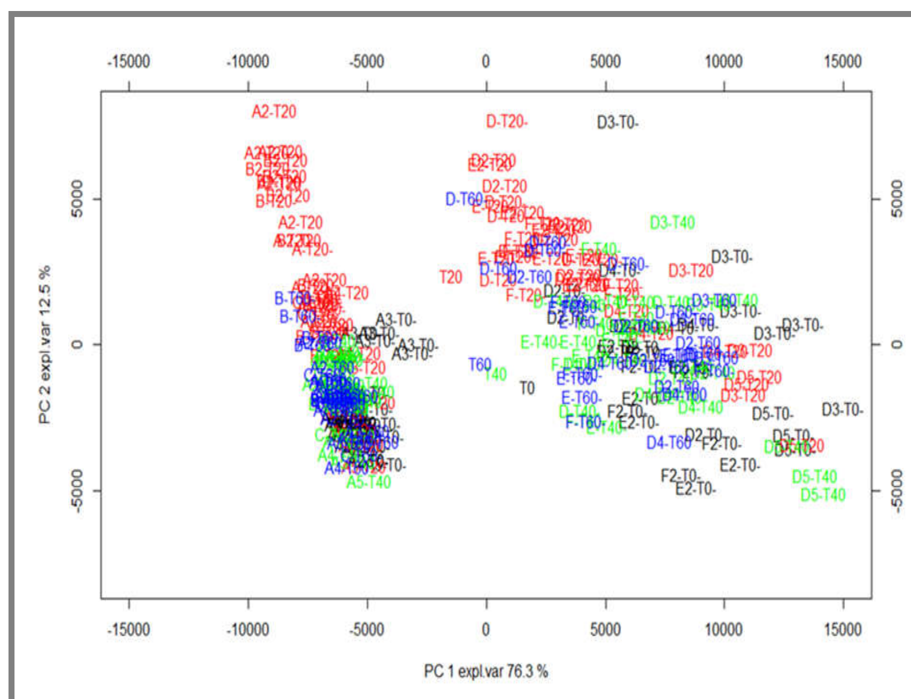


Fig. 9: PCA calculated on the NOESY1D NMR spectra of the samples after the removal of T0 samples from batch 1. Samples are coloured according to time points: T0 in black, T20 in red, T40 in green and T60 in blue.

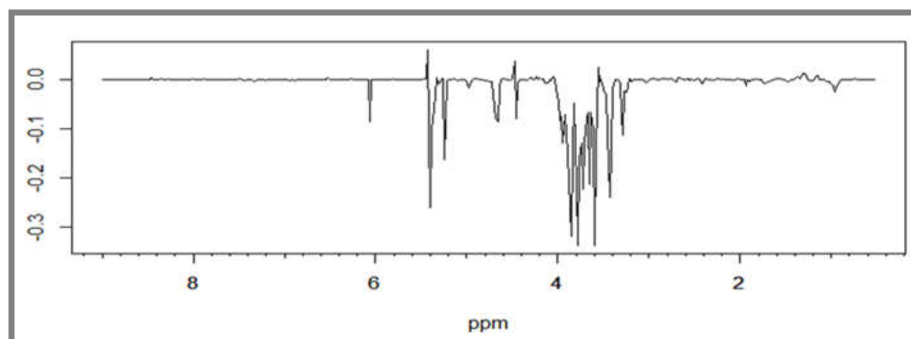


Fig. 10: Loadings for PC2, showing mostly sugar signals responsible for the variation along that principal components, with samples at low PC2 values having higher sugar concentrations.

For this reason, it was decided to eliminate the enrichment bias by removing the bins who were mostly responsible for this discrimination. As stated above, this was done by calculating a PLS-DA for BEF vs. PLA. The removed bins were 264 out of 453 and the remaining area corresponded on average to the 41% of the original spectral area. After the selection of the bins through PLS-DA, correlation analysis was performed to observe the clusters in which bins would group. After careful consideration, a threshold of $p>0.7$ was selected to generate 14 clusters of bins, forming a total of 92 selected bins and 92% of the total area of the bins selected through PLS-DA.

These clusters would represent the spectral regions with most information on the samples and on their differences along time-points. Some sample outliers were removed, thus the matrix was

reduced to 14 clusters in 220 samples (spectra).

ANOVA was then performed on the selected clusters in order to assess possible differences due to:

- time-points amongts the two separate pancake types (BEF and PLA)
- enrichment (BEF vs. PLA) at each different time-point
- enrichment (BEF vs PLA)

Results are shown in Table 2. It was found that most of the variance was found in cluster 6, which consisted of 5 bins from the sugar region, mainly corresponding to an anomeric signal at 4.2 ppm (sucrose). In effect, this signal is higher for PLA pancakes and seems to have differences also along time-points in some batches, resulting in a significant difference between T0 and T20, T20 and T40 and T0 and T40. At the same time-points, BEF vs PLA differences are also mostly significant, whilst at T60 ANOVA reuslts in not-statistically significant differences between the two types of pancakes. The regions where organic acids and amino acids signals are present are also statistically different between BEF and PLA (Clusters 12, 13 and 14) and show also some differents among each group between time-points. In cluster 13, BEF samples seem to slightly decrease area during time, again showing lower concentrations in this region at T20, especially for batches 2 and 3. PLA samples, instead, have a greater area at later time-points for the first batch, whilst the other ones only show a slightly similar trend to BEF spectra. Cluster 14, appears to have a significant difference between T40 and T60, though when investigating the signals, this difference becomes less evident and the batches show different behaviour for the areas at these two time-points. The differences seen in cluster 4, appear only from a signal that arises at 6.07 ppm at T60 for the fourth batch and in all the fifth batch. This signal might be a sample impurity. In addition, also the differences seen in Cluster 1 appear not interesting, since the cluster is mainly composed by signal tails. The trend that was visible from the first PCA, where sugars seemed to be lower in concentration at T20, is still visible in the selected clusters, though the statistical significance is mostly masked by single batches effect. In effect, by visual inspection of the sample areas, in the last batches this trend is not present.

In effect, the large difference present in the pancake batches seemed to be the main issue in this study and proves how important reproducibility is in food production, also in matters of food stability, due to the different behavior experienced in pancake samples in relation to their batch.

Table 2: Results of the ANOVA calculated on the 14 separate clusters for different combinations.

	Cl1	Cl2	Cl3	Cl4	Cl5	Cl6	Cl7	Cl8	Cl9	Cl10	Cl11	Cl12	Cl13	Cl14
BEF vs PLA	1,00	0,482	0,920	0,227	0,494	0,000	0,178	0,878	0,959	0,854	0,052	0,006	0,000	0,036

BEF	Cl1	Cl2	Cl3	Cl4	Cl5	Cl6	Cl7	Cl8	Cl9	Cl10	Cl11	Cl12	Cl13	Cl14
T0 vs T20	0,757	0,632	0,313	0,951	0,966	0,187	0,633	0,259	0,533	0,243	0,398	0,348	0,517	0,609
T0 vs T40	0,008	0,382	0,141	0,733	0,425	0,008	0,796	0,893	0,931	0,063	0,946	0,146	0,452	0,931
T0 vs T60	0,823	0,203	0,866	0,574	0,513	0,456	0,007	0,309	0,723	0,243	0,057	0,919	0,929	0,894
T20 vs T40	0,216	0,706	0,189	0,145	0,205	0,379	0,933	0,777	0,587	0,994	0,103	0,218	0,012	0,380
T20 vs T60	0,897	0,693	0,034	0,213	0,668	0,807	0,027	0,796	0,980	0,693	0,016	0,706	0,029	0,116
T40 vs T60	0,069	0,620	0,361	0,396	0,485	0,450	0,193	0,840	0,929	0,697	0,281	0,524	0,109	0,002

PLA	Cl1	Cl2	Cl3	Cl4	Cl5	Cl6	Cl7	Cl8	Cl9	Cl10	Cl11	Cl12	Cl13	Cl14
T0 vs T20	0,839	0,901	0,960	0,003	0,128	0,000	0,179	0,525	0,241	0,231	0,498	0,206	0,133	0,723
T0 vs T40	0,827	0,722	0,099	0,134	0,418	0,008	0,576	0,168	0,097	0,199	0,215	0,032	0,048	0,758
T0 vs T60	0,349	0,499	0,588	0,434	0,944	0,927	0,094	0,172	0,982	0,563	0,661	0,669	0,613	0,856
T20 vs T40	0,025	0,364	0,294	0,005	0,870	0,013	0,899	0,461	0,192	0,371	0,441	0,694	0,000	0,267
T20 vs T60	0,932	0,249	0,430	0,825	0,479	0,078	0,347	0,792	0,705	0,592	0,601	0,613	0,217	0,891
T40 vs T60	0,042	0,574	0,567	0,228	0,108	0,163	0,582	0,817	0,371	0,814	0,298	0,710	0,009	0,068

BEF vs PLA – Time points	Cl1	Cl2	Cl3	Cl4	Cl5	Cl6	Cl7	Cl8	Cl9	Cl10	Cl11	Cl12	Cl13	Cl14
T0 BEF vs PLA	0,923	0,098	0,872	0,486	0,009	0,007	0,369	0,532	0,494	0,474	0,138	0,489	0,133	0,295
T20 BEF vs PLA	0,296	0,949	0,069	0,513	0,850	0,001	0,432	0,634	0,454	0,701	0,500	0,341	0,043	0,602
T40 BEF vs PLA	0,352	0,860	0,376	0,641	0,969	0,042	0,449	0,415	0,242	0,526	0,694	0,536	0,000	0,064
T60 BEF vs PLA	0,078	0,800	0,271	0,223	0,074	0,096	0,614	0,560	0,425	0,798	0,067	0,661	0,085	0,051

BEF and PLA samples had differences in their profiles, though these were not that evident along the storage, proving that the enrichment did not effectively affect the shelf-life of the product. The trend observed for some signals, where the area decreases at T20 to later increase again, might be due to the reorganization of carbohydrates and sugar molecules during storage (i.e. staling processes, starch retrogradation, etc.), though no literature is present on 1H-NMR study of this specific behaviour.

Conclusion

The proposed method for the evaluation of the food matrix stability is based on the molecular profile of its aqueous extract, acquired through NMR spectroscopy. In this study two different

types of pancakes, enriched with two bioactive ingredients or controls, were evaluated during their shelf-lives. The method employed consisted in the acquisition of NMR spectra of samples and the evaluation of their profile variations during storage. It was seen that PLA and BEF pancakes had different molecular profiles, though the effect of storage was masked by differences in the production of the pancake batches. Further studies are needed for the evaluation of the shelf-life of the products, though no major impact on the matrix was seen in the selected conditions and period of storage, proving that the enrichment did not cause any major stress to the food matrix. The study proposes an effective analytical pipeline for this kind of evaluations and can be employed in further stability assessments.

5.3 Final conclusions

The present studies on various enriched food products represent the first step towards the devlopment of a specific analytical method employing NMR for the investigation of the stability of a food matrix. These studies have shown the capability of NMR spectroscopy to capture a lot of information on what occurs in a food items in relation to many concurrent factors (enrichment, bacth production, storage) and the consequences of these factors and their interaction on the molecular profile of the food item analyzed. In effect, the non-target approach of NMR analysis allows to define a more complex metabolite profile for the food product and observe even unforeseen and otherwise not visible changes in its matrix. This is particularly important in enriched products which have to undergo a great variety of testing before being ready for the market. It is in fact fundamental that the bioactive molecules added in the product are stable during shelf-life but can easily become available when the product is consumed and the capability of this method to picture the interaction between the food matrxi and the added molecule is of great help in this kind of assay.

REFERENCES CHAPTER 5

Augustin, Mary Ann, and Luz Sanguansri. "Challenges and Solutions to Incorporation of Nutraceuticals in Foods. "Annual review of food science and technology" 6 (2015): 463-477.

Ciampa et al., 2012 "Changes in the Amino Acid Composition of Bogue (Boops boops) Fish during Storage at Different Temperatures by ^{1}H-NMR Spectroscopy" – Nutrients (2012), 4(6), 542-553.

Dave N. et al., 2014, "Antioxidant Activity, Textural Properties and Storage Stability of Honey Fortified Biscuits." (2014).

Dieterle, F., Ross, A., Schlotterbeck, G., & Senn, H. (2006). Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. Analytical chemistry,78(13), 4281-4290.

Edel A. L. et al., 2015, "Stability of bioactives in flaxseed and flaxseed-fortified foods." Food Research International.

Fundo J. F. et al., 2015, "Molecular Dynamics and Structure in Physical Properties and Stability of Food Systems".

Ganesh V. et al., A Review: Supplementation of Foods with Essential Fatty Acids – Can It Turn a Breeze without Further Ado?, 2015, Critical Reviews in Food Science and Nutrition.

Heertje I., Structure and function of food products: A review, 2014, Food Structure, 1, 3-23.

Kilcast, D., & Subramaniam, P. (2000) "The Stability and Shelf-Life of Food" Elsevier, ISBN: 978-1-85573-500-2

Laghi et al., 2014 - Laghi, Luca, Gianfranco Picone, and Francesco Capozzi "Nuclear magnetic resonance for foodomics beyond food analysis." TrAC Trends in Analytical Chemistry 59 (2014): 93-102.

Li-Chan, Eunice CY. "Bioactive peptides and protein hydrolysates: research trends and challenges for application as nutraceuticals and functional food ingredients. "Current Opinion in Food Science" 1 (2015): 28-37.

López-Córdoba et al., Yerba mate antioxidant powders obtained by co-crystallization: Stability during storage, Journal of Food Engineering 124 (2014) 158–165.

Mc Clements D. J., 2015, "Enhancing nutraceutical bioavailability through food marix design", Current Opinion in Food Science, 4:1-6.

Moro A. et al., 2015, "Dairy matrix effect on the transference of rosemary (Rosmarinus officinalis) essential oil compounds during cheese making", J. Sci. Food Agric., 95: 1507-1513.

Norton J. et al., 2014, "Designing Food Structures for Nutrition and Health Benefits", Annu. Rev. Food Sci. Technol., 5:177-95.

Picone et al. (2011) "Metabolomics as a Powerful Tool for Molecular Quality Assessment of the Fish Sparus aurata." Nutrients 3.2.

Salcedo-Sandoval L. et al., Shelf-life of n-3 PUFA enriched frankfurters formulated with a konjac-based oil bulking agent, LWT – Food Science and Technology 62 (2015) 711-717.

Samuelsen A. et al., Effects of orally administered yeast-derived beta-glucans: A review, Mol. Nutr. Food Res. 2014, 58, 183–193.

Sinn N. et al., Effects of n-3 fatty acids, EPA v. DHA, on depressive symptoms, quality of life, memory and executive function in older adults with mild cognitive impairment: a 6-month randomised controlled trial, British Journal of Nutrition (2012), 107, 1682–1693.

Turgeon S. L. et al., 2011, "Food matrix impact on macronutrients nutritional properties", Food hydrocolloids, 25 1915-1924.

Yang M. et al., 2011, "Food Matrix Affecting Anthocyanin Bioavailability: Review", Current Medicinal Chemistry, 18:291-300.

Yousouf et al., Health Benefits of Anthocyanins and Their Encapsulation for Potential Use in Food Systems: A Review, 2015, Critical Reviews in Food Science and Nutrition.

CHAPTER 6: FOODOMICS TO INVESTIGATE THE DIGESTION PROCESS

INTRODUCTION

Nutrition is nowadays recognized as key factor in human health. This has spurred an increasing number of food and nutritional studies aiming at the assessment of the real nutrient intake and at the optimization of food items matching human needs. One of the aspects that is generally forgotten in these kinds of investigation is the so-called “matrix-effect”, the complex interaction between specific nutrients/molecules and the food matrix itself (see Chapter 5). This regards the supra-molecular organization of foods and makes food items with similar composition actually different. The matrix effect, as mentioned in the previous chapter, can greatly affect nutrient bioavailability and bio-accessibility. Bio-accessibility regards the extraction of the molecules made accessible to the solvent through the digestion and their diffusion to the absorption site. Bioavailability is instead the amount of a nutrient that can be used for normal physiological functions and it therefore concerns digestion and absorption of molecules in the guts.

Many different factors can impact on the food composition, the matrix organization and the complex network of molecular interactions and the compartmentalization of nutrients, thus modifying nutrient bioavailability and nutritional value of a food product. For these reasons, it is necessary in nutritional evaluations to study the behaviour of a food product during the various digestive steps and the influence of particular processing procedures on the final nutritional quality. In order to investigate on these qualities, in vitro digestion is a fundamental tool, allowing the study of the processes occurring in the human body during digestion with lower costs and without all the implications and difficulties of in vivo studies. Proton NMR spectra registered on liquid extracts may represent an extremely versatile tool to have an overview of digestion, especially for what concerns proteins, since it allows the simultaneous observation of free amino acids, peptides and proteins, also giving for these latter qualitative information about their dimension [Bordoni et al. ,2011; Bordoni et al., 2014]. This research work has investigated digestion in many different food products employing an in vitro model developed and perfected in the last few years, in order to assess the impact of different food characteristic on their final nutritional value and on the bioavailability of specific important nutrients. More focus was put on protein-rich foods, since the digestive step can play a fundamental role in their digestibility and nutritional value. In fact, through digestion, protein-based food can liberate important molecules

such as aminoacids and bioactive peptides. It is thus of great interest to understand whether these high-value metabolites are made accessible to the human body through digestion and how different factors can impact on this. In this chapter, the digestibility of different products is evaluated, here intended as the result of solubilisation and hydrolysis of molecules which are thus made ready for absorption.

6.1 Genetic impact on digestibility – Investigation on two cultivars of peppers (*Capsicum annuum*)

Materials and methods

In this study two different cultivars of peppers (Cornelio and Lamuyo) in two different colours (yellow and red) were analysed. This was a preliminary study so only a total of 6 samples for each cultivar were given (2 red and one yellow). Digestion was applied in duplicate employing the method developed by Minekus et al. (2014), therefore a total of 12 samples were obtained. For each sample of pepper, 4 grams were weighted and diluted in a 7% TCA (trichloroacetic acid). Samples were then weighted and mixed at 150 rpm at room temperature. 1 mL of the obtained sample was taken and centrifuged for 5 minutes at 4° C and 14000 rpm. 900 µL of the centrifuged sample were then mixed with 160 µL of deuterated water (D_2O) with the internal standard (TSP - Trimethylsilylpropanoic acid). The pH was then adjusted to 7, the sample was centrifuged again for 10 minutes at 14000 rpm and finally it was inserted in the NMR tube for the analysis. Spectra were acquired through the NOESY1GPPR1D sequence, with acquisition time of 2.27 s and 256 scans for a total time of 31min 44s. 12 spectra for digested peppers and 12 spectra for non-digested peppers were acquired. In addition, the spectrum of all the reagents and enzymes used for digestion was acquired ("blank"), in order to be able to subtract the signal from the digestion system from the NMR spectra of digested samples. Spectra were automatically baseline and phase corrected on the software TopSpin, then saved as ASCII files and imported on R for further processing and statistical analysis. Spectral chemical shifts were referred to the internal standard TSP at 0.0 ppm. Then peripheral regions and the signal of the solvent were removed. Normalization was carried out on the anomeric proton of lactose, since this molecule was contained in equal measure in all samples (including the "blank"). A binning step was also carried out in order to reduce matrix dimension and avoid misalignment. After this, the spectral difference between spectra of digested peppers and the "blank" was calculated. In this way, only the signals belonging to the molecules originating from digestion or from the pepper itself were left and ready for statistical analysis. Spectra of

digested samples were then normalized through the PQN algorithm (Dieterle et al., 2006). Correlation between the spectral results and traditional nutritional assays was calculated for the values of vitamin C, phenols and antioxidant activity. Signals (bins) with a correlation higher than 0.8 were selected and evaluated.

Student's t test was carried out in the the following six spectral regions through the software Graphpad Prism, in order to test for differences between the two cultivars:

- Acids/Aliphatic region: 0.5-2.5 ppm;
- Sugars: 3-4.5 ppm;
- Aromatic molecules: 7-8.5 ppm;
- Phenols: 7.225-7.475 ppm;
- Vitamin C: 4.49-4.56 ppm – correlated with antioxidant activity;
- Phenol signal: 7.29 ppm – correlated with the phenolic content;

Results

Non digested peppers gave a spectrum with an area that was lower in comparison to their digested counterpart and the “blank”, simply composed by the digestive fluids used in the protocol (Figure 1). Many enzymes and reagents are used for digestion, thus this is why the spectrum for the “blank” is particularly rich of signals.

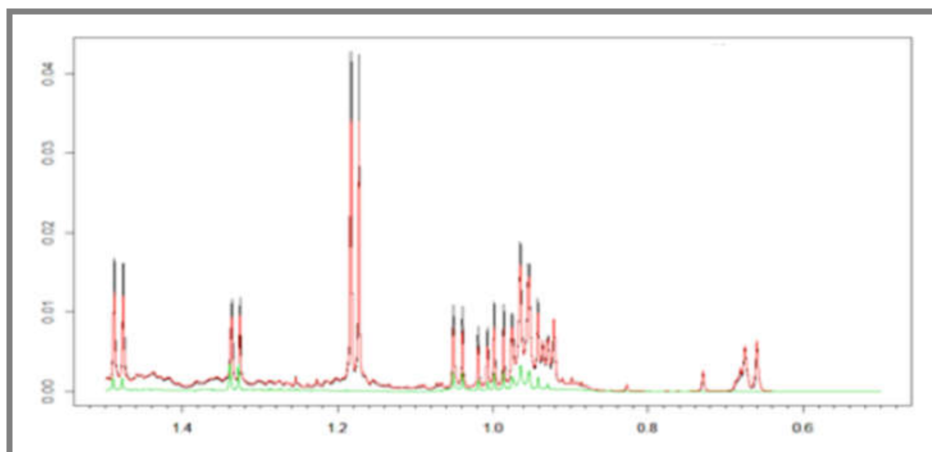


Fig. 1: NOESY1D NMR spectra for the “blank” (black), composed by the digestive fluids employed in the *in vitro* digestion method, a non-digested pepper (green) and its digested counterpart (red) after referencing to the internal standard TSP. As visible, some small signals derive from the non-digested pepper, whilst the majority seem to be caused by the “blank” due to the high concentration of the digestive fluid employed.

For this reason, the “blank” spectrum was subtracted from each spectrum of digested peppers. After this step, the spectra are more comparable to the non-digested sample spectra in the carbohydrate region (fig. 2).

Digestion clearly liberates a great amount of aliphatic and aromatic molecules. This is evident from the increase in intensity of these regions and the appearence of new signals (Fig. 2).

In fact, on average the digested samples spectra have an area 27% greater than their non-digested counterparts. This is limited to 15% in the carbohydrate regions, whilst it is a lot higher in the aromatic and aliphatic regions (73 and 79% respectively).

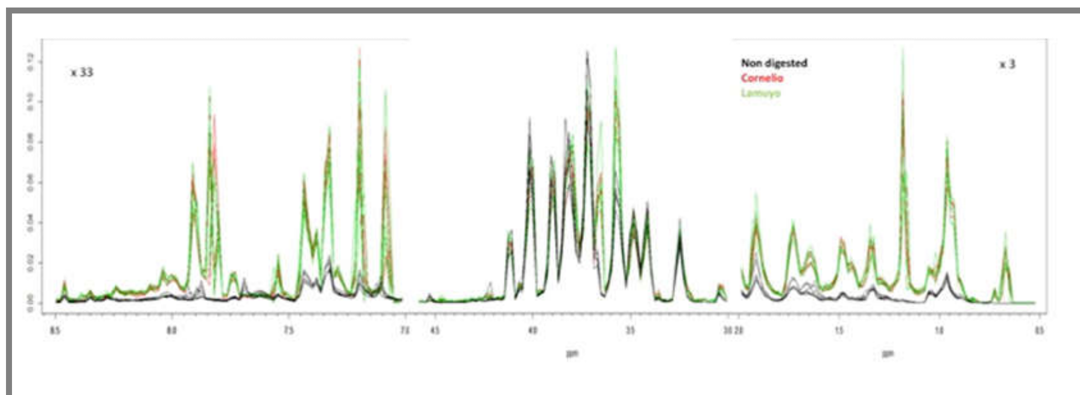


Fig. 2: NOESY1D NMR spectra for non digested peppers (in black) and digested peppers (Cornelio in red and Lamuyo in green) after binning and PQN normalization. In the carbohydrate region displayed in the middle (3-4.5 ppm) spectra are very similar. On the other hand, some signals are higher after digestion, especially in the aromatic region (on the left, zoomed 33 times) and the aliphatic region (on the right, zoomed 3 times). Moreover, some signals like the one at 7.29 ppm appear only after digestion. This signal is highly correlated to the phenolic content.

Correlation between NMR spectral data and spectrophotometric data

For what concerns phenols, many spectral bins are highly correlated with the quantities measured with a spectrophotometer. The region mostly correlated is the one between 7.26 and 7.44 ppm. Moreover, the signal mostly related to the phenolic content is the one at 7.29 ppm for which the correlation is reported in Figure 3.

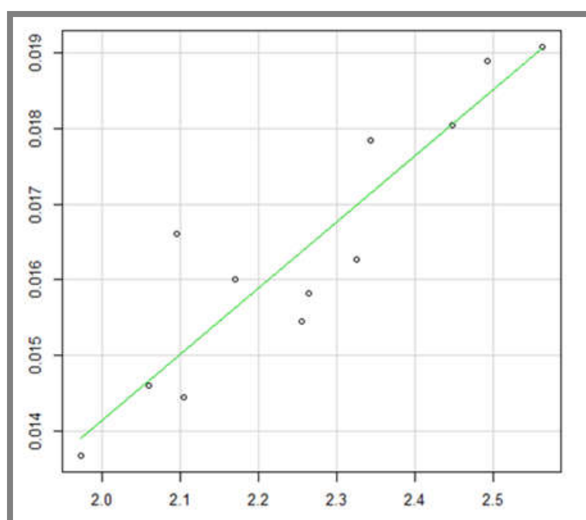


Fig. 3: Correlation between the area of the signal at 7.29 ppm (y-axis) and the phenol values measured by spectrophotometry (x-axis). The regression line is shown.

A similar correlation is experienced between the same spectrophotometric values and the area of the NMR spectral region between 7.26 and 7.44 ppm, as reported in Figure 4. This region is the one where signals from phenolic molecules appear. When the whole aromatic region is considered, though, the R2 between its area and the phenolic content measured traditionally is only 0.29%. This proves that the traditional technique only measures a part of the phenolic molecules, whilst the NMR method is capable of picturing aromatic molecules not detectable through the traditional assay. This is due to the fact that the traditional assay is based on the bond between the measured metabolites and the specific reagents which will give out a determined colour, measured by a spectrophotometer, thus possibly leaving out of the quantitation some small molecules. In contrast, the NMR method is capable of observing such metabolites.

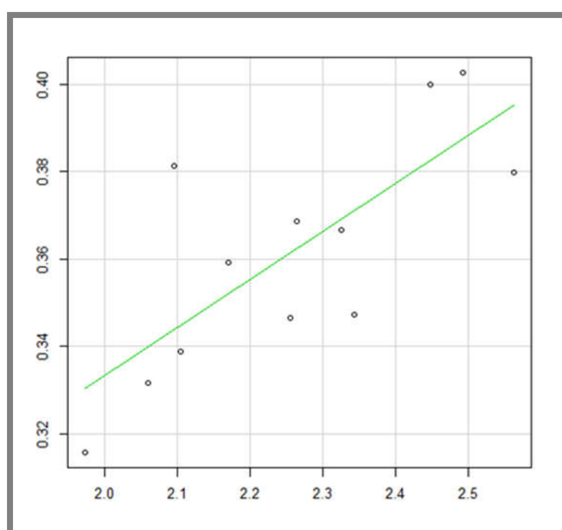


Fig. 4: Correlation between the area of the aromatic region between 7.26 and 7.44 ppm (y-axis) and the phenol content measured by spectophotometric assay (x-axis). The regression line is shown.

For what concerns the antioxidant power of pepper, signals which mostly correlates with the values from traditional analysis are the ones at 8.02, 7.86, 4.55, 4.49, 4.4, 4.29 e 2.08 ppm.

Among these, the best result is obtained for the spectral region between 4.5 and 4.56 ppm, where the signal from ascorbic acid is present (Fig. 5).

Univariate statistical analysis was carried out on the area of six different spectral regions to understand if differences exist between the different cultivars. Table 1 reports the mean values with standard deviations for the selected areas in the two cultivars.

As shown, no significant difference was experienced.

Table 1: Mean values and standard deviation for the various spectral region selected in the two different cultivars.

	Aromatic region	*Sugar region*	*Aliphatic region*	*Phenolic region*	*Vitamin C*	*Phenolic signal*
CORNELIO n=6	113.78±8.28[a]	5921.67±436.20[a]	1544.30±77.55[a]	36.30±3.29[a]	16.72±1.73[a]	1.59±0.17[a]
LAMUYO n=6	114.35±11.19[a]	6016.33±553.51[a]	1520.83±127.30[a]	36.11±3.56[a]	17.02±2.48[a]	1.62±0.21[a]

Number of samples in each class is reported (n).

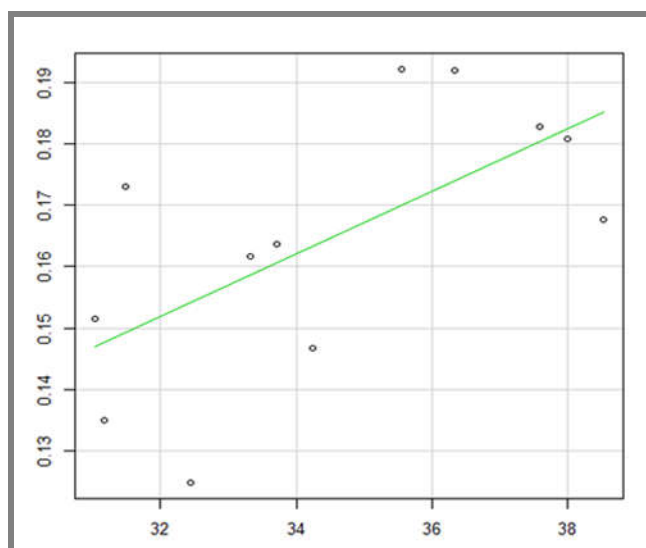


Fig. 5: Correlation between the area of the region between 4.49 and 4.55 ppm (y-axis) and the antioxidant activity measured by traditional assays (x-axis). The regression line is shown (R2= 0.66).

Conclusion

To conclude, the two cultivars of pepper did not show any statistical diffeence in the NMR spectra. This, though, can be due to the fact that only a small number of samples was available. The study needs to be carried out with a larger sample cohort in order to obtain better and more robust statistical results. The NMR method has proved to be fast and highly reproducible since it confirmed the results from traditional assays of spectrophotometry. The capability of NMR spectroscopy to capture a great range of molecule proved to be fundamental for this kind of assessment, helping in the understanding of the digestibility.

6.2 Processing impact on digestibility – Investigation on a traditional cured meat product (Bresaola)

Materials and Methods

This study investigated the digestibility of Bresaola, a cured meat typical of Northern Italy, depending on its ripening time. Bresaola samples were cured for three different time-lenght: “Time 0” (after 10 days of salting), “Time 2” (after 2 weeks of ripening) and “Time 4” (after 4

weeks of ripening). These samples were then sent from Cesena, University of Bologna, to Teagasc institute for the simulation of gastric and intestinal digestion following the protocol described in figure 6 for a static in vitro digestion method suitable for food. In total, fourteen digestions were carried out, both with and without substrates as control. The enzymes assays were also carried out following the described digestion protocol. The samples of cured meat were digested in triplicate according to Bordoni et al. (2014), following the scheme reported in Figure 6.

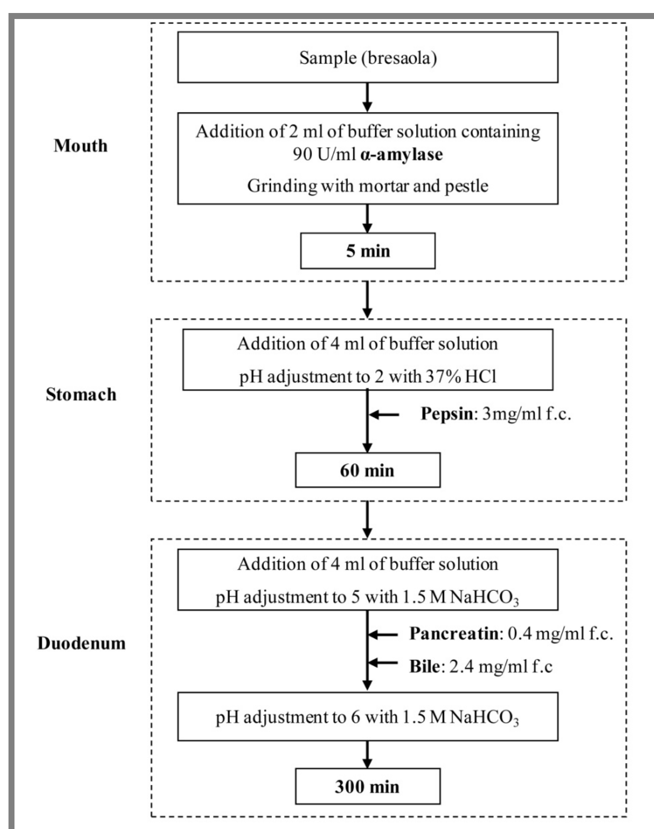


Fig. 6: Scheme of the *in vitro* digestion protocol employed (from Bordoni et al., 2014).

Human digestion was simulated *in vitro* inside a 100 mL flask, kept at 37 °C in a water bath on a magnetic stirrer equipped with a heating plate. Chemical composition of the digestive fluid, pH and residence periods were adjusted to mimic the physiological conditions. Two samples were collected during digestion: P1, at the end of gastric phase, and P2, at the end of the duodenal phase. In samples P1 the pH was increased to 8 with 35% NaOH to prevent possible modifications induced by acidic conditions. Samples at P2 were acidified to pH 2 with 37% HCl to stop pancreatic hydrolysis and to avoid bias caused by different pH values. P1 and P2 samples were stored at -80 °C before ^{1}H-NMR experiments. A mock digestion without the enzymes was also carried out on meat samples in order to investigate the simple effect of pH on the meat and to compare the digested samples with non-digested ones. Even in this case, P1 and P2 samples were collected. Samples

were prepared for ^{1}H-NMR by adding to 1 mL of each sample 160 µL of 100 mM phosphate buffer in deuterium oxide (D_2O), containing 10 mM 3-TrimethylSilyl-Propanoic-2,2,3,3-d4 acid sodium salt (TSP) as internal standard. After adjusting the pH to 7.00, the samples were centrifuged at 14000 rpm for 5 min in order to further remove impurities. All ^{1}H-NMR spectra were recorded at 25 °C on a Bruker US+ Avance III spectrometer operating at 600 MHz, equipped with a BBI-z probe and a B-ACS 60 sampler for automation (Bruker BioSpin, Karlsruhe, Germany). The spectra were collected with a 90° pulse of 14 µsec with 10 W of power, a relaxation delay of 5 sec and an acquisition time of 2.28 sec. The spectra were registered by means of the first increment of a NOESY sequence, designed to suppress the residual signal of the solvent, while giving, for each proton, peaks proportional to the concentration of the substance they belong.

Results and discussion

The acquired NMR spectra were investigated in order to assess the variation occurring in samples along ripening and after digestion. Spectra were referenced to the TSP signal at 0.0 ppm, solvent signal and noise were removed from the spectrum and the matrix was aligned with reference to the signal of lactate, also employed as a signal for normalization, after careful consideration. The spectra of digestive enzymes were also acquired and these were subtracted from the spectra of the digested samples, in order to remove signals belonging just to the digestive fluids employed in the protocol. Variance of non-digested samples was considered between ripening time-points, so between T0 and T2 and T0 and T4, calculated as the average and standard deviation of this variance:

$(T_2-T_0)/(\text{area } T_2)*100$ or $(T_4-T_0)/(\text{area } T_4)*100$

For digested samples, instead, the variance was calculated for each time-point and sample as:

$(T_{0\ dig}-T_{0\ non\ dig})/(\text{area } T_{0\ non\ dig})*100$ or $(T_{2\ dig}-T_{2\ non\ dig})/(\text{area } T_{2\ non\ dig})*100$

or $(T_{4\ dig}-T_{4\ non\ dig})/(\text{area } T_{4\ non\ dig})*100$

therefore considering the spectra of non-digested meat as a reference.

This was done in several spectral regions chosen for their informative power:

- Total spectra
- Aromatic region (6-7.85 ppm)
- Alpha region (3.6-5 ppm)
- Amides region (7.9-9 ppm)
- Gamma-1 region (0.72-1.1 ppm)

- Gamma-2 region (0.56-0.72 ppm)

The results are shown in the tables below (Table 2 and 3)

Table 2: Results of the variance for the non-digested samples among different ripening points. * are samples for which the variance was negative. T0-T2 represents the difference between the starting ripening point T0 and the ripening time-point T2. T0-T4 represents the difference between the start of the ripening (time point T0) and the end of ripening (time-point T4).

	Gastric T0-T2	Gastric T0-T4	Intestinal T0-T2	Intestinal T0-T4
Total area	11.84±0.14	10.99±0.63	8.44±2.14	7.97±0.42
Aromatic area	9.43±6.67	3.23±3.01	3.21±1.97	5.98±4.18*
Alpha area	15.45±2.81	14.40±1.31	11.32±1.73	8.71±1.80
Amides Area	9.48±4.79*	11.76±5.52*	7.80±6.27	9.57±8.19*
Gamma 1 Area	30.86±6.45	37.71±16.48	8.27±7.61	13.47±10.54*
Gamma 2 Area	11.67±12.14	12.21±5.49	12.80±12.01	10.51±6.38

Table 3: Results of the variance for the digested samples among different ripening points. * are samples for which the variance was negative. NI is a non-informative region for that type of spectra. T0 is the time-point of the start of the ripening, T2 and T4 are a mid-point and the end of ripening of Bresaola, respectively.

	Gastric T0	Gastric T2	Gastric T4	Intestinal T0	Intestinal T2	Intestinal T4
Total area	29.43±9.83	26.06±2.56	20.89±13.44*	778.25±121.26	845.08±46.86	701.46±230.57
Aromatic area	36.13±4.08*	30.80±5.05*	30.61±3.01*	346.46±51.70	346.23±18.16	330.88±97.75
Alpha area	12.91±8.10*	12.00±3.23*	11.41±0.90*	178.1±30.26	145.10±5.76	135.07±80.55
Amides Area	47.32±8.85*	31.95±4.85*	35.04±3.45*	138.66±40.54	197.81±31.65	213.78±71.58
Gamma 1 Area	NI	NI	NI	919.21±509.28	629.83±80.39	1740.52±567.22
Gamma 2 Area	14.64±8.51*	11.87±7.79*	22.55±3.12*	500.26±178.25	532.83±85.45	483.17±201.85

As it can be seen from Table 2, in the gastric phase most regions increase in area, apart from the amide region. In the intestinal phase, samples at the end of ripening show an area lower than T0 for the aromatic, amides and gamma 1 area. This proves that the effect of pH can simply modify the capability of the matrix of liberating molecules, though it shows how there is a tendency for the product to be more digestible along the ripening time. For what concerns the amide area, which seems to decrease during ripening time, this effect can be due to various reasons: molecules present in the samples can get broken up into smaller molecules and their signal show up in other parts of the spectra, the broad signals from big proteins or peptides are altered into sharper and smaller amino acid signals or the molecules with signals in this spectral region become less available during ripening due to modifications in the matrix structure. In Table 3 we can see that in the gastric phase, the non-digested spectra are generally higher than the digested one in the selected regions, whilst the opposite is for intestinal samples. This shows that is mainly during the duodenal phase that molecules are liberated from the matrix from enzyme action and can therefore be available for absorption and, later, for their functions.

Conclusions

This preliminary study proved the investigative power of the NMR technique. In this research the digestibility of Bresaola was tested in relation to the ripening time. It was seen that the process of curing can increase the bioavailability of molecules such as aminoacids and small peptides, which are released mostly during the duodenal phase of digestion. Through the analysis of NMR spectra it was in fact possible to understand first the impact of the curing process on the capability of the matrix of releasing these molecule. Moreover the digestion of the matrix was studied, in order to understand whether the product was easily digested and in which phase the molecules of interest where made available for absorption.

6.3 Cooking impact on digestibility – Investigation on different meat-based recipes.

Materials and methods

Four different meat preparation were studied. Each type of meat was either cooked through a traditional method (TR) or with sous-vide (vacuum) cooking (SV). Three samples for each cooking method were analyzed for each meat type, so to have a total of 6 samples for 4 products. Each sample was digested in vitro employing the method described in 13.2. Digested samples were centrifuged and filtered. Before digestion, the protein content of each sample was measured through the Kjeldhal method. Digested samples were analyzed through NMR spectroscopy.

In a 1.5 mL eppendorf, 160 µl of phosphate buffer in deuterated water (D_2O) were added to 900 µL of digested samples. pH was adjusted to 7.0 by means of addition of 1M solutions of either HCl or NaOH. Samples were then centrifuged at 14000 rpm and 4°C for 10 minutes. 800 µl of surnatant were then transferred in NMR tubes for the analysis. NMR spectra were acquired through a 600 MHz spectromoter using an acquisition time of approximately 30 minutes and a NOESY1GPPR sequence. Spectra were referred to the internal standard (TSP) at 0.0 ppm. Peripheral spectral regions and the signal from the solvent were removed from the spectra. First, spectra of the digested samples and the “blank” (only digestive fluids and enzymes) were normalized on the TSP signal to be compared. After a binning procedure the spectrum of the blank was subtracted from all spectra, in order to just observe the signals from the food matrix. The area of the region between 4.15 and 4.66 ppm (“alpha region”) was then calculated, in order to measure the protein release after digestion. To have the real value of the aminoacid in the alpha region, the obtained values were corrected through this equation (Eq. 1):

$$R\,N\,C(std)(Vs+Vb)\frac{1}{Vs}Vt\,PM$$

where:

R = Area of the Alpha region/Total protein content (measured by Kjeldahl method);

N = number of protons in the internal standard (N=9);

C(std) = concentration of the internal standard (TSP);

Vs = sample volume;

Vb = buffer volume;

Vt = total volume;

PM = average molecular weight of amino acids (PM=110 g/mol).

Means and standard deviations of the obtained values were calculated and compared for each product in the two cooking methods.

Results and discussion

As visible in Figure 7, samples did not show much difference in the amino acid content between the two types of cooking: only slight differences are visible in the average values and when the standard deviation is considered the difference is not statistically significant.

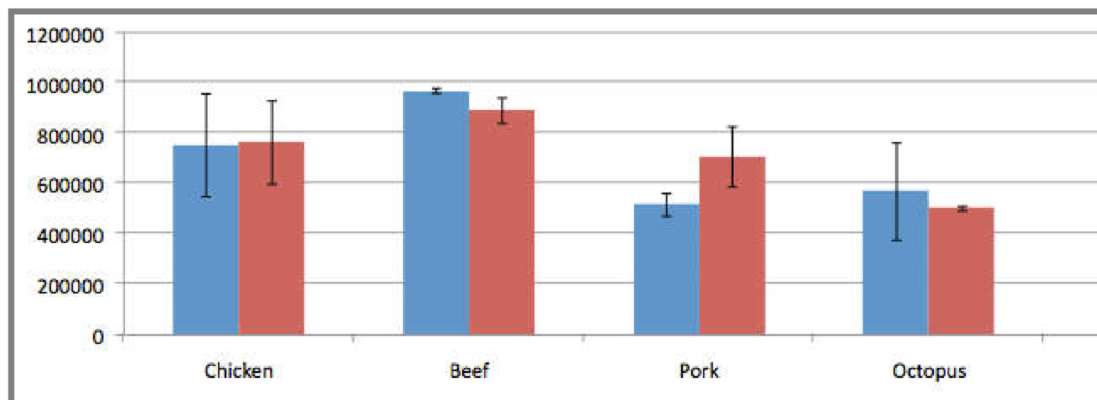


Fig. 7: Mean values with standard deviation for the aminoacid content of digested samples cooked with the traditional method (blue) or in vacuum (red), measured through NMR spectroscopy and calculated applying equation 1.

The only product showing an increased protein digestibility is pork, where the new cooking method in vacuum proved to increase the product’s protein digestibility. It is clear though that there might exist a particular matrix-related response to the cooking method. Recent studies have confirmed that vacuum cooking is an efficient technique for the improvement of food digestibility. Sangsawad and colleagues [Sangsawad et al., 2016] showed that chicken breast cooked in vacuum at 70°C for 30 minutes had a greater protein digestibility and antioxidant power compared to

samples cooked at higher temperatures and longer.

In fact, cooking at higher temperatures could cause the formation of indigestible protein aggregates, which could reduce digestion proteolysis and which, in turn, could be fermented by the gut microbiota into potentially damaging molecules (i.e. ammonia) [Oberli et al., 2015]. In this case only one product showed greater protein digestibility, measured as the amount of aminoacids made available through digestion, when the new cooking technique was employed, though this can be due to the fact that a small number of samples was available. For this reason, more research is necessary with greater sample numerosity, in order to further test the possible difference in protein digestibility between the two cooking techniques. With a larger number of samples a lower value of standard deviation can be obtain and significant difference could be found among samples. Moreover, different cooking time and temperature combinations should be evaluated, in order to understand the influence of both factors in the final meat quality and aminoacid realease, as carried out in other studies [Roldan et al., 2014; Oberli et al., 2015]. In fact, both the temperature and the cooking time can impact on meat quality, for example affecting the production of free radicals and thus protein and lipid oxidation. For this reason, it is important to find the best combination of these factors in order to gain greater protein digestibility.

Conclusions

The nutritional value of protein-based food is deeply related to their digestibility and thus to the amount of aminoacid that becomes available during digestion. Vacuum cooking seemed to improve protein digestibility in just one of the four analysed samples (pork meat), showing how the technique can improve the nutritional value of a meat product but the matrix itself can influence the results. Further research with more samples is necessary to prove these findings and gain more robust results. In addition, different combinations of time and temperature of cooking should be tested to understand the impact of each factor and to find the best possible cooking procedure to obtain a higher protein digestibility for each product.

6.4 Digestibility of food supplements – Lowpept Study

Materials and methods

Lowpept is a powdered food supplement of dairy peptides with the aim of lowering and stabilizing blood pressure. This product is obtain through the hydrolisis of casein with food-grade pepsin [Recio et al., 2006; Contreras et al., 2009] and contains antihypertensive peptides RYLGY and

AYFYPEL [Anadón et al., 2010]. In this study Lowpept and casein were digested through the in vitro protocol described in Figure 8 in order to assess their behaviour during digestion and the real bioaccessibility of the peptides contained in Lowpept. Samples were collected in triplicate after each digestive phase (oral, gastric and duodenal) for both products (casein and Lowpept). 700µl of 100mM phosphate buffer in deuterium oxide (D_2O), containing 10mM 3-TrimethylSilyl-Propanoic-2,2,3,3-d4 acid sodium salt (TSP) as internal standard was added to all the samples. After pH adjustment to 7.00, all samples were centrifuged at 14,000g for 5 minutes for further particulate removal. ^{1}H-NMR spectra were recorded at 300K on a Bruker US+ Avance III spectrometer operating at 600MHz, equipped with a BBI-z probe and a B-ACS 60 sampler for automation (Bruker BioSpin; Karlsruhe, Germany). Spectra were collected with a 90° pulse of 14µs with 10W of power, a relaxation delay of 5s and an acquisition time of 2.28s [Bordoni et al., 2011]. Two different NMR experiments were carried out on the samples: NOESY and JRES.

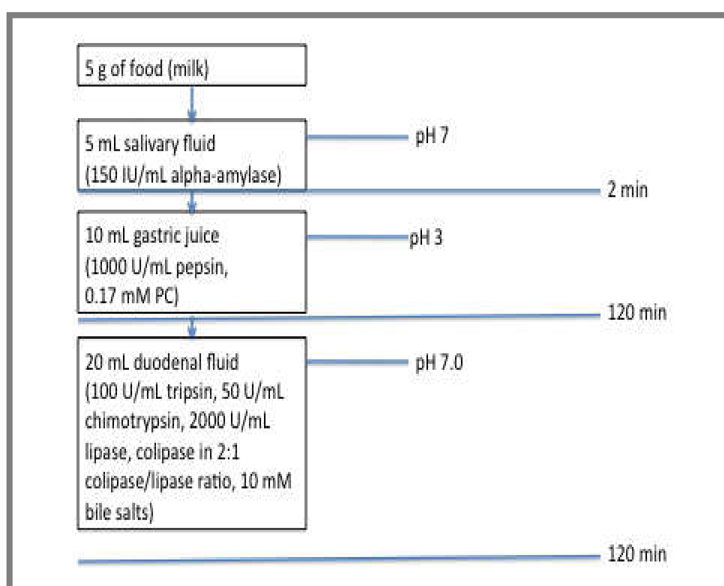


Fig. 8: Scheme of the digestion protocol employed, adapted from the COST Infogest protocol (Minekus et al., 2014).

Noesy. Spectra were obtained by collecting 64 scans into 32K data points, covering a 20 ppm spectral width. The NOESYGPPR1D sequence (a standard pulse sequence in the Bruker library) was applied to suppress the residual signal from HOD, this sequence incorporates the first increment of the NOESY pulse sequence and a spoil gradient. Phase and baseline corrections were performed automatically using TopSpin version 3.0 (Bruker BioSpin; Karlsruhe, Germany). The chemical shifts were internally referenced to the TSP signal at 0.00 ppm [Bordoni et al., 2011].

Jres. 2D J-resolved 1H spectra were acquired using a double spin echo sequence 9,26, with 2 scans per increment, collecting 8k data points for 512 increments. The first 128 increments were used for

further processing based on T2*-relaxation rates. Spectral width was 7.2 kHz for direct dimension and 80 Hz for indirect (spin-spin coupling) dimension. The interpulse delay was 2.6 s. Data were multiplied with a nonshifted sine-bell function in both dimensions, then Fourier-transformed followed by calculation of the magnitude spectra, which were tilted and symmetrized using functions in Topspin (ver. 3.0, Bruker Biospin, Karlsruhe, Germany). 1D 1H NMR spectra were acquired using a sequence corresponding to one increment of the NOESY pulse sequence with homospoil-gradients and presaturation of the water signal during relaxation (2.0 s) and mixing period (0.1 s). The number of scans was 256, with 64k data points over a spectral width of 12 kHz [Yilmaz et al., 2011]. After the Fourier transform, spectra were normalized so that spectral amplitude was adjusted for solute concentration and for the p1 parameter, to wit the reciprocity coefficient [Hoult, 2000; Tropp et al., 2013] In this way, spectra were comparable according to their absolute intensity. All the pre-processing steps were performed using R-project® platform version 2.15.1. Signal assignment was carried out on the base of literature [Lamanna et al., 2008; Consonni & Cagliani, 2008; Bordoni et al., 2011] and by using Amix software (version 2.1.3, Bruker). Statistical analysis of spectral areas was carried out using R. The analysis of variance (ANOVA) was performed by the command 'anova'. In this way a two-way analysis of variance, comparing both the type of sample and the digestion step, was carried out. This analysis was validated by means of the post-hoc LSD-test (Fisher's Least Significant Difference test), comparing the mean of one group with the mean of another using a set of individual t-test. In this way, it shows in which way samples are different from one other.

Results and discussion

To follow the solubilization of molecules as a result of digestion, spectra were registered by means of the NOESY pulse sequence. This sequence is designed to suppress the water residual signal, while giving for each kind of proton of each substance peaks proportional to the concentration of the substance itself. Through HR-NMR spectroscopy, only soluble molecules can be seen. Insoluble molecules, even if present in suspension in centrifuged sample, will not be shown; they will eventually generate very broad and flat signals. For this reason, only some rare milk proteins, the soluble part such as lactoglobulin, will be seen by NMR before digestion. The rest of the protein content will appear only after the enzymatic disruption and solubilization. HR-NMR spectra of the casein digested in the oral phase, shows mainly broad peaks due to soluble proteins which are too big or denatured to give clear signals.

The only sharp signals appearing in the oral spectra are organic acids (citric acid, acetic acid, formic acid), ethanol, acetone and sugars (lactose). After gastric digestion, the broad signals increase, as a result of a higher solubilization of big molecules (as seen in Figure 9).

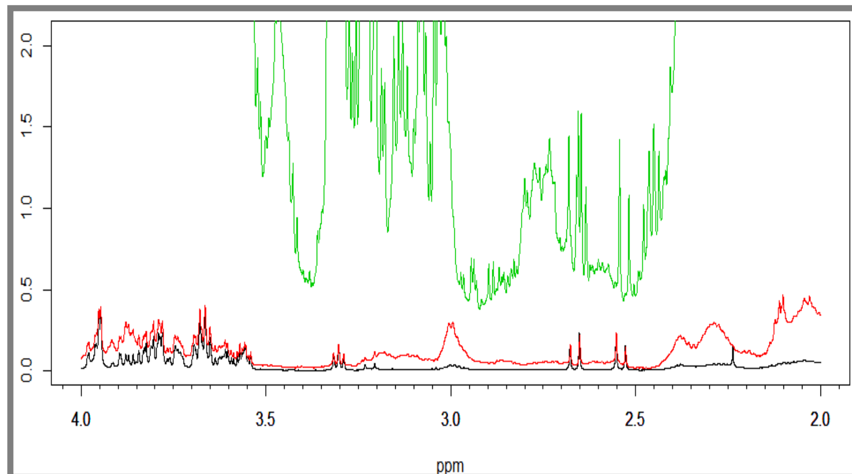


Fig. 9: NOESY1D NMR spectra for the oral (black), gastric (red) and duodenal (green) phase digestates of casein in the sugar region.

Duodenal digestion gives a further increase in those signals, but results also in the appearance of other sharp peaks, belonging to free amino acids. The only observable difference between casein and Lowpept spectra is present after the oral phase of digestion (Fig. 10), since Lowpept is already undergone a peptic digestion. Thus, Lowpept oral spectra show a greater presence in broad signals, belonging to soluble proteins and big peptides.

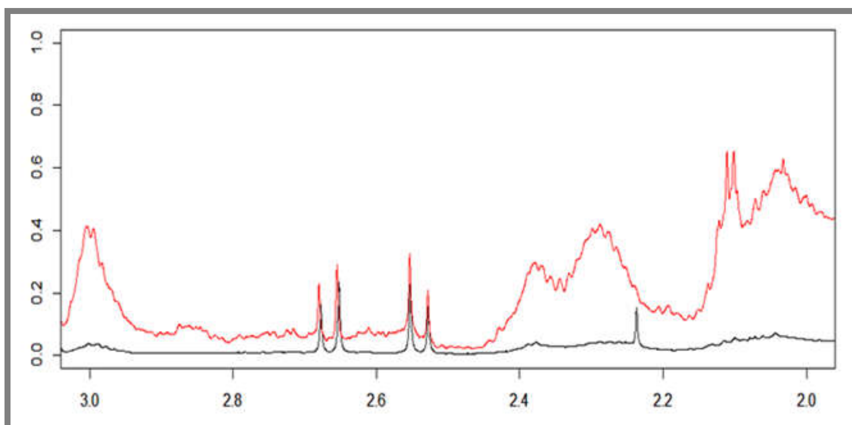


Fig. 10: NOESY1D NMR spectra of oral digestates for casein (black) and Lowpept (red).

Observing total spectral areas it is possible to highlight how, already after gastric digestion, the quantity of small and big peptides increases. In effect, endopeptidases (pepsin) start digesting proteins in the stomach, forming smaller fragments, which can be, according to their dimensions,

visible by NMR. A greater increase in spectral areas is seen, as stated, after the duodenal step of digestion, since the multiple action of the enzymes trypsin and chymotrypsin allows the formation of free aminoacids and small peptides, which can be easily observed by NMR spectroscopy. These molecules, in effect, and aminoacids in particular, can generate sharp peaks, which are observed greatly in the area of the duodenal digested samples.

With the observation of specific spectral areas where aromatic and branched amino acids and amidic residues can be found, it is possible to find out what happens in food products during digestion, especially after the gastric and duodenal steps. We decided to focus our attention on three areas, which show a great variation during these digestive phases. The chosen regions are:

- AROMATIC AMINO ACIDS: from 6.78 to 6.88 ppm especially between oral and gastric phase, since a large signal is present and from 6.88 to 6.92 ppm in particular for the duodenal phase, since sharp peaks appear, like the aminoacid tyrosine.
- BRANCHED AMINO ACIDS: from 0.84 to 0.94 ppm for the gastric digestion, since here again broad peaks appear, and between 0.94 and 1.06 ppm for the duodenal phase since sharp signals occur after this digestion step, such as valine, leucine and isoleucine.
- AMIDIC NH: between 7.3 and 8.9 ppm

Concentrating the analysis on these regions allows the comparison of the two different products digested, to understand their specific profiles and behaviour during and after digestion, in order to underline possible differences between the two.

These regions were compared both for the spectra obtained with the NOESY sequence than for the ones registered with the JRES technique. In both cases, as reported, the Lowpept samples show greater areas than casein, either for gastric and for duodenal phase, even if in a lower measure.

After oral digestion, Lowpept is obviously showing greater spectral areas, since it is already undergone through pepsin hydrolysis. In order to validate the hypothesis of a true difference between the samples, statistical analysis was performed. The technique employed was a two-way analysis of variance (ANOVA) and a post-hoc LSD validation test. In this way, the ANOVA should prove if there are statistically significant differences in samples depending either on the product (Casein or Lowpept) or on the digestive step (oral, gastric or duodenal). Results, reported in Table 4, show significant difference just concerning the phases of digestion.

Through the LSD-test validation, it is demonstrated how these differences are evident only between duodenal phase and the earlier two steps, whilst there's no significant difference between oral and gastric steps.

Table 4: Mean and standard deviation of the spectral areas calculated for casein and Lowpept, both for duodenal and gastric phase. Letters a and b indicate the significance of the difference after LSD post-hoc test comparing the two products, first, and the digestive step, secondly.

	NOESY		**JRES**	
areas	Casein	Lowpept	Casein	Lowpept
gastric	3.66±0.30 [a,a]	3.99±0.24 [a,a]	3.33±0.05 [a,a]	3.82±0.05 [a,a]
duodenal	106.02±0.71 [a,b]	100.00±5.67 [a,b]	107.41±14.73 [a,b]	100.00±9.03 [a,b]

These notions are validated by the observation of NMR spectra obtained through the NOESY technique. Visualizing the lowfield region, which shows proton amidic signals (7.3-8.9 ppm), it is possible to gain information regarding water accessibility in exchange sites in peptides. This chemical exchange determines the disappearance, under the water suppressed signal, of amidic signals when peptides are small enough to make all their amidic proton accessible to the solvent. The aromatic region (6.5-7.5 ppm), instead, shows information on non-exchangeable protons belonging to aromatic amino acids, thus approximately 10% of the entire protein composition. The diagnostic power of this two region is clear observing the evolution of spectral profiles during digestion. Initially, the lowfield area is basically flat before gastric digestion, for what concerns casein, and becomes crowded with broad signals after this digestive step, increasing even more the area after the duodenal phase. In Lowpept, this area seems decreasing after the gastric phase, but it shows another great increase after the intestinal digestion. This can be explained by the fact that, after gastric phase, NH2 show that peptones go into solution more slowly than the rate in which peptones become peptides. In the duodenal phase, the solubilization of peptones starts being preponderant again, therefore the net result is a great increase in NH2 signals. To resume the digestive process of these two different samples some spectral parameters where chosen: the ratio between the area of the two food matrices in the regions corresponding to aromatic (AaA) and branched (AaR) aminoacids and the lowfield region where amidic NH_2 (ANH) can be found. The chosen regions where found to be very informative and among the most variable between the digestive steps. By comparison of the areas of these regions it is possible to assess which sample is greatly digested, generating a higher value in the spectral area considered. This can mean that this type of sample could generate a greater quantity of compounds easily accessible for our bodies, such as free aminoacids. Table 5 and 6 show the values of these areas both for casein and Lowpept after gastric and duodenal digestion. The ratio between casein and Lowpept is usually less than 1 in value (Table 7 and 8). Lowpept seems to generate a greater amount of digestive products such

as peptides and aminoacids.

Table 5 and 6: Mean values and standard deviations for the areas of samples in the different regions. AaA: aromatic aminoacids region, AaR: branched aminoacids region, ANH: amidic NH2 region. CN stands for casein and LP for LowPept. Letters a and b indicate the significance of the difference after LSD post-hoc test comparing the two products, first, and the digestive step, secondly.

NOESY	AaA	AaA	AaR	AaR	ANH	ANH
Area	Casein	Lowpept	casein	Lowpept	casein	Lowpept
Gastric	0.0451±0.0043$^{a, a}$	0.3746±0.0370$^{b, a}$	0.2474±0.0186 $^{a, a}$	7.1613±0.1265 $^{a, a}$	0.1813±0.0233 $^{a, a}$	2.8566±0.0539 $^{a, a}$
Duodenal	0.0518±0.0021$^{a, b}$	0.3904±0.0196 $^{b, b}$	0.3046±0.144 $^{a, b}$	6.9446±0.3468 $^{a, b}$	0.1860±0.0197 $^{a, b}$	2.9810±0.1171 $^{a, b}$

JRES	AaA	AaA	AaR	AaR	ANH	ANH
Area	casein	Lowpept	casein	Lowpept	casein	Lowpept
Gastric	0.0337±0.0050 $^{a, a}$	0.0719±0.0098 $^{a, a}$	0.6387±0.0829 $^{a, a}$	4.7137±0.5895 $^{a, a}$	0.2725±0.0077 $^{a, a}$	2.4819±0.3280 $^{a, a}$
Duodenal	0.0497±0.0012 $^{a, b}$	0.1033±0.0011$^{a, b}$	0.7411±0.0384 $^{a, b}$	5.0639±0.3447 $^{a, b}$	0.2358±0.0015 $^{a, b}$	2.1098±0.1221 $^{a, b}$

Table 7: Results from the comparison of spectral areas in the region chosen for aromatic (AaA) and branched (AaR) aminoacids using NOESY spectra.

Type of AA	Phase	ppm (start)	ppm (end)	Casein/Lowpept ratio
AaA	Oral	6.78	6.88	0.05
AaR	Oral	0.84	0.94	0.11
AaA	Gastric	6.78	6.88	0.87
AaR	Gastric	0.84	0.94	0.96
AaA	Duodenal	6.88	6.92	0.81
AaR	Duodenal	0.94	1.06	1.03

Table 8: Results from the comparison of spectral areas in the region chosen for aromatic (AaA) and branched (AaR) aminoacids using JRES spectra

Type of AA	Phase	ppm (start)	ppm (end)	Casein/Lowpept ratio
AaA	Oral	6.78	6.88	0.23
AaR	Oral	0.84	0.94	0.12
AaA	Gastric	6.78	6.88	0.66
AaR	Gastric	0.84	0.94	0.69
AaA	Duodenal	6.88	6.92	0.86
AaR	Duodenal	0.94	1.06	0.93

Comparing the values of areas in these regions, both for branched and aromatic aminoacids, through a two-way Analysis of Variance, the results were the one represented in Tables 5 and 6, the first letter (a or b) shows the difference between casein and Lowpept, whilst the second letter (a or b) shows the difference between gastric and duodenal phase. The difference between digestive steps is always significant, in specific between the duodenal step and the previous two, in

every region and with every NMR method performed. In addition, NOESY technique, both in the amidic and the aromatic region, shows significant differences among the type of samples, therefore between casein and Lowpept. This is particularly important to understand how real differences between the samples are possible and how this distinction can have an effect on human nutrition. Tyrosine, in effect, lying in the aromatic region, is an aminoacid with many positive functions in the human organisms, such as the synthesis of neurotransmitters like dopamine and norepinephrine [O'Brien et al., 2007], and it has been proven how its assumption is particularly beneficial in conditions of stress or fatigue [Deijen et al., 1999; O'Brien et al., 2007; Mahoney et al., 2007]. In addition to that, other studies have demonstrated how tyrosine can reduce blood pressure in rats [Sved et al., 1979; Yamori et al., 1980]. The branched amino acids analysed, instead, thus leucine, isoleucine and valine, show very similar levels in casein and Lowpept, especially observing NOESY spectra, whilst in JRES spectra, the ratio between the areas of the two samples, shows again a greater value for the Lowpept sample. These amino acids are used for their role in protein synthesis and to produce energy thanks to their aliphatic portion [Brestenský et al., 2015]. They can also contrast fatigue and preserve the immune defense system [Blomstrand, 2006; Calder, 2006], therefore resulting to be very important for our bodies. From these results, it is possible to understand how Lowpept has potentially better nutritional characteristics than casein. Moreover, the fact that a higher quantity of tyrosine is potentially released during digestion from Lowpept further proves the anti-hypertensive properties of this product.

Conclusions

Observing the spectral behaviour of the two samples during digestion, it is possible to see how the area of NMR spectra shows a great increase from the original food product (casein), already in the gastric phase, thanks to protein digestion. Through the study of spectral areas shows it was assessed that the most important step for the further assimilation and absorption of nutrient is the duodenal phase, during which proteins and previously-formed peptides are cut into smaller peptides and free aminoacids. It is evident, therefore, how the digestion of these food matrices is a double step: initially food is digested into big molecules and then, these are split into dimensionally smaller molecules, capable of being absorbed by the intestines. 1H-NMR technique is very useful for the observation of the various steps in protein digestion, allowing to see what happens in every digestive phase and during the entire digestion. As shown, the analysis of determined spectral

areas, can give many informations on the specific hydrolysis of metabolites of interest, such as protein, in order to understand how a sample can be more digested than another and if molecules of greater nutritional and health interested are formed. In this case, it was demonstrated how Lowpept shows, once again, better nutritional properties, in specific resulting more digestible and assimilable than casein. In this way Lowpept can perform more effectively its anti-hypertensive effect, also thanks to the liberation of a greater quantity of beneficial aminoacids such as tyrosine.

REFERENCES CHAPTER 6

Anadón, A., Martínez, M. A., Ares, I., Ramos, E., Martínez-Larrañaga, M. R., Contreras, M. M., Ramos, M., & Recio, I. (2010). "Acute and repeated dose (4 weeks) oral toxicity studies of two antihypertensive peptides, RYLGY and AYFYPEL, that correspond to fragments (90–94) and (143–149) from α s1-casein." Food and chemical toxicology,*48*(7), 1836-1845.

Blomstrand, E. (2006) "A role for branched-chain amino acids in reducing central fatigue." The Journal of nutrition, 136(2), 544S-547S.

Bordoni, A., Picone, G., Babini, E., Vignali, M., Danesi, F., Valli, V., Di Nunzio, M., Laghi, L., & Capozzi, F. (2011). "NMR comparison of in vitro digestion of Parmigiano Reggiano cheese aged 15 and 30 months." Magnetic Resonance in Chemistry, 49(S1), S61–S70.

Bordoni, A., Laghi, L., Babini, E., Di Nunzio, M., Picone, G., Ciampa, A., Valli, V., Danesi, F., & Capozzi, F. (2014) "The foodomics approach for the evaluation of protein bio-accessibility in processed meat upon in vitro digestion" Electrophoresis, 35(11), 1607-1614.

Brestenský, M., Nitrayová, S., Patrás, P., Heger, J., & Nitray, J. (2015). Branched chain amino acids and their importance in nutrition. The Journal of Microbiology, Biotechnology and Food Sciences, 5(2), 197.

Calder, P. C. (2006) "Branched-chain amino acids and immunity." The Journal of nutrition, 136(1), 288S-293S.

Consonni, R., & Cagliani, L.R. (2008) "Ripening and geographical characterization of Parmigiano Reggiano cheese by 1H NMR spectroscopy." Talanta, 76(1), 200–205.

Deijen, J. B., Wientjes, C. J. E., Vullinghs, H. F. M., Cloin, P. A., & Langefeld, J. J. (1999) "Tyrosine improves cognitive performance and reduces blood pressure in cadets after one week of a combat training course." Brain research bulletin, 48(2), 203-209.

Hoult, D.I. (2000) "The Principle of Reciprocity in Signal Strength Calculations - A Mathematical Guide" Concepts in Magnetic Resonance, 12(4), 173-187.

Lamanna, R., Piscioneri, I., Romanelli, V., & Sharma, N. (2008) "A preliminary study of soft cheese degradation in different packaging conditions by 1H-NMR" Magnetic Resonance in Chemistry, 46(9), 828–831.

M.M. Contreras, R. Carrón, M.J. Montero, M. Ramos, I. Recio (2009) "Novel casein-derived peptides with antihypertensive activity" International Dairy Journal, 19(10), 566–573.

Mahoney, C. R., Castellani, J., Kramer, F. M., Young, A., & Lieberman, H. R. (2007) "Tyrosine supplementation mitigates working memory decrements during cold exposure." Physiology & behavior, 92(4), 575-582.

Minekus, M., Alminger, M., Alvito, P., Ballance, S., Bohn, T., Bourlieu, C., Carrière, F., Boutrou, R., Corredig, M., Dupont, D., Dufour, C., Egger, L., Golding, M., Karakaya, S., Kirkhus, B., Le Feunteun, S., Lesmes, U., Macierzanka, A., Mackie, A., Marze, S., McClements, D.J., Ménard, O., Recio, I., Santos, C.N., Singh, R.P., Vegarud, G.E., Wickham, M.S.J., Weitschies, W., & Brodkorb, A. (2014). "A standardised static in vitro digestion method suitable for food–an international consensus." Food & function, 5(6), 1113-1124.

Oberli, M., Marsset-Baglieri, A., Airinei, G., Santé-Lhoutellier, V., Khodorova, N., Rémond, D., Foucault-Simonin, A., Piedcoq, J., Tomé, D., Fromentin, G., & Benamouzig, R. (2015). "High true ileal digestibility but not postprandial utilization of nitrogen from bovine meat protein in humans is moderately decreased by high-temperature, long-duration cooking." The Journal of nutrition, 145(10), pp.2221-2228.

O'Brien, C., Mahoney, C., Tharion, W. J., Sils, I. V., & Castellani, J. W. (2007) "Dietary tyrosine benefits cognitive and psychomotor performance during body cooling."Physiology & behavior, 90(2), 301-307.

Recio, I., López, I., Quirós, A., Hernández, B., Gómez, J.A., Miguel, M., Amigo, L., Ramos, M., Aleixandre de Artiñano, A.,

(2006) "Bioactive peptides identified in enzymatic hydrolyzates of milk casein and method of obtaining same". International Patent Application WO2006/131586 A1.

Roldan, M., Antequera, T., Armenteros, M., & Ruiz, J. (2014). Effect of different temperature–time combinations on lipid and protein oxidation of sous-vide cooked lamb loins. Food chemistry, 149, 129-136.

Sangsawad, P., Kiatsongchai, R., Chitsomboon, B., & Yongsawatdigul, J. (2016). "Chemical and Cellular Antioxidant Activities of Chicken Breast Muscle Subjected to Various Thermal Treatments Followed by Simulated Gastrointestinal Digestion." Journal of food science, 81(10), C2431-C2438.

Sved, A. F., Fernstrom, J. D., & Wurtman, R. J. (1979) "Tyrosine administration reduces blood pressure and enhances brain norepinephrine release in spontaneously hypertensive rats." Proceedings of the National Academy of Sciences, 76(7), 3511-3514.

Tropp, J. (2013) "The Reciprocity Principle in NMR Reception" - Proceedings of the International Society of Magnetic Resonance in Medicine, 21.

Yamori, Y., Fujiwara, M., Horie, R., Lovenberg, W. (1980) "The hypotensive effect of centrally administered tyrosine." European Journal of Pharmacology, 68(2), 201– 204.

Yilmaz, A., Nylberg, N.T., & Jaroszewski, J.W. (2011) "Metabolic Profiling Based on Two-Dimensional J-Resolved 1H-NMR Data and Parallel Factor Analysis." Anlytical Chemistry, 83(21), 8278-8285

CHAPTER 7: FOODOMICS IN THE INVESTIGATION OF THE ANIMAL METABOLOME

Animals can be studied through metabolomics just as humans. In effect, the investigation of animal biofluids through metabolomic techniques is now more and more employed for many reasons. First of all, the holistic approach of metabolomics can be of great help in veterinary studies, allowing the full profiling of animals together with genomics and proteomics assays. Moreover, the foodomics/metabolomics approach can be applied in animal studies in order to investigate on the later quality of animal-derived food products. As stated before, the metabolome is an inventory of small molecules, thus the animal metabolome can represent potential human nutrients and health-promoting compounds and can be associated to a nutrient profile. [Harnly et al., 2007]. Moreover, the animal metabolome can be studied as a model of the human metabolome due to similarities in animal and human metabolism [Dessì et al., 2014]. For all these reasons, the animal metabolome is of great importance in many research field and of great interest. One of the most employed animal model in metabolomics studies is the pig. The pig has been employed as an animal model for many different reasons. In fact, it has been defined as the best “non-primate” model for the investigation of human nutrition [Miller & Ullrey, 1987]. In addition, being a large single-stomached omnivore mammal, the pig's gut physiology is very similar to that of humans. For these reasons it has been employed in recent years in many studies to investigate risk factors for various diseases such as stress, obesity or diabetes and to study the effectiveness and metabolism of drugs [Bustad & McClellan, 1966]. Moreover, the fact that the pig represent an important food source and is employed for development of surgery and for transplants has boosted the research on this animal in many fields, ranging from genomics to immunology [Merrifield et al., 2011]. Therefore, the metabolic characterization of pigs represents the next step to fully profile porcine metabolism and to better understand also the human metabolome and its variations. In this research work, two studies were carried out on pigs. The first one regarded pig colostrum (Chapter 7.1). The scope of this research was to investigate the composition of this peculiar biofluid and assess any possible relation between its profile, sow breeds and piglets' mortality and growth rates. This assessment is particularly important both for for veterinary reasons, in order to help the growing conditions of piglets, and for the profiling of this particular biofluid. The second study (Chapter 7.2) investigated on piglets' plasma and urine. This study investigated the differences in the metabolome of newborn piglets due to the effect of antibiotics and to the presence or absence

of a particular pathology: necrotizing enterocolitis. This disease affects both human and animal premature newborns and therefore the study is very useful also for human medicine. In addition, since the study focused on both the direct effect of the antibiotics (i.e. presence of antibiotic metabolites) and also on the indirect consequences (e.g. absence of disease, thus alteration of metabolic pathways), it is of great value for human application, since it allows to understand the influence of specific factors on different metabolome characteristics.

This study was also particularly useful for the development and improvement of an analytical workflow on the metabolome of plasma and urine and for the assignment of NMR signals.

REFERENCES CHAPTER 7 (INTRODUCTION)

Harnly, J. M., Bhagwat, S., & Lin, L. Z. (2007) "Profiling methods for the determination of phenolic compounds in foods and dietary supplements." Analytical and bioanalytical chemistry, 389(1), 47-61.

Dessì, A., Marincola, F.C., Masili, A., Gazzolo, D., & Fanos, V. (2014) "Clinical Metabolomics and Nutrition: The New Frontier in Neonatology and Pediatrics" BioMed Research International, 2014.

Miller, E. R., & Ullrey, D. E. (1987). "The pig as a model for human nutrition." Annual review of nutrition, 7(1), 361-382.
Bustad, L.K., & McClellan, R.O. (1966) "Swine in Biomedical Research" Science, 152(3728), 1526-1528.

Merrifield, C.A., Lewis, M., Claus, S.P., Beckonert, O.P., Dumas, M.E., Duncker, S., Kochhar, S., Rezzi, S., Lindon, J.C., Bailey, M., Holmes, E., & Nicholson, J.K. (2011) "A metabolic system-wide characterisation of the pig: a model for human physiology" Molecular BioSystems, 7(9), 2577-2588.

7.1 Foodomics to investigate pig colostrum

STATUS: Submitted to the Journal of Animal Science

Title

Metabolomics characterization of colostrum in three sow breeds and its influences on piglets' survival and litter growth rates[1]

Authors

G. Picone,* M. Zappaterra,* D. Luise,* A. Trimigno, F. Capozzi, V. Motta, R. Davoli, P. Bosi, P. Trevisi

Department of Agricultural and Food Sciences (DISTAL), Alma Mater Studiorum - University of Bologna, Viale Fanin 46, Bologna, 40127, Italy;

[*] Contributed equally to the research.

[1]We acknowledge Società Agricola La Fortezza s.s. for the samples of sows colostrum and in particular we gratefully acknowledge the support and assistance of veterinary John Forlenza.

[2]Corresponding author: paolo.trevisi@unibo.it

Role of the PhD candidate: statistical analyses of the NMR spectra, writing of the Materials and Methods and Results and Discussion together with Gianfranco Picone.

Abstract

Colostrum differs from milk and is the first secretion of mammary gland produced in late pregnancy and during the hours immediately preceding and succeeding parturition. This secretion represents an essential vehicle of passive immunity, prebiotic compounds and growth factors involved in intestinal development. Its composition has been investigated mainly in human and cow, but very little is known about pig colostrum metabolome and how it varies between pig breeds and different farrowing parity. Thus, the aim of the present research is to provide new information about pig colostrum composition and the associations between some metabolites, the breed of the sow and the survival and growth rates of their litters. Colostrum samples were gathered from 58 parturitions of sows belonging to three different breeds chosen for their importance in the Italian heavy pig production: 31 Large White, 15 Landrace and 12 Duroc, respectively. Farrowing was not induced, and colostrum samples were collected after the first birth and before the last piglet was born, defatted and centrifuged in a 10kDa cut-off membrane. The

eluted was analysed using ^{1}H-NMR spectroscopy. The Principal Components Analysis (PCA) was assessed on the obtained spectra. In addition, using a Stepwise Regression and a Linear Regression analyses the metabolites named after the signals assignment were tested for their associations with piglets' performances. 25 metabolites were identified, comprehending monosaccharides, disaccharides (such as lactose), organic acids (lactate, citrate, acetate and formate), peptides (such as creatine) and others compounds, including nucleotides. PCA results evidence a clustering due to breed and season effects. Lactose was the main compound determining the assignment of the samples into different clusters according to the sow breed. The amount of dimethylamine identified in colostrum was associated with the piglets' mortality at three days after birth (P = 0.004). This is the first study characterizing swine colostrum metabolome using the ^{1}H-NMR spectroscopy technology and the results obtained will contribute to improve the knowledge on colostrum's deeper composition and variability. Furthermore, this work may help understanding the compounds that influence piglets' survival and growth in addition to the best- known immunoglobulin colostrum fraction.

Key words: metabolome; pig breeds; colostrum; ^{1}H-NMR spectroscopy; lactose; piglet survival

Introduction

The pre-weaning litter environment has been proven to affect the pigs development and performances during later life (Vallet et al., 2016) and in particular colostrum intake, coupled with birth weight, was found to influence piglets' growth and mortality (Devillers et al., 2011; Ferrari et al., 2014; Decaluwé et al., 2014b). Colostrum provides new-borns with energy and passive immunity (Noblet et al., 1997; Rooke and Bland, 2002): in particular, most of the literature concerns the effects of the different immunoglobulins on piglets' health and survival capacities . Studies assessed on human and bovine colostrum suggested important roles in new-borns' health also for other bioactive molecules, such as nucleotides, oligosaccharides, organic acids and peptides (Gopal and Gill, 2000; Schlimme et al., 2000; Korhonen, 2013; He et al., 2014), but little is still known about the presence of these metabolites in sows' colostrum and their association with piglets' performances. Furthermore, to date little or no information about pig breed influence on colostrum composition is available and most of the knowledge about metabolites composition of swine colostrum was produced on samples gathered after farrowing induction, fact that may alter colostrum composition (Foisnet et al., 2011). In this study, 58 colostrum samples were collected during a natural parturition with the aims i) to analyse through a NMR-based metabolomics

approach the colostrum compounds with a maximum 10 kDa molecular weight in three pig breeds, ii) to evaluate which factors mostly affect the colostrum composition, iii) to test the associations between the identified metabolites, the sow maternal attitude, and the piglets' survival and growth rates.

Materials and methods

The procedures complied with Italian law pertaining to experimental animals and were approved by the Ethic-Scientific Committee for Experiments on Animals of the University of Bologna, Italy.

Animals and sampling

Fifty-eight colostrum samples were collected from 58 different farrowing of pure breed sows: 12 Duroc (D), 15 Landrace (L) and 31 Large White (LW). All sows were raised in the same farm and were not treated with antibiotics and medical products during gestation. Farrowing was not induced, and the colostrum sampling was carried out during natural parturition, after the birth of the first piglet and before the parturition of the last, across all teats. All samples were immediately frozen at -20 °C and then stored at -80 °C until the samples preparation for NMR analysis. For each sow, parity and data related with the reproductive performances were recorded: the number of piglets alive and the litter body weight (LBW) were recorded at birth and at day 3, cleansed from the weight of the piglets dead. The litter weight gain (BWG) was then calculated for the period from birth to day 3. Furthermore, the number of weaners per litter was recorded as well as the occurrence of diarrhoea during suckling (1= presence of diarrhoea events from piglets' birth until weaning, 0= absence of diarrhoea event).

Colostrum preparation for ^{1}H-NMR analysis

Colostrum was de-frozen, carefully mixed by inversion, and 15 ml of each colostrum sample were diluted 1:1 with pure water. To each diluted sample, 0.02% of sodium azide was added, to inhibit bacterial growth during the sample preparation. Then the sample was defatted through a centrifugation at 4 °C for 30 minutes at 1500 x g. The aqueous phase was transferred to a clean falcon avoiding the outer layer of fat, and centrifuged again; this procedure was repeated three times. 5 ml of the obtained aqueous phase were then transferred in Amicon Ultra 10 kDa membrane centrifugal filters (Merck Millipore, Merck KGaA, Darmstadt, Germany) and filtered by centrifugation at room temperature for 90 minutes at 5500 x g. This step was needed to eliminate immunoglobulins and other proteins with high molecular weight. The eluted sample was then weighted and lyophilized.

^{1}H-NMR measurements

At the time of sample processing, for each gram of lyophilized sample 250 µl distilled water was added. Eighty µl of thawed sample was centrifuged at 14,000 × g for 5 min (Scilogex D3024 High Speed Micro-Centrifuge, Rocky Hill, CT, USA) and then added to 720 µl of distilled water and 100 µl of a D_2O solution of 3-(trimethylsilyl)-propioniate-2,2,3,3-d_4 (TMSP) (Cambridge Isotope Laboratories Inc, Tewksbury, MA, USA) with a final concentration of 6.25 mM. ^{1}H-NMR spectra were recorded at 298 K with an AVANCE spectrometer (Bruker BioSpin, Karlsruhe, Germany) operating at a frequency of 600.13 MHz, equipped with an autosampler with 60 holders. The HOD residual signal was suppressed by applying the NOESYGPPR1D sequence (a standard pulse sequence included in the Bruker library) incorporating the first increment of the NOESY pulse sequence and a spoil gradient. The HOD residual signal was suppressed by applying the first increment of the NOESY pulse sequence and a spoil gradient. Each spectrum was acquired using 32 K data points over a 7211.54 Hz spectral width (12 ppm) and adding 256 transients. A recycle delay of 5 s and a 90° pulse of 11.4 µs were set up. Acquisition time (2.27 s) and recycle delay were adjusted to be 5 times longer than the longitudinal relaxation time of the protons under investigation, which has been not longer than 1.4 s. The data were Fourier transformed and phase and baseline corrections were automatically performed using TopSpin version 3.0 (Bruker BioSpin, Karlsruhe, Germany). Signals were assigned through a combination of literature assignments and by the use of a multimedia library included in Chenomx NMR suite 8.2 professional software (Chenomx, Edmonton, Alberta, Canada).

Data analysis

The collected data were aggregated to create homogenous classes. The farrowing parity (from 1 to 3 were classified as 1, parity order > 4 were considered as 2), the parturition season (from 1 to 4). The seasons were assigned as follows: 1= parturition between the 1st of December and the 28th of February; 2= between the 1st of March and the 31st of May; 3= between the 1st of June and the 31st of August; 4= between the 1st of September and the 30th of November. Among the studied animals, 6 sows gave birth during season 1, 19 during season 2, 21 during season 3 and 12 during season 4. Statistical analyses on spectra data were performed using R computational language (ver. 3.1.2) and MATLAB (ver R2014b, MathWorks Inc.). Each NMR spectrum was processed by means of scripts developed in-house as follows: spectra baseline was adjusted by employing the signals identification algorithm named “baseline.peakDetection” from R (version 3.1.2) package “Baseline” (https://cran.r project.org/web/packages/baseline/index.html). Chemical shift referencing was

performed by setting the TMSP signal to 0.00 ppm. The following spectral regions were removed prior to data analysis: the regions including only noise (the spectrum edges between 11.00 and 8.65 ppm and between 0.15 and -1.00 ppm), the NMR signal which is strongly affected by the residual solvent signals (water, between 4.90 and 4.50 ppm) and the glycerol's signals from 3.82 and 3.76 ppm, from 3.69 and 3.63 ppm and from 3.60 and 3.54 ppm**.** Spectra were then normalized by means of probabilistic quotient normalization method (PQN) and binned. The first operation is aimed at removing possible dilution effects. The second one avoids the effect of signals misalignments among different spectra due to variations in chemical shift of signals belonging to some titratable acids. The binning operation is performed by subdividing the spectra into 369 bins, each integrating 120 data points (0.0219 ppm each). In order to focus on the real information contained in the spectra, bins that an average higher value than noise were selected. In this way, a total of 201 bins were kept. The spectra obtained were then analysed through an unsupervised multivariate approach using Principal Component Analysis (PCA). The PCA was conducted on the 201 bins matrix to identify the outlier samples, and test the existence variables contributing to samples clustering. The multivariate models were calculated and the results were visualized on both scores and loadings' plot. This multivariate analysis is the predominant linear dimensionality reduction technique used when dealing with scientific dataset and it is defined as an unsupervised method as it does not use class labels for discriminating between groups. After reducing the dimensionality of the dataset, it produces new linear combinations of the originals variables which can be plotted in a score plot (Bailey et al., 2003). In order to determine the spectral regions encompassing most of the discriminative information, bins with a loading value greater than 1% of the overall standard deviation of all loading values were selected. The identified metabolites included in the significant bins emerged from the loadings' plot and additional metabolites relevant for their biological function were selected and grouped in a new dataset named C-dataset. The C-dataset was used to conduct an analysis of variance (ANOVA) with the aim to confirm if the amounts of the identified compounds were influenced by the effects of breed and farrowing season identified with the PCA and parity order. The model utilized for this analysis was:

$y = \beta 0 + \beta p * b + \beta p * s + \beta p * o + E$

Where:

$\beta 0$ was the intercept;

βp was the corresponding regression coefficient;

y was the amount of each identified metabolite;

b was the sow breed (LW; D; L);

s was the farrowing season (1; 2; 3; 4);

o was the parity order (1; 2);

E was the error.

This first part was conducted to test if sows breed influences colostrum profile, and if in addition to breed there are other "environmental" variables affecting colostrum quality (in this case the farrowing season and the parity order were tested).

Then, a stepwise regression analysis was used to select, among the metabolites included in the C-dataset and sows' reproductive performances, the variables that had to be included with the breed, the farrowing season and the parity order in the final GLM model for the identification of the metabolites related to piglets' performances. This statistical analysis involves starting with no variables in the model and adding gradually each metabolite and sow reproductive parameter (the litter weight and the number of alive piglets at birth) to evaluate which one of the colostrum identified compounds and sows' reproductive abilities most influenced the piglets' survival and growth. The results obtained from the stepwise regression analysis were then confirmed through General Regression Analysis (GLM), considering as *y* variables the BWG, the number of weaned piglets, the number of dead piglets from birth until day 3 or the number of piglets dead from day 3 to weaning; as independent variables were considered the sows breed, the farrowing season, the parity order of the sow and the significant factors identified through the stepwise regression analysis. The utilized GLM model was:

$y = \beta 0 + \beta p * b + \beta p * s + \beta p * o + \beta p * p + \beta p * q + \beta p * r + \ldots + E$

Where:

y was BWG, the number of weaned piglets, the number of piglets dead from the birth day until day 3 or the number of dead from day 3 to weaning;

β0 was the intercept;

βp was the corresponding regression coefficient;

b was the sow breed (LW; D; L);

s was the farrowing season (1; 2; 3; 4);

o was the parity order (1; 2);

p; *q*; *r* were the significant metabolites identified through the stepwise regression analysis;

E was the error.

Finally, all the variables that did not show an effect on the dependent variables were removed

from the model and only the significant effects were maintained. The *prcomp* function of R environment was used to perform the PCA analysis on bins matrix (R Core Team, 2015). The ANOVA analysis, the stepwise regression analysis and the regression model were carried out on SAS software using PROC REG and PROC GLM respectively (SAS® 9.4, SAS Inst. Inc., Cary, NC).

Results

Dataset description

In Table 1 the complete dataset is detailed Duroc sows had on average a lower number of piglets at birth (8.92 ± 2.28) respect to Landrace (12.60 ± 1.72) and Large White (11.90 ± 2.26), while the newborns of Landrace and Large White breeds presented a lower weight at birth (1.38 ± 0.15 kg and 1.43 ± 0.16 kg, respectively) compared to Duroc piglets (on average 1.59 ± 0.23 kg).

Colostrum spectra

In Figure 1 a NMR molecular profile of sow colostrum is represented. The ^{1}H spectrum is mainly dominated by the carbohydrate signals overlapping in the midfield region between 3.49 and 4.49 ppm (Figure 1B). Those belong to lactose and nucleotides sugars such as UDP-glucose and UDP-galactose and nucleotide as UMP. In this area, also signals from creatine and its products arise (3.04-3.05 ppm). Amino acids mainly fall in the upfield region, between 0.99 and 3.49 ppm, together with signals from organic acids (Figure 1A). In this part of the spectrum fall also signals from threonine (1.33 ppm) and alanine (1.49 ppm), lactic acid (1.33 ppm), acetic acid (1.92 ppm), succinic acid (2.41 ppm) and citric acid (2.54 and 2.67 ppm). Finally, in the downfield region (Figure 1C) signals of different phenolic compounds can be observed, but in this case, only formic acid was assigned (8.4 ppm), together with signals from the nucleotide sugars UDP-glucose and UDP-galactose (5.5-6 ppm, 7.9-8 ppm) and UMP (8.1 ppm, 5.98-5.99 ppm, 4.42 ppm) as listed in table 2. The 25 compounds have been identified through a combination of literature assignments and by the use of a multimedia library included in Chenomx NMR suite 8.2 professional software (Chenomx, Edmonton, Alberta, Canada).

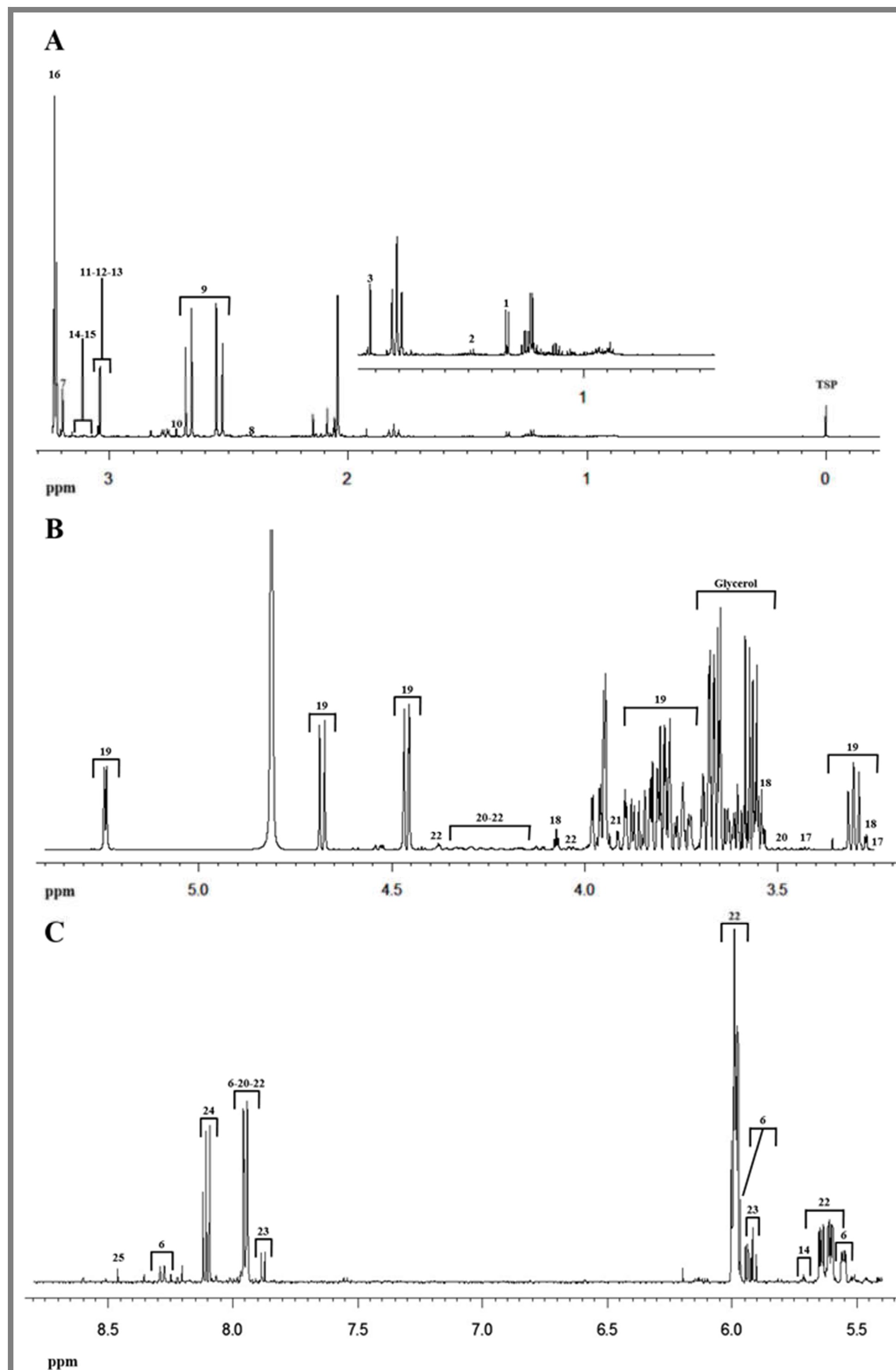


Figure 1. Typical ^{1}H-NMR spectrum of aqueous extract of colostrum. ^{1}H-NMR spectrum registered on a colostrum sample. The spectrum has been split into three parts for the sake of clarity. Some resonances have been assigned by using Chenomx software and listed in table 2: A) Aliphatic or upfield region; B) Carbohydrate or midfield region, characterized by the presence of signals belonging to sugars and glycerol and C) Aromatic region or downfield region.

Factors affecting colostrum composition

After alignment, normalization and binning, the dataset contained a total of 58 colostrum spectra characterized by 201 bins and PCA was applied on it to investigate differences on the metabolome

between groups. For sow's parity order, in the total colostrum spectra no PCA clustering was identified (data not showed). While Figure 2A and B shows that samples clustered on PC1-PC2 due to the effects of the sow breeds (Figure 2A) and on PC2-PC3 due to the farrowing seasons (Figure 2B).

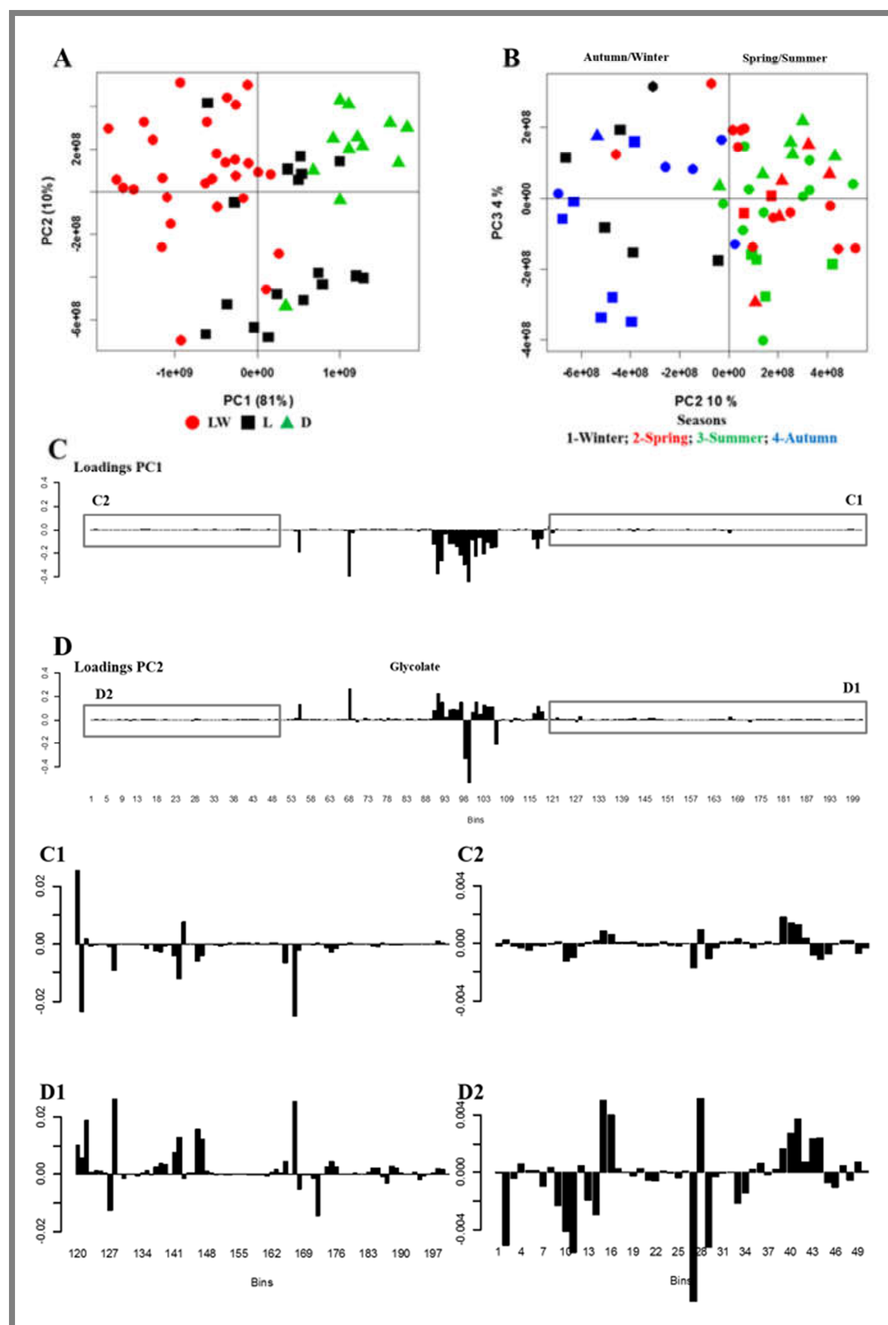


Figure 2. Scoreplots of PCA on ^{1}H-NMR binned spectra of colostrum obtained from different sow breed: A) PC1 vs PC2 and B) PC2 vs PC3. The first two PCs represent the 91% of the total variance. C-D) Loadings bar-plot for spectral bins along PC 1 and 2 respectively. Downfield (C1 and D1) and upfield (C2 and D2) regions of C and D loadings bar-plot were expanded on the vertical scale to appreciate the presence of small bar plot.

The PC1 explained the 81% of the total variance and separated the colostrum spectra of D and LW, while PC2 (10% of the variance) discriminated the L colostrum composition from the ones of LW

and D sows. The PC2-PC3 plot highlighted the season effect, in particular along PC2 the differences in the colostrum spectra due to seasons 1 and 4 (winter-autumn) against season 2 and 3 (spring-summer). The weighting of each variable (bin) is represented by the loadings plot in Figure 2C and 2D in which are displayed the loadings from PC1 and PC2 respectively as a bar plot, where each bar corresponds to a single spectral variable (bin). The main bins accounting for the spectral differentiation and their relative chemical shift were listed in the Supplementary Table 1 (SS1). As emerging from SS1 table, most of the signals included in these discriminant bins were assigned to the corresponding metabolites. The C-dataset, which was used for the following statistical analyses, resulted to be composed of 25 metabolites, listed in Table 2. The parity, breed and season effects on colostrum composition were then confirmed through the ANOVA analysis on the identified metabolites described in the C-dataset, and the results are reported in Table 3. For sow's parity order, not significant data are obtained. Parity order showed only trends nearing statistical significance for succinate (P = 0.097), creatine phosphate (P = 0.091), creatinine (P = 0.061) and UDP-glucose (P = 0.061) (data not show). Breed and season resulted to be the major factors affecting the assigned compounds. Indeed, the amounts of alanine (P = 0.004; P = 0.004), citrate (P < 0.0001; P = 0.006), succinate (P < 0.0001; P = 0.024), dimethylamine (P = 0.030; P = 0.0001), creatine (P <0.0005; P < 0.0005), creatine phosphate (P = 0.003; P < 0.0001), cis-aconitate (P = 0.030; P < 0.0001), taurine (P = 0.002; P = 0.001), glycolate (P < 0.0001; P = 0.001) and UMP (P = 0.001; P = 0.009) were affected by both breed and the season of the farrowing respectively. Season affects significantly (P < 0.0001) the amount of acetate, creatinine and formate, where a higher level was registered for all the metabolites during the cold months. On the other hand, the amounts of o-acetylcholine (P < 0.0001), sn-glycerophosphocholine (P = 0.036), UDP-n-acetylglucosamine (P = 0.001), lactose (P < 0.0001), myo-inositol (P = 0.001) and UDP-glucose (P < 0.0001) were affected only by sow breed. In particular, the colostrum of L samples showed upper signals for UDP-Glucose, UDP-galactose and sn-glycerophosphocholine compared to the other two breeds, while LW colostrum was characterized by a major quantity of lactose, taurine, myo-inositol and glycolate.

Factors affecting litter performances

The stepwise regression analysis revealed that, in addition to the influence of sows' reproductive performances (the litter weight and the number of alive piglets at birth) some peculiar metabolites can contribute to piglets' survival and growth parameters (Table 4). In particular, the litter weight at birth and the concentration of acetate significantly entered in the model for BWG (P < 0.0001

and *P* = 0.003, respectively); the higher number of alive piglets at birth and the increased concentration of colostrum cis-aconitate resulted to be important variables affecting positively the number of weaned piglets (*P* < 0.0001 and *P* = 0.019, respectively), while dimethylamine (*P* = 0.0002) and taurine (*P* = 0.013) entered as variables in the model for the number of dead piglets per litter at day 3. There was no influence of farrowing season and parity order on BWG, the number of weaned pigs or the number of dead piglets at day 3. The outcomes of the stepwise regression analysis were then tested with the GLM, and the results reported in Table 5. Both the higher average piglets' weight at birth (*P* < 0.0001) and the colostrum acetate concentration (*P* = 0.003) affected positively BWG (Table 5). The number of dead piglets at day 3 was mainly influenced by the concentration of dimethylamine (*P* = 0.001) and taurine (*P* = 0.027) in colostrum and partially by the litter size at birth (*P* < 0.1). In addition, the litter size at birth (*P* = 0.001) and a lower level of cis-aconitate in colostrum (*P* = 0.010) and the sow breed showed a trend (*P* = 0.021) were significantly associated with the number of weaned piglets.

Discussion

This is the first study based on ^{1}H-NMR metabolomics approach describing in three pig breeds the colostrum metabolome profile, the factors underlying its composition and the associations between colostrum metabolites and litter's fitness during suckling. The three breeds showed different reproductive abilities in accordance with literature (Blasco et al., 1995; Sonderman and Luebbe, 2008), with L and LW sows exhibiting a higher average number of piglets alive at birth compared to D sows. These differences between breeds are also visible at the colostrum composition level (Simmen et al., 1990): considering the whole spectrum, the three breeds display clustering tendency, with the colostrum lactose amount explaining most of the colostrum composition differences between breeds. In particular, L and LW breed samples presented higher values of lactose. Lactose rate in cow milk is commonly associated with the health status of mammary gland, as higher lactose concentrations are positively correlated to healthier mammary glands and low amounts of this compound indicate the existence of intrammamary infections (Park et al., 2007). Considering the data available for the present work, it is not possible to support the same consideration in lactating sows, due to the absence of reference value for sow milk. Furthermore, differences on UDP-n-acetylglucosamine. UDP-glucose and UDP-galactose were observed between breeds. UDP-sugars are intermediate products in cellular protein glycosylation and in the synthesis of lactose and other sugars, and are known to have also autocrine/paracrine

signalling functions (Lazarowski and Harden, 2015). UDP-sugars are detected by the purinergic receptor P2Y14, a G Protein Coupled Receptor, found in many epithelia and in immune and inflammatory cells (Lazarowski and Harden, 2015). In weaned pigs, P2Y14 is expressed in the gastro-intestinal tract, particularly in the pyloric area (Colombo et al., 2014). We are not aware of the entity of the gastrointestinal expression of this receptor in piglets included in the present study; however it can be hypothesized that the presence of UDP-sugars may be involved in the activation/maturation of the neonate pig gastrointestinal immunity. The breed effect on colostrum composition is not exclusively confined to sugars profile, which for sure are the most abundant metabolites between the identified, but also to other compounds, including alanine, citrate, succinate, creatine phosphate, creatinine, dimethylamine, cis-aconitate, myo-inositol and o-acetilcholine. Furthermore, the obtained colostrum spectra were affected also by the farrowing season: the samples gathered during winter and autumn exhibited differences in colostrum compositions respect to colostrum secreted during spring and summer. These differences could be ascribed to the environmental conditions affecting sows' performances: compounds such as acetate, which showed to be more abundant during cold seasons, may reflect the nutritional state of sows during cold months. Acetate in sows, typically fed high fibre diets in gestation, is the main product of hindgut fermentations; acetate is also a precursor for mammary synthesis of fat milk. Thus the season effect may reflect a reduced presence of fibre in diet during the warmer seasons or a different mammary usage of acetate in favour of the more efficient use of glucose (Linzel et al., 1969). Indeed, during cold seasons, the higher feed intake makes available larger amounts of energy and nourishing compounds respect to the lower daily feed intake characterizing sows living at higher environmental temperatures (Gourdine et al., 2006). In addition, farrowing season affected also the creatine pathway: in particular, creatine and creatine-phosphate amounts during the period ranging from September to February were significantly lower than in spring and summer; on the contrary creatinine was higher during the same period. Creatine is an important nutrient for the newborn (Brosnan and Brosnan, 2007), thus variations in creatine content of colostrum may have nutritional relevance. In mice, it has been shown that milk creatine is extracted from the circulating plasma by the mammary gland, which conversely has little or no capacity to synthesize creatine (Lamarre et al., 2010). No research data is available for sow colostrum, but it can be assumed that also in this case variations in colostrum may reflect variations in blood creatine concentration. Here the variations in the ratio creatine and creatine-phosphate to creatinine may have resulted from a higher degradation of the first two compounds

into creatinine, but there is no evidence that this reaction occurs in the mammary gland. The increasing amount of creatinine level is in general associated with a higher mobilization of stored proteins and indirectly with fat and lean levels in the body mass (Van Niekerk et al., 1963). A recent study of Decaluwé et al. (2014a) associated an increased amount of blood creatinine on 1st day of lactation with lower feeding levels in sows during late gestation period. However, we could not control feed intake in the days before farrowing and do not know if it changed with season. Thus further research is necessary to explain variations of creatine and related compounds in colostrum.

Some of the identified compounds were associated with litter weight gain during the first three days of life and to piglets' survival rates at day 3 and at weaning. In particular, we suppose that the positive effect of acetate on BWG was linked to the role of this compound in *de novo* synthesis of lipids and glucose (den Besten et al., 2013) and adipogenesis stimulation (Hong et al., 2005). Additionally, taurine colostrum concentration showed a positive correlation with piglets' survival rate at three days of life. Taurine was already proven to play a critical role in neonatal development, including the development of central nervous system and other tissues (Bryson et al., 2001; Aerts and Van Assche, 2002). Due to the essential role of this compound in neonatal period, it can be easily understood why a higher concentration of taurine in sows' colostrum exerted positive effects on piglets' survival during first days of life, independently from sows' breed. As regards the number of dead piglets at three days of life, this performance was associated with the concentration of dimethylamine secreted in colostrum. Dimethylamine is a biogenic amine, synthesised by bacterial action, known for its mutagenic, irritative and barrier-disrupting properties (Galli et al., 1993; Fluhr et al., 2005). The irritative effect of this compound can explain the significant negative association found in the present research between the colostrum dimethylamine amount and the piglets' survival capacity. The observed increase in dimethylamine secretion could be a direct effect of the higher occurrence of bacteria fermentations in animal feeds (Juszkiewicz et al., 1980), then ingested by sows and secreted in colostrum and milk.

Similarly to dimethylamine, also cis-aconitate was negatively associated to piglets' survival capacity from birth to weaning. As cis-aconitate is an intermediate compound synthesized by several enteric bacteria, comprising *Salmonella* enterica (Lewis and Escalante-Semerena, 2006) and *Escherichia coli* (Shimizu, 2013), the increased amount of this tricarboxylic acid in colostrum may be a marker of the presence of pathogenic bacteria strains infecting maternal gut and mammary gland. Maternal microbiome is proved to affect newborn digestive tract, and pathogenic strains can easily pass from sow's gut to piglets' enteric tract. Thus, the increase of cis-aconitate in colostrum

secretion might be the first sign of gut dysbiosis, which may then affect also piglets' microbiome and digestive tract homeostasis.

In conclusion, this study demonstrates that colostrum metabolome is greatly affected by breed and, in particular, Duroc sows showed colostrum compositions unlike any other. This result agrees with the generally accepted view that the differences among Duroc and white coated pig breeds may originate from distinct genetic origins, and consequently suggests that further genetic studies may help explaining the variations found among breeds in colostrum compositions. From the observation of the results obtained it can be suggested that the different temperatures occurring during the year affect sows' metabolism and, in turn, can also affect colostrum composition. Among the identified metabolites, acetate and taurine showed their positive effects on piglets' performances and survival rate, while dimethylamine and cis-aconitate exerted a negative influence on new-borns capacity to survive. This research represents a preliminary step towards the knowledge of pig colostrum composition and it is one of the first studies focusing on the associations between different swine colostrum compositions and litter performances using the ^{1}H-NMR technique. Further investigations are needed to extend the identification of the different compounds in swine colostrum and to elucidate their effects on new-borns. Furthermore, the possible interaction between sows' feeding and microbiota in the modulation of colostrum metabolome deserves further investigations.

TABLES

Table 1. Mean and standard deviation of the measured parameters for each breed and for the total population.

	[1] D	[2] L	[3] LW	Total
Number of sows	12	15	31	58
Order of parturition	2.750	4.067	4.161	3.845
Number of piglets born alive per litter				
Mean	8.917	12.600	11.903	11.466
SD	2.275	1.724	2.256	2.494
[4]Average LBW at birth, kg				
Mean	14.033	17.207	16.890	16.381
SD	3.679	2.081	2.828	3.060
Average piglet's weight at birth, kg				
Mean	1.588	1.376	1.432	1.450
SD	0.226	0.153	0.155	0.184
Number of alive piglets per litter at 3 days				
Mean	8.250	12.133	10.871	10.655
SD	1.865	1.767	2.291	2.453
Number of dead piglets per litter at 3 Days				
Mean	1.600	1.750	1.583	1.619
SD	1.342	0.500	0.996	0.974
[4]Average LBW at 3 days, kg				
Mean	16.400	21.167	19.758	19.428
SD	3.993	3.059	4.264	4.211
Average piglet's weight at 3 days, kg				
Mean	2.001	1.761	1.826	1.846
SD	0.275	0.240	0.213	0.244
[5]BWG, kg				
Mean	3.085	4.453	4.103	3.983
SD	0.864	1.461	1.771	1.598
Number of weaned piglets				
Mean	7.333	11.133	10.032	9.759
SD	1.875	1.552	2.316	2.423
Incidence of diarrhoea				
Mean	3	1	4	8

[1] D stands for Duroc.
[2] L stands for Landrace.
[3] LW stands for Large White.
[4] LBW stands for litter Body Weight.
[5] BWG stands for the average litter Body Weight Gain from birth to day 3.

Table 2. Assignment table of the identified metabolites present in the ^{1}H-NMR spectra of colostrum at pH 7.420 and considered in the C-dataset. Chemical shift values are referenced to TMSP proton signals at 0.00 ppm. Chemical shift values are referenced to TMSP proton signals at 0.00 ppm. Glycerol (3.568, 3.661 and 3.793 ppm has not been listed as it has not been included in the PC analysis).

Assigned number	^{1}H chemical shift (ppm) [a]	Compound
1	1.332 (d)	Lactate
2	1.486 (d)	Alanine*
3	1.923 (s)	Acetate
4	2.028 (s)	N-Acetylglutamate*
5	2.063 (s)	N-Acetylglucosamine*
6	2.089 (s) - 5.552(dd) - 5.967 (d) - 7.944 (d) - 8.287 (d)	UDP-N-Acetylglucosamine
7	2.147 (s) -3.222 (s)	O-Acetylcholine
8	2.408 (s)	Succinate*
9	2.539 (d) - 2.667 (d)	Citrate
10	2.720 (s)	Dimethylamine
11	3.039 (s)	Creatine
12	3.046 (s)	Creatine phosphate
13	3.050 (s)	Creatinine
14	3.119 (d) - 5.712 (m)	cis-Aconitate
15	3.204 (s)	Choline
16	3.231(s) - 4.330 (m)	sn-Glycerophosphocholine
17	3.272(t) - 3.532 (dd) - 4.073 (t)	Myo-Inositol
18	3.259 (t) - 3.428 (t)	Taurine
19	3.302 (t) -3.684:3.906 (m), 3.980 (d) 4.461 (d) - 4.679 (d) - 5.243 (d)	Lactose
20	3.480 (s) - 4.142:4.278 (m) -5.607 (dd) -5.967 (m) - 7.940 (d)	UDP-glucose
21	3.935 (s)	Glycolate
22	4.142:4.278 (m) - 4.379 (m) - 5.664 (dd)- 5.990 (m) - 7.942 - 7.995(d)	UDP-Galactose
23	5.917 (d) - 7.879 (d)	Uridine
24	8.406 (s)	Formate
25	4.423 (t) - 5.990 (m) - 8.102 (d)	UMP

[a] d, doublet; dd, doublet of doublets; m, multiplet; s, singlet; t, tripleTable 3. Effects of sow breed and season on identified colostrum metabolites.
* These compounds were included in the C- dataset for their specific biological role.

Table 3. Effects of sow breed and season on identified colostrum metabolites

Metabolite	Breed[1]			SEM	*P*-value	Season[2]				SEM	*P*-value
	D	L	LW			1	2	3	4		
Acetate	9.57	11.17	9.90	0.91	0.673	13.59	7.55	5.95	13.77	0.89	<0.0001
Lactate	5.38	6.70	8.88	1.85	0.456	4.93	10.18	8.71	4.13	1.81	0.137
Alanine	1.77	2.20	2.44	0.17	0.004	1.65	2.50	2.51	1.88	0.17	0.004
Citrate	209	301	257	13	<0.0001	246	286	265	228	12	0.006
Succinate[3]	2.21	3.26	3.50	0.19	<0.0001	2.59	3.26	3.55	2.55	0.21	0.024
Dimethylamine	2.86	4.44	4.51	0.49	0.029	2.30	5.19	5.39	2.85	0.48	0.0001
Creatine	39.8	59.9	58.7	3.4	<0.0005	40.9	59.0	64.7	46.6	3.3	<0.0005
Creatine Phosphate	3.40	7.99	8.01	0.88	0.003	1.90	8.50	10.78	4.69	0.86	<0.0001
Creatinine	13.6	16.7	16.5	1.1	0.400	19.9	11.8	11.6	19.2	1.1	<0.0001
Cis-Aconitate	1.41	1.93	1.74	0.14	0.030	1.10	2.14	2.28	1.26	0.14	<0.0001
O-Acetylcholine	77.1	196.9	156.6	13.9	<0.0001	101.7	148.4	171.7	152.3	13.6	0.093
sn-Glycerophosphocholine	446	543	414	38	0.018	430	457	507	477	37	0.816
Choline	7.82	10.62	9.91	1.19	0.455	8.86	11.64	9.43	7.87	1.16	0.340
N-Acetilglutamate	6.35	9.90	14.03	3.77	0.134	7.94	9.69	8.53	14.21	3.69	0.683
N-Acetilglucosamine	10.9	15.4	11.7	1.6	0.072	10.6	14.0	13.8	12.3	1.6	0.621
UDP-N-Acetilglucosamine	22.4	34.4	33.9	2.2	0.001	26.2	33.4	33.8	27.4	2.1	0.072
Lactose	458	579	811	30	<0.0001	535	644	667	619	30	0.168
Taurine	1.57	4.10	6.05	0.87	0.002	0.98	6.23	6.19	2.21	0.85	0.001
Myo-Inositol	63.53	76.95	82.86	3.76	0.001	62.58	73.17	81.13	80.91	3.68	0.070
UDP-Glucose	6.07	9.78	6.23	0.54	<0.0001	6.08	7.30	8.09	7.97	0.54	0.280
Glycolate	28.1	39.8	45.8	2.4	<0.0001	28.4	41.8	45.9	35.6	2.3	0.001
Uridine	3.14	3.72	3.23	0.38	0.551	3.05	3.55	3.49	3.37	0.38	0.936
UDP-Galactose	32.6	74.0	42.5	4.0	<0.0001	39.3	54.2	61.4	44.0	3.9	0.006
Formate	4.49	4.43	4.03	0.35	0.380	6.27	3.04	2.29	5.67	0.34	<0.0001
UMP	13.30	24.20	21.60	1.70	0.001	21.82	16.04	16.84	24.08	1.91	0.009

Mean of the identified metabolites are expressed as absolute area

[1] The Breed is assigned as D for Duroc, L for Landrace and LW for Large White.

[2] The seasons were assigned as follows: 1= if the parturition was included in the period between the 1st of December and the 28th of February; 2= between the 1st of March and the 31st of May; 3= between the 1st of June and the 31st of August; 4= between the 1st of September and the 30th of November.

[3]Succinate is the only metabolite showing an effect of the parity order (*P* value = 0.039), in particular with a parity order less than 4 succinate has a mean of 2.77, while the same metabolite with a parity order more than or equal to 4 has a mean of 3.20, a SEM of 0.15.

Table 4. Results of the stepwise regression analysis.

Variables	Coefficient	SE coefficient	F value	*P*-value
Model for BWG[1] (R^2 = 0.4286; C(p) = 0.8735)				
Intercept	0.715	0.533	1.79	0.186
Acetate	0.10433	0.03348	9.71	0.0029
Average piglet's weight at birth	0.00846	0.00164	26.63	<.0001
Model for the number of weaned piglets (R^2 = 0.4343; C(p) = 2.0849)				
Intercept	4.801	1.385	12.01	0.001
Cis- Aconitate	-0.90181	0.37157	5.89	0.0185
The number of alive piglets at birth	0.58395	0.09863	35.05	<.0001
Model for the number of piglets dead per litter at day 3 (R^2 = 0.2304; C(p) = 29.1881)				
Intercept	-0.333	0.27	1.44	0.2352
Dimethylamine	0.33082	0.08318	15.82	0.0002
Taurine	-0.11423	0.04432	6.64	0.0127

[1]BWG stands for the litter Body Weight Gain from birth to day 3.

Table 5. Results of the GLM analysis with the significant variables affecting litter body weight gain at day 3, the number of dead piglets at day 3 and the number of weaned pigs.

Variables	Coefficient	SE	*P*-value
GLM for BWG[1]			
Acetate[2]	0.104	0.033	0.003
Average piglet's weight at birth	0.008	0.002	<.0001
GLM for Number of dead piglets at day 3			
Intercept	-1.219	0.557	0.033
Dimethylamine[2]	0.296	0.084	0.001
Taurine[2]	-0.100	0.044	0.027
Number of alive piglets at birth	0.084	0.046	0.074
GLM for Number of weaned pigs			
Breed:		0.46	0.021
LW[3]	9.92		
L[4]	10.59		
D[5]	8.31		
Cis-Aconitate[2]	-0.952	0.354	0.010
Number of alive piglets at birth	0.413	0.111	0.001

[1]BWG stands for litter Body Weight Gain from birth to day 3 of life.
[2]Metabolites concentrations were considered in area arbitrary unit.
[3]LW stands for Large White.
[4]L stands for Landrace.
[5]D stands for Duroc

REFERENCES CHAPTER 7.1

Aerts, L., and F. A. Van Assche. 2002. Taurine and taurine-deficiency in the perinatal period. J. Perinat. Med. 30:281–286. doi: 10.1515/JPM.2002.040.

den Besten, G., K. Lange, R. Havinga, T. H. van Dijk, A. Gerding, K. van Eunen, M. Müller, A. K. Groen, G. J. Hooiveld, B. M. Bakker, and D.-J. Reijngoud. 2013. Gut-derived short-chain fatty acids are vividly assimilated into host carbohydrates and lipids. Am. J. Physiol. Gastrointest. Liver Physiol. 305:G900–10. doi:10.1152/ajpgi.00265.2013.

Bailey, N. J., M. Oven, E. Holmes, J. K. Nicholson, and M. H. Zenk. 2003. Metabolomic analysis of the consequences of cadmium exposure in Silene cucubalus cell cultures via 1 H NMR spectroscopy and chemometrics. Phytochemistry, 62(6), 851-858.

Blasco A., J.P. Bindanel, and C.S. Haley. 1995. Genetic and neonaltal survival. In: Varley, M.A. editor, The neonatal pig: Development and survival. CAB International,Wallingford, UK, p. 17-37.

Brosnan, J.T., and M. E. Brosnan. 2007. Creatine: endogenous metabolite, dietary, and therapeutic supplement. Annu Rev Nutr. 2007;27:241–61. doi: 10.1146/annurev.nutr.27.061406.093621.

Bryson, J. M., S. C. Jackson, H. Wang, and W. L. Hurley. 2001. Cellular uptake of taurine by lactating porcine mammary tissue. Comp. Biochem. Physiol. - B Biochem. Mol. Biol. 128:667–673.

Colombo, M., D. Priori, P. Trevisi, and P. Bosi. 2014. Differential Gene Expression in the Oxyntic and Pyloric Mucosa of the Young Pig. PLoS One 9:e111447. doi:10.1371/journal.pone.0111447.

Decaluwé, R., D. Maes, A. Cools, B. Wuyts, S. De Smet, B. Marescau, P. P. De Deyn, and G. P. J. Janssens. 2014a. Effect of peripartal feeding strategy on colostrum yield and composition in sows. J. Anim. Sci. 92:3557–3567. doi:10.2527/jas2014-7612.

Decaluwé, R., D. Maes, B. Wuyts, A. Cools, S. Piepers, and G. P. J. Janssens. 2014b. Piglets' colostrum intake associates with daily weight gain and survival until weaning. Livest. Sci. 162:185–192. doi: 10.1016/j.livsci.2014.01.024.

Devillers, N., J. Le Dividich, and A. Prunier. 2011. Influence of colostrum intake on piglet survival and immunity. Animal 5:1605–12. doi:10.1017/S175173111100067X.

Dieterle, F., A. Ross, G. Schlotterbeck, and H. Senn. 2006. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in ^{1}H NMR metabonomics. Anal. Chem. 78:4281-4290. doi: 10.1021/ac051632c.

Ferrari, C. V., P. E. Sbardella, M. L. Bernardi, M. L. Coutinho, I. S. Vaz, I. Wentz, and F. P. Bortolozzo. 2014. Effect of birth weight and colostrum intake on mortality and performance of piglets after cross-fostering in sows of different parities. Prev. Vet. Med. 114:259–266. doi:10.1016/j.prevetmed.2014.02.013.

Fluhr, J. W., D. Kelterer, S. Fuchs, M. Kaatz, R. Grieshaber, P. Kleesz, and P. Elsner. 2005. Additive impairment of the barrier function and irritation by biogenic amines and sodium lauryl sulphate: A controlled in vivo tandem irritation study. Skin Pharmacol. Physiol. 18:88–97. Original Paper. doi: 10.1159/000083709.

Foisnet, A., C. Farmer, C. David, and H. Quesnel. 2011. Farrowing induction induces transient alterations in prolactin concentrations and colostrum composition in primiparous sows1. J. Anim. Sci. 89:3048–3059. doi:10.2527/jas.2010-3507.

Galli A., M. Paolini, G. Lattanzi G, G. Cantelli-Forti, and G Bronzetti. 1993. Genotoxic and biochemical effects of dimethylamine. Mutagenesis. 8(3):175-8.

Gopal, P. K., and H. S. Gill. 2000. Oligosaccharides and glycoconjugates in bovine milk and colostrum. Br. J. Nutr. 84

Suppl 1:S69–74.

Gourdine, J. L., J. P. Bidanel, J. Noblet, and D. Renaudeau. 2006. Effects of season and breed on the feeding behavior of multiparous lactating sows in a tropical humid climate. J. Anim. Sci. 84:469–480.

He, Y., S. Liu, S. Leone, and D. S. Newburg. 2014. Human colostrum oligosaccharides modulate major immunologic pathways of immature human intestine. Mucosal Immunol. 7:1–14. doi: 10.1038/mi.2014.20.

Hong, Y. H., Y. Nishimura, D. Hishikawa, H. Tsuzuki, H. Miyahara, C. Gotoh, K. C. Choi, D. D. Feng, C. Chen, H. G. Lee, K. Katoh, S. G. Roh, and S. Sasaki. 2005. Acetate and propionate short chain fatty acids stimulate adipogenesis via GPCR43. Endocrinology 146:5092–5099. doi: 10.1210/en.2005-0545.

Juszkiewicz T., B. Kowalski, E. Pastucha.1980. Occurrence and formation of nitrosamines in animal feeds. IARC Sci Publ. 31:589-94.

Korhonen, H. J. 2013. Production and properties of health-promoting proteins and peptides from bovine colostrum and milk. Cell. Mol. Biol. (Noisy-le-grand). 59:12–24.

Lamarre, S. G., Edison, E. E., Wijekoon, E. P., Brosnan, M. E., & Brosnan, J. T. 2010. Suckling rat pups accumulate creatine primarily via de novo synthesis rather than from dam milk. The Journal of nutrition, 140(9), 1570-1573. doi: 10.3945/jn.110.125476.

Lazarowski, E. R., and T. K. Harden. 2015. UDP-Sugars as Extracellular Signaling Molecules : Cellular and Physiologic Consequences of P2Y 14 Receptor Activation. Mol Pharmacol. 2015 Jul;88(1):151-60. doi: 10.1124/mol.115.098756.

Lewis, J. A., and J. C. Escalante-Semerena. 2006. The FAD-dependent tricarballylate dehydrogenase (TcuA) enzyme of Salmonella enterica converts tricarballylate into cis-aconitate. J. Bacteriol. 188:5479–5486. doi: 10.1128/JB.00514-06.

Van Niekerk, B. D., J. T. Reid, A. Bensadoun, and O. L. Paladines. 1963. Urinary creatine as an index of body composition. J. Nutr. 79:463–473.

Noblet, J., J. Y. Dourmad, M. Etienne, and J. Le Dividich. 1997. Energy Metabolism in Pregnant Sows and Newborn Pigs. J. Anim. Sci. 75:2708–2714.

Ogawa, S., T. Tsukahara, T. Imaoka, N. Nakanishi, K. Ushida, and R. Inoue. 2016. The effect of colostrum ingestion during the first 24 hours of life on early postnatal development of piglet immune systems. Anim Sci J. 87(12):1511-1515. doi: 10.1111/asj.12573.

Park, Y. K., H. C. Koo, S. H. Kim, S. Y. Hwang, W. K. Jung, J. M. Kim, S. Shin, R. T. Kim, and Y. H. Park. 2007. The analysis of milk components and pathogenic bacteria isolated from bovine raw milk in Korea. J. Dairy Sci. 90:5405–5414. doi: 10.3168/jds.2007-0282.

Rooke, J. A., and I. M. Bland. 2002. The acquisition of passive immunity in the new-born piglet. Livest. Prod. Sci. 78:13–23.

Schlimme, E., D. Martin, and H. Meisel. 2000. Nucleosides and nucleotides: natural bioactive substances in milk and colostrum. Br. J. Nutr. 84 Suppl 1:S59–68.

Shimizu, K. 2013. Metabolic Regulation of a Bacterial Cell System with Emphasis on Escherichia coli Metabolism. ISRN Biochem. 2013:1–47. doi: 10.1155/2013/645983.

Simmen, F. A., K. Y. Whang, R. C. M. Simmen, G. A. Peterson, M. D. Bishop, and K. M. Irvin. 1990. Lactational variation and relationship to postnatal growth of insulin-like growth factor-I in mammary secretions from genetically diverse sows. Domest. Anim. Endocrinol. 7:199–206.

Sonderman,J.P., and J. J. Luebbe. 2008. Semen production and fertility issues related to differences in genetic lines

of boars. Theriogenology 70: 1380–1383. doi:10.1016/j.theriogenology.2008.08.009.
Vallet, J. L., J. A. Calderon-Diaz, K. J. Stalder, C. Phillips, R. A. Cushman, J. R. Miles, L. A. Rempel, G. A. Rohrer, C. A. Lents, B. A. Freking, and D. J. Nonneman. 2016. Litter of origin trait effects on gilt development. J. Anim. Sci. 94:96–105. doi: 10.2527/jas.2015-9644.

Vallet, J. L., J. R. Miles, L. A. Rempel, D. J. Nonneman, and C. A. Lents. 2015. Relationships between day one piglet serum immunoglobulin immunocrit and subsequent growth, puberty attainment, litter size, and lactation performance. J. Anim. Sci. 93:2722–2729. doi: 10.2527/jas.2014-8535.

Vallet, J. L., J. R. Miles, and L. A. Rempel. 2013. A simple novel measure of passive transfer of maternal immunoglobulin is predictive of preweaning mortality in piglets. Vet. J. 195:91–97. doi: 10.1016/j.tvjl.2012.06.009.

R Core Team. 2015. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Wu, J., M. Domellöf, A.M. Zivkovic, G. Larsson, A. Öhman, and M.L. Nording. 2016. NMR-based metabolite profiling of human milk: A pilot study of methods for investigating compositional changes during lactation. Biochem. Biophys. Res. Commun. 469: 626-632. doi:10.1016/j.bbrc.2015.11.114.

7.2 Foodomics for the investigation of the pig metabolome

DATA SOON TO BE PUBLISHED – Here only the NMR findings are reported

Title

Untargeted metabolomics reveals effects of oral anti-biotic intake to prevent pre-term neonatal piglets from necrotizing enterocolitis

Authors

Ping-Ping Jiang,†, Alessia Trimigno,‡, Jan Stanstrup,¶, § Bekzod Khakimov,‡ Søren Balling Engelsen, ‡ Per T. Sangild,† Lars O. Dragsted ¶,*

† Department of Veterinary and Animal Sciences, ¶ Department of Nutrition, Sports and Exercise, ‡ Department of Food Science, University of Copenhagen, Frederiksberg, Denmark; ‡ Department of Agricultural and Food Sciences, University of Bologna, Campus di Scienze degli Alimenti, Cesena, Italy.
§Present address: Jan Stanstrup, Steno Diabetes Center, Gentofte, Denmark. E-mail: jpzs@steno.dk
*Address correspondence to Prof. Lars O. Dragsted, Tel: +45-35 33 26 98, Fax: +45-35 33 24 69, E-mail: ldra@nexs.ku.dk

These authors contributed equally to the manuscript.

Competing Interests: The authors declare no conflict of interest.
This work is supported by the Innovation Fund Denmark (the NEOMUNE project, #0603-00774B and project #5195-00001B).

Role of the PhD candidate: analysis of the NMR spectra of piglets' biofluids, statistical analysis and writing of the main draft of the paper, together with Ping Ping Jiang (in charge for the LC-MS part and sample collection).

ABSTRACT

Prophylactic oral and systemic antibiotics prevent necrotising enterocolitis (NEC) in premature neonates, which is a severe microbiome-involving gut disorder. It is hypothesised that prophylactic antibiotics can affect the metabolism of preterm neonates, reflected as change in the plasma and urinary metabolome. In the current study, plasma and urinary metabolomes of formula-fed preterm pigs, serving as a model for NEC, that received antibiotics immediately after birth for four days were profiled by LC-MS and 1H-NMR in parallel, and were compared with those of pigs that did not receive antibiotics. In total, 44 metabolites were detected with significantly differential abundance in plasma and/or urine by both technologies. In antibiotics-

treated pigs, tryptophan metabolism is inclined towards the kynurenine pathway over the serotonin pathway shown by specific metabolites. Metabolites associated with the gut microbiome, including 3-phenyllactic acid, 4-hydroxyphenylacetic acid and phenylacetylglycine, all from phenylalanine, and three bile acids showed lower levels in AB pigs where the gut microbiome was extensively attenuated. The appearance of microbial metabolites in body fluids may serve as early signs of NEC, which is important for timely clinical interventions. Besides the evident NEC-preventing effect, longer-term effects of antibiotics on nutrient homeostasis in preterm neonates warrant further investigations.

Keywords: antibiotics, metabolomics, necrotising enterocolitis, UPLC-MS, NMR, preterm pigs

Introduction

Necrotizing enterocolitis (NEC) is a very common disease occurring in pre-term newborns. Its causes and treatments are still currently discussed, though some risk factors have been proved. In effect, NEC is believed to be triggered by prematurity, firstly, type of nutrition (i.e. enteral) and gut microbiota [Schnabl et al., 2008]. Nowadays, many studies have in fact linked the microbiome as one of the main causes of NEC [Mai et al., 2011], though it is not proved to be one determined pathogen causing NEC alone, but more a generic bacterial overgrowth [Lee & Polin, 2003; Lin & Stoll, 2006]. This can be caused by many characteristics of a newborn intestine, such as weak immunological response, diet malabsorption, reduced gut motility and function; these factors can increase the immobility of gut contents and both things can induce a greater microbial fermentation, which will impact on the already weak mucosal epithelium of the infant intestines [Neu et al., 2005]. As a consequence, this barrier can be even more weakened by the growing bacterial population, causing a greater and spreading inflammation, which could results in other medical conditions [Neu & Walker, 2011]. In last four decades antibiotics have been used for the treatment of NEC [Bell et al., 1973]. Some doubts on their employment, though, still exist, since they might cause resistance and later intestinal diseases, while long-term effects are not known. Most studies have shown that employment of antibiotics for the treatment of NEC greatly reduces the disease incidence in premature and low-birth-weight infants [Bury & Tudehope, 2001; Downard et al., 2012]. Jensen and co-workers investigated the impact of the prophylaxis with three antibiotics on the health status and the incidence and pathogenesis of NEC in pre-term delivered piglets. The hypothesis of a positive impact of the broad-spectrum antibiotics was proven, since no treated piglet developed NEC,

whilst the contrary was experienced in the control group. In this study we investigated blood plasma and urine samples derived from the same piglets and applied untargeted metabolomics approach prior to investigating metabolic effects of the treatment on the piglets and to better understand the effect of the antibiotics.

Materials and methods

Experimental design

Through caesarean section, twenty-four piglets (Landrace X Large White X Duroc) were delivered pre-term on day 105 of gestation from two healthy sows. Piglets were immediately transferred to thermally-regulated incubators with oxygen supply until respiration was stable. Then, they were fitted with umbilical and orogastric catheters and immunized with maternal plasma. All these procedures and further experimental details are described in a previously published work [Jensen et al., 2014] and were approved by the National Committee on Animal Experimentation in Denmark (protocol number 2004/561-910). Piglets were stratified in relation to birth weight and sex and divided into two groups: control and intervention. The intervention group received three oral and systemic broad-spectrum antibiotics: Ampicillin, Gentamycin and Metronidazole, commonly used for the treatment of NEC and/or sepsis in intensive care units for newborns. The antibiotics were given both orally and by intramuscular injection to ensure good systemic and intraluminal concentrations (Table 1). The treatment was given right after feeding with an oral bolus. Control pig received instead a corresponding quantity of saline solution. Pigs were checked for signs of discomfort or weakness and euthanized if necessary. Blood samples were taken 24h and 48h after the first feed after an in vivo absorption test with galactose, useful for the investigation of the galactose absorption capacity by the apical sodium/glucose cotransporter 1 [Jensen et al., 2014]. After five days, pigs given a bolus of lactulose and mannitol for in vivo intestinal permeability tests. Then, they were all euthanized and tissue and urine samples were collected [Jensen et al., 2014].

NMR sample preparation

Urine samples were stored at room temperature for 20-30 minutes, then centrifuged at 6000 rpm for 5-10 mins. 540 microliters were taken and inserted into 1.5 mL Eppendorf tubes and 60 microliters of 7.4 pH phosphate buffer 1.5 M KH2PO4 in deuterated water with 0.1% internal reference trimethylsylilpropionic acid (TSP) were added and NaN3. Samples are then strongly mixed and ready for spectral acquisition. In total, 12 samples of urine were available: 7 controls

and 5 treated. EDTA Tubes from BD (Becton, Dickinson and Company) are used for containing 500 microliters of plasma and 500 microliter of buffer B. 500 mL of buffer is so composed: 380 mL of water, 0.4 g of TSP, 10.05 g of NA2HPO4*7H20, 5 mL 4% NaN3. The buffer is set at pH 7.4 by addition of 1M solutions of either NaOH or HCl. The remaining volume to 400 mL is water, then 100 mL of deuterated water are added. Plasma samples are not shaken nor centrifuged not to form bubbles or foam, thus are ready for spectral acquisition. A total of 21 (11 controls and 10 treated) plasma samples were available.

Table 1: Dosages, Reasons of Use and Treatment time for the three antibiotics employed.

Antibiotic	Dosage, mg/kg		Target Bacteria	Treatment time
	Oral	Systemic		
Ampicillin	50	50	Gram + (-)	Twice a day
Gentamycin	2.5	2.5	Gram – (+), Aerobes, Mycoplasma	Once a day
Metronidazole	10	10	Anaerobes	

NMR acquisition

Data Analysis

12 urine and 21 plasma spectra were ready for data analysis after phase and baseline correction on Topspin (Bruker Topspin 3.5, Bruker BioSpin). The assignment of metabolites' resonances was carried out through literature, web databases and the Chenomx software (Chenomx NMR Suite 8.2, Chenomx Inc.). Further data analysis was carried out on MatLab (MatLab R2014b 8.4.0, MathWorks . The internal standard TSP was employed as a chemical shift reference in both cases. Spectral alignment was performed only in regions of interest where found necessary, through the employment of the icoshift algorithm [Savorani et al.,2010]. Solvent regions and uninteresting regions (water, urea, TSP) were removed, as well as signals from impurities. Samples were then normalized through the probabilistic quotient algorithm [Dieterle et al., 2006] and analysed by means of visual inspection and principal component analysis (PCA), applied both on the whole spectrum and on fixed intervals (iPCA). Pareto scaling was applied as a pre-processing tool before the calculation of the PCA models. After the first inspection, informative regions were selected for their discriminative power and analyzed through Analysis of Variance (ANOVA).

Results and discussion

Urine spectra

Already by careful visual inspection, urine spectra show some peculiar markers of both the

disease and the intervention. These signals were later investigated by statistical analysis. Nonetheless, the urine NMR spectra of anti-biotic treated piglets were dominated by the signals corresponding to mannitol, which was one of the major metabolite fed to piglets together with antibiotics. In addition, several signals of antibiotics including ampicillin and metronidazole were exclusively detected in treated piglets only (Fig. 1). The mannitol region was tendenciously higher in the antibiotics group (Fig. 2), as also found by LC-MS (data not shown). The piglets are fed with mannitol and lactulose 3-5 h prior to euthanasia in order to test for intestinal permeability. Lactulose/mannitol ratio is in fact an index of gut permeability: mannitol is a small molecule, thus index of small intestinal permeability, whilst lactulose, slightly bigger, is an index of colonic permeability [Camilleri et al., 2010]. Lactulose employs paracellular transport, whilst mannitol transcellular transport, though they are both absorbed passively and not metabolized. In patients with Chron's disease or cirrhosis the L/M ratio is higher [Dastych et al., 2008], meaning that lactulose is more absorbed. The ratio has been used also to investigate the permeability in prenatal babies, to test whether the use of human milk and not formula was better [Taylor et al., 2009] and it was found that when babies were fed human milk, the ratio was lower.

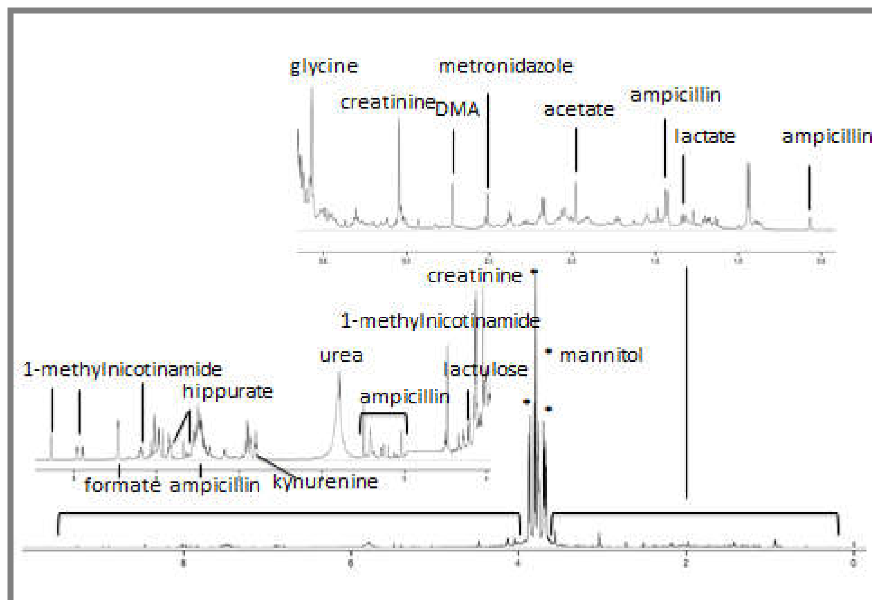


Fig. 1: Representative NOESY1D ^{1}H NMR spectrum obtained on a piglet urine sample.

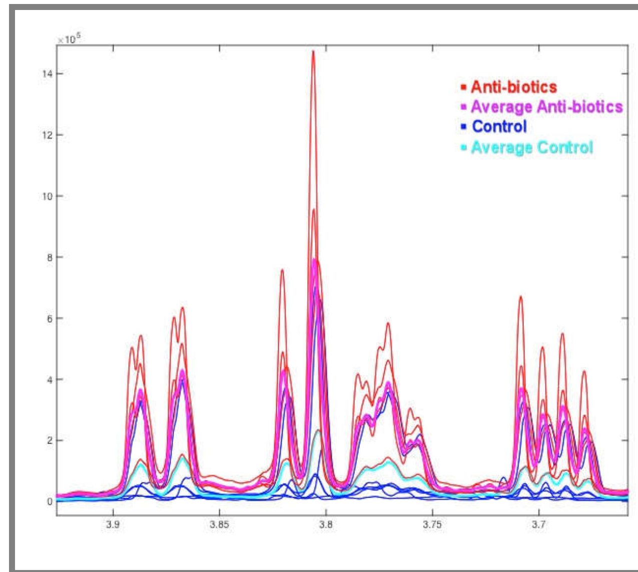


Fig. 2.: Mannitol signals recorded in urine NOESY1D NMR spectra. In red are the signals from piglets treated with antibiotics, with the average represented in purple, and in blue the spectra from control piglets, with the average represented in light blue.

The ratio of L/M was found to be higher in rats with NEC than compared to healthy rats [Zani et al., 2010]. In this case, the level of mannitol appear generally higher in treated piglets: in fact, if the surface area is reduced, like in the case of NEC, the output of mannitol is reduced, thus its urinary concentration will decrease [Pearson et al., 1982]. Likewise, in this study it was not surprising that mannitol was the major metabolite discriminating the two groups of piglets (data not shown). It is worth mentioning that the level of mannitol measured in the urine of piglets by NMR and enzymatic assay depicted high correlation (R2=0.86), as shown in Figure 3.

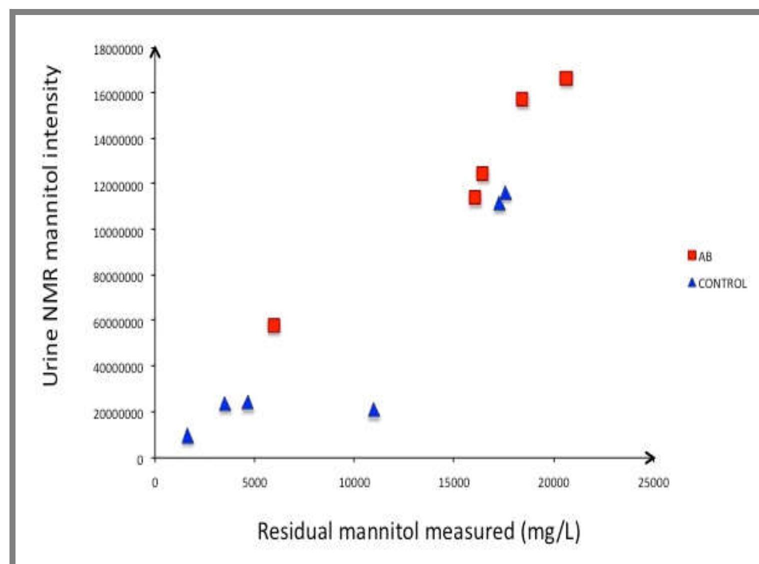


Fig. 3: Correlation between the mannitol measured by enzymatic assay and the mannitol seen through NMR spectroscopy in urine.

In order to investigate metabolic effects other than mannitol, PCA was perfomed on the pareto-scaled NMR spectra after removal of the chemical shift regions that corrrespond to mannitol (Fig. 4).

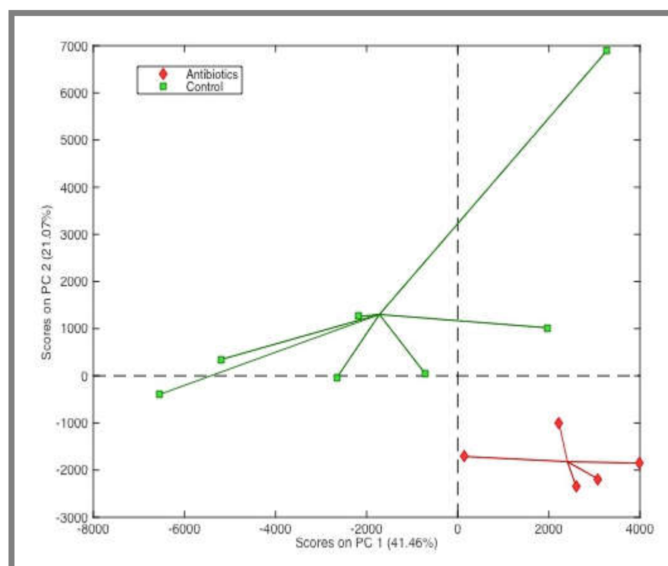


Fig. 4: PC1 vs PC1 scoreplot of the PCA model calculated for urine NOESY1D NMR spectra without the region containing mannitol signals. A separation is seen between control (green) and treated (red) samples.

PC1 loadings of the newly calculated PCA model was mostly dominated by the signals corresponding to antibiotics including Ampicillin and Metronidazole. Interval PCA (iPCA) calculated on 100 consecutive regions of the entire NMR spectra for whole dataset confirmed these findings, and these regions were further inspected. A metabolite heatmap which represents correlation coefficients between iPCA based selected NMR spectral regions over samples showed positive and/or negative significant correlation of spectral regions. This enabled ready interpretation of spectral regions and interactions of identified biomarker metabolites. The results of this calculation are shown in Figure 5.

It is clear how the signals from the antibiotics are highly correlated. Signals from ampicillin at 1.44 ppm and 7.4-7.5 ppm are highly correlated, and also correlated with the signal at 0.6 ppm which is a penicilloic acid from ampicillin–also seen by LC-MS (data not shown)-, and the metronidazole signals at 2.5 ppm. Signals from ampicillin in the region between 5 and 5.5 ppm are also highly correlated (highlighted in red). When only the antibiotic treated piglets are selected (data not shown), the correlations are even stronger between the 0.6 ppm signal, 1.4 ppm signal and 7.5 ppm signal, and the same occurs for the two regions between 5 and 5.5 ppm.

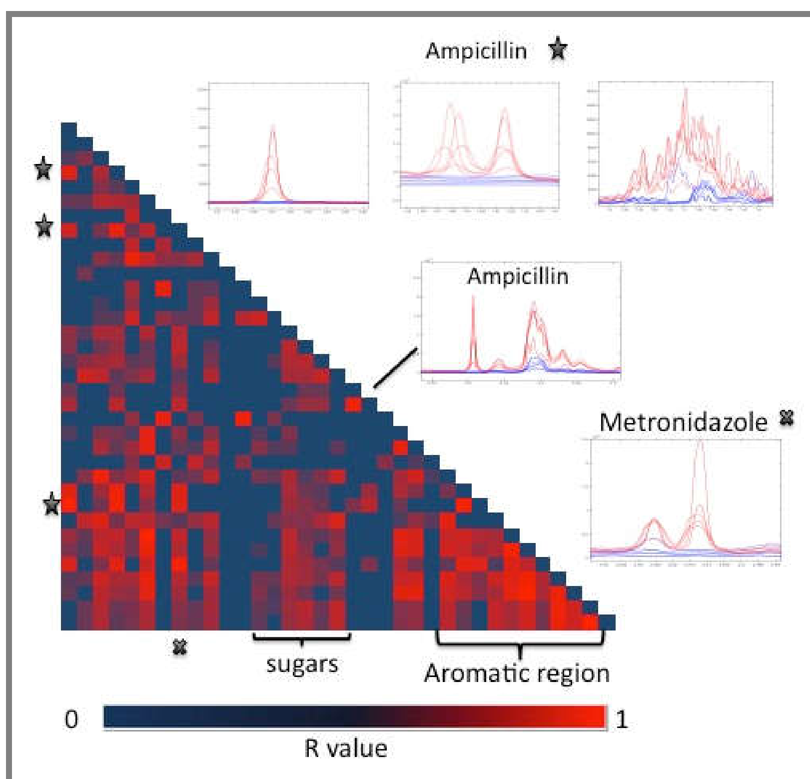


Fig. 5: Metabolite heatmap for the whole spectral set, showing only the significant ($p<0.05$) r values. Signals highly correlated and important for discrimination between the two groups are shown, with the antibiotic group represented in red and the control group in blue.

The antibiotic metabolites' resonances have been assigned differently and variously in literature, but in our assessment the signals that can be defined are listed in Table 2, with some references. Signals from the third antibiotic, gentamycin, and its metabolites have not been found either in urine or in plasma spectra. This is due to the fact that this molecule is not metabolized and has a very fast wash-out time of 6 hours. For this reason, since samples have been drawn at least 12 hours later, it is clear how no sign of this metabolite can be found in the biofluid samples, as also seen by LC-MS (data not shown).

Table 2: Chemical shifts and information on the antibiotic signals found in the spectra.

Chemical shift (ppm)	Multiplicity	Molecule	References
0.57	s	Ampicillin Penicilloic acids	Shamsipur et al., 2002; Connor et al., 1994
1.44	d	Ampicillin (methyl alpha and beta)	Tung et al., 2000
2.52	s (shape of doublet due to metabolism)	Metronidazole	Allars et al., 1985
4.96	d	Ampicillin	NMRDB
5.35	d	Ampicillin	NMRDB
5.5	d	Ampicillin	NMRDB
7.4-7.5	m	Ampicillin	NMRDB

These signals, and other informative regions, were tested through ANOVA to investigate where

there was significant difference between the two groups. Other regions that showed significant differences between the two groups were a doublet at 4.2-4.4 ppm, which could arise from the resonance of some carbohydrate, and the region between 1.45 and 1.95 ppm, where the two groups display very different profiles. In this region, aminoacids such as lysine, leucine, glutamine, alanine, arginine and threonine are found. These aminoacids have been found to decrease with the occurrence of NEC in serum of pre-term infants [Becker et al., 2000]. The observed spectra tend to confirm this difference (Fig. 6).

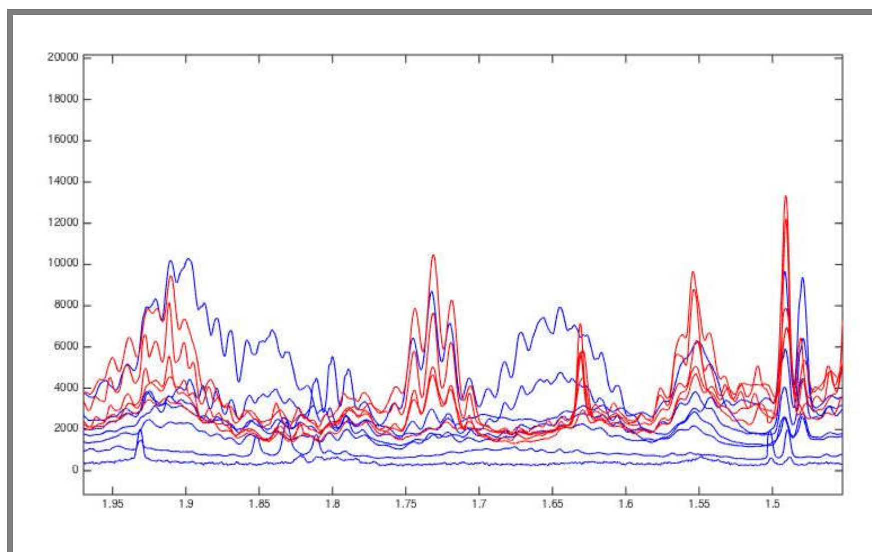


Fig. 6: Urine NOESY1D NMR spectra of the region between 1.45 and 1.95 ppm, displaying significant difference between the two groups due to the presence of very different spectral profile. Amino acid signals are in fact higher in treated samples (red), whilst two controls (blue) show signals from 2-aminoadipate.

In addition, by visual inspection, it is possible to see that 1-methylnicotinamide is higher in the urine of the antibiotic group, whilst the control group showed no signal for this molecule. The LC-MS data have found differences in urinary nicotinamide (data not shown). The methylated metabolite can be found in urine.

Plasma spectra

Plasma NOESY1D spectra show very broad signals from lipids and proteins (Fig. 7). When analyzed through PCA, also in this case some separation was seen between the two groups (data not shown). To understand the regions responsible for this discrimination, iPCA was performed, together with visual inspection, and some regions were then investigated, also through ANOVA, to assess independent discriminative power of separate spectral regions. Again, the signals corresponding to antibiotics were found to be major discriminating variables. Signals at 2.52 and 8.07 ppm derived from metronidazole appeared to be significantly higher in

anti-biotic treated piglets. The ratio between these signals (Fig. 8) was calculated to further prove the metabolite assignment and corresponded to a 1:3.5 ratio between the singlet at 8.07 ppm and at 2.5 ppm which corespond to H-4 and 2-CH3, respectively, that are in approximately a 1:3 ratio in metronidazole.

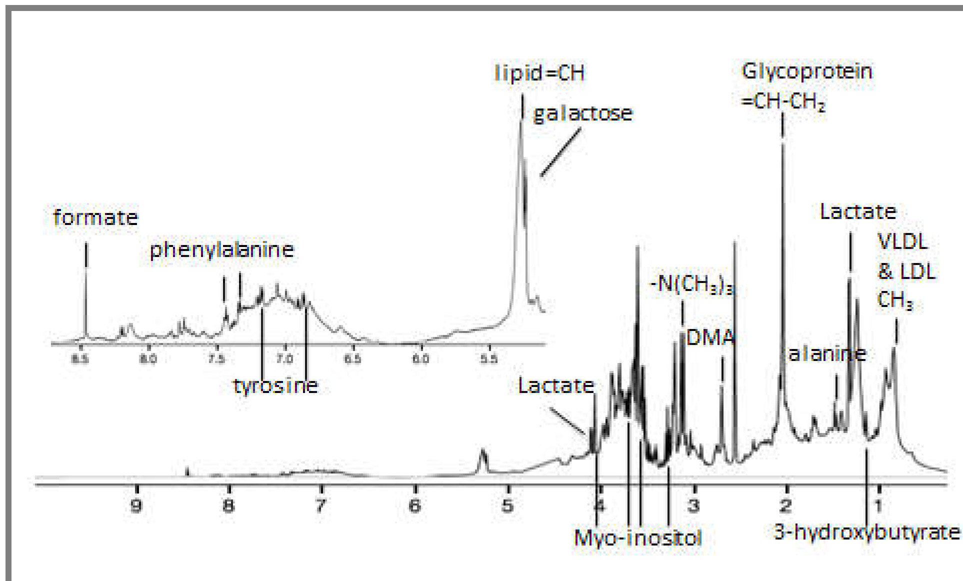


Fig. 7: Plasma NOESY1D NMR spectrum with metabolites found from databases and literature.

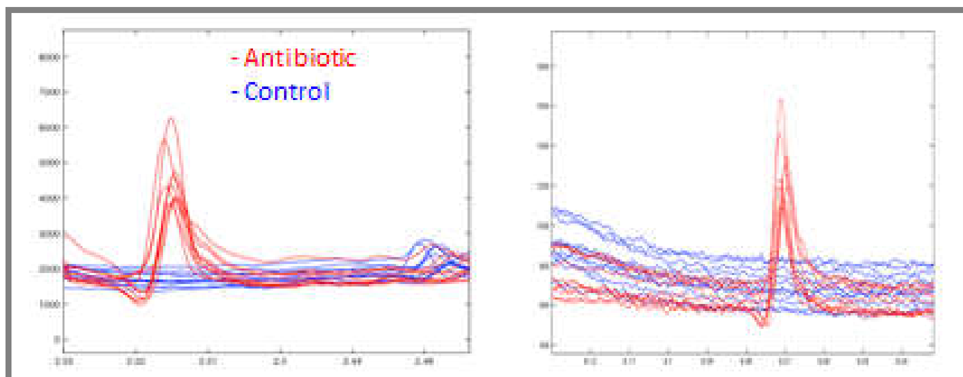


Fig. 8: Plasma NOESY1D NMR signals from metronidazole, appearing in antibiotic-treated samples (red) and not in control samples (blue)

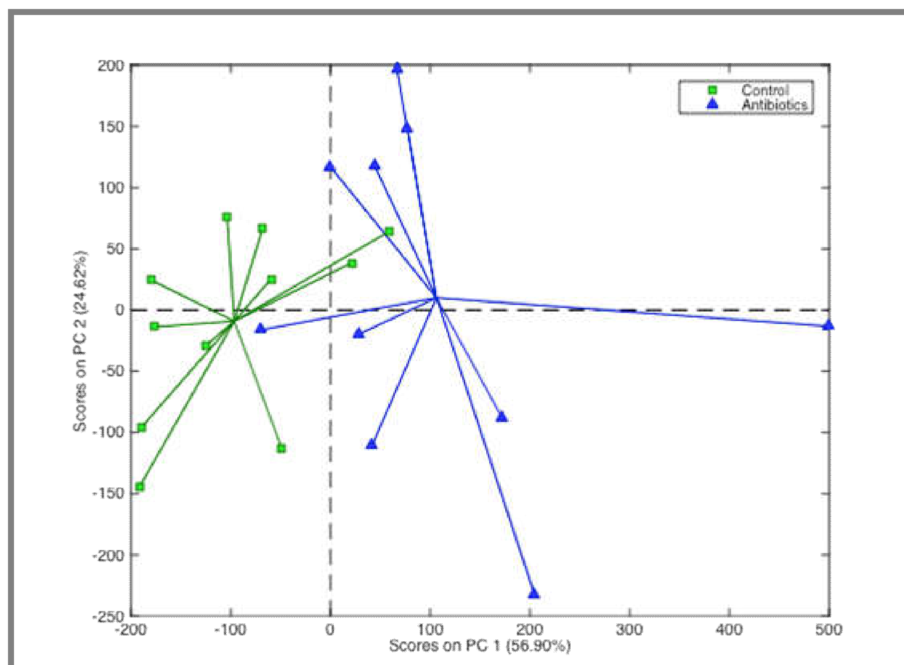


Fig. 9: Scoreplot of PC1 vs PC2 for the plasma NOESY1D NMR signals of galactose at 5.25 ppm. Samples in green are controls and in blue are treated.

Another signal that appeared different was a doublet at 5.25 ppm. As seen in Figure 9, separation can in fact be seen in this region when PCA is calculated on it. The region of the anomeric signal of either glucose or galactose, in fact, appears higher in the antibiotic cohort. Galactose was given to the pigs through an orogastric catheter 24 and 48 hours after the first fed to test for its absorption capacity by the apical Na+/glucose cotransporter. Blood test did not show statistically significant differences [Jensen et al., 2014]. The plasma here employed was collected at time of death, therefore should not come from this feeding, though it might arise from the formula fed to the pigs, containing lactose. For this signal, plasma spectra show some difference with a greater galactose/glucose absorption in the treated cohort. Again, the mannitol region also appears discriminative, with the treated samples having higher concentration of this metabolite, as seen in urine. Regions where the signals from lipids are found also appear to discriminate among the two groups: the antibiotic-treated cohort shows higher concentration in both cholesterol-related and lipid-related protons. As seen also by LC-MS (data not shown), nicotinamide appears higher in treated piglets, its significance, though, lacks, being in very noisy regions of the spectra. Some signals, this time, appear just for control samples, though, due to regions in which they appear, they do not arise as statistically significant through ANOVA. Doublets from 4-phenylacetate appear just in the control samples, as also reported by the LC-MS results (data not shown). This metabolite of phenylalanine

derives from the gut microbiota and its lower levels might be related to the effect of the antibiotics on the bacterial population. The other metabolites from phenylalanine bacterial metabolism found in higher levels in control by LC-MS (data not shown) are not clearly visible in NMR spectra due to the fact that they are overruled by bigger signals, such as the ones from the antibiotics. Another signal appearing just for control samples is shown in Figure 10.

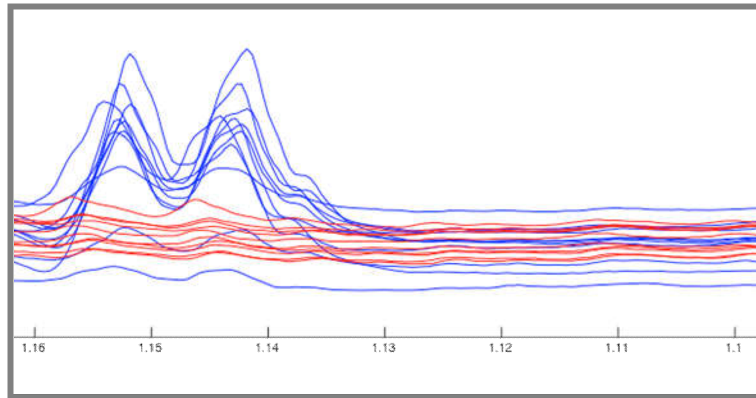


Fig. 10: Plasma NOESY1D NMR signal appearing at 1.15 ppm only in control (blue) samples and not in treated (red) ones.

A doublet at 1.15 ppm, in fact, clearly appears for control samples. The chemical shift might indicate that this resonance belongs to ß-hydroxybutyrate. Elevated plasma/serum levels of ß - hydroxybutyrate have been linked to diseases such as colorectal cancer [Ni et al., 2014] or cirrhosis [Qi et al., 2012]. In addition, this molecule was also found to be more elevated in pre-term piglets fed a carbohydrate formula that was more likely to have them develop NEC [Call et al., 2014]. Ketone bodies are generally mildly higher in newborns and they can be employed for brain and lung development [Yeh & Sheehan, 1985]. This signal was also observed by Brown and colleagues [Brown et al., 1989] in the urine of a baby that was born too small and had prenatal brain damage. In their paper the signal was though attributed to propilen-glycol assuming a contamination in the injection of an anticonvulsant. It might be the case that this metabolite is either coming from higher carbohydrate fermentation (which induces NEC) or is less employed in brain development, a condition also found to be correlated with NEC.

Summary of NMR results

To sum up, NMR spectroscopy was able to confirm some of the findings of LC-MS analysis (here not shown) and highlight other metabolic shifts related with both the antibiotic treatment or the presence of NEC. Nicotinamide was also linked to the treatment by NMR, being found as it is or methylated in both urine and plasma spectra. The same correspondence between the two methods was seen for plasma kynurenine, whilst glutamate was seen to be typical of control in

plasma samples by both techniques. Other metabolites were only spotted significantly different by NMR, such as aliphatic aminoacids for treatment and beta-hydroxybutyrate for control. The findings of both metabolomics approaches are summarized in the following Table 3.

Table 3: Metabolites found to be discriminant between the two groups in either one or both of the metabolomics approaches.

Discriminative metabolita	Group with higher concentration	Biofluid	Technique
3-Hydroxykynurenine	AB	urine; plasma	LC-MS
3-Phenyllactate	C	urine;plasma	LC-MS
4-Hydroxyphenylacetate	C	urine;plasma	LC-MS
5-Hydroxyindoleacetate	C	urine; plasma	LC-MS
Aliphatic aminoacids	AB	Urine	NMR
Ampicillin	AB	urine; plasma	LC-MS; NMR (urine)
ß-Hydroxybutyrate	C	Plasma	NMR
Cholate	C	Plasma	LC-MS
Galactose/Glucose	AB	Plasma	NMR
Glutamate	C	Plasma	LC-MS; NMR
Glycochenodeoxycholate	C	Plasma	LC-MS
Hydroxyphenylacetylglycine	C	Urine	LC-MS
Hyodeoxycholate/Taurodeoxycholate	C	Plasma	LC-MS
Hypoxanthine	C	Urine	LC-MS
Inosine	C	Urine	LC-MS
Kynurenine	AB	urine; plasma	LC-MS; NMR (plasma - Not Significant)
Lipids (=CH); VLDL; LDL; Cholesterol	AB	Plasma	NMR
Mannitol	AB	urine; plasma	LC-MS; NMR
Metronidazole	AB	urine;plasma	LC-MS; NMR
N-Acetylglutamate	C	Plasma	LC-MS
N-Acetylmethionine	AB	Plasma	LC-MS
Nicotinamide	AB	urine; plasma	LC-MS; NMR (1-methylnicotinamide in urine)
N-Phenyl-L-Phenylalanine	C	Urine	LC-MS
Phenylacetylglycine	C	Urine	LC-MS
Pseudouridine	C	Urine	LC-MS
Pyroglutamate	C	Plasma	LC-MS

These results prove how the two metabolomics techniques can go hand in hand in this type of investigations, both convalidating and completing findings. In effect, LC-MS and NMR, due to their different characteristics –sensitivity for the former and simplicity and non-specificity for the latter- can help in the identification of many metabolic markers of different chemical natures, as was the case in this study.

REFERENCES CHAPTER 7.2

Schnabl, K.L., Van Aerde, J.E., Thomson, A.B., & Clandinin, M.T. (2008) Necrotizing enterocolitis: A multifactorial disease with no cure. Wold Journal of Gastroenterology, 14, 2142-2161.

Mai, V., Young, C.M., Ukhanova, M., Wang, X., Sun, Y., Casella, G., Theriaque, D., Li, N., Sharma, R., Hudak, M., & Neu, J. (2011) Fecal Microbiota in Premature Infants Prior to Necrotizing Enterocolitis. PloS one, 6(6), e20647.

Lee, J.S., & Polin, R.A. (2003) Treatment and prevention of necrotizing enterocolitis. Seminars in Neonatology, 8, 449-459.

Lin, P.W., & Stoll, B.J. (2006) Necrotising enterocolitis. Lancet, 368, 1271-1283.

Neu, J., Chen, M., Beierle, E. (2005) Intestinal innate immunity: How does it relate to the pathogenesis of necrotizing eneterocolitis. Seminars in Pediatric Surgery, 14, 137-144.

Neu, J., & Walker, W.A. (2011) Necrotizing Enterocolitis. The New England Journal of Medicine, 364, 255-264.

Bell, M.J., Kosloske, A., Benton, C., Benton, C., Martin, L.W. (1973) Neonatal necrotizing enterocolitis in infancy: prevention of perforation. Journal of Pediatric Surgery, 8, 601-605.

Bury, R.G., & Tudehope, D. (2001) Enteral antibiotics for preventing necrotizing enterocolitis in ow birthweight or preterm infants. Cochrane Database of Systematic Reviews, 1, CD000405.

Downard, C.D., Renaud, E., St. Peter, S.D., Abdullah, F., Islam, S., Saito, J.M., Blakely, M.L., Huang, E.Y., Arca, M.J., Cassidy, L., Aspelund, G. (2012) Treatment of necrotizing enterocolitis: an America Pediatric Surgical Association Outcomes and Clinical Trials Committee systematic review. Journal of Pediatric Surgery, 47, 2111-2122.

Jensen M.L., Thymann, T., Cilieborg, M.S., Lykke, M., Mølbak, L., Jensen B.B., Schmidt, M., Kelly, D., Mulder, I., Burrin, D.G., & Sangild, P.T. (2014) Antibiotics modulate intestinal immunity and prevent necrotizing enterocolitis in preterm neonatal piglets. American Journal of Physiology – Gastrointestinal and Liver Physiology, 306(1), G59-G71.

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.

Dieterle, F., Ross, A., Schlotterbeck, G., & Senn, H. (2006) Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. Analytical Chemistry, 78(13), 4281-4290.

Camilleri, M., Nadeau, A., Lamsam, J., Linker Nord, S., Ryks, M., Burton, D., Sweetser, S., Zinsmeister, A.R., & Singh, R. (2010) Understanding measurements of intestinal permeability in healthy humans with urine lactulose and mannitol excretion. Neurogastroeneterology & Motility, 22(1), e15-e26 et al., 2009, Neurogastroenterol Motil.

Dastych, M., Dastich Jr, M., Novotna, H., & Cihalova, J. (2008) Lactulose/Mannitol Test and Specificity, Sensitivity, and Area Under Curve of Intestinal Permeability Parameters in Patients with Liver Cirrhosis and Chron's Disease. Digestive Diseases and Science, 53(10), 2789-2792.

Taylor, S.N., Basile, L.A., Ebeling, M., & Wagner, C. (2009) Intestinal Permeability in Preterm Infants by Feeding Type: Mother's Milk Versus Formula. Breastfeeding medicine, 4(1), 11-15.

Zani, A., Ghionzoli, M., Lauriti, G., Cananzi, M., Smith, V.V., Pierro, A, De Coppi, P., & Eaton, S.(2010) Does intestinal permeability lead to organ failure in experimental necrotizing enterocolitis? Pediatric Surgery International, 26(1), 85-89.

Pearson, A.D.J., Eastham, E.J., Laker, M.F., Craft, A.W., & Nelson, R.(1982) Intestinal permeability in childern with Chron's disease and coeliac disease. British Medical Journal, 285(6334), 20-21.

Ni, Y., Xie, G., & Jia, W. (2014). Metabonomics of Human Colorectal Cancer: New Approaches for Early Diagnosis and Biomarker Discovery. Journal of Proteome Research, 13(9), 3857-3870.
Qi, S., Tu, Z., Ouyang, X., Wang, L., Peng, W., Cai, A., & Dai, Y. (2012) Comparison of the metabolic profiling of hepatitis B virus-infected cirrhosis and alcoholic cirrhosis patients by using 1H NMR-based metabonomics. Hepatology Research, 42(7), 677-685.

Call, L., Stoll, B., Garcia, S., Bauchart-Thevret, C., Donnelly, J., Sheikh, F., Akinkuotu, A., Olutoye, O., Wittke, A., & Burrin, D. (2014), PO-0607b Metabolomic Determinants of Necrotizing Enterocolitis in Preterm Piglets. Neonatal Nutrition and Gastroenterology, 99, A451.

Yeh, Y., & Sheehan, P.M. (1985) Prefential utilization of ketone bodies in the brain and lung of newborn rats. Federation proceedings, 44(7), 2352-2358.

Brown, J.C.C., Mills, G.A., Sadler, P.J., & Walker, V. (1989) 1H-NMR studies of urine from premature and sick babies. Magnetic Resonance in Medicine, 11(2), 193-201.

CHAPTER 8: THE HUMAN METABOLOME

The human metabolome is the set of metabolites in a human, more specifically in determined biofluids or tissues. The metabolome was defined as "the complete complement of all small molecule (<1500 Da) metabolites found in a specific cell, organ or organisim" [Wishart et al., 2007]. The metabolome is a very useful tool for medical and biological sciences since it is directly linked to the physiological status of an individual. [Beecher, 2003]. In fact, the metabolome represents the final product of the chain reactions occurring biologically from genes to transcripts to proteins to finally metabolites, constituting together with the relative 'omic' science, the four main blocks of system biology. For this reason, the biochemical knowledge that can be gained through the investigation of metabolomic profiles is of great dimensions. Many metabolomics studies started with diagnostic purposes, such as inborn errors, cancer, Chron's disease, etc. For many diseases it is now possible to identify specific biomarkers, though it is still of great interest how to specifically draw the line between a disease and healthy status, especially since many factors affect the metabolome. For this reason, most diagnostic metabolomics studies are carried out with subjects in a fasted state. [Krug et al., 2012] Though it would be more interesting to understand all the mechanisms involved in the characterization of the human metabolome and identify their possible interaction in the definition of the global health status.

8.1 Urine

Urination is the way our bodies eliminate hydrosoluble waste products. In effect, urine is the product of the extraction of kidneys of soluble wastes and excess water, sugars and other compounds from the bloodstream. This transparent and sterile biofluid will therefore consists mainly of urea – generated by the metabolism of aminoacids -, salts, creatinine, ammonia, organic acids, water-soluble toxins and pigments from the break-down of haemoglobin – i.e. urobilin, which gives the amber colour to urine. Despite being considered a waste product, the diagnostic value of urine is great and started from the ancient Egyptians. In fact, urine physical characteristics, starting from the colour, were considered a tool for disease diagnostics also during the Byzantine era and up to the Middle ages. Up to now, many assays done on urine are normally employed to define the health state, especially due to the fact that it a really accessible biofluid. In recent years, science has been more and more interested in the generation of the whole metabolic profile of urine, employing many different techniques, from

GC-MS to HPLC to NMR. It was demonstrated that up to 294 different metabolites can be identified in urine, though the quantification can be done to sligthly less than 100, being far more difficult. In addition to this number, many other metabolites were also found and quantified in specifically “targeted” studies or researches involving single metabolites. Up to now, the Urine Metabolome Database (UMDB: http://urinemetabolome.ca) contains information on approximately 3 thousand metabolites or metabolites species, all detected in human urine using various techniques and technologies. The majority are of exogenous origin (diet, medicines, exposure, etc.), many are endogenous and some have microbial origin. Most of these metabolites are hydrophilic molecules. In fact, urine mostly contains amino acids, carbohydrates and their derivatives and conjugates, and only very low concentration of lipids, though lipids and fatty acids contribute to many chemicals present in urine. Other molecules that are found in high concentration are, as stated, ammonia, urea, creatinine, hippuric acid and hydroxy acids with their derivatives. In effect, the most abundant metabolites in urine are urea, creatinine, hippurate and citrate (deriving from hydroxy acids). Some other metabolites might be found with concentrations spanning in great ranges and those are generally due to dietary intake. All the metabolites found belong to 230 different chemical classes – and 25 “super classes” - depicting the great chemical diversity of this biofluid. The differences between urine and other biofluids are many and the reasons why urine is generally preferrable are many. Saliva and cerebrospinal fluid (CSF) have far less compounds (5 to 10 times less) and chemical diversity (2 to 3 times lower). Serum is richer in lipids than urine and every compound found in urine can be also found in blood, being th urine metabolome a subset of the serum one. In spite of this, many compounds found in urine were not found in blood and this occurs because kidneys concentrate metabolites from blood, where they are below detection limits, into urine. In fact, some compounds may have 1000-fold concentration differences between these two biofluids. For all these reasons, urine is particularly useful for medical diagnostics and drug or diet monitoring, since kidneys will be able to concentrate specific metabolites correlated to diseases and to filter toxins and xenobiotics [Bouatra et al., 2013].

8.2 Plasma and serum

Plasma is another commonly used matrix for both biological and clinical study. Plasma is one of the two components of blood, it represents the liquid carrier in which blood cells (platelets, erythrocytes, leukocytes) are suspended, accounting for 50-55% of blood volume [Psychogios et al., 2011]. To obtain plasma, blood is treated with an anticoagulant, such as heparin or EDTA,

then the sample is centrifuged and the non-cellular portion is removed or decanted [Yu et al., 2011]. It is often analysed together with serum, though they each can be preferred in specific cases. Plasma, for example, works best when testing oral glucose tolerance in diabetic subjects [Yu et al., 2011]. Serum, instead, is the supernatant fluid obtain from blood when no anticoagulant is added. Serum has no clotting proteins and is less viscous than plasma. Serum and plasma are both aqueous solution (95% water approximately) and they are composed by many different substances, such as proteins and peptides (albumins, enzymes, hormones, lipoproteins, globulins, etc.), electrolyties, nutrients (amino acids, lipids and carbohydrates) and other molecules from waste or other origins [Psychogios et al., 2011]. For what concerns small molecules, it was demonstrated that most molecules present in plasma are the same as those in serum, though some differences could be seen, especially in their concentration [Liu et al., 2010], which was generally higher in serum [Yu et al., 2011]. The main differences is found in the compounds related to clotting, which stimulates eicosanoid biosynthesis, thus related metabolites are generally lower in serum [Psychogios et al., 2011]. In addition, recent studies [Yu et al., 2011] have proved that plasma has a better reproducibility, possibly due to the fact that its collecting procedure it is easier, not requiring coagulation time, as plasma instead does. The importance of the analysis of blood is due to the fact that it acts as liquid carrier for all the metabolites discarded, secreted and exctreted by tissues and that altered state or lesions in organs and tissues will result in plasma and serum chemical alteration. For these reasons, blood serum and blood plasma are often used for clinical and diagnostic analysis, also due to the fact that it is a very accessible biofluid. Nowadays, many modern techniques such as HPLC or MS or NMR are employed for blood profiling. The Serum Metabolome Database (SMDB: http://www.serummetabolome.ca) now contains more than 4 thousand analytes found in serum and/or plasma and their relative informations (origin, concentration, etc.). Most of these molecules are hydrophobic or related to lipids, such as di- and tri-glycerides, fatty acids, phospholipids, steroids and their derivatives. In addition, amino acids, lactate, glycerol, glucose and other metabolic byproducts can be found. Generally the concentration of metabolites can vary in a range of +/- 50%, though some can reach even +/- 100% variation (e.g. Lactic acid, glycine, glucose), due to various effects like diet, gender, age, genetics or lifestyle.

8.3 Faeces

Another biofluid that is on the rise in metabolomics research is stool. In fact, due to the great interest in the interaction between the gut microbiota, the host and metabolism, the faecal

metabolome can be a great tool for the assessment of this sort of interaction. The microbiota found in the gut is demonstrated to have a huge influence in many the host's physiological functions, such as the regulation of the immune system or the response to pathogens and diseases [Smirnov et al., 2016]. Most of the studies carried out up to now were focused on rats and mice, though the same interactions should occur into human systems. Stool samples are thus the easier sample to collect and analyse for metabolomics analysis investigating microbiota-host interactions. Faecal microbiota is only a subset of the gut population, though it could help in the interpretation of metabolic patterns, together with genetics and protein expression. Faecal water will be able to give information on the gut microbial activity and also on the digestive processes, being a solution of the metabolites originating from diet and digestive biochemistry. A database of the human faecal metabolome is currently under construction (http://www.fecalmetabolome.ca/) and it is a subset of the human metabolome database.

8.4 Research work overview

As stated in Chapter 1, numerous factor impact on the human metabolome and in different measure. In this research work, at first a general investigation on the genetical and environmental factors mostly impacting on the metabolome, such as gender, was carried out (Chapter 8.5), together with the evaluation of the possible existance of disease-related markers (Chapter 8.7). Moreover, the search of dietary biomarkers was pursued, both in observational and intervention studies (Chapter 8.6). As mentioned in the first chapter, it is in fact fundamental for current nutritional and health studies to be able to identify the specific biomarkers of food consumption and diets. In addition, it is interesting to understand whether the adherence to a particular diet can help in the human capability of contrasting particular diseases (Chapter 8.7). This will help in the definition of the markers of wellness state and to better understand the shifts it undergoes caused by the various external and internal factors impacting on it.

8.5 Foodomics For The Investigation Of The Human Urinary Metabolome – The CHANCE Project

8.5.1 Impact of different factors on the human urinary metabolome

After the assessment of the food fingerprint, it is necessary to understand the different dietary patterns in which the analysed food are consumed and the impact of these specific diets on the human body. As stated at the beginning of this chapter, though, it is fundamental to separate

the effect of the different factors impacting on the human metabolome and characterizing groups of subjects. In order to do this, it is of great help to have the availability of large sample cohorts, in order to have the greatest possible variation of conditions and peculiarities. Part of this research work was thus carried out on the spectral information collected throught the EU FP7 Study CHANCE ("Low cost technologies and traditional ingredients for the production of affordable, nutritionally correct foods improving health in population groups at risk of poverty"). The aim of the CHANCE project was to identify the relative existing nutritional deficiency and find nutritional strategies for the prevention of malnutrition in population groups at risk of poverty, through the development of affordable but nutritionally-rich food products. According to 2007 Eurostat statistics, there is a robust need to better understanding the existing nutritional barriers to healthy nutrition of 79 million EU-27 citizens considered "at-risk-of-poverty" (ROP). There are different ways to describe this economical condition, though all these European subclusters and ethnic populations have in common low purchasing power, limited education and the highest risk of diet-related diseases due to sub-optimal nutrition. Being this a large population set, it is clear how this project could have a great impact. One of the objectives of CHANCE was to evaluate the actual impact of the assessed nutritional criticalities on the metabolic profile by using innovative methodologies based on multivariate statistical analysis of NMR spectra acquired on urine samples. A representative sample of volunteers was recruited and surveys on the eating habits and food intake were collected together with urine samples. Samples were collected in five recruitment centres and sent to two analytical laboratories for storage and processing according with European standard operative procedures (SOP). The volunteers belonged to the selected subpopulation that was considered mostly at-risk-of-poverty in the relative country. A number of volunteers considered affluents but within the same age and gender classes were considered, too, as a control population. Having such a complex and large dataset was thus a key advantage in this study, allowing to investigate all the underlying complex information contained in the metabolic profile of urine samples. In effect, different biometric and anthropometric characteristics of the selected subjects - gender, ethnicity, presence of diseases, lifestyle and typical diet, etc. - generate specific metabolic patterns and those factors could cover the important information researched in experimental studies. It is thus very interesting to understand which metabolic patterns and/or the metabolites are related to these peculiar attributes and to verify the existing literature on this topic. To do this, a metabolomics approach was employed, involving

the use of ^{1}H-NMR spectroscopy and chemometric tools. This combination has been proven of great potential for this kind of investigation on human biofluids. In this case, the power of NMR spectroscopy combined with multivariate statistical tools is employed to assess which factors have the most effect in the metabolic profile of the analysed subjects and to find specific metabolic patterns for those factors. The pros of having a dataset this large were the possibility to not simply stick to the assessment of the eventual differences between ROP and affluents, but also to ivnestigate on many different characteristics impacting on the human metabolome, as mentioned above. In addition, the large amount of meta-data was also of incredible help in the investigation of the power and robustness of the dietary tools usually employed in this kind of investigation: food frequency questionnaires and 24 hour dietary recalls. The work done in the CHANCE project was therefore very useful for the determination of the best analytical protocols for the investigation on human diet, human metabolome and their relationship.

8.5.2 Materials and Methods

Five recruiting centers were selected to collect samples: Finland, Italy, Lithuania, Serbia and UK. In each center a different population was selected according to which population range was considered at risk of povery in that country, thus the selected populations were:

- Finland: males and females > 65 years old
- Italy: females from 25 to 65 years old
- Lithuania: males and females from 25 to 65 years old
- Serbia: males and females from 25 to 65 years old
- UK: males and females > 65 years old

In each center both ROP subjects and control (affluent, AFF) counterpart from the same age and gender subgroup were selected and the samples were thus as described in table 1.

For each subject two urine samples were collected, together with informaiton on their anthropometric measures and physical activity. The nutritent intake and dietary patterns were assesd by means of two 24 hour dietary recalls and one food frequency questionnaire per subject. The urine samples were analysed by 1H-NMR spectroscopy.

630 µl of urine sample were centrifuged to remove debris, then 540 µl of surnatant were placed in a clean eppendorf containing 60 µl of D2O-based phosphate buffer containing also tri-methyl-silyl propionate (TSP) as internal standard and sodium azide (NaN3) as an antibacterial agent. A total of 450 µl of the mixture was transferred into 4.25-mm-outer-diameter NMR tubes

(for Bruker Match holder).

Table 1: Number of ROP and AFF (control) samples for each recruiting center.

Recruiting center	N° ROP subjects	N° AFF subjects
Finland	74	163
Italy	200	100
Lithuania	146	103
Serbia	200	100
UK	161	43

The 1H-NMR spectra were acquired using a Bruker 600 MHz spectrometer (Bruker BioSpin) operating at 600.13 MHz proton Larmor frequency and equipped with a 5 mm CPTCI 1H-13C/31P-2H cryo-probe including a z-axis gradient coil, an automatic tuning-matching (ATM) and an automatic sample changer. A PT 100 thermocouple provided temperature stabilization at the level of approximately 0.1 K for the sample. In order to obtain temperature equilibration (303.1 K), before measurement, samples were kept for at least 3 minutes inside the NMR probe-head. For each urine sample, a one-dimensional NMR spectrum was acquired with water peak suppression, using a standard pulse sequence (NOESYPRESAT; Bruker), with 64 free induction decays (FIDs), 64k data points, a spectral width of 12.019 Hz, an acquisition time of 2.7 s, a relaxation delay of 4 s, and a mixing time of 100 ms. Free induction decays were multiplied by an exponential function. Data processing followed this procedure:

- data cleaning: exclusion of non-informative NMR regions
- data organization: uniformity of labels, sorting of data, use of class identifiers
- pre-processing: signal alignment, 2-Norm normalization and binning (each bin = 10 spectral points)
- averaging: calculation of the average of the two existing time points per subject to avoid noise
- scaling: mean centering and Pareto scaling
- exploratory data analysis: removal of outliers (diabetic subjects, etc.)

The final spectral matrix consisted of 1289 spectra of 5800 points. All data processing and data mining was carried out on the Matlab software (Mathworks).

Results and discussion

The analysis of the spectra consisted firstly in the assessment of any kind of grouping among samples. This was particularly useful since it is possible to prove information found in literature

having such a large data cohort. The first thing that was observed through simple spectral visualization and principal component analysis (PCA) was the presence of clear outliers. These outliers were characterized by specific signals that were easily assigned.

Most of the cases found could be grouped in three types of outliers:

- diabetic subjects, displaying signals from glucose
- subjects using medications i.e. Paracetamol
- subjects consuming alcohol, showing the ethanol signals

This information was particularly useful for many reasons: firstly it proved that the information declared by the subject was not always accountable, since these subjects found to be outliers did not declare use of drugs, alcohol or the presence of diabetes. In addition, it proved the efficiency of the NMR technique and multivariate data analysis for the detection of molecules in urine. After the removal of these outliers, spectra were analysed in order to assess any possible gender bias. This was done, again, by means of PCA. A gender bias was clearly seen in the Finnish cohort, especially when considering PC1 and PC5 (Fig. 1). By inspection of the PCA loadings, it was possible to prove that female subjects were characterized by higher concentrations of citrate, hippurate and TMAO, whilst males showed higher concentration of creatinine. This was confirmed by literature [Psihogios et al., 2008], since citrate is believed to be regulated by sex hormones and creatinine is correlated with muscle mass [Kochhar et al., 2006.]. Other clusterings that were visible were due to the recruitment center. Most of the variation was seen between Finland and UK clusters, especially when considered alone and without the influence on the PCA model of the other clusters (Fig. 2).

In this case, separation was caused by hippuric acid. This separation can be generated by different factors: firstly, the geographical origin can cause some genetical differences among subjects, and this is also stressed by the fact that the British cohort was composed by a specific ethnicity (subjects from India, Pakistan or Bangladesh). Moreover, the difference can also be due to specific diets: hippurate is believed to increase in urine due to increased consumption of tea [Daykin et al., 2005; van Dorsten et al., 2006], fruit juices or wine [Jacobs et al., 2012] and other authors have associated it with food preservatives [Zuppi et al., 1997]. The main aim of the study, though, was to find differences between ROP and AFF subjects.

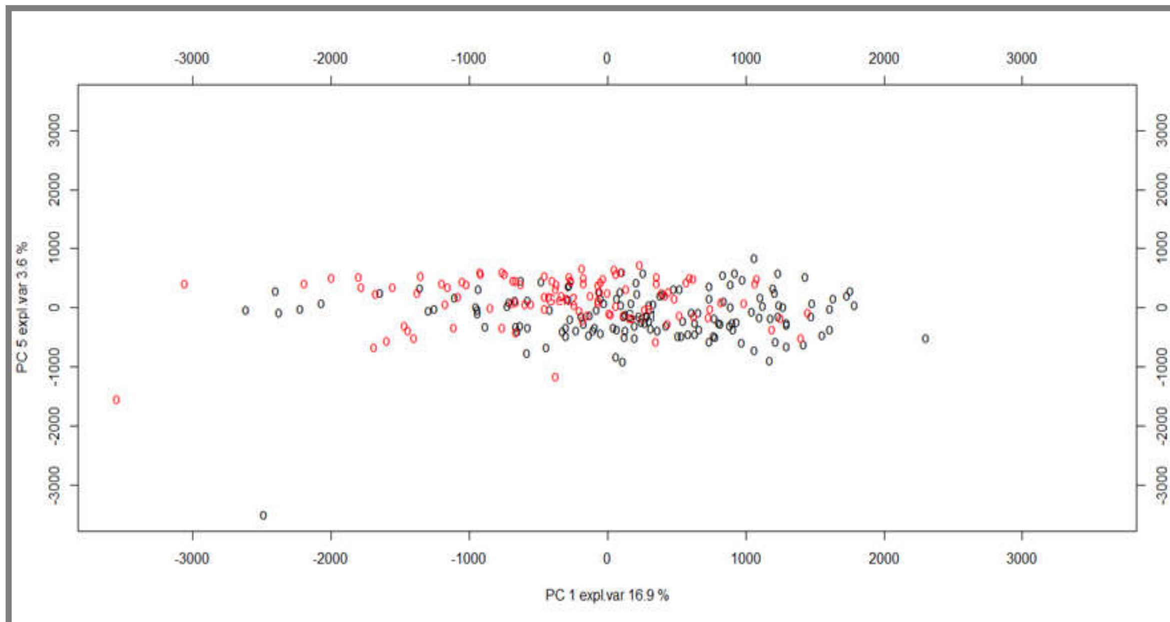


Fig. 1: PCA scoreplot of PC1 vs PC5 for the Finnish cohort. Female subjects are represented in black and male subjects in red.

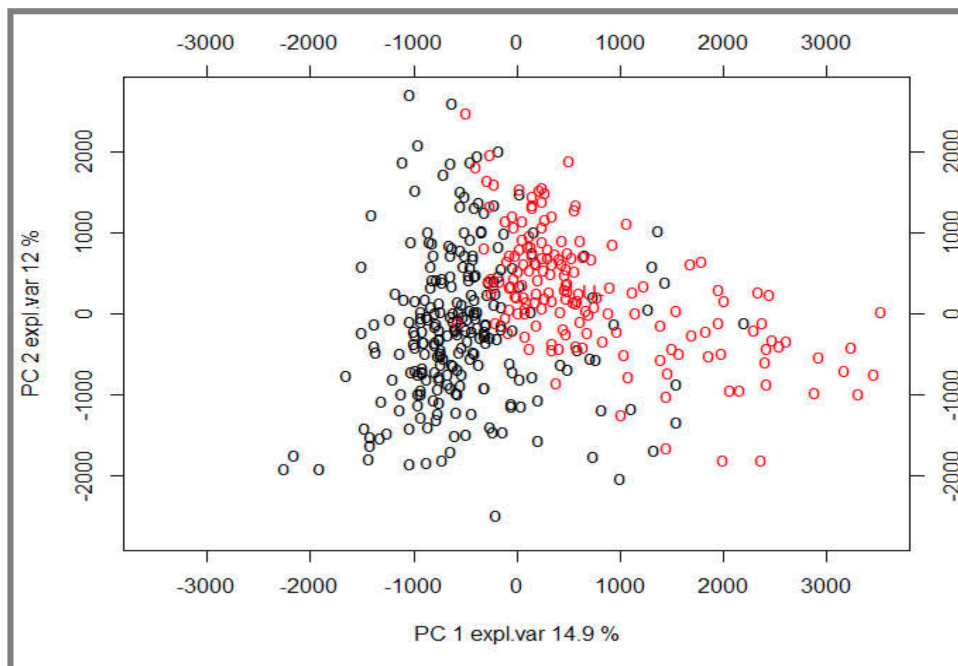


Fig. 2: PCA calculated only with the Finnish (black) and British (red) cohort.

This was done firstly by simple PCA and the cohorts with less biases (i.e. Italian cohort – only female subjects) were considered to see whether removing other factors could show some separation among the two classes, though this was not the case. Interval Partial Least Square Discriminant Analysis was also perfomed, leaving one recruitment center out at the time as cross validation, to assess whether some signals could be responsible of some separation among the two classes, but no results were found. As a final option, median spectra for the two

classes were calculated and subtracted from each other using the formula: median ROP – median AFF. This was done to check whether some signals appeared to be very different between the two groups. The median and not the average spectrum was employed, in order to avoid outliers' biases. The spectral difference thus obtained was then investigated: the signals that seemed to be appearing most different between the two medians were analysed by PLS regression. These signals, though, appeared not to be informative of any possible difference between the two populations.

Conclusions

The CHANCE project allowed the investigation on a large cohort of metabolome samples with many different peculiarities. This allowed the investigation of some spectral characteristics of the human urine metabolome that were described in literature and were confirmed from this study: gender is one of the main factors influencing the human metabolome and ethnicity and diet can impact greatly on the spectral features and subject clustering, too. All of this proved that the human metabolome is a great source of information and that much more can be found after “peeling off” the layers of the overlapping bias of other factors such as the one described above. For this reason, some differences due to the economic status might be present in the spectrum, though it was not found in this preliminary study. Moreover, there might even be no real differentiation between the two selected classes since their general diet - and therefore its chronic effect on the metabolome - might, in the end, be more similar than expected and not as related to the economic status as thought.

8.6 Foodomics for the search of food intake biomarkers

The importance of food intake biomarkers in nutritional and health studies is nowadays fully recognized by the increasing number of related studies and international research projects and joint initiatives. One of the reason for the increasing interest in biomarkers is to obtain robust compliance markers, in order to simplify the interpretation of both observational and clinical studies [Esko et al., 2017]. Moreover, the identification of dietary biomarkers can help in the validation of measurements of diet quality and in the identification of specific metabolic pathways that it could influence. [Playdon et al., 2016] In addition, the role of food biomarkers is fundamental when there is no or little knowledge on food composition (i.e. Bioactive compounds and food contaminanta) and in recent researches correlating specific biomarkers with increased risk of diseases [Scalbert et al., 2014]. For all these reasons, it is necessary to

develop a foodomic approach capable of identifying specific food and diet biomarkers both in intervention cohorts and in observational studies in free-living population. In this research work, at first an attempt at the development of an algorithm for automatic biomarker identification was carried out on the CHANCE cohort (see section 8.6.1), then, thanks to the involvement in an acute intervention study, specific food biomarkers were searched (see section 8.6.2). The approach that is currently being developed and the biomarkers that were and will be found could later be applied to other observationals studies.

8.6.1 The quest for dietary biomarkers in observational studies

The CHANCE dataset mentioned in chapter 8.5 was employed for further research. Having such a large dataset at hand, it was in fact possible to further investigate the spectral and nutritional information collected, in order to outline an analytical pipeline capable of extracting information on food biomarkers from the analysis of both dietary records and metabolome spectra. After much trial and implemention, a specific algorithm, capable of extracting information on the dietary data and comparing this information with the signals composing the spectra, was developed in order to spot possible food intake biomarkers. The algorithm was developed on the software R and runs as described in Figure 3 below. Firstly, the algorithm makes sure that both the nutritional and food information from dietary questionnaires and the spectra are ordered in the same way and that subject names are coherent. After that, the algorithm starts analysing the information on the nutritional and dietary data, assessing the dimension of the subgroup of subjects declaring 0 consumption/intake for that food/nutrient. This is done in order to understand if there is a sufficiently big “zero intake” subgroup, otherwise the subgroup declaring the highest intake (fourth quartile) will be compared with the first quartile subgroup, ordered by intake. Having selected these two subgroups to be compared, the same subjects will be selected in the spectral matrix. In order to have a comparable measure of the two subgroups, the larger one will be redimensioned to match the size of the smaller one and the subjects will be randomly selected. After this, the mean, median and standard deviation spectra for each group will be calculated and the points matching the required threshold will be selected. In addition, a plot of the two subgroups spectra (mean +/- standard deviation) and of the selected thresholds in different spectral regions will be generated and saved, in order to be later inspected. This makes the analysis quite fast and easy, approachable for any kind of researcher. The problems faced, though, are mostly related to the

choice of subgroups and of threshold (as highlighted by the arrows in Figure 3).

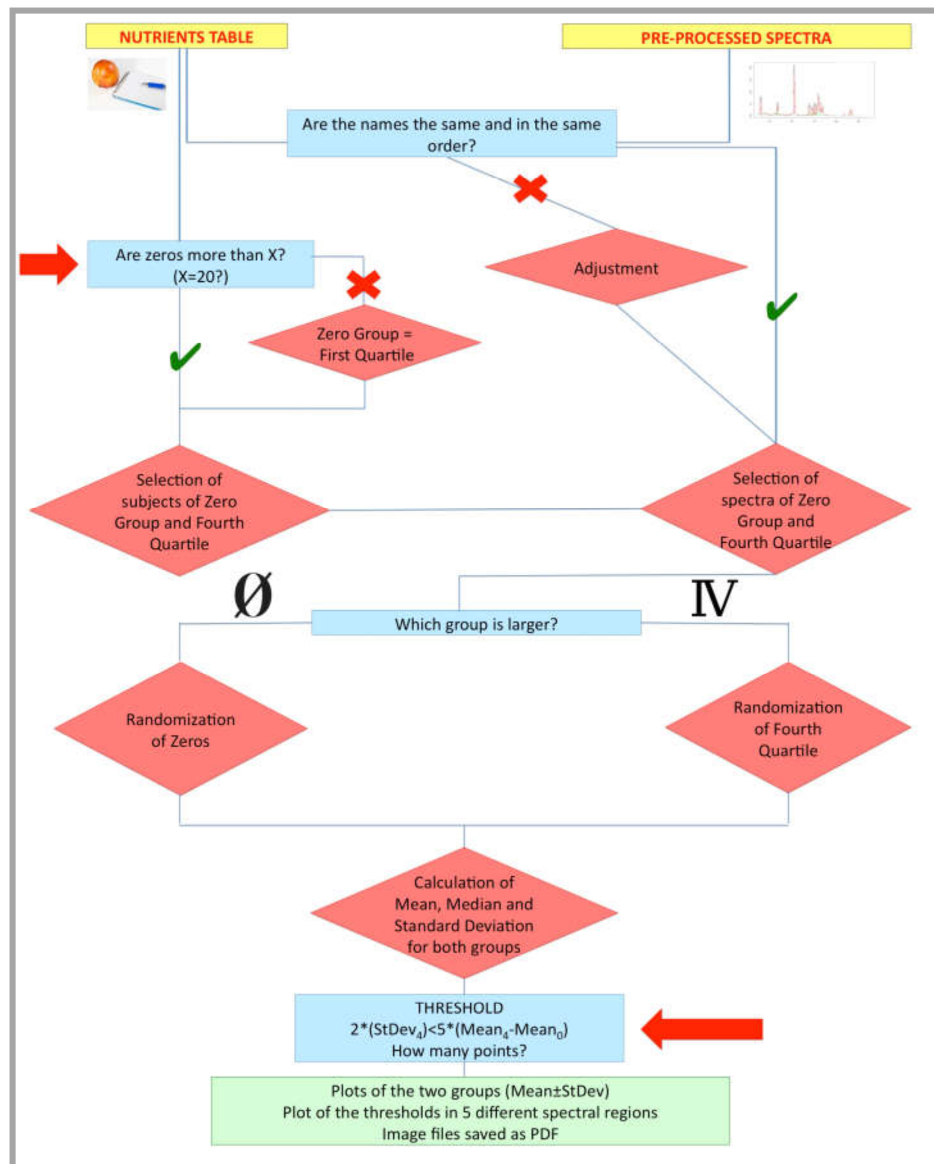


Fig. 3: Algorithm developed for the automatic detection of food intake biomarkers.

When employed in the CHANCE dataset, the algorithm did not perform perfectly and this can be ascribed to many factors:

- firstly, as stated, the reported thresholds might not be optimal and should still be improved
- the information reported in the dietary questionnaires is not always accurate, since subjects tend to forget or over/underestimate their intakes [Trimigno, 2013]
- when nutrients are used to select subgroups a greater mistake can be included in the analysis since a lot of error is made when transforming the dietary information of food consumption into intake of nutrients [Trimigno, 2013]

- the signals reported to be biomarkers for specific foods might actually vary due to the intake of other products and/or their metabolism might depend on the individual subject.

Future research will be carried out to implement and improve the algorithm and test it on other data matrices, where nutritional data is more robust.

8.6.2 Search of food intake biomarkers in an intervention study

As stated, the search of food intake biomarkers has given rise to many international collaborations and research projects. One of these is the Joint Programming Initiative called FoodBAll, which is still ongoing. FoodBAll stands for Food Biomarker Alliance and aims at identifying and quantifying specific food biomarkers in the human metabolome. This goal will aid the nutrition scientists in the assessment of human diets and their actual impact on the human body and in the development of tailor-made dietary guidelines. The main bullet-points that are and will be covered by the project are:

- discovery of the food intake biomarkers
- assessment of nutritional status biomarkers
- classification and validation of the identified food biomarkers
- correlate food biomarkers to health status and disease risk modulation

In addition, the project aims at compiling various online databases and standard operating procedures (SOPs) that could be useful and employed in any other metabolomic assessment. My collaboration with the FoodBAll project consisted and consists in the analysis of specific foods and the human serum and urine of subjects that have consumed those products and controlled standard meals. In this way, the acute effect of the selected food items and the metabolic kinetics can be assessed in the different biofluids and the specific food biomarkers can be spotted and quantified.

Materials and methods

The samples analysed for now for the FoodBAll Project had the following study design: it was a randomized controlled cross-over study, including 12 subjects and with a maximum 20% drop-out rate. Participants were healthy females and males from 18 to 40 years old and with a normal BMI (18.5<x<30 kg/m2). Eleven subjects completed the study. Subjects were randomized in order to consume a different food item at each time-point. Three were the food items considered: full fat pasteurized milk, Swiss Cheese and soy milk. For each intervention two run-in days were considered, during this period subjects had to avoid the analysed food

and products related and keep a food diary. In the last run-in day they had a standardized dinner and then had to fast. Then, during test day, the food product was given, then standardized lunch and dinner were served. During this day urine samples were collected at fasted state and in the intervals: 0-1h, 1-2h, 2-4h, 4-6h, 6-12h, 12-24h from the test meal. The wash out period then lasted from 96 hours to 8 weeks. After the wash out, the cycle described above started again with a different test product, until all subjects had consumed all three test meals. Urine samples were kept at 4°C between collection, then 1 mL aliquots were taken (10 maximum), after centrifugation at 1800g for 10 mins. These aliquots were then stored at -80°C until NMR analysis. 630 µl of urine sample were centrifuged to remove debris, then 540 µl of surnatant were placed in a clean eppendorf containing 60 µl of D_2O-based phosphate buffer containing also tri-methyl-silyl propionate (TSP) as internal standard and sodium azide (NaN_3) as an antibacterial agent. A total of 450 µl of the mixture was transferred into 4.25-mm-outer-diameter NMR tubes (for Bruker Match holder). The ^{1}H-NMR spectra were acquired using a Bruker 600 MHz spectrometer (Bruker BioSpin) operating at 600.13 MHz proton Larmor frequency and equipped with a 5 mm CPTCI 1H- 13C/31P-2H cryo-probe including a z-axis gradient coil, an automatic tuning-matching (ATM) and an automatic sample changer. A PT 100 thermocouple provided temperature stabilization at the level of approximately 0.1 K for the sample. In order to obtain temperature equilibration (303.1 K), before measurement, samples were kept for at least 3 minutes inside the NMR probe-head. For each urine sample, a one-dimensional NMR spectrum was acquired with water peak suppression, using a standard pulse sequence (NOESYpresat; Bruker), with 64 free induction decays (FIDs), 64k data points, a spectral width of 12.019 Hz, an acquisition time of 2.7 s, a relaxation delay of 4 s, and a mixing time of 100 ms. Prior to Fourier transformation, each FID was apodized by Lorentzian line broadening of 1 Hz and the corresponding spectra were phase and baseline corrected manually using TopSpin (TopSpin 3.5, Bruker).

The same software was used to export data in ASCII file format. Spectra were then imported into R software. Chemical shift referencing was performed by setting the TSP signal to 0.0 ppm. The spectral regions including only noise (the spectrum edges below 0.5 and above 10 ppm), the NMR signals which are strongly affected by the residual solvent peak of water (between 4.95 and 4.7 ppm) and the urea signal (5.45-6.1 ppm) were removed prior to data analysis. Normalization was carried out using the PQN algorithm (Dieterle et al., 2006).

The number of spectral data points was then reduced by a binning operation, choosing a bin

size of 100 points. Multivariate data analysis was then carried out on the reduced data matrix.

Preliminary results

Firstly, multilevel PLSDA was carried out on the dataset in order to select the variables (bins) mostly responsible for the separation between treatments (cut-off=0.5) and calculate PCA employing these bins. In this way a nice separation is achieved (Figure 4), though not many bins are selected in this way and it is difficult to define the variables mostly responsible for this grouping.

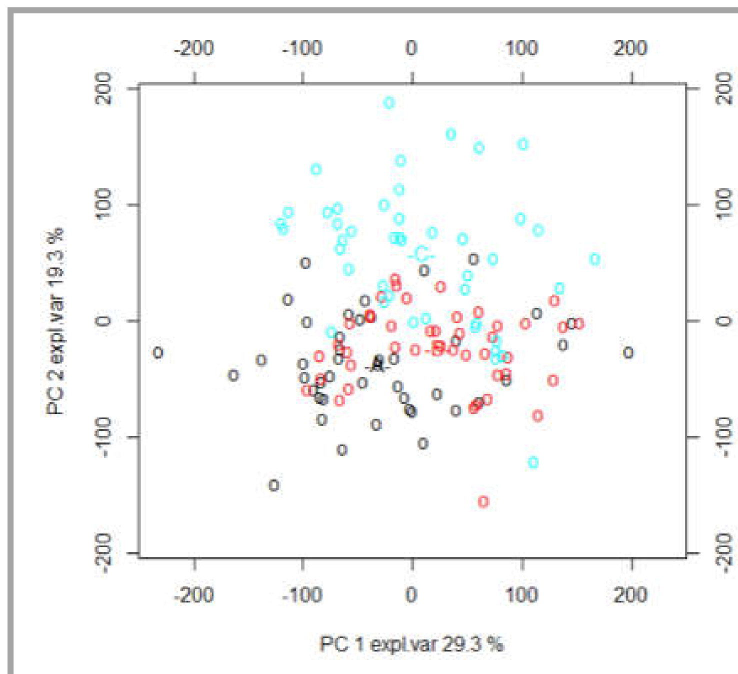


Fig. 4: PCA calculated on the variables selected by ML-PLSDA. Samples in black were collected after the subjects received milk as test-meal, in red during cheese intervention and in cyan during control (soy).

The main differences observed in the PCA were ascribable at signals originating for the standardized meals (chicken and rice – data not shown -), therefore a new approach had to be followed in order to focus on the real kinetics of the test meal metabolism. For this reason, only four data points were selected: T0-1, T1-2, T2-4 and T4-6 and OPLSDA was carried out on the matrix comparing two food interventions at the time. This allowed to focus more on the metabolome differences caused by the single food items and eliminating noise and bias from other factors (i.e. gender, individual responses, etc.). To better understand the variables causing separation, the OPLSDA was carried out on a reduced data matrix composed by the integrals of identified metabolites or clear unidentified signals. This spectral matrix consisted therefore in 44 variables. Autoscaling was performed before carrying out OPLSDA. The best separation was obtained in the comparison between milk and soy intervention. In this case, as expected, sugars

were mostly responsible for the discrimination between the two reported treatments (Fig. 5)

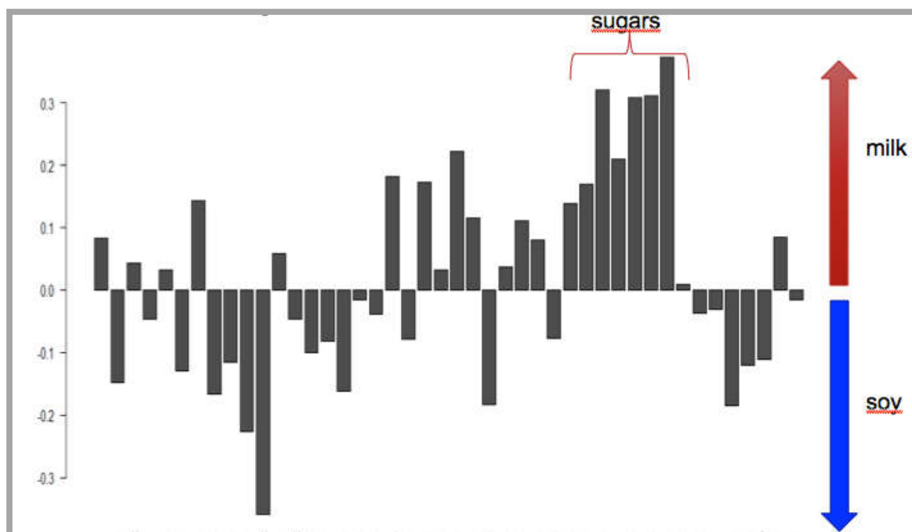


Fig. 5: OPLS-DA loadings for the comparison of soy and milk treatments. The sugar signals are the one mainly characterizing the spectra of subjects who consumed milk as the test-meal. The signal which mainly described the milk treatment was the one related to galactose. This metabolite, in fact, is strictly correlated with milk consumption, since it derives from the metabolism of lactose, the main sugar in milk.

In order to investigate better the kinetics of galactose metabolism, the integral of its signal is measured in the different time-points and compared between subjects and with the same region in spectra measured after the soy treatment.

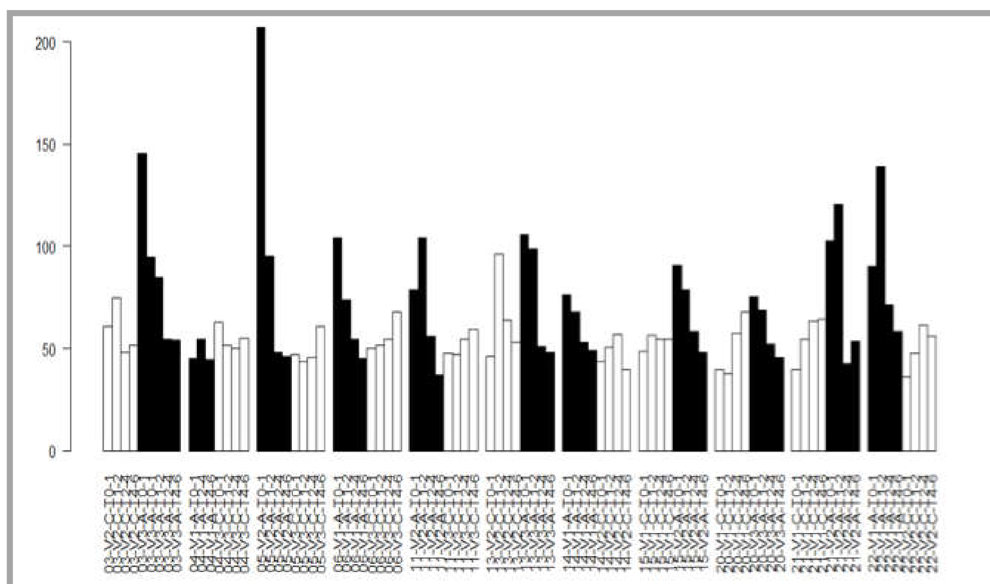
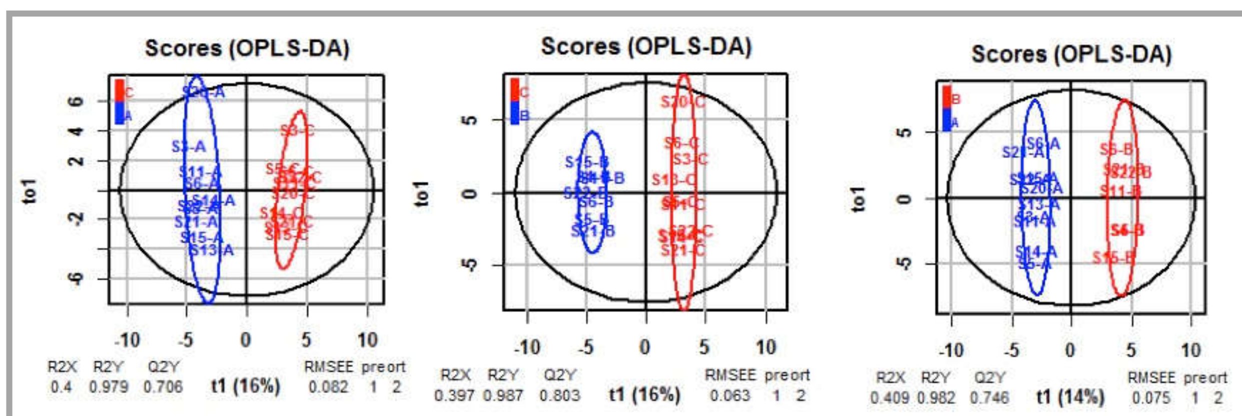


Fig. 6: Area of galactose for each subject during the different time-points (ordered) and for the two different treatments: soy in white and milk in black.

As shown in Figure 6, for most subjects the kinetic shows an increase of the galactose concentration in the first hour after the meal, which then rapidly decrease and stabilize. It is interesting to observe how galactose kinetics and concentration are though highly variable

among the individual subjects, proving how even in a simple investigation such as this one, the individual factor can be of great impact and thus how that effect must always be kept in mind in metabolomics investigations. After this first investigation, for each subject, spectra from samples at the following time-points were summed, in order to mimick a 6 hours urine pool: T0-1, T1-2, T2-4 and T4-6. Only samples with 4 time-points were kept, therefore obtaining a data-set consisting of 10 spectra for milk and soy intervention and 7 spectra for the cheese intervention. This spectral matrix was then imported in MatLab (R2014b, MathWorks). 95 bins with significant signals were then selected manually after visual inspection of the spectra. These signals were then imported into R software again, where orthogonal partial least squares discriminant analysis (OPLSDA) was carried out through the “biocLite” package (https://www.bioconductor.org/). Treatment were compared 2 by 2 in order to be able to calculate OPLS-DA (Fig. 7).

Fig. 7: OPLS-DA scoreplots for the three different comparisons: milk (A) vs soy (C), cheese (B) vs soy and milk vs cheese.



Signals characterising most of the variance of the model were then selected: the loadings with greater values than the standard deviation were identified. These consisted in 60 variables (signals). These signals, most significant for the discrimination, were then analysed through univariate statistical analysis by means of Kruskal Wallis test. 30 metabolites were statistically different with $p<0.05$ and among these, 23 had a p value lower than 0.01. Four signals appeared to be biomarkers for cheese intake, fourteen for milk and twelve for soy. These are listed in Table 2. Biomarkers for milk intake mainly correspond to the typical sugars contained in milk: lactose and galactose. A slight trend for higher concentration after milk treatment can be seen also in dimethylamine and hippurate, the latter though seems significant only in one of its signals, probably since this chemical shifts overlaps with signals from lactose and galactose.

Therefore, these metabolites do not really qualify as specific milk biomarkers. Moreover, a signal which was assigned to galactonate appeared to be present only in milk-treated samples. This is an intermediate compound originating from galactose metabolism, thus this is why it only comes up during milk intervention.

Table 2: list of discriminative signals for each product, with the corresponding p-value calculated with ANOVA and Kruskal-Wallis. * are two peaks from the same triplet, divided because the signal from succinate fell over the mid-peak and altered the results for the whole signal. ** have a p-value < 0.01

Chemical shift	Multiplicity	Molecule	p-value	Food
1.22 ppm	s	Unassigned	0,041222316	Milk
2.05-2.07 ppm	-	Glycine+Glutamine derivatives	0,044324702	Milk
2.72 ppm	s	Dimethylamine	0,048483555	Milk
3.36 ppm	s	Unassigned	0,044204254	Milk
3.48-3.52 ppm	-	Galactose	0,001077383 **	Milk
3.64-3.67 ppm	-	Lactose	0,001461257 **	Milk
3.74-3.78 ppm	-	Galactose	0,002072213 **	Milk
3.98 ppm	d	Hippurate	0,035156028	Milk
4.28 ppm	s	Galactonate	0,000132142 **	Milk
4.47 ppm	d	Lactose	0,000103439 **	Milk
4.6 ppm	d	Galactose	0,000353543 **	Milk
5.25 ppm	d	Lactose	0,024962944	Milk
5.28 ppm	d	Galactose	5,21052E-05 **	Milk
5.39 ppm	s	Allantoin	9,14346E-05 **	Milk
1.06 ppm	t	Propionate	0,045377279	Cheese
1.49 ppm	d	Alanine	0,003255039 **	Cheese
2.39 ppm	m	Pyroglutamate*	0,00198528 **	Cheese
2.42 ppm	m	Pyroglutamate*	0,001269751 **	Cheese
2.48 ppm	s	Pyridoxine	0,000110658 **	Soy
3.59 ppm	s	Unassigned	0,000193435 **	Soy
4.44 ppm	s	Trigonelline	0,000100879 **	Soy
6.56 ppm	d+d	Unassigned	8,78076E-05 **	Soy
6.59 ppm	s	Trans-aconitate	0,008365213 **	Soy
7.02 ppm	s	Unassigned	0,001000726 **	Soy
7.03 ppm	s	Unassigned	0,000520769 **	Soy
8.05 ppm	t	Trigonelline	8,4E-05 **	Soy
8.17-8.22 ppm	-	Unassigned	0,000371796 **	Soy
8.29 ppm	s	Unassigned	0,002891678 **	Soy
8.84 ppm	dd	Trigonelline	0,000221678 **	Soy
9.13 ppm	s	Trigonelline	0,006814776 **	Soy

In the region at 2.05-2.07 ppm where glycine and glutamine N-acetyl derivatives show signals, a particular singlet can be seen for some samples after milk-treatment. These metabolites could derive from the amino-acid metabolism and the reason why only some subjects show these higher concentration might be due to the employment of different metabolic pathways. The signal from allantoin at 5.39 ppm is also observed as a biomarker of milk intake. This metabolite originates from purine metabolism and is also found in milk (Gonda & Lindberg, 1997). A clear singlet is also visible after milk intake at 1.218 ppm. In this region signals from aliphatic molecules such as amino acids and fatty acids are observed. Therefore, this signals could most possibly belong to a molecule from one of these classes. For what concerns cheese intake, the main biomarkers found are propionate, alanine and pyroglutamic acid. Propionic acid was found in increased fecal concentration after cheese consumption by Zheng et al. (2015). The increased concentration of this SCFA might be due to a higher concentration of this metabolite in cheese and, moreover, to the effect of the gut microbiota (St-Onge et al., 2000). The greater concentration of alanine after cheese consumption might be due to the high level of this amino acid in Gruyere cheese. Pyroglutamic acid, instead, is a typical compound of cheese, especially ripened ones. The production of this metabolite is highly dependent on the whey starter microflora and not on milk (Mucchetti et al., 2000) and this is why it can be found as a cheese intake biomarker. The main biomarker found for soy intake is trigonelline. This metabolite is a typical compound accumulating in legumes and coffee seeds and is being studied for many possible beneficial effects, such as inhibition of the proliferation of cancer cells (Ashihara et al., 2015). Pyridoxine is vitamin B6 and soy products are a good source of this metabolite, although sometimes soy-derived products show much lower concentrations in respect to soybeans. Anyhow, soybean and soy-products could represent a good source of vitamin B6 and this is why it was found that there could be a higher risk of B6-deficiency in non-vegetarians (Lebiedzińska et al., 2006). Trans-aconitate is another plant metabolite occurring in wheat and soybean (Yuhara et al., 2015), therefore this organic acid could be identified as a soy intake biomaker. Other signals found to be discriminative for soy consumption could not be assigned. The chemical shifts of these signals, though, hint at possible isoflavone-like molecules. Caligiani et al. (2010), characterized soybean extracts by 1H-NMR spectroscopy and observed how many signals from isoflavones were characterizing the aromatic region of the spectrum (6-8 ppm). Further investigation could prove that these signals belong to compounds deriving from isoflavone metabolism. When the selected signals are used to calculated a PCA on the samples,

a good discrimination is thus obtained, as shown in Figure 8.

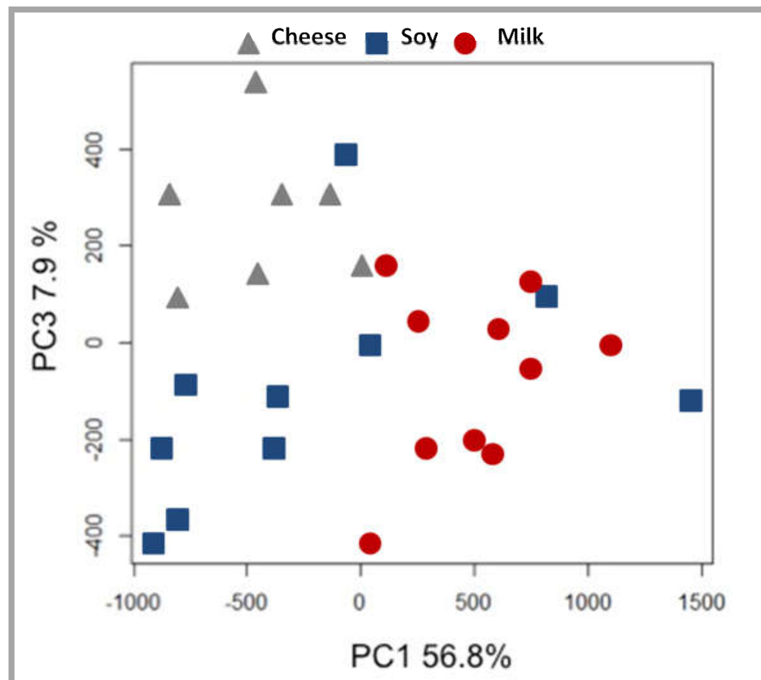


Fig. 8: PCA calculated on the 30 most discriminative signals.

The metabolites found through both multivariate and univariate statistical analysis could therefore be good candidate biomarkers of food intake for the three selected products: milk, cheese and soy.

Conclusions

The search for food intake biomarkers is one of the main challenges in current nutritional and metabolomics studies. This research profiled two possible analytical pipelines that could be perfected and then employed in both intervention and observational studies. Firstly, the developed algorithm for the automatic detection of biomarkers of food consumption can be improved by a better definition of the employed threshold. This could be done by the application of the algorithm on more robust data. Moreover, the biomarkers that have been and will be discovered through the targeted approach applied in the FoodBAll project can be then looked for, more specifically, in larger cohorts also from observational studies, in order to obtain validation.

8.7 Foodomics for the definition of the health status

A human being can be seen as a complex system, made by the genetic make-up, the lifestyle, the diet, the microbiome and, as a results, the physiology. This complex system defines how the single person responds to treatments and other external effects. For all this reasons, modern science has started to analyse more individuals than groups, employing each subject as his/her own control. Moreover, the knowledge that each specific disease can be derived from different pathway alterations, has made clear that the reduction of a disorder to a simple factor or marker cannot be employed anymore. Omics sciences have therefore come in use for the development of holistic and high-throughput approaches capable of understanding all the interactions between nutrient intakes, lifestyle, genetics, environmental factors and human physiology. Modern health science, therefore, is on the quest for the discovery of specific early on-set disease biomarkers. Traditionally, the molecoles associated to diseases were found at clinical end-points and thus not much helpful for early discovery. Nowadays, though, modern omics, especially metabolomics, are focused on the research of specific early alterations of patterns of metabolites that could hint at disease risk due to the related modification of metabolic pathways [Kaput, 2016]. It is thus fundamental to be able to define the specific health status of an individual and discover the shifts occurring when specific diseases (or risk of those) are involved. In order to do so, it is necessary to understand both the individual metabolome as the snap-shot of the health-status and to discover all the possible shifts that could occur in a population leading to an increased disease risk or, on the contrary, to a reduction of it and an improvement of the wellness state. In the future then, the few (chemical) analysis now employed in medicine will give place to the complex definition of the "metabolic signature", capable of describing all the alterations from the state of wellness, i.e. type of disease occurring, stage of progression. This signature will therefore be a great tool for disease prediction, prognosis and diagnosis. Moreover, it will be possible to understand underlying disease mechanisms and stratify patients in relation to their different metabolic and molecular alteration and drug-response [Beger et al., 2016]. In this way, the old "treatment-failure" approach, based on different trials and follow-ups with different treatments until the wanted results was obtained, is substituted by a tailor-made medicine approach (Figure 9), where each patient is firstly described by its metabotype and peculiarities (clinical data, lifestyle information, etc.) and predictively assigned to a profile which should respond to a specific

treatment, reducing the chances of failure.

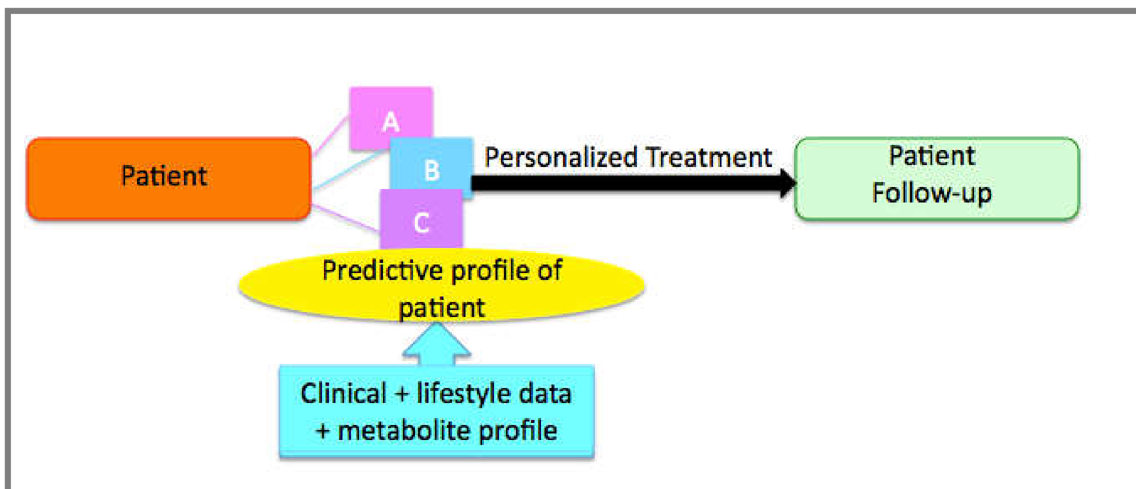


Fig. 9: new personalized approach for treatment in medicine (adapted from Beger et al., 2006).

In this research work, two studies were involved in the definition of the disease biomarkers for different chronic pathologies (leukemia and cancer), unfortunately nowadays widespread. In addition, the possibility of investigating the specific diet of the analysed subjects has paved the way to the later investigation of possible diet-disease interaction in recovery and, eventually, prevention.

8.7.1 Foodomics for the search of disease specific biomarkers

Materials and methods

Samples of urine and serum from 73 acute myeloid leukemia (AML) and 127 healthy controls were collected as the first part of the study. For many patients samples were collected in duplicate, in order to avoid possible biases related to external factors such as daily diet or medication. Together with the biological samples, information on anthropological parameters, clinical data, lifestyle and diet were collected for each analyzed subject. 540 µL of sample (urine or serum) were adjusted with 60 µL of 1.5 M KH2PO4 potassium phosphate buffer (pH 7.4) in deuterium oxide (D2O) containing 0.1% TSP (3-TrimethylSilyl-Propanoic-2,2,3,3-d4 acid sodium salt) . The samples were then centrifuged for 10 minutes at 4° C and 14k rpm. 590 µL were then collected and transferred in a 5mM NMR tube. NMR spectra were recorded at 300 K on a Bruker US+ Avance III spectrometer operating at 600 MHz, equipped with a BBI-z probe and a B-ACS 60 sampler for automation (Bruker BioSpin). Urine samples were acquired employing a 1D NOESY-presat pulse sequence, with 64 scans, whilst serum spectra were acquired using both a 1D-NOESY-presat pulse sequence, with 4 scans and a CPMG sequence with 256 scans. Each spectrum was automatically processed adjusting phase and baseline with the command

apk0.noe using TopSpin version 3.0 (Bruker BioSpin) and applying a line broadening factor of 1 Hz. The chemical shifts were internally referenced to the Trimethylsilyl propanoic acid (TSP, Cambridge Isotope Laboratories) at 0.00 ppm (part per million of frequency). After the Fourier Transformation (FT) and prior to multivariate analysis, data underwent pre-processing using the R 3.2.2 software (http://www.r-project.org/). Signal assignment was carried out by comparing their chemical shift, multiplicity and J-coupling values with literature [Wang et al., 2013] and by using Chenomx software (Chenomx Inc.). Spectral regions without consistent information, including peripheral regions, regions of the solvent signal and, in the case of urine, also the urea signal, were removed from the spectra. Urine spectra were then binned in order to reduce misalignments among signals, while serum spectra did not need this step. The last pre-processing step regarded normalization through the Probabilistic Quotient (PQN) algorithm [Dieterle et al., 2006].

Results and discussion

After normalization, the duplicates from the same subject were compared, in order to check whether some major difference was presence. Moreover, to check for possible outliers, a first principal component analysis (PCA) was calculated. Some AML patients immediately resulted as peculiar: the spectra from these subjects were characterized by the presence of signals of a particular metabolite, not found in databases and libraries and probably belonging to a drug metabolite. These subjects were thus excluded from the following multivariate data analysis. Urine samples were then evaluated for potential spectral clustering due to gender, which was confirmed. For this reason, it was chosen to separate female samples from male samples and analyze them independently. This analysis revealed some differences, between AML (both in the female and male cohort) and control spectra (see Fig. 10). In particular, control samples showed higher intensities of the signals for citrate and hippurate. In serum samples, we confirmed the presence of a higher signal for 2-hydroxyglutarate (2HG) in some patients carrying mutated *IDH2* (isocitrate dehydrogenase) [Wen et al., 2015] (Fig. 11).

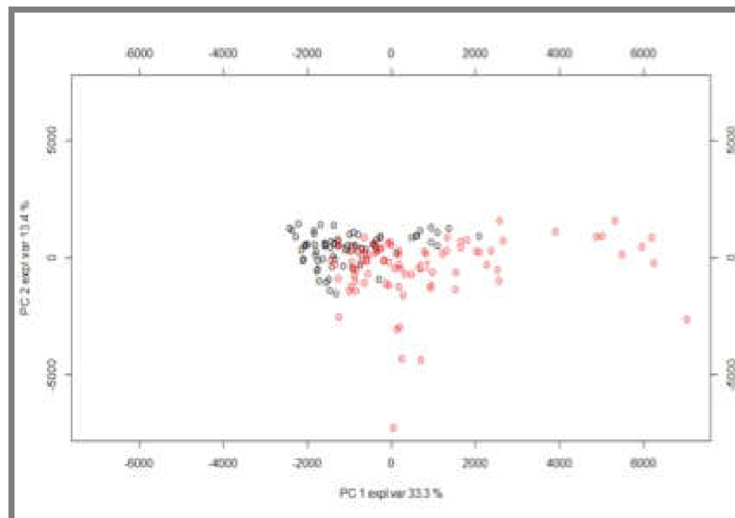


Fig. 10: PCA scoreplot of female subjects calculated on urine NOESY1D NMR spectra, showing a separation between AML patients (black) and healthy controls (red).

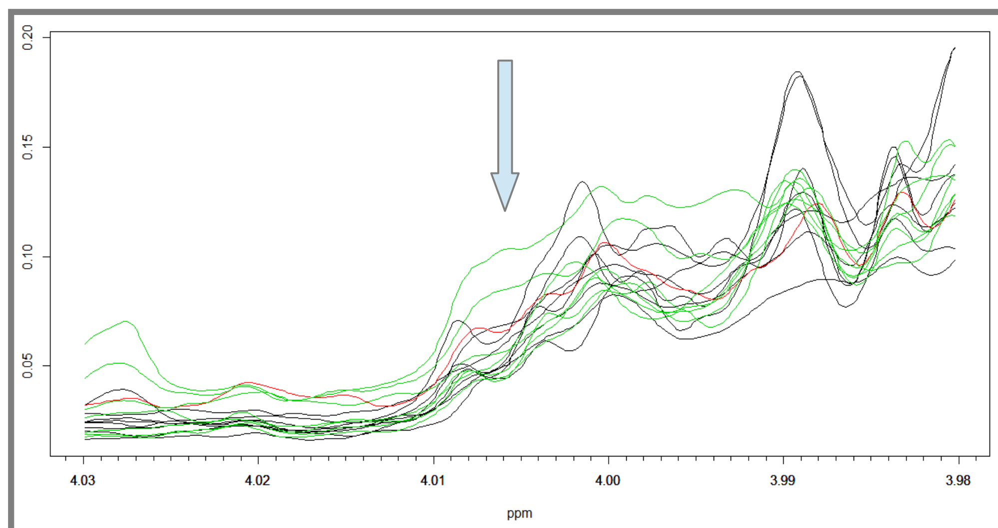


Fig. 11: Region of the serum NMR CPMG spectra where the signal from 2-HG is found (arrow). In black are samples with wild-type gene, in red a IDH-1 mutant and in green IDH-2 mutants. It is visible how two IDH-2 mutant subjects show a higher signal intensity.

Female and male serum samples were analyzed separately as well, in order to remove any potential gender-bias. PCA revealed differences between patients and control spectra, which were mainly due to increased lactate and glucose signals and lower alanine signals for males and mainly increased lactate concentration in females, compared with controls. The reported difference in glucose and alanine concentrations were also found in literature [Wang et al., 2013].

8.7.2 Foodomics for the investigation of the dietary changes in breast cancer patients

The aim of this research was to investigate the possible link between dietary habits and the metabolic profile of a population of women affected by premature stage breast cancer with respect to a control cohort of healthy women. This study was a preliminary attempt to develop a protocol in light of a much bigger investigation done in collaboration with an important center for Cancer Research and Cure (IRST – IRCCS, Meldola (FC)). The following step in the research will be to evaluate the presence of clustering among subjects in relation to their prognosis and clinical information and to relate with dietary data.

Materials and methods

The pilot study recruited 16 subjects from 45 to 74 years old. All subjects were female and included 6 healthy controls and 6 patients. The study was approved by the Ethical Committee IRST IRCCS-AVR (CEIIAV). Patients were characterized by histological diagnosis of operable mammary carcinoma. They were all non-smokers, did not have any previous cancer and had no other important condition. Controls were recruited from the breast cancer screening program and were healthy non-smoker subjects with similar demographical characteristics to the patients (i.e. age). Anthropometric data were collected, together with dietary information (food frequency questionnaires and 24-hour dietary recall). Five urine and blood samples were collected from patients at fast and in the morning in the days -14, -7, -1, +14, +28 from the day of surgery. Controls had instead only one sample of urine and blood collected, again in the morning and fasted state. The 24-hour dietary recall (24HDR) was filled the day before sampling. 5 mL of blood and 5 mL of urine were centrifuged at 2500 rpm for 15 minutes. 2 aliquots of serum of 2 mL and 1 aliquot of urine of 2 mL were then transferred in criovials and frozen at -80°C. 540 µL of sample (urine or plasma) were adjusted with 60 µL of 1.5 M KH2PO4 potassium phosphate buffer (pH 7.4) in deuterium oxide (D2O) containing 0.1% TSP (3-TrimethylSilyl-Propanoic-2,2,3,3-d4 acid sodium salt). The samples were then centrifuged for 10 minutes at 4° C and 14k rpm. 590 µL were then collected and transferred in a 5mM NMR tube. NMR spectra were recorded at 300 K on a Bruker US+ Avance III spectrometer operating at 600 MHz, equipped with a BBI-z probe and a B-ACS 60 sampler for automation (Bruker BioSpin). Urine samples were acquired employing a 1D-NOESY-presat pulse sequence, with 64 scans, whilst serum spectra were acquired using both a 1D-NOESY-presat pulse sequence, with 4 scans and a CPMG sequence with 256 scans. Each spectrum was automatically processed adjusting

phase and baseline with the command apk0.noe using TopSpin version 3.0 (Bruker BioSpin) and applying a line broadening factor of 1 Hz. Each spectrum was automatically processed adjusting phase and baseline with the command apk0.noe using TopSpin version 3.0 (Bruker BioSpin) and applying a line broadening factor of 1 Hz. The chemical shifts were internally referenced to the Trimethylsilyl propanoic acid (TSP, Cambridge Isotope Laboratories) at 0.00 ppm (part per million of frequency). After the Fourier Transformation (FT) and prior to multivariate analysis, data underwent pre-processing using the R 3.2.2 software (http://www.r-project.org/). Signal assignment was carried out by comparing their chemical shift, multiplicity and J-coupling values with literature and by using Chenomx software (Chenomx Inc.). Spectral peripheral regions, solvent signal and, in the case of urine, also the urea signal, were removed from the spectra. Spectra were then then normalization through the Probabilistic Quotient (PQN) algorithm [Dieterle et al., 2006] and reduced in dimension through binning (100 points width).

The information collected from dietary recalls and food frequency questionnaires (FFQ) was reported in an Excel sheet as food dose (S/M/L) and frequency (per day/per week). The quantity for each dose of each food item was calculated using the Metadieta software (Meteda S.r.l.). The same software was employed for the calculation of nutritional values for 100 g of each food item. A tailor-made algorithm developed on Qbasic was then employed to elaborate all data in order to gain information on the daily intake of food and nutrients for each subjects, both from the FFQs and 24HDR.

Results and discussion

First analysis of NMR spectra was carried out through Principal Component Analysis (PCA), in order to detect possible outliers and clusterings. Immediately one sample resulted as outliers and, after removal of this one, a group of subjects was spotted as completely different from the other samples due to an abnormal high intesity of the signal for trimethylamine-N-oxide (TMAO) (data not shown). The removal of these outliers showed the presence of further outlying subjects, in this case due to the signals from mannitol, a molecule that is usually found in candies or chewing-gums [Lenz et al. 2004]. After the final removal of these outlying samples, a PCA without outliers was calculated. In the scoreplot of this PCA, though, no clear separation or clustering among samples is visible, neither due to the belonging to the patients or control groups nor to the sampling time-point. Anyhow, some patients show a clear distinction from patients, therefore to better investigate this possible difference and avoid a big influence of

inter-day variation, two approaches were selected. In the first one, the average PC scores for patients before and after surgery were calculated and re-plotted on the score-plot (Fig. 12).

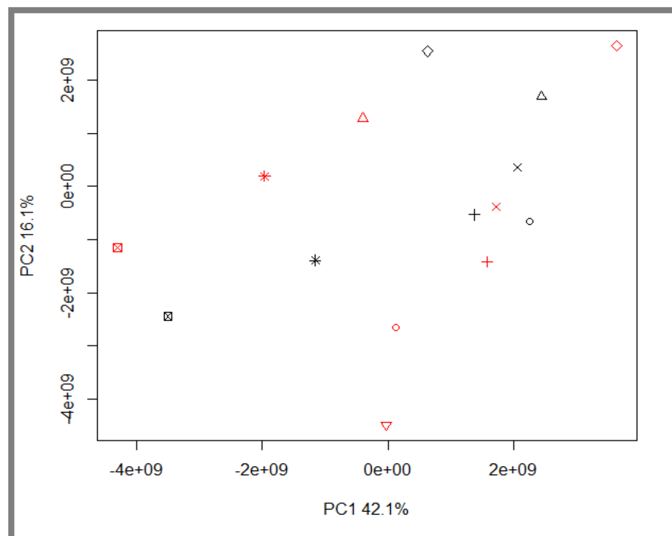


Fig. 12: Average scores for each patient before (black) and after (red) surgery. One subject (downward facing triangle) only has post-surgery average, since all the pre-surgery spectra were spotted as outliers.

Different behaviours can be spotted: most subjects shift towards the left of the scoreplot, some also move towards low PC 2 levels, whilst others just move downwards. Two subjects, instead, move upwards, in contrast with all the other ones. A similar analysis was carried out considering the median spectra for pre and post-surgery spectra and the median spectrum for controls, since generally the median value is more robust than the average, especially when outliers are present. A PCA was then calculated on these median values (Fig. 13).

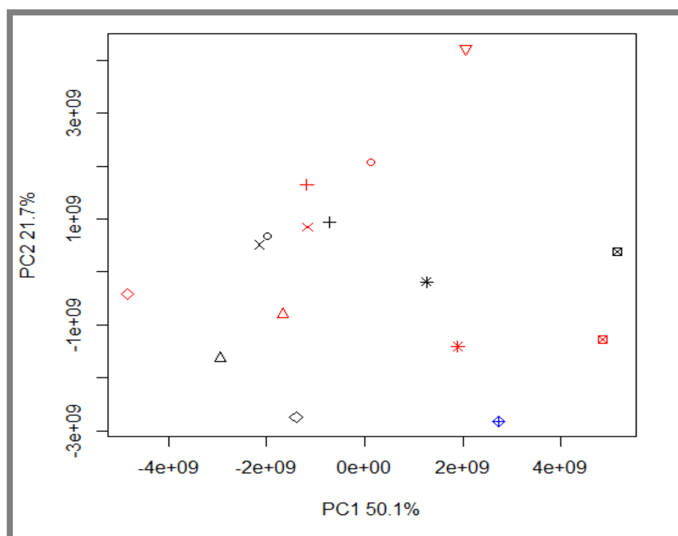


Fig. 13: PCA calculated on the median spectra for controls (blue) and for each patient before (black) and after (red) surgery. One subject (downward facing triangle) only has post-surgery average, since all the pre-surgery spectra were spotted as outliers.

This PCA showed results extremely similar to the ones previously described (fig. 12). Controls seem to be on the lower-right side of the score-plot, the direction to which the two patients going in the opposite directions from all others move. Anyhow, it can not be affirmed that this simply means a return to a “normal” and healthy state. The metabolites mostly responsible for the movements along PC1 and PC2 are creatinine and hippurate, together with TMAO and citrate. Controls show higher concentrations for these metabolites. This information will have to be proved with a larger cohort and then, the reasons for these metabolic shifts will be more throroughly investigated. After this initial investigation on the NMR spectra, the dietary patterns that could be observed from the nutritional questionnaires were investigated. Modern nutrition is in fact now focused on the relation between these dietary patterns and the occurrance of specific chronic diseases. Dietary patterns in this cohort were thus obtained from the multivariate analysis of the dietary and nutritional information. At first, macro-categories of food were selected through recurring principal component analysis until all foods were grouped. The selected categories were:

A: olive oil, oil, butter and mayonnaise

B: cheese and nuts

C: pulses, whole-grains and offal meat

D: cereal-based foods (pizza, bread, egg-pasta, cake, etc.)

E: fish and beef

F: eggs, poultry, pork and seafood

G: breakfast food

H: potatoes, fries, pasta

I: vegetables

L: fruits

M: red wine and beer

N: milk

O: jam, juice, sodas, caffeinated drinks

These food groups will then be employed for the analysis of dietary patterns in the larger cohort of this project. Information about the intake of these macro-food categories will be summed up and subjects will be then clustered in relation to their specific consumptions. In this way it will be possible to assess whether there is a link between the metabolic behaviour and the dietary habits.

Conclusions

The two studies here reported defined an initial approach into the definition of specific metabolomes related to different disease (leukemia or cancer) and the relation between the health state and nutrition. In the first study, metabolite biomarkers for leukemia were found and further study could also help in the stratification of AML patients in relation to other characteristics (i.e. Genomics). The continuation of the study on more sample could improve the characterization of AML through metabolomics. This could be of help in other research, since the definition of disease-specific metabolotypes could aid in the description of the human metabolome. The second study helped in the definition of an analytical pipeline with the scope of defining the metabolic state of cancer patients before and after surgery and later relate this information to dietary patterns found through the statistical analysis of the nutritional and dietary data. This pipeline will be employed in the larger sample cohort that will be available from this project, but could also be of use in other investigations of the same kind.

REFERENCES CHAPTER 8

Ashihara, H., Ludwig, I. A., Katahira, R., Yokota, T., Fujimura, T., & Crozier, A. (2015). "Trigonelline and related nicotinic acid metabolites: occurrence, biosynthesis, taxonomic considerations, and their roles in planta and in human health." Phytochemistry Reviews, 14(5), 765-798.

Beecher, C.W.W. (2003) "The Human Metabolome" Metabolic Profiling: Its Role in Biomarker Discovery and Gene Function Analysis, Springer US, 311-319.

Beger, R. D., Dunn, W., Schmidt, M. A., Gross, S. S., Kirwan, J. A., Cascante, M., Brennan, L., Wishart, D.S., Oresic, M., Hankemeier, T., Broadhurst, D. I., Lane, A.N., Suhre, K., Kastenmüller, G., Sumner, S.J., Thiele, I., Fiehn, O., & Kaddurah-Daouk, R. (2016). "Metabolomics enables precision medicine:"a white paper, community perspective""Metabolomics,12(10), 149.

Bouatra, S., Aziat, F., Mandal, R., Guo, A.C., Wilson, M.R., Knox, C., Bjorndahl, T.C., Krishnamurthy, R., Saleem, F., Liu, P., Dame, Z.T., Poelzer, J., Huynh, J., Yallou, F.S., Psychogios, N., Dong, E., Bogumil, R., Roehring, C., & Wishart, D.S. (2013) "The Human Urine Metabolome" PloS one, 8 (9) 1-28.

Daykin, C.A., Van Duynhoven, J.P.M., Groenewegen, A., Dachtler, M., Van Amelsvoort, J.M.M., & Mulder, T.P.J. (2005) "Nuclear Magnetic Resonance Spectroscopic Based Studies of the Metabolism of Black Tea Polyphenols in Humans" Journal of Agricultural and Food Chemistry, 53(5), 1428−1434.

Dieterle, F., Ross, A., Schlotterbeck, G., & Senn, H. (2006). "Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics." Analytical chemistry, 78(13), 4281-4290.

Esko, T., Hirschhorn, J.N., Feldman, H.A., Hsu, Y.H.H., Deik, A.A., Clish, C.B., Ebbeling, C.B., & Ludwig, D.S. (2017) "Metabolomic profiles as reliable biomarkers of dietary composition" American Journal of Clinical Nutrition, ajcn144428.

Gonda, H. L., & Lindberg, J. E. (1997). "Effect of diet on milk allantoin and its relationship with urinary allantoin in dairy cows." Journal of Dairy Science, 80(2), 364-373.

Jacobs, D.M., Spiesser, L., Garnier, M., de Roo, N., van Dorsten, F., Hollebrands, B., van Velzen, E., Draijer, R., & van Duynhoven, J.P.M. (2012) "SPE–NMR metabolite sub-profiling of urine", Analytical and Bioanalytical Chemistry, 404(8), 2349–2361.

Kaput, J. (2016) "Systems-Level Nutrition Approaches to Define Phenotypes Resulting from Complex Gene-Environment Interactions" In Next-Generation Nutritional Biomarkers to Guide Better Health Care, 84,1-13 Karger Publishers.

Kochhar, S., Jacobs, D.M., Ramadan, Z., Berruex, F., Fuerholz, A., & Fay, L.B. (2006) "Probing gender-specific metabolism differences in humans by nuclear magnetic resonance-based metabonomics," Analytical Biochemistry, 352(2), 274–281.

Krug, S., Kastenmüller, G., Stückler, F., Rist, M.J., Skurk, T., Saile, M., Raffler, J., Römisch-Margl, W., Adamski, J., Prehn, C., Frank, T., Engel, K.H., Hofmann, T., Luy, B., Zimmermann, R., Moritz, F., Schmitt-Kopplin, P., Krumsiek, J., Kremer, W., Huber, F., Oeh, U., Theis, F.J., Szymczak, W., Hauner, H., Suhre, K., & Daniel, H. (2012) "The dynamic range of the human metabolome revealed by challenges" The FASEB Journal, 26(6), 2607-2619.

Lebiedzińska, A., & Szefer, P. (2006). "Vitamins B in grain and cereal–grain food, soy-products and seeds." Food chemistry, 95(1), 116-122.

Lenz, E. M., Bright, J., Wilson, I. D., Hughes, A., Morrisson, J., Lindberg, H., & Lockton, A. (2004). "Metabonomics, dietary influences and cultural differences: a 1 H NMR-based study of urine samples obtained from healthy British and Swedish subjects." Journal of pharmaceutical and biomedical analysis,36(4), 841-849.

Liu, L., Aa, J., Wang, G., Yan, B., Zhang, Y., Wang, X., Zhao, C., Cao, B., Shi, J., Li, M., Zheng, T., Zheng, Y., Hao, G., Zhou, F., Sun, J., & Wu, Z. (2010) “Differences in metabolite profile between blood plasma and serum” Analytical Biochemistry, 406(2), 105–112.

Mucchetti, G., Locci, F., Gatti, M., Neviani, E., Addeo, F., Dossena, A., & Marchelli, R. (2000). “Pyroglutamic acid in cheese: presence, origin, and correlation with ripening time of Grana Padano cheese.” Journal of dairy science, 83(4), 659-665.

Playdon, M.C., Moore, S.C., Derkach, A., Reedy, J., Subar, A.F., Sampson, J.N., Albanes, D., Gu, F., Kontto, J., Lassale, C., Liao, L.M., Männistö, S., Mondul, A.M., Weinstein, S.J., Irwin, M.L., Mayne, S.T., & Stolzenberg-Solomon, R. (2016) “Identifying biomarkers of dietary patterns by using metabolomics” The American Journal of Clinical Nutrition, ajcn144501.

Psihogios, N.G., Gazi, I.F., Elisaf, M.S., Seferiadis, K.I., & Bairaktari, E.T. (2008) “Gender-related and age-related urinalysis of healthy subjects by NMR-based metabonomics”, NMR in Biomedicine, 21(3),195-207.

Psychogios N., Hau, D.D., Peng, J., Guo, A.C., Mandal, R., Bouatra, S., Sinelnikov, I., Krishnamurthy, R., Eisner, R., Gautam, B., Young, N., Xia, J., Knox, C., Dong, E., Huang, P., Hollander, Z., Pedersen, T.L., Smith, S.R., Bamforth, F., Greiner, R., McManus, B., Newman, J.W., Goodfriend, T., & Wishart, D.S. (2011) “The human serum metabolome” PloS one, 6(2), e16957.

Scalbert, A., Brennan, L., Manach, C., Andres-Lacueva, C., Dragsted, L.O., Draper, J., Rappaport, S.M., van der Hooft, J.J., & Wishart, D.S. (2014) “The food metabolome: a window over dietary exposure.” American Journal of Clinical Nutrition, 99(6), 1286-308.

Smirnov K. S., Maier, T.V., Walker, A., Heinzmann, S.S., Forcisi, S., Martinez, I., Walter, J., Schmitt-Kopplin, P. (2016) “Challenges of metabolomics in human gut microbiota research” International Journal of Medical Microbiology, 306(5), 266-279.

St-Onge, M. P., Farnworth, E. R., & Jones, P. J. (2000). “Consumption of fermented and nonfermented dairy products: effects on cholesterol concentrations and metabolism.” The American journal of clinical nutrition, 71(3), 674-681.

Trimigno, A. (2013) “Indagine sulle relazioni tra stato nutrizionale e profile metabolic umano attraverso la meta-analisi degli spettri NMR di urine e dei questionari FFQ di frequenza dei consumi alimentari” (Master’s thesis)

van Dorsten, F., Daykin, C.A., Mulder, T.P.J., & van Duynhoven, J.P.M. (2006) “Metabonomics Approach To Determine Metabolic Differences between Green Tea and Black Tea Consumption” Journal of Agricultural and Food Chemistry, 54(18), 6929-6938.

Wang, Y., Zhang, L., Chen, W. L., Wang, J. H., Li, N., Li, J. M., Mi, J. Q., Zhang, W. N., Li, Y., Wu, S. F., Jin, J., Wang, Y. G., Huang, H., Chen, Z., Chen, S. J., & Tang, H. (2013). “Rapid diagnosis and prognosis of de novo acute myeloid leukemia by serum metabonomic analysis.” Journal of proteome research, 2(10), 4393-4401.

Wen, H., Cho, H. R., Yun, T., Kim, H., Park, C. K., Lee, S. H., Choi, S. H., & Park, S. (2015). “Metabolomic comparison between cells over-expressing isocitrate dehydrogenase 1 and 2 mutants and the effects of an inhibitor on the metabolism” Journal of neurochemistry, 132(2), 183-193.

Wishart, D.S., Tzur, D., Knox, C., Eisner, R., Guo, A.C., Young, N., Cheng, D., Jewell, K., Arndt, D., Sawhney, S., Fung, C., Nikolai, L., Lewis, M., Coutouly, M.A., Forsythe, I., Tang, P., Shrivastava, S., Jeroncic, K., Stothard, P., Amegbey, G., Block, D., Hau, D.D., Wagner, J., Miniaci, J., Clements, M., Gebremedhin, M., Guo, N., Zhang, Y., Duggan, G.E., Macinnis, G.D., Weljie A.M., Dowlatabadi, R., Bamforth, F., Clive, D., Greiner, R., Li, L., Marrie, T., Sykes, B.D., Vogel, H.J., & Querengesser, L.(2007) “HMDB: the Human Metabolome Database”, Nucleic Acids Research, Vol. 35(Database Issue), D521–D526.

Yu, Z., Kastenmüller, G., He, Y., Belcredi, P., Möller, G., Prehn, C., Mendes, J., Wahl, S., Roemisch-Margl, W., Ceglarek, U., Polonikov, A., Dahmen, N., Prokisch, H., Xie, L., Li, Y., Wichmann, H.E., Peters, A., Kronenberg, F.,

Suhre, F., Adamski, J., Illig, T., Wang-Sattler, R. (2011) “Differences between Human Plasma and Serum Metabolite Profiles” PloS One, 6 (7), 1-6.

Yuhara, K., Yonehara, H., Hattori, T., Kobayashi, K., & Kirimura, K. (2015). “Enzymatic characterization and gene identification of aconitate isomerase, an enzyme involved in assimilation of trans-aconitic acid, from Pseudomonas sp. WU-0701.” FEBS journal, 282(22), 4257-4267.

Zheng, H., Yde, C. C., Clausen, M. R., Kristensen, M., Lorenzen, J., Astrup, A., & Bertram, H. C. (2015). “Metabolomics investigation to shed light on cheese as a possible piece in the French paradox puzzle.” Journal of agricultural and food chemistry, 63(10), 2830-2839.

Zuppi, C., Messana, I., Forni, F., Rossi, C., Pennacchietti, L., Ferrari, F., & Giardina, B. (1997) “1H NMR spectra of normal urines" Reference ranges of the major metabolites” Clinica Chimica Acta, 265(1), 85-97.

CHAPTER 9: CONCLUSIONS

The research work carried out during this PhD was focused on the application of the foodomics approach in many different, though correlated, topics. Foodomics is a new omic science researching the relations between food and nutrition in order to allow tailor-made dietary recommendation and help in the understanding of the individual wellness state.

Of course, in order to reach such knowledge, the starting point is the definition of food itself. In this research work, literature on the subject was investigated to define all the possible characteristics and factor affecting food quality and the foodomics platforms and technologies applied and applicable for this kind of investigation. Moreover, some of these factors and their influence on specific food parameters were investigated. It is in fact fundamental to understand how different farming and transforming techniques, such as organic cultivation or food enrichment with bioactive compounds impact on the final nutritional values of products. It is not always simple to define how specific factors alter the final content in nutrients of a food product, since transformation processes and storage can greatly influence this. This is why the storage behaviour of food products was investigated, in order to understand how the food matrix itself could protect bioactive molecules added in the food product. An analytical approach employing NMR spectroscopy was thus developed and could be applied in similar investigation. This approach will be further tested in more batches of the analysed enriched foods, in order to strengthen the algorithm and validate the obtained results. In the future, the developed algorithm can be employed on different food products, both traditional and novel, so that a greater understanding of the interaction between the food matrix and specific nutrients and non-nutrients during shelf-life can be gained. This is particularly useful for the knowledge on the real nutritional value of foods, which is of particular interest for enriched products, containing specific bioactive molecules with health benefits and which need to be able to carry out the desired function when consumed. For this reason, it is fundamental to know whether these molecules are protected during the shelf-life of the product and then released during digestion. A more thorough approach investigating the whole path from shelf-life matrix behavior to digestion will be employed in future studies, so that it is possible to follow the whole destiny of specific molecules of interest.

In effect, even when a product is found to contain beneficial molecules and those can be protected during the product's shelf-life, their actual bioaccessibility and bioavailability can be

very different than that expected. This is why in vitro digestion studies are carried out. During this research work, many different food products (from simple bell peppers to cooked meat-based meals) were investigated after in vitro digestion to assess whether the most important and beneficial molecules (such as amino acis) are actually more available for human use after specific modification in the original product.

Further studies will be carried out to better understand the role of specific production and processing steps, such as cooking, in the final fate of the molecules of interest. It would be of particular use, in fact, to understand the relationship between all the different variables impacting on the production of a food item and the bioavailability of beneficial molecules through digestion, especially since nowadays many food companies are trying to claim greater digestibility for their products and are developing enriched food items.

As stated, a thorough investigation of the whole route of a metabolite from the food to its metabolism by the human body and the microbiota would be the goal of foodomics.

The next step for this foodomics goal, is thus the understanding of the human metabolome and all the different factors impacting on it. One small step towards this can be done using animal models, such as the pig. Pigs are believed to have a very similar in their metabolism to humans, therefore the research on their specific metabolome can be of great help for further human studies. Moreover, since many pig-derived food products exist, the study of the impact of nutrition or diseases on the pig metabolome is of great importance also for the food industry and for veterinary research. In this research work, two studies were carried out on pigs, investigating colostrum, urine and plasma. In this way the impact of diseases, drug-treatments and genetic was investigated. In addition, more knowledge on the molecular profiling of these biofluids was gained.

Finally, the human metabolome was investigated. Single factors such as gender or geographical origin were found to greatly impact on the human metabolic profile. Moreover, a foodomic approach was defined in order to search for specific dietary and food intake biomarkers. This will be particularly helpful for nutritional and health studies where it is necessary to define the compliance of investigated subjects and understand possible relationships between diet and wellness state. In this sense, it is also necessary, thus, to understand how different diseases impact on the human metabolome, and this is why two studies with this goal were included in this research work.

Forthcoming studies will continue these investigation, since more knowledge on the human

metabolome is still needed. The aspiration is to obtain a full view on the factors impacting on the metabolome and the derived patterns and to compile a related database, which will render more straightforward the identification of the causes of specific metabotypes and sample clustering in observational studies and the assessment of the metabolic alterations happening in analysed cohorts.

To sum up, this research work helped in the profiling of a full foodomic approach from the food item to a tailor-made nutrition capable of helping subjects in prevention or recovery from diseases. The studies carried out have helped in the definition of specific and robust analytical pipelines and metabolomics standard operating procedures that can be of great help in future foodomics studies. Moreover, the great power of NMR spectroscopy was once again underlined. The capability of NMR spectroscopy of quickly picturing a great range of molecules with very little sample preparation and without destroying samples makes it possible to analyse easily many different samples.

ANNEX 1 – MATERIALS AND METHODS (SOPs)

Equipment

Laboratories supplies

- Pipettes Standard – Gilson's Pipetman® P (P10, P20, P100, P200, P1000, P5000 and P10000. Range of volumes from 5 µl to 10000 µl) with suitable tips (Diamond ® precision tip)
- Eppendorf ® safe lock microcentrifuge tubes volumes 0.5, 1.5 and 2 ml
- Bottle in polypropylene (PP) from 100-250-500 and 1000 ml
- Laboratory ceramic mortar grinder with pestle
- Laboratory glassware
- Beakers, low form, with spout (50-250-500 ml) by Simax
- Beute from 50 ml (SCHORR DURAN)
- Cylinders from 50-1000 ml (PIREX)
- Quartz cuvets from 3 ml (EXACTA)
- Corning® Disposable Pasteur Pipettes, Bulk Pack, Non sterile (SIGMA ALDRICH®)
- Graduated pipette from 5 to 20 ml (class A)
- Inox steel Spatulas
- Inox steel Scalpels
- Syringe filter in cellulose with diameter 0.20 µm
- Pipette controller
- AMPOL NMR sample tuber for use up to 700 MHz NMR (203, round bottom)

Safety and protection supplies

- Safety Eyewear Glasses
- Natural Latex Powdered and Powder – Free Exam Gloves
- White lab coat

Reagents

- Trichloroacetic acid (CCl_3COOH, 163.39 g/mol, TCA) 6.1 N, SIGMA ALDRICH
- Potassium hydroxide (KOH, 56.11 g/mol), SIGMA ALDRICH
- Sodium sulfate (Na_2SO_4, 142.04 g/mol) granular anhydrous, SIGMA ALDRICH
- Potasssium dihydrogen phospate (KH_2PO_4, 136.09 g/mol), PANREAC
- 3-(trimethylsilyl)-propionic-2,2,3,3-d4 acid sodium salt, ($(CH_3)_3SiCD_2CD_2CO_2Na$, 127.27 g/mol TSP), 98 atom % D, BRUKER
- Potassium hydrogen phosphate anhydrous (K_2HPO_4, 174.18 g/mol), PANREAC
- Hydrochloric Acid (HCl, 36.01 g/mol) 37% purity, SIGMA ALDRICH
- Milli-Q demineralized water, SIGMA-ALDRICH
- Deuterium oxide (D_2O, 20.04 g/mol, 99.9% purity), BRUKER
- Deuterated Chloroform ($CHCl_3$, 99.96% purity) SIGMA ALDRICH
- Pancreatin (Sigma-Aldrich, P7545)
- Human salivary α-amylase (Sigma-Aldrich, 1031)
- Porcine pepsin (Sigma -Aldrich, P7012)
- Porcine trypsin (Sigma-Aldrich, T0303)
- Bile (Sigma-Aldrich, P8631)
- $CaCl_2(H_2O)_2$ (Merck 2382)
- NaOH (Merck 9141)
- HCl (J. T. Baker 6081)
- KCl (Merck 4936)
- KH_2PO_4 (J. T. Baker 0240)
- $NaHCO_3$ (Merck 6329)
- $MgCl_2(H_2O)_6$ (Merck 5833)
- $(NH_4)_2CO_3$ (Sigma-Aldrich, 207861)
- Pefabloc SC (4-(2-Aminoetyl)benenesulfonyl fluoride) (Fluka, Sigma-Aldrich, Ref: 76307)
- Ultrapure type I water, generated by a Milli-Q system (referred in text as water)

Annex 1: *Materials and Methods (SOPs)*

Solutions – Solutions preparation

- KOH (25% w/w): 25g KOH dissolved carefully and with stirring in 75 ml of distilled water
- Trichloroacetic acid (7.5%): 7.5 g of TCA dissolved in 92.5 ml of distilled water
- Solution HCl (1 M): 82.85 ml of HCl (37% or 12.07 M) dissolved in 917.15 ml of distilled water
- 500 mM Pefabloc: 119.85 mg of Pefabloc weighted and diluted in 1 mL of Milli-Q water
- 0.3M CaCl2(H2O)2: 2.205 g of CaCl2•2 H2O weighted and diluted in 50 mL of Milli-Q water
- 0.5M KCl: 7.456 g of KCl weighted and diluted in 200 mL of Milli-Q water
- 0.5M KH2PO4: 6.804 g of KH2PO4 weighted and diluted in 100 mL of Milli-Q water
- 1M NaHCO3: 16.802 g of NaHCO3 weighted and diluted it in 200 mL of Milli-Q water
- 2M NaCl: 23.378 g of NaCl weighted and diluted in 200 mL of Milli-Q water
- 0.15M MgCl2(H2O)6: 1.525 g of MgCl2(H2O)6 weighted and diluted in 50 mL of Milli-Q water
- 0.5M (NH4)2CO3 : 2.4025 g of (NH4)2CO3 weighted and diluted it in 50 mL of Milli-Q water
- 1M NaOH and 1M HCl: employed for pH adjustment of stock solutions of simulated digestion fluids

Buffer Solution

- Phosphate Buffer, 0.1 M and at pH 7.00

Instruments

- Eletronic digital tecnica Balance (max 2200 g, d=0.01 g), SCALTEC (SBA 52)
- Eletronic digital analytical Balance (max 220g, d=0.0001g), SCALTEC (SBA 31)
- IKA ® ULTRA-TURRAX® homogenizer T18, basic, AC input 115 V
- JENWAY Model 3310 pH Meter with glass bodied combination electrode swing arm electrode holder & ATC
- Heating magnetic stirrer mod. ARE, VELP Scientifica®
- Beckman Coulter TM Microfuge® 18 Microcentrifuge (max 14000 rpm adjustable in 500 increments)
- Spectrophotometer UV-Vis (UV-1601-Shimadzu)
- FT-NMR Avance Bruker AvIII (600 MHz) spectrometer Ultra Shield Plus equipped with:
  - Electronic 3-channel RF consisting amplifiers from 100 - Watt 1H and broadband for X 300 Watt
  - Control unit gradients GCU
  - Control system of the temperature BVT3000
  - Probe 5 mm with Z grad 1H-13C-15N
  - Autosampler with 60 holders
  - Software Topspin 3.0

Software

NMR data processing

- MESTREC

  MestRe-C (http://mestrelab.com) is a software suite designed for the processing, visualization and analysis of HR-NMR data. It can display and manipulate single or multiple spectra, allowing their analysis and interpretation. It can handle various spectral formats, from different instrument vendors and it is capable to recognize different data format and convert them into MestRe-C format in an automatic fashion.
- TOPSPIN

  Topspin is an integrated software from Bruker (https://www.bruker.com/products/mr/nmr/nmr-software/software/topspin/overview.html) employed both for the acquisition of NMR spectra through the Bruker Avance spectrometer and consequent spectral processing and analysis. The software allows the access to a great experimental library, containing both standard Bruker pulse sequences and user-defined experimental libraries. Through Topspin it is possible to set up many different experiments. Topspin also allows the import and export of NMR data and spectra in various formats. The software is particularly useful for spectral processing through steps like transforms (i.e. Fourier Transform), phase correction, baseline correction, peak-picking or signal integration.

Chemometrics and data processing

Multivariate statistical analysis was carried out employing various softwares, such as R, MatLab, or Latentix depending on the purpose of the analysis and the chosen statistical technique.

- R

R (http://www.r-project.org) is a GNU project offering a free language and software for graphics and statistical computing. Through the R software it is possible to carry out a vast range of statistical analyses, such as classification or clustering, and graphical displays. It can effectively handle and storage data, operate with matrices and allows users to add further functions and implement statistical and graphical techniques.

- MATLAB
  MatLab (Matrix Laboratory, http://www.mathworks.com) is a platform employing the so-called MatLab language, based on matrix operations. The software offers a great variety of toolboxes, allowing the performance of many algorithms and graphical displays. This computing environment is particularly useful, therefore, for the manipulation and plotting of data and for the creation of new algorithms and user interfaces, permitting the cross-link with programs written in languages such as Fortran, Java, Python, C or C++.
- LATENTIX
  Latentix (http://www.latentix.com/) is a software built for the examination and plotting of raw data and data analysis through PCA or PLS. It allows a user-friendly variable selection and validation, together with data pre-processing tools and the possibility to export the generated analysed matrices and plots.

Spectral assignment

- CHENOMX
  Chenomx NMR Suite is a software for the identification and quantitation of metabolites in NMR spectra. The Chenomx software contains a vast reference library with hundreds of compound models, together with information about their pH sensitivity, line shapes, chemical shifts and width. In addition, Chenomx offers information about the presence of these compounds in biofluids, displaying compound cards with chemical, biological and clinical data.

Main SOPs employed

- *Collection and storage of Urine samples (S.O.P.)*
  A schematic work flow for the management of urine samples is shown in Figure 1. Urine will be collected only from donors who have freely given their informed consent.

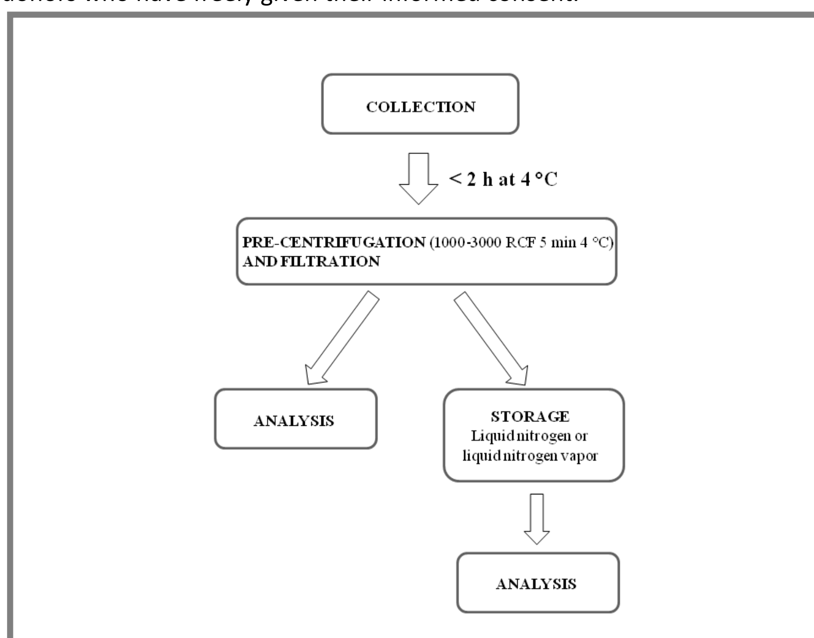


Fig. 1. Work flow for the optimal processing and management of urine samples (from Bernini et al., 2011)

Health and safety warnings

Due to the fact that it is often impossible to know what might be infectious, all specimens are treatedas infectious and handled according to "standard precautions". Urine must be processed only within a level 2 containment laboratory and only by designated and trained staff. White coat, gloves, safety glasses and other individual protection devices must always be worn while collecting and handling samples. All the operations made by operator on the biological samples must be performed using a 4 microbiological safety cabinet. Working with liquid nitrogen and other chemical products is hazardous therefore all procedures must comply with safety rules and regulations.

Recommendations for the subjects

In order to avoid any alteration of the urine sample, the donor must fast overnight before morning pre-prandial collection. Thoroughly clean the genitals with normal saline, if possible. If not indicated in an explicit way, female subjects must not collect urine samples during menstruation. The urine must be collected in a labeled sterile container, after discarding the first urine flow.

Equipment and Supplies needed

- Sterile plastic cup with lid (of 50 ml or more, to be given to the donors)
- Gauze pads
- Soap and clean water (or normal saline) if possible
- Microbiological safety cabinet
- Sterile 50 ml tubes
- Centrifuge with sealed buckets or sealed rotor
- 1 ml pipettor
- Sterile tips for pipettor
- Sterile cryovials (2 ml) to aliquot urine
- Cryoboxes to store the aliquots
- At least -20°C refrigerators

Quality control

The instruments must be checked, calibrated, cleaned and disinfected regularly according to the manufacturer recommendations. All controls must be registered.

Specimen identification

The donor's specimen must be unambiguously identified at the time of collection. Specimens should be labeled and handled respecting patient privacy in accordance with regulations. Each specimen must be labeled with an identifier that links it to the donor unique identification number; this ensures traceability of the specimen and separation of personal and clinical data.

Urine collection

1) Subject must collect the first urine of the morning on empty stomach.
2) The patient must receive clear instructions to discard urine for a few seconds and then to hold the cup in the urine stream for a few seconds to catch a midstream urine sample. This should decrease the risk of contamination from microorganisms living in the urethra;
3) The subject must collect at least 15 ml of urine (for particular analytical demand this minimum value could be increased). To decrease the risk of contamination from skin microorganisms, the patient should be trained to avoid touching the inside or rim of the plastic cup with the skin of the hands, legs or external genitals. The cap must be firmly tightened when the patient has finished;
4) For hospitalized or debilitated patients, it may be necessary to wash the external genitals with soapy water to reduce the risk of contamination. If soap and clean water are not available, the area may be rinsed with normal saline. The area must be dried thoroughly with gauze pads before collecting the urine;

5) Donors identity must be confirmed verbally by checking that name and signature are correct on the informed consent and that name, address and date of birth match those on the specimen accompanying sheet;
6) The specimen receptacle must be labeled with a simple and self-explained 15 digits sample code e.g.
XXX Y 00001 UR 001 A(B,C,..) , attributed as following:
XXX identify the research project
Y identify the recruitment center
00001 identify the donor (e.g. donor No. 1)
UR identify the sample type (in this case UR = urine)
001 identify the sample given by the donor No. 1 on the collection date
A, B, C, D, E,... identify the sample aliquot corresponding to a specific sample
7) For each sample the donor is going to donate, he must fill the appropriate form concerning data about the diet the day before collection, and attach it to the urine sample;
8) The urine sample can be stored at room temperature no longer than 3 hours.
9) The personnel of the collection center takes the sample with respective form which must contain: identifier and anamnesis of donor; copy or reference of informed consent; date and time of collection and shipping; number and type of specimens; approximate temperature of transport; notes.
10) The specimen must be kept refrigerated at 2-8°C maximum and must not be frozen prior to storage. Keeping the specimen refrigerated will decrease the risk of overgrowth of contaminating microorganisms.
11) The specimen should be carried to the laboratory for aliquoting as soon as possible and certainly within 6 hours from morning pre-prandial collection. When in transit, urine collections should be maintained refrigerated. Sample receptacles must be leak-proof and tightly sealed. The laboratory must be informed about arrival of the sample.

Processing procedures

1) Check that all specimens and respective documentation are present; if something is missing, contact collection center; in case of failure, discard the specimen according to safety rules and regulations.
2) Data of each specimen must be registered in electronic and paper archive.
3) The urine should be centrifuged for 5 minutes at 810 Å~ g prior aliquoting. Supernatant should be transferred in a new sterile tube and split at least into five aliquots (1.0 ml each), using the appropriate labeled cryovials;
4) Samples must be stored at -80°C. If it is possible, for long term storage is better to keep the samples in liquid nitrogen or liquid nitrogen vapor.
5) Check that all the data regarding the sample (including processing procedures) and the subject have been collected and registered in an electronic database.

Transportation

Transport must be performed following bio-safety guidelines.
The specimens should be enclosed a sheet containing: number and type of aliquots with their uniqueidentification codes, anamnesis of donor; date and details of production; date of shipping; temperature of transport; notes; instructions for opening the packaging and the sample receptacle; negative results of microbiological tests (if available); eventually, presence of cryogenic gas. Special handling requirements should be provided to couriers responsible for the transport of the specimens.

Transport of frozen samples

It is critical to maintain the cold chain all the times during transport, storage and delivery. The tubes must be transported upright and secured in a leak-proof secondary packaging. Dry ice shall be placed around the secondary packaging or alternatively in an overpack with one or more complete packaging.
The external container must bear: the name, phone number and address of the recipient; the person to contact in case of problems; the inscriptions “biological substance category B” in letters at least 6 mm high, “handle with care”; the diamond-shaped mark (50 mm by 50 mm) reporting inside the UN3373; transport indications; eventually, presence of cryogenic gas.

Annex 1: *Materials and Methods (SOPs)*

Sample handling

Equipment and Supplies needed

- Microbiological safety cabinet
- Centrifuge with sealed buckets or sealed rotor
- 1 ml and 200 ul pipettor
- Sterile eppendorf
- Sterile H (2 ml) to aliquot urine
- NMR tubes

Preparation of solvents

Reagents

- KH_2PO_4 (sodium dihydrogen phosphate) Mw = 119.98
- NaN_3 (sodium azide) Mw = 65.01
- TSP (sodium 3-trimethylsilyl [2,2,3,3 2H4] propionate) Mw = 172.27
- H_2O (HPLC-grade or MilliQ water)
- D_2O (deuterium oxide)
- Concentrated KOH (around 20 M)

Solution 1: Buffer for samples (100 mL 0.3 M phosphate buffer).
a) 1.5M KH_2PO_4/D_2O, pH 7.4 (with KOD/D_2O)
b) 2mM NaN_3.
c) 0.1% TSP (=3-(trimethyl-silyl)propionic acid-d4, Aldrich 269913)
first step: dissolve 10.2g KH2PO4 in 40ml d2O
second step: dissolve 50mg TSP in 3-5ml d2O
third step: mix both solution very well (ultra sonic)
fourth step: adjust pH to 7.4 with KOH-tablets or strong KOD-solution (KOH-tablets in D2O)
The buffer dissolves completely at pH >5
fifth step: fill it up to 50ml and mix very well
Add 10% of this buffer to the urine sample, mix very well

Sample preparation

The sample preparation procedure is described using relative measures in order to make it more general. Adjust these volumes to the NMR probe used.
1. Frozen samples are defrosted at room temperature for 30 min. After thawing the sample preparation procedure has to be completed within 2 hours. The samples have to be gently shaken before use.
2. Centrifuge 630 µL of urine at 14000 RCF (Relative centrifugal force) for 5 min, in order to eliminate the residual cells.
3. Add 540 µL of the supernatant to 60 µL of potassium phosphate buffer (see Section 2.4.2). Mix by inversion or gentle vortex stirring.
4. Put 540 µL of the mixture into a 4.25 mm disposable NMR tube. Store at 4°C until ready for analysis.
3. NMR Analysis and Data storage

Sample handling

- Before placing new samples, remove any other samples from previous run, in order to avoid
- mistakes
- Place the NMR in the autosampler
- Carefully take note of the position in which the tubes are placed

Remember to always randomize samples in order to not introduce spurious effects: if the analysis contemplate comparison between control and treated groups, be sure to not acquire all samples from one group first and then (for example after several days and/or with a new stock of buffer) all the samples form the other groups. In a such situations, it is hard to understand whether or not some differences are due to the metabolic state or to different operative conditions.

Preacquisition procedures

Temperature

- Calibrate temperature using e.g., a standard methanol sample (optional)
- Perform experiments at constant temperature, 300 K. Edit temperatures using the command
- [edte]
- Equilibrate
- Load the sample into the probe and allow the sample to equilibrate in the spectrometer (4-5 min in tube probes)
- Tune and match ([wobb])
- Automatically tune and match every biofluid sample using the command [atma]
- Control tuning and matching using the command [wobb]
- If tune and match is not sufficiently good, tune and match a representative sample (QC sample or first biofluid sample) manually using the command [atmm]

Shimming

- Run the 'topshim gui'
- Within the user interface: TUNE Z only, then TUNE X only, then TUNE Y only
- Pulse length([p1, pl1])
- Determine 90° pulse length for a representative sample using the automatic 'pulsecal' routine.
- Receiver gain ([rg])
- Receiver gain is not adjusted for each sample. It is fixed to a value of 36.
- RF carrier frequency offset for water ([O1])
- Determine the offset of the water signal for a representative sample (e.g., QC sample or the first biofluid sample)
- This parameter is used for the whole data set. Use the 'gs' command to optimize the O1. A fine
- Tuning of o1 can be also done using an iterative optimization procedure described below:
- Run a simple zg with 1 scan on the first sample
- Transform the FID (fp)
- Phase all signal right from the residual water peak, do not care about water phase
- Increase or decrease the O1 depending on the shape of the water signal: if the water signal has a negative dip on the left side that means that O1 is too far left, i.e. too big. Similarly, if the negative dip is on the right hand side O1 is too small. Change accordingly the O1 value and repeat the above procedure until the O1 is optimized

Transfer

- Transfer all settings to the experiments to be acquired

Automation setup:

- Start ICON-NMR by using the command [iconnmr]. Choose [automation]
- Prepare the firs data set
- Give a name to the data set corresponding to the sample's name
- Chose the experiments and the solvent parameters
- Set a title reporting the project, the date, the temperature and the nature of sample. Use the title field to add any significant information, such as the sample code
- Use the "copy" command to create the data sets for all the other samples to be acquired
- Double check everything and start the run
- Do not forsake your samples during the acquisition, especially in case of long run: from time to time come back to the instrument and check if everything is working fine

NMR acquisition

Acquire all NMR spectra using the following parameters

- The 1D NOESY-presat pulse sequence: [-RD-90°-t-90°-tm-90-ACQ]
- Create a new data set with edt
- Read the parameter set rpa
- Set the correct solvent eda�solvent (H_2O + D_20)
- eda|Pulprog = noesygppr1d.comp (double presaturation 1D NOESY version)

- eda|TD = 64k (Number of time domain data point)
- eda|SW = 20 ppm (Spectral width)
- eda|NS = 64 (Number of scans)
- eda|DS = 4 (Dummy scans)
- eda|AQ = 2.72 s (Acquisition time)
- eda|D "array" [D1] = 4.0 s (Relaxation delay - RD)
- eda|D "array" [D8] = 100 ms (Mixing time (tm)) (presaturation quality not highly influenced by this value)
- eda|D "array" [D16] = 200 µs)
- Total acquisition time ~ 4-5 min for 64 scans

NMR processing

Processing of 600 MHz 1H NMR spectra

- Reference the spectra to TSP at δ = 0.00 ppm
- edp|LB = 1 Hz (Line broadening)
- edp|SI = 128k (double the number of Fourier domain points (TD))
- A target line width of not more than 0.5 Hz for the TSP peak at half height is desirable
- Automatically phase the spectra using the command [apk]. If the phasing is not sufficiently good, phase the spectra manually
- Baseline correction using the command [abs]

NMR data storage

- Before storing acquired data, do a visual inspection of the spectra in order to detect any macroscopic problem in the dataset.
- FID (Free Induction Decay) must be stored as it is.
- Raw NMR data should be stored in ASCII format prior to normalization (vertical alignment) and alignment (horizontal alignment) and solvent removal.
- The spectral region containing the water peak and urea must be discarded (e.g. 6-4.5 ppm).

Before any statistical analysis, spectra must be normalized in order to compensate for different dilution factors. The most used scaling techniques are total area scaling, which scales all spectra by the total sum of the intensities, assumes that the total concentration of metabolites in a sample is more or less constant, or probabilistic quotient normalization (PQN), which calculates the most probable normalization coefficients.

Troubleshooting

Check carefully the quality of your spectra. Large peaks width or difficulties in shimming can be due to the presence of air micro-bubbles in the samples. In this case a good solution is to vigorously spin the tube with an hand centrifuge immediately before the analysis. Imidazole proton peaks of histidine are good indicators of the pH of urine samples. If too large misalignments of these pecks are found, some error in the buffer preparation is likely. If the shift between samples due to pH differences is a problem, the application of a bucketing step with size 0.02 to 0.04 ppm can be a solution. If some spectra show extremely higher peaks than the average, it is better to exclude that peaks from the normalization algorithm.

Collection and storage of serum samples (S.O.P.)

A schematic work flow for the management of serum samples is depicted in Figure 2. Serum will be collected only from donors who have freely given their informed consent.

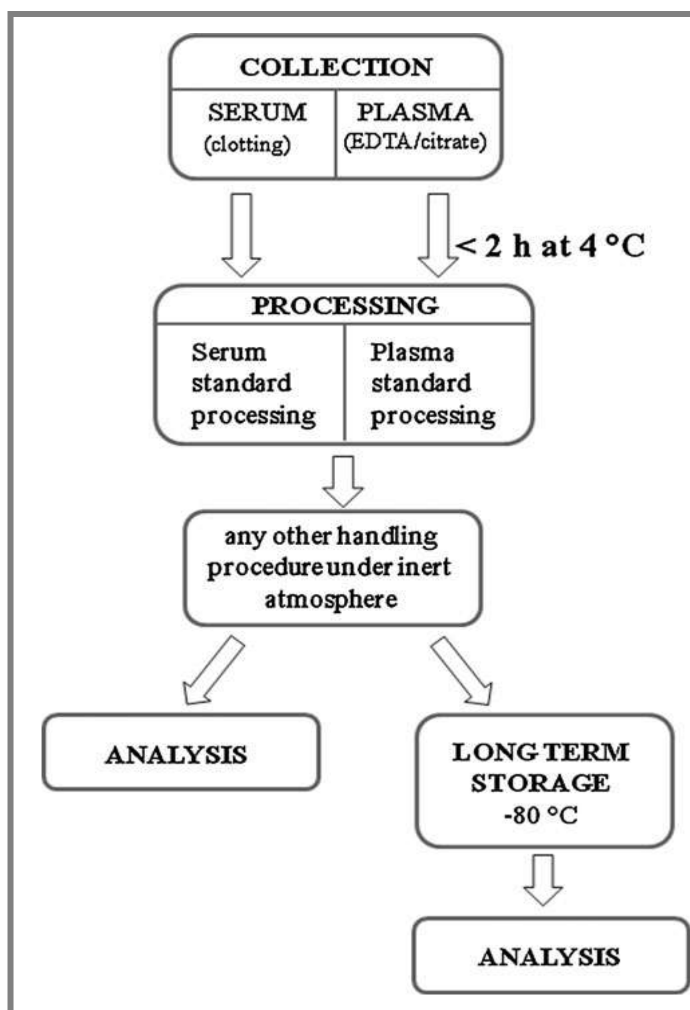


Fig. 2: Work flow for the optimal processing and management of serum samples (from Bernini et al.,2011)

Health and safety warnings

Due to the fact that it is often impossible to know what might be infectious, all specimens are treated as infectious and handled according to "standard precautions". Serum must be processed only within a level 2 containment laboratory and only by designated and trained staff. White coat, gloves, safety glasses and other individual protection devices must always be worn while collecting and handling samples. All the operations made by operator on the biological samples must be performed using a microbiological safety cabinet. Working with liquid nitrogen and other chemical products is hazardous therefore all procedures must comply with safety rules and regulations.

Recommendations for the subjects

In order to avoid any unwanted source of variation in the serum sample, the donor must observe the recommendation for the inclusion in the study and his compliance to the whole experimental protocol must be checked by appropriate investigation. The serum must be collected in sterile conditions.
Equipment and Supplies needed

- Refrigerated centrifuge with sealed buckets or sealed rotor
- 1 ml pipettor
- Sterile tips for pipettor
- Sterile cryovials (2 ml) to aliquot serum
- Cryoboxes to store the aliquots
- At least -20°C refrigerators, better -80°C Refrigerators

Quality control

The instruments must be checked, calibrated, cleaned and disinfected regularly according to the manufacturer recommendations. All controls must be registered.
Specimen identification
The donor's specimen must be unambiguously identified at the time of collection. Specimens should be labelled and handled respecting patient privacy in accordance with regulations. Each specimen must be labelled with an identifier that links it to the donor unique identification number; this ensures traceability of the specimen and separation of personal and clinical data.

Serum collection

- Donors identity must be confirmed verbally by checking that name and signature are correct on the informed consent and that name, address and date of birth match those on the specimen accompanying sheet
- Blood samples are withdrawn from healthy donors using different SSTTM II Advance BD Vacutainers®, or equivalent, for serum isolation
- At least 2 aliquots of 5 ml serum must be collected for each subject
- The vacutainer must be labelled with a simple and self-explained sample code, identifying the research project, the recruitment centre, the donor, the sample given by the donor on the collection time and date, the sample aliquot
- For each sample the donor is going to donate, he must fill the appropriate form concerning data about the diet the day before collection, and attach it to the serum sample
- The serum sample is incubated for clotting at room temperature for 30 min
- After the incubation, vacutainers for serum collection are centrifuged at 1,500 RCF (Relative Centrifugal Force) for 10 min at 25°C

Processing procedures

- Supernatant should be transferred in a new sterile tube and split at least into five aliquots (1.0 ml each), using the appropriate labelled cryovials
- Samples must be stored at -80°C. If it is possible, for long term storage (>1 year) is better to keep the samples in liquid nitrogen or liquid nitrogen vapor
- Check that all the data regarding the sample (including processing procedures) and the subject have been collected and registered in an electronic database

Transportation

Transport must be performed following bio-safety guidelines. The specimens should be enclosed a sheet containing: number and type of aliquots with their unique identification codes, anamnesis of donor; date and details of production; date of shipping; temperature of transport; notes; instructions for opening the packaging and the sample receptacle; negative results of microbiological tests (if available); eventually, presence of cryogenic gas. Special handling requirements should be provided to couriers responsible for the transport of the specimens.
At the arrival, the operators in the NMR lab must:

- Check that all specimens and respective documentation are present; if something is missing, contact collection centre; in case of failure, discard the specimen according to safety rules and regulations.
- Data of each specimen must be registered in electronic and paper archive.

Transport of frozen samples

It is critical to maintain the cold chain all the times during transport, storage and delivery. The tubes must be transported upright and secured in a leak-proof secondary packaging. Dry ice shall be placed around the secondary packaging or alternatively in an overpack with one or more complete packaging. The external container must bear: the name, phone number and address of the recipient; the person to contact in case of problems; the inscriptions "biological substance category B" in letters at least 6 mm high, "handle with care"; the diamond-shaped mark (50 mm by 50 mm) reporting inside the UN3373; transport indications; eventually, presence of cryogenic gas.

Sample handling

Equipment and Supplies needed

- Microbiological safety cabinet
- Refrigerated centrifuge with sealed buckets or sealed rotor
- 1 ml and 200 uL pipettor
- Sterile 2 mL eppendorf
- NMR tubes

Preparation of solvents

- Na_2HPO_4 (sodium dihydrogen phosphate) Mw = 141.96
- NaN3 (sodium azide) Mw = 65.01
- d6-DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid) Mw = 224
- H_2O (HPLC-grade or MilliQ water)
- D_2O (deuterium oxide)
- Concentrated KOH (around 20 M)
- 2-chloropyrimidine-5-carboxylic acid (2CLPYR5CA) Mw =158.54

Solution 1: Buffer for samples (10 mL 0.07 M phosphate buffer)*

- 0.07 M Na_2HPO_4/D_2O, pH 7.4 (with KOD/D_2O)
- 0.038 M NaN_3
- 0.02 M d6-DSS (4,4-dimethyl-4-silapentane-1-sulfonic acid)
- 0.02 M 2-chloropyrimidine-5-carboxylic acid (2CLPYR5CA)

first step: dissolve 0.0993g Na2HPO4 + 0.022g DSS +0.0247g NaN3 + 0.0216g 2C5CA in 10ml D_2O
second step: mix solution very well (ultra sonic)
third step: adjust pH to 7.4 with KOH-tablets or strong KOD-solution (KOH-tablets in D_2O)
The buffer dissolves completely at pH >5.
Add 3⁄5 of this buffer to the serum sample, mix very well (400 µL of serum + 240 µL of buffer)
* The inclusion of 2-chloro pyrimidine-5-carboxylic acid provides a down field signal that greatly improves phasing result during automatic spectral processing. Poorly phased spectra give rise to quantification errors due to erroneous area calculations under each peak. It is stable at room temperature for more than 1 week and shifts at 8.773 ppm at pH 7.40, which does not overlap with any biological metabolites we have been able to identify in serum.

Sample preparation

The sample preparation procedure is described using relative measures in order to make it more general. Protocols that can be considered, depending on the laboratory routine used to build their current database and on the software adopted for assignments, are the following:

- Using the serum sample as it is, without any preliminary processing, then mixing with the buffer system at pH=7.4 containing d6-DSS and 2CLPYR5CA (described above) and applying the CPMG pulse sequence
- Filtrate the serum sample through 3kDa molecular weight cut off, then mixing with the buffer system at pH=7.0 containing d6-DSS (0.5mM) and 2CLPYR5CA (0.5mM) and applying the 1D-NOESY pulse sequence
- Elute the serum sample through SPE plate (Ostro 96-well plates, Waters) and dry, then reconstitute with D2O and the buffer system at pH=7.0 containing TSP and applying the NOESY pulse sequence
- Precipitate the serum sample with methanol (2:1 solvent/sample) and dry, then reconstitute with D_2O and the buffer system at pH=7.0 containing TSP and applying the NOESY pulse sequence
- Adjust these volumes to the NMR probe used

Protocol a)

- Frozen samples are defrosted at room temperature for 30 min. After thawing the sample preparation procedure has to be completed within 2 hours. The samples have to be gently shaken before use
- Centrifuge 500 µL of serum at 14000 RCF (Relative centrifugal force) for 5 min, in order to eliminate possible precipitates
- Add 400 µL of the supernatant to 240 µL of sodium phosphate buffer. Mix by inversion or gentle vortex stirring
- Put 500 µL of the mixture into a 5 mm (or 4.25 mm, depending on the probe bore) disposable NMR tube.
- Store at 4°C until ready for analysis

Protocol b)
- Pre-rinse the 3 kDa filter (washed 7X with sterile double-distilled water)
- Filter sample through 3 kDa filters at 4 °C using a centrifuge (12,000 g force)
- Adjust pH to 7.0 with 20 mM K_2HPO_4
- Add d6-DSS for chemical shift referencing
- For optimal automated phase correction add 2-chloropyrimidine-5-carboxylic acid
- Check the pH of each sample prior to spectrum acquisition to ensure the pH is about 7.0

Protocol c)
- Put thawed samples into Ostro 96-well plates followed by the forceful addition of 1% formic acid in acetonitrile (3:1 solvent/sample) for in-well protein precipitation
- Hold the plate at 4° C for 10 min for further protein precipitation
- Insert the collection plate at the bottom of the Ostro plate and place the entire plate above the Positive-Pressure Manifold (Waters)
- Elute in the collection plate and dry the solvent completely
- Reconstitute with D2O containing the buffer system
- Put the mixture into a NMR tube
- Store at 4°C until ready for analysis
- Protocol d)
- Mix thawed samples with methanol in 1:2 ratio (v/v), vortex, and incubate at −20 °C for 20 min.
- Centrifuge at 13 400 rcf for 30 min to pellet proteins
- Decant supernatants to fresh vials and dry
- Reconstitute dried samples with D2O containing the buffer system.
- Put the mixture into a NMR tube
- Store at 4°C until ready for analysis

NMR Analysis and Data storage

Sample handling
- Before placing new samples, remove any other samples from previous run, in order to avoid mistakes.
- Place the NMR in the autosampler.
- Carefully take note of the position in which the tubes are placed

Remember to always randomize samples in order to not introduce spurious effects: if the analysis contemplate comparison between control and treated groups, be sure to not acquire all samples from one group first and then (for example after several days and/or with a new stock of buffer) all the samples from the other groups. In a such situations, it is hard to understand whether or not some differences are due to the metabolic state or to different operative conditions.

Preacquisition procedures

Temperature
- Calibrate temperature using e.g., a standard methanol sample (once a week)
- Perform experiments at constant temperature, 300 K. Protocol a) and b) use 298K

Equilibrate

Load the sample into the probe and allow the sample to equilibrate in the spectrometer (5 min in tube probes)

Tune and match
- Automatically tune and match every biofluid sample
- Control tuning and matching
- If tune and match is not sufficiently good, tune and match a representative sample (QC sample or first biofluid sample) manually

Shimming
- Run the shimming routine (for Bruker ‘topshim gui’)
- Within the user interface: TUNE Z only, then TUNE X only, then TUNE Y only
- Pulse length

- Determine 90° pulse length for a representative sample
- Receiver gain ([rg])
- Receiver gain is not adjusted for each sample but it is set to a constant value (in our case 36).
- RF carrier frequency offset for water ([O1])
- Determine the offset of the water signal for a representative sample (e.g., QC sample or the first biofluid sample). This parameter is used for the whole data set. Use the standard routine of the laboratory to optimize the O1. A fine tuning of o1 can be also done using an iterative optimization procedure described below:
- run a simple zg with 1 scan on the first sample
- transform the FID (fp)
- phase all signal right from the residual water peak, do not care about water phase
- increase or decrease the O1 depending on the shape of the water signal: if the water signal has a negative dip on the left side that means that O1 is too far left, i.e. too big. Similarly, if the negative dip is on the right hand side O1 is too small. Change accordingly the O1 value and repeat the above procedure until the O1 is optimized

Transfer

Transfer all settings to the experiments to be acquired. Automation setup (only for Bruker)

- Start ICON-NMR by using the command [iconnmr]. Choose [automation]
- Prepare the first data set
- Give a name to the data set corresponding to the sample's name
- Chose the experiments and the solvent parameters
- Set a title reporting the project, the date, the temperature and the nature of sample. Use the title field to add any significant information, such as the sample code
- Use the "copy" command to create the data sets for all the other samples to be acquired
- Double check everything and start the run

Do not forsake your samples during the acquisition, especially in case of long run: from time to time come back to the instrument and check if everything is working fine.

NMR acquisition

Protocol a)

Acquire all NMR spectra using the following parameters:

- The Carr-Purcell-Meiboom-Gill (CPMG) spin-echo sequence [-RD-90x°-(t-180y°-t)l4-ACQ]
- Create a new data set with edt
- Read the parameter set rpa
- Set the correct solvent eda|solvent (H2O + D2O)
- eda|Pulprog = cpmgpr1d.comp (presaturation CPMG version)
- eda|TD = 32k (Number of time domain data point)
- eda|SW = 12 ppm (Spectral width)
- eda|NS = 8 (Number of scans per block)
- eda|TD0 = 32 (Number blocks)
- eda|DS = 16 (Dummy scans)
- eda|AQ = 2.28 s (Acquisition time ACQ)
- eda|D "array" [D1] = 4.0 s (Relaxation delay - RD)
- eda|D "array" [D20] = 30 µs (spin echo delay t)
- eda|L "array" [L4] = 128 (loop counter)
- Total acquisition time ~ 28 min 59 s for 256 scans

NMR processing

Processing of 600 MHz 1H NMR spectra

- Reference the spectra to d6-DSS at δ = 0.00 ppm
- edp|LB = 1 Hz (Line broadening)
- edp|SI = 64k (double the number of Fourier domain points (TD))
- A target line width of not more than 1.5 Hz for the 2CLPYR5CA peak at half height is desirable (with LB=0).
- Automatically phase and baseline correction the spectra using the command [apk0.noe]. If the phasing is not sufficiently good, phase the spectra manually.

NMR data storage
Before storing acquired data, do a visual inspection of the spectra in order to detect any macroscopic problem in the dataset.

- FID (Free Induction Decay) must be stored as it is.
- Raw NMR data should be stored in ASCII format prior to normalization (vertical alignment) and alignment (horizontal alignment) and solvent removal.

The spectral region containing the water peak must be discarded (e.g. 5.1-4.7 ppm).
Before any statistical analysis, spectra must be normalized in order to compensate for different dilution factors. The most used scaling techniques are total area scaling, which scales all spectra by the total sum of the intensities, assumes that the total concentration of metabolites in a sample is more or less constant, or probabilistic quotient normalization (PQN), which calculates the most probable normalization coefficients.

Troubleshooting

Check carefully the quality of your spectra. Large peaks width or difficulties in shimming can be due to the presence of air micro-bubbles in the samples. In this case a good solution is to vigorously spin the tube with an hand centrifuge immediately before the analysis. Imidazole proton peaks of histidine are good indicators of the pH of serum samples. If too large misalignments of these peaks are found, some error in the buffer preparation is likely. If the shift between samples due to pH differences is a problem, the application of a bucketing step with size 0.02 to 0.04 ppm can be a solution. If some spectra show extremely higher peaks than the average, it is better to exclude that peaks from the normalization algorithm. Quality Check parameters @ LB=0 Hz Signal/Noise ratio: Signal at 0.00 ppm (2CLPYR5CA) / Noise: region between -0.5 and -0.8 ppm SI/NO > 1000 Line-width at half height for 2CLPYR5CA < 1.5 Hz (including 0 LB and fitting to Lorentzian).

REFERENCES ANNEX 1

Bernini, P., Bertini, I., Luchinat, C., Nincheri, P., Staderini, S., & Turano, P. (2011) "Standard operating procedures for pre-analytical handling of blood and urine for metabolomic studies and biobanks." Journal of biomolecular NMR, 49(3-4), 231-243.

ANNEX 2 – LIST OF ABBREVIATIONS

2-DE = 2-Dimensional Electrophoresis
24HDR = 24-Hour Dietary Recall
2HG = 2-Hydroxy Glutarate
AaA = Aromatic Amino acids
AaR = Branched Chain Amino acids
AB = Anti-Biotics
AC = Anthocyanins
ADC = Analogue to Digital Converter
AFF = Affluent
AML = Acute Myeloid Leukemia
ANH = Amidic NH2
ANN-MLP = Artificial Neural Network with MultiLayer Perceptrons
ANOVA = Analysis Of Variance
APCI = Atmospheric Pressure Chemical Ionization
ASAP = Atmospheric Solids Analysis Probe
ASCA = ANOVA Simultaneous Component Analysis
BD = Bio-Dynamic
BEF = Bioactive Enriched Food
BG = Beta-Glucans
BMRB = Biological Magnetic Resonance Bank
BSE = Bovine Spongiform Encephalopathy
BWG = Body Weight Gain
C = Controls
CE = Capillary Electrophoresis
COSY = Correlated Spectroscopy
COW = Correlation Optimized Warping
CPMG = Carr-Purcell Meiboom-Gill
CSF = Cerebro-Spinal Fluid
D = Duroc
DART = Direct Analysis in Real Time
DESI = Desorption ElectroSpray Ionization
DFD = Dark-Firm-Dry
DHA = DocosaHexaenoic Acid
DSM = DeSinewed Meat
DTW = Dynamic Time Warping
(i)ECVA = (interval) Extended Canonical Variate Analysis
ELISA = Enzyme-Linked Immuno-Sorbent Assay
ERETIC = Electronic Reference To access In vivo Concentrations
FA = Factor Analysis
FFQ = Food Frequency Questionnaire
FFT = Fast Fourier Transform
FI-ESI-IT-MS = Flow Injection ElectroSpray Ionization Ion Trap Mass Spectrometry
FI-ESI-TOF-MS = Flow Injection ElectroSpray Ionization Time-Of-Flight Mass Spectrometry
FID = Free Induction Decay
FT = Fourier Transform
GABA = Gamma Amino Butyric Acid
GC = Gas Chromatography
GI = GastroIntestinal
GLM = General Regression Analysis
GM = Genetically Modified
GMO = Genetically Modified Organism
HDM = Hand Deboned Meat
HILIC = Hydrophilic Interaction Chromatography
HMDB = Human Metabolome Database

Annex 2: *List of Abbreviations*

HPLC = High-Performance Liquid Chromatography
HR-MAS = High Resolution Magic Angle Spinning
ICP-MS = Induced Couple Plasma Mass Spectrometry
IDH = Isocitrate DeHydrogenase
L = Landrace
L/M = Lactulose/Mannitol
LBW = Litter Body Weight
LDA = Linear Discriminant Analysis
LSD = Least Significant Difference
LV = Latent Variable
LW = Large White
MALDI = Matrix-Assisted Laser Desorption/Ionization
MRM = Mechanically Recovered Meat
MS = Mass Spectrometry
MVA = MultiVariate Analysis
NAD+/NADH = Nicotinamide Adenine Dinucleotide/ + Hydrogen
NEC = Necrotizing Enterocolitis
NMR = Nuclear Magnetic Resonance
NOE = Nuclear Overhauser Effect
OPLS = Orthogonal Partial Least Squares regression
OPLSDA = Orthogonal Partial Least Squares Discriminant Analysis
ORG = Organic
PC = Principal Component
PCA = Principal Component Analysis
PCR = Polymerase Chain Reaction
PH0 = zero-order Phase Correction
PH1 = first-order Phase Correction
PLA = Placebo
(i)PLS = (interval) Partial Least Squares regression
(i)PLS-DA = (interval) Partial Least Squares Discriminant Analysis
PNN = Probabilistic Neural Network
PQN = Probabilistic Quotient Normalization
PSE = Pale-Soft-Exudative
PUFA = Poly Unsaturated Fatty Acids
QUANTAS = Quantification by Artifical Signal
RF = Radio Frequency
RH = Relative Humidity
RMSECV = Root Mean Square Error of Cross Validation
ROP = at Risk Of Poverty
RP = Reverse Phase
SDS-PAGE = Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis
SE = Substantial Equivalence
SIL = Silcora
SMDB = Serum Metabolome Database
SPI = Soy Protein Isolate
SV = Sous-Vide (Vacuum)
TA = Titrable Acidity
TD-NMR = Time Domain Nuclear Magnetic Resonance
TFQ = Total Food Quality
THS = Thompson Seedless
TMA = Tri-Methyl Amine
TMAO = Tri-Methyl Amine N-Oxide
TOF-MS = Time-Of-Flight Mass Spectrometry
TR = Traditional
TSP/TMSP = 3-(Trimethylsilyl)propionic-2,2,3,3-d4 acid sodium salt
TTS = Total Soluble Solids
UDP = Uridine DiPhosphate

Annex 2: *List of Abbreviations*

UMDB = Urine Metabolome Database
UMP = Uridine MonoPhosphate
UPLC = Ultra high-Pressure Liquid Chromatography
VIP = Variable Importance in Projection
WET = Water suppression Enhanced through T1 effects
WHC = Water Holding Capacity
WT = Wild Type
YAN = Yeast Assimilable Nitrogen