




**University of Cagliari**

**Ph. D. IN CHEMICAL SCIENCES AND TECHNOLOGIES**

**XXVIII cycle**

# Application of NMR-based metabolomics techniques to biological systems: a case study on bivalves

**DISCIPLINARY SECTOR CHIM/01**

**Presented by:** **Violetta Aru**

**Ph.D. Coordinator:** **Prof. Mariano Casu**

**Tutor:** **Dr. Flaminia Cesare Marincola**

**2014-2015**

# Acknowledgements

The present Ph.D. thesis is written based on three-years research at the University of Cagliari (Italy). It is submitted as a requirement for obtaining the Ph.D. degree with the further mention of Doctor Europaeus. The whole research has been financed by Sardinia Regional Government (P.O.R. Sardegna F.S.E. Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2007-2013 - Axis IV Human Resources, Objective l.3, Line of Activity l.3.1.).

First of all, I would like to thank my supervisor, Dr. Flaminia Cesare Marincola who gave me the opportunity to perform this Ph.D. She has introduced me to the research world providing me encouragement and wise advices throughout my time as her student. Thank you for your trust during these three years.

I also would like to express my sincere gratitude to Prof. Søren Balling Engelsen and Prof. Francesco Savorani which have been my supervisors during the eight months I spent in the Spectroscopy & Chemometrics section of the Department of Food Science (University of Copenhagen). Thanks for your guidance and for have been such a knowledgeable and nice people to work with.

I wish to thank all of my coauthors who, with their expertise, have helped me to sharpen the ideas presented in our papers. The motivating discussions with you gave an important contribution to the success of my work.

I also take the opportunity to thank the colleagues and friends I met in the University of Copenhagen. Thank you for your kind support, enthusiasm and friendship. You are the best colleagues a Ph.D. student could wish for.

Last but not least, I want to thank all my family for their encouragement, patience and understanding. In particular, I dedicate this thesis to the memory of my grandparents Ninetta Dore and Albano Pavanetto, whose role in my life was, and remains, immense. I have saved the last words of acknowledgment for one person who, with endless love, supported me since I was born. Grazie mamma!

Cagliari, March 31, 2016

Viola

# Abstract

In the past years, metabolomics has progressed greatly, providing a reliable and high-throughput approach particularly feasible for the study of complex biological systems. Indeed, thanks to the development of powerful analytical methods capable of screening a large number of chemical compounds in a sample, nowadays metabolomics plays an important role in gaining biological insights toward the influence of internal (genetic and developmental) and external (environmental) factors on phenotypes.

The work presented in this Ph.D. thesis shows examples of applications of NMR-based metabolomics to the study of bivalves, facing challenges of interest in both food and environmental sciences. The outcome of such studies yield insights, at molecular level, into several aspects concerning the impact of different storage conditions on shellfish quality and the effect of natural and anthropogenic environmental stressors on bivalves' metabolic profiles.

Firstly, the effect of different cold storage conditions on the hydrosoluble chemical components of *Mytilus galloprovincialis* (Lamarck, 1819) was investigated for the first time by NMR-based metabolomics. The data revealed substantial time-related changes in the metabolic profiles of mussels stored at 0 °C and 4 °C. The observed biochemical modifications were in good agreement with the microbiological quality of samples, reflecting changes in their microbial loads. These results confirmed the potential use of metabolomics as a reliable method to assess seafood freshness.

Secondly, a metabolomic approach was also applied to study the effect of short-term exposure to heavy metals on two different clams' species: *Ruditapes decussatus* (Linnaeus, 1758) and *Ruditapes philippinarum* (Adams & Reeve, 1850). Heavy metals are considered to be among the most harmful pollutants that can contaminate marine environments. The toxicity of trace metals gives rise from their persistent nature which lead to environmental accumulation. Given the high accumulation rates associated with their filter feeding attitude, bivalves mollusks are considered to be feasible monitoring organisms and are widely used in biomonitoring programs. *R. decussatus* and *R. philippinarum*, two bivalves' species widely distributed along the Italian coasts, were selected in this Ph.D. project for assessing lead and zinc effects on their metabolic

profiles. The results evidenced a sensible short-time metabolic response upon metal exposure, pointing out a main variability in the content of amino acids and organic osmolytes in relation to both metal nature and bivalve species. These findings show that NMR-based metabolomics has the required sensitivity and specificity to gain insights into the biochemical consequences arising upon heavy metals exposure, providing thus a useful tool for the identification of putative biomarkers as fast and sensitive indicators of contaminant-induced stress.

Finally, $^{1}$H NMR-based metabolomics was applied with the aim to assess the effects of seasonal change on *Ruditapes decussatus* metabolic profile. The results of the present study demonstrated that the combined use of advanced multivariate statistical techniques with NMR spectroscopy is a feasible approach to discriminate specimens of *R. decussatus* according to the sampling season. Moreover, the sensitivity of this analytical tool allowed the individuation of those metabolites whose relative amount significantly varied according to seasonal change (alanine and glycine), paving the way for further investigations that would contribute to achieve additional insights on bivalves' bio-ecological framework.

# List of publications

## ***Paper I***

"Changes in the $^{1}$H NMR metabolic profile of mussels (*Mytilus galloprovincialis*) with storage at 0°C". **V. Aru**, M. B. Pisano, P. Scano, S. Cosentino, and F. Cesare Marincola (2015). In Capozzi, F., Laghi, L., & Belton, P. S (Eds), Magnetic Resonance in food science (pp. 181-198). RSC publishing.

## ***Paper II***

"Metabolomics analysis of shucked mussels' freshness". **V. Aru**, M. B. Pisano, F. Savorani, S. B. Engelsen, S. Cosentino and F. Cesare Marincola (2016). Food Chemistry, 205, 58–66.

Supplementary material has been submitted as Data in Brief: "Data on the changes of the mussels' metabolic profile under different cold storage conditions" by **V. Aru**, M. B. Pisano, F. Savorani, S. B. Engelsen, S. Cosentino and F. Cesare Marincola.

## ***Paper III***

"Metabolic responses of clams, *Ruditapes decussatus* and *Ruditapes philippinarum,* to short-term exposure to lead and zinc". **V. Aru**, G. Sarais, F. Savorani, S. B. Engelsen and F. Cesare Marincola. Marine Pollution Bulletin, Submitted.

## ***Paper IV***

"Seasonal variations in the metabolic profile of *Ruditapes decussatus* from Santa Gilla lagoon". **V. Aru**, S. B. Engelsen, F. Savorani, J. Culurgioni, G. Sarais, G. Atzori, S. Cabiddu and F. Cesare Marincola. Manuscript Draft.

# Table of contents

# Preface

"Omics" sciences are considered valuable and complementary analytical tools giving relevant contributions to the study of biological systems. Metabolomics, in particular, is a powerful approach which provides an integrated overview on the functional status of a biological system in relation to environmental stimuli or genetic perturbations [1].

Initially mostly applied in physiology and disease diagnosis [2], currently metabolomics has been successfully applied also in several others fields such as environmental and food sciences [3], [4]. "Environmental metabolomics" is a scientific discipline aimed at analysing the interactions of organisms with their environment [5]. The application of the metabolomic approach in environmental sciences is particularly useful to understand, at molecular level, the toxic effect of environmental stressors, especially when the mode of action is unknown [6]. "Food metabolomics" is a high-throughput technique actually applied in several fields of food science. It aims at analysing the molecular features of food related to aspects such as quality, authenticity, diet effects, taste, and aroma characteristics [7].

High-field proton ($^1$H) Nuclear Magnetic Resonance (NMR) spectroscopy is an extremely versatile technique that covers a remarkable role in metabolomics. This analytical platform offers different advantages with respect to traditional techniques, including the ease of sample preparation and the possibility, with a single experiment, to gain chemical information on a wide range of compounds present in a sample [8]. If coupled with chemometric analysis, it becomes a suitable tool in the analysis of complex biological samples, making NMR-based metabolomics a powerful approach for assessing biochemical responses arising from typical/atypical physiological conditions.

In the work presented in this Ph.D. thesis, a "NMR-based metabolomics" approach has been carried out to address issues of interest in environmental and food sciences. The objective of this study was to characterize the metabolic profile of several bivalves' species and to assess the significant of the biochemical changes observed in relation to cold storage conditions, heavy metal pollution and seasonal change.

This work has resulted in two publications (Paper I and Paper II), one paper submitted to an international journal (Paper III) and one manuscript under finalization (Paper IV).

The chapters of this thesis provide a conceive presentation of the theoretical and practical aspects behind the applied scientific methods. In *Chapter 1*, a general overview on metabolomics is given, with a focus on the utility of this approach in environmental and food science and on the workflow behind each metabolomic study. *Chapter 2* provides a brief discussion on some theoretical aspects of the multivariate statistical methods used in this thesis for the analysis and interpretation of metabolic data. In *Chapter 3* the principles of NMR spectroscopy, the analytical platform used in this Ph.D. work, and its utility in metabolomics are presented. *Chapter 4* introduces a general overview on some relevant bioecological aspects related to bivalve mollusks with a focus on bivalve omics studies. *Chapter 5* presents a summary of the results of this Ph.D. project and, finally, *Chapter 6* includes conclusions and future perspectives.

# References


[1] U. Roessner and J. Bowne, "What is metabolomics all about?," *BioTechniques*, vol. 46, no. 5, pp. 363–365, 2009.

[2] M. Bahadori, F. Mohammadi, "Metabolomics in medicine," vol. 4, pp. 13–22, 2005.

[3] J. G. Bundy, M. P. Davey, and M. R. Viant, "Environmental metabolomics: a critical review and future perspectives," *Metabolomics*, vol. 5, no. 1, pp. 3–21, 2009.

[4] A. Trimigno, F. C. Marincola, N. Dellarosa, G. Picone, and L. Laghi, "Definition of food quality by NMR-based foodomics," *Current Opinion in Food Science*, vol. 4, pp. 99–104, 2015.

[5] B. P. Lankadurai, E. G. Nagato, and M. J. Simpson, "Environmental Metabolomics: an emerging approach to study organisms responses to environmental stressors.," *Environmental Review*, vol. 21, no. July, pp. 180–205, 2013.

[6] R. S. Viant, M.; Rosenblum, E. S.; Tjeerdema, "NMR-based metabolomics: a powerful approach for characterizing the effects of environmental stressors on organism health.," *Environ. Sci. Technol.*, vol. 37, no. 21, pp. 4982–4989, 2003.

[7] J. M. Cevallos-Cevallos, J. I. Reyes-De-Corcuera, E. Etxeberria, M. D. Danyluk, and G. E. Rodrick, "Metabolomic analysis in food science: a review," *Trends in Food Science & Technology*, vol. 20, no. 11–12, pp. 557–566, 2009.

[8] H. Winning, F. H. Larsen, R. Bro, and S. B. Engelsen, "Quantitative analysis of NMR spectra with chemometrics.," *Journal of magnetic resonance (San Diego, Calif. : 1997)*, vol. 190, no. 1, pp. 26–32, 2008.

# 1 Metabolomics

## 1.1 General remarks

The omics sciences are readily increasing disciplines aimed at the study of biological systems [1]. They include, among the others, genomics, transcriptomics, proteomics, and metabolomics. Whereas genomics, transcriptomics, and proteomics are based on the analysis of the genome, gene expression and proteins, respectively, metabolomics is deemed as the end point of the "omics cascade" [2].

In 2002 metabolomics was defined by Fiehn as "the qualitative and quantitative study of the metabolome[1] in a biological system" [3]. It is focused on the study of overall dynamic set of the endogenous and exogenous low molecular weight metabolites (<1.5 KDa) which includes various organic species such as amino acids, fatty acids, nucleic acids, carbohydrates, organic acids, vitamins, polyphenols and lipids (metabolome). Being the downstream product of gene expression, as well as the ultimate response to disease or environmental influences, the metabolome can be regarded as a reliable snapshot of the molecular phenotype of an organism. Therefore, by the study of the global metabolic profile of a biological sample in relation to genetic variations or external stimuli, metabolomics aims at providing comprehensive insights into the metabolic state of an organism.

Basically, metabolomic studies can be divided in *targeted* and *untargeted* analyses [4]. In untargeted approaches, significant metabolites are, by definition, unknown prior to analysis, while in targeted analysis, the physico-chemical characteristics of the metabolites are known and an exhaustive separation of them from the matrix is usually required for quantification.

A further differentiation of metabolomic analyses can be done based on the scientific application: *metabolic profiling*, *metabolic fingerprinting*, *metabolic footprinting*, and *metabolomics* [5]. *Metabolic profiling* is the quantitative analysis of a group of pre-

[1] The term "metabolome" was first used in 1998 by Olivier et al. [46] to describe the set of low molecular metabolites synthesized by an organism.

defined metabolites, like members of a particular pathway. *Metabolic fingerprinting* and *metabolic footprinting* are metabolomic studies focused on the classification of samples by analysing their intracellular metabolites (endometabolome) and extracellular metabolites (exometabolome), respectively. Lastly, *metabolomics* can be defined as the complete analysis of the entire cellular metabolome, in which all the metabolites are quantified and identified. While *target analysis*, *metabolic profiling*, and *metabolomics* are all quantitative approaches that require unique identification of all metabolites, *metabolic fingerprinting* and *metabolic footprinting* are semi-quantitative approaches and even unknown metabolites can be used to get deeper insights into samples metabolic profiles (untargeted).

Finally, based on the specific objective of the analysis and data manipulation, metabolomic studies can be also distinguished in *discriminative*, *informative*, and *predictive*. Discriminative analysis aims at finding significant differences among sample populations, informative metabolomics concerns the development and continuous update of metabolite databases (i. e. the Human Metabolome DataBase-HMBD), while predictive metabolomics is focused on the creation of statistical models to predict a variable that is difficult to quantify by other means [6].

## 1.2 Environmental and food metabolomics

Allowing the measurement of hundreds or even a thousand metabolites, metabolomics is nowadays considered to be a powerful approach in several scientific areas where it provides an integrated overview on the functional status of a biological system [2]. It covers a wide range of applications, from phytochemistry [7] to clinical medicine [8]. A large number of metabolomics studies has been performed in the medical field, where metabolomics has an immense potential as an early diagnostic tool [9], [10], as well as in pharmacology and toxicology to understand the mechanisms of drug action and for drug discovery [11]-[14]. Over the past few years, recent breakthroughs in metabolomic technologies have led to an increased number of applications of this analytical platform also in other fields among which environmental and food sciences (Fig. 1.1).

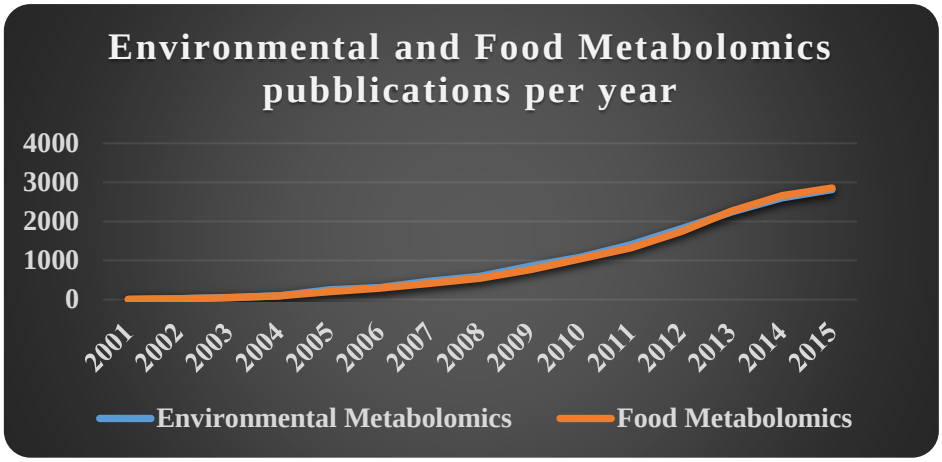



**Figure 1.1.** Environmental and food metabolomics publications per year (2001-2015). Web of Science Scopus was used to search for metabolomics publications using the terms "environmental" "metabolomics" in "all fields" and "food" "metabolomics" in "all fields". All document files were considered in the search criteria.

"Environmental metabolomics" is a steadily-developing research field among metabolomics sub-disciplines. It is viewed as the application of metabolomics techniques for the characterization of the organism responses to one or more environmental stressors such as temperature variations, pollution, or biotic interactions [15] - [17]. Such investigations can be performed either on free-living organisms and on those reared under laboratory conditions to mimic scenarios encountered in the natural environment [18]. The studies performed in this field demonstrate the potential of this technique to detect and elucidate the impact of the environmental stressors in a variety of aquatic and terrestrial organisms [19]. Due to their capacity to accumulate organic and inorganic harmful compounds from the surrounding environment, a wide attention has been focused on the study of invertebrates [20] such as earthworms and bivalve mollusks. Earthworms, in particular, have been used as sentinel organisms to assess soil health [21], [22], while marine bivalves have been used mainly for the assessment of heavy metal pollution [23]-[26].

Undeniably, one of the main applications of metabolomics, in the recent years, has covered the field of food sciences where discriminative and predictive metabolomics is used for addressing multiple and systemic issues related to safety, quality, processing, and nutritional properties of food [27]. The possibility to chemically characterize foodstuff has led to the description of the molecular features related to taste, texture,

aroma, or colour of food and beverages. Thus, this 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. A variety of food matrices has been object of investigation. Virgin olive oil [28], meat [29], milk [30], diary products [31], fish [32], and alcoholic beverages [33] are just an example of those analysed by metabolomics.

Considerable information on the chemistry and biological properties of food constituents has been collected and stored in electronic databases so that they can be automatically updated and retrieved. Among these, the Food DataBase (FooDB) is the world's largest and most comprehensive resource of both macronutrients and micronutrients [34].

## 1.3 Metabolomics workflow

Independently of the field of application, metabolomic studies have to pass through numerous steps in order to achieve profitable and reliable results. These aspects are summarized in Figure 1.2.

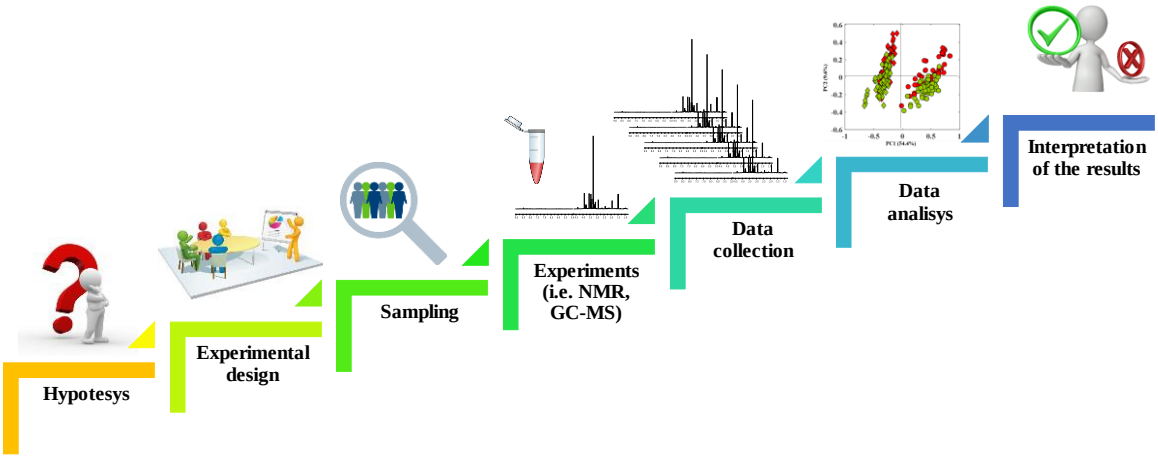



**Figure 1.2.** Typical workflow of a metabolomic study: hypothesis, experimental design, sampling, analytical platforms, data collection, multivariate data analysis and interpretation of the results represent the basic steps in metabolomics.

In the following subsection, an overview on the practical aspects concerning metabolomic analysis is given, with a focus on sample preparation, analytical platforms, and multivariate statistical analysis of data.

### 1.3.1 Sample preparation

In metabolomic experiments, sample selection and preparation is a crucial point. A correct sampling provides a real snapshot of the metabolome at a certain point in time, hence the necessity to adopt procedures fostering an unbiased sampling.

Strategies of sampling and sample preparation vary according to the experimental setup. Different strategies for the metabolites sampling can be performed. Extracellular metabolites present in human or animal biofluids are sampled using either non-invasive (urine) or invasive (serum, plasma, cerebrospinal fluid) methods. The process of sampling can change the metabolome composition [35].

Samples storage is another crucial point in metabolomic analysis as the continued freezing/defrosting of samples could damage their molecular stability. The inhibition of the enzymatic activity, aimed to preserve sample biochemical composition, is normally achieved through the freeze clamping or freezing in liquid nitrogen followed by storage at -80°C [35].

Concerning metabolite extraction, the adopted procedures dictate the nature and levels of the extracted metabolites. For non-targeted approaches, the objective is to extract the maximum number of metabolites from many chemical classes in a quantitative and non-biased manner with minimal losses of metabolites. For metabolic profiling, extraction is generally performed by the disruption of cell walls and subsequent distribution of metabolites into polar (methanol, water) and non-polar (chloroform, hexane, ethyl acetate) solvents followed by the removal of the cellular residue.

The preparation of samples for the analysis is also dependent on the metabolomics strategy employed. Targeted analyses requires separation of the metabolome into chemical classes. For metabolic profiling and fingerprinting analyses, samples are mainly analysed directly without further separation of metabolites into subclasses.

In spite of the great variability related to the sampling procedure, it is generally recognized that the impact of unpredictable biological variability is much higher than that related to the analytical one.

### 1.3.2 Analytical platforms

Metabolites separation and identification is made possible thanks to several advance analytical techniques. Nuclear Magnetic Resonance (NMR) spectroscopy, Mass Spectrometry (MS), and chromatographic methods (such as Gas Chromatography (GC) and High Performance Liquid Chromatography (HPLC)) are the most commonly used platforms [35]. All techniques have advantages and drawbacks and there is not existing analytical technique completely suitable for metabolomic studies. In particular, NMR and MS have been demonstrated to be complementary and powerful analytical approaches for the complete characterization of the metabolome [36].

By virtue of its numerous advantages, NMR spectroscopy is largely used in metabolomic studies for the analysis of bulk metabolites. NMR is not hindered by difficulties arising from the detection of different compound classes displaying dissimilar chemical properties, thus offering a remarkable analytical advantage over other spectroscopic techniques. Furthermore, due to the non-selectivity of the method, no prior knowledge of the samples is required. In spite of the high number of advantages, NMR spectroscopy has its main drawback in sensitivity, with limits of detection on the order of 10 μM or a few nmol at high fields using new cryoprobes [36]. Differently from NMR spectroscopy, the high sensitivity of MS detection makes it an important method for metabolites quantification even at extremely low concentration (typically pg level) [36]. MS is often coupled with chromatographic techniques, as Liquid Chromatography-Mass Spectrometry (LC-MS) and Gas Chromatography-Mass Spectrometry (GC-MS). LC-MS is particularly suitable for the analysis of a wide range of non-polar compounds, while GC-MS is applied for the analysis of volatile compounds. Nevertheless, in GC-MS the detection and quantification of many compounds requires a pre-analysis derivatization and a pre-selection of the "expected" metabolites. This represents an obvious limitation of the methodology, as many non derivatized chemical classes will be lost to the analysis.

### 1.3.3 Data analysis

#### 1.3.3.1 Multivariate statistical analysis (MVA)

Another key element in the metabolomic workflow is the data analysis. The complexity of the metabolic profiles, due to the presence of signals arising from hundreds or even thousands of metabolites, and the inherent variability in each sample makes the use of efficient data mining techniques a pre-requisite for maximizing the recovery of relevant metabolic signatures or biomarkers. This is achieved by using several mathematical and statistical tools aimed to maximize the information that can be extracted from complex chemical data consisting of many hard-to-identify or unknown components. Those methods are mainly based on multivariate statistical approaches which have been extensively described in the literature [37]. Depending on the multivariate method, the statistics utilized in metabolomics can be divided into non-supervised and supervised techniques. Non-supervised approaches such as Principal Component Analysis (PCA) [38] are generally applied to explore the overall statistical variance with the goal of clustering the samples, according to the different metabolic features, and detecting outliers. Supervised statistical techniques, such as Partial Least Square Discriminant Analysis (PLS-DA) [39] and methods combining further data filtering, are often needed in metabolomics to detect class-related biomarkers. Further details on the utility of MVA in metabolomic studies are given in Chapter 2.

#### 1.3.3.2 Correlation analysis in metabolomics

Besides MVA, the data obtained from metabolomic experiments can be analysed in terms of metabolic correlations, built upon the interdependencies between metabolites. Correlation analysis is widely applied in the study of systems biology and has been demonstrated to be a valuable tool also in the field of metabolomics where it is commonly applied to explore the relationships between the concentration of different metabolites. Indeed, it has been demonstrated that the analysis of the metabolic maps offers a way to obtain additional information about the physiological state of an organism [40]. The observed correlations can be considered thus as a global fingerprint

of the biological system. Based on the pair-wise correlation between their concentrations in a sample, metabolites can be also integrated into metabolic correlation networks usually characterized by a remarkable degree of complexity [41].

The measure of associations between different metabolites is usually referred as Pearson's correlation (linear correlation between variables). Nevertheless, other types of correlation (i.e. the non-linear Spearman correlation) can be performed on metabolomic datasets [42]. Independently from the type of correlation, the concomitant alteration in the metabolite concentrations presupposes a variability source. In this context, different scenarios can be outlined: specific perturbations (the change in the metabolite levels results from a specific intervention), global perturbations (i.e. changes in the metabolite levels in response to environmental stressors) and intrinsic variability (the change in the metabolite levels arises from the intrinsic variability of cellular metabolism) [41].

Although extensively applied in the field of metabolomics for the assessment of metabolite-metabolite associations, the link between metabolites and biological functions or biochemical pathways is still poorly understood [43]. In particular, the understanding of unknown metabolic pathways, based on the observed correlations, actually represents a challenge and thus the discrimination between different metabolic phenotypes is often limited to relatively pragmatic approaches that involve no recourse to the actual "biological meaning" of the data [40].

The metabolite-metabolite correlation analysis is actually applied in different scientific areas where it provides an additional and complementary information on the pathophysiological status of a biological system [41]. In particular, several metabolomics studies, based on the assessment of the degree of metabolite-metabolite associations, have been performed in the plant biology field. Among these, Ribeiro et al. have applied a metabolite-metabolite correlation analysis to achieve deeper insights into the physiological and biochemical responses of *Ricinus communis* seedlings to different temperatures [44], while Hu et al. have used non-targeted metabolomics to study the metabolic variation between different rice cultivars, highlighting the presence of correlations between the metabolic phenotype and geographic origin of the rice seeds [45].

# References


[1] D. P. Berry, K. G. Meade, M. P. Mullen, S. Butler, M. G. Diskin, D. Morris, and C. J. Creevey, "The integration of 'omic' disciplines and systems biology in cattle breeding," *Animal*, vol. 5, no. 04, pp. 493–505, 2011.

[2] B. D. Dettmer, K.; Hammock, "Metabolomics — A New Exciting Field within the ' omics ' Sciences," *Environ. Health Perspect.*, vol. 112, no. 7, pp. A396–397, 2004.

[3] O. Fiehn, "Metabolomics – the link between genotypes and phenotypes," *Plant Molecular Biology* , vol. 48, pp. 155–171, 2002.

[4] M. A. Fernández-Peralbo and M. D. Luque de Castro, "Preparation of urine samples prior to targeted or untargeted metabolomics mass-spectrometry analysis," *TrAC Trends in Analytical Chemistry*, vol. 41, pp. 75–85, 2012.

[5] M. Oldiges, S. Lütz, S. Pflug, K. Schroer, N. Stein, and C. Wiendahl, "Metabolomics: Current state and evolving methodologies and tools," *Applied Microbiology and Biotechnology*, vol. 76, no. 3, pp. 495–511, 2007.

[6] D. S. Wishart, "Metabolomics: applications to food science and nutrition research," *Trends Food Sci Technol*, vol. 19, pp. 482–493, 2008.

[7] H. Kyong, S. Khan, E. G. Wilson, S. D. Prat, A. Meissner, S. Goraler, A. M. Deelder, Y. Hae, and R. Verpoorte, "Phytochemistry Metabolic classification of South American Ilex species by NMR-based metabolomics," *Phytochemistry*, vol. 71, no. 7, pp. 773–784, 2010.

[8] J. L. Spratlin, N. J. Serkova, and S. G. Eckhardt, "Clinical Applications of Metabolomics in Oncology: A Review," *Clinical Cancer Research*, vol. 15, no. 2, pp. 431–440, 2009.

[9] A. H. M. Emwas, R. M. Salek, J. L. Griffin, and J. Merzaban, "NMR-based metabolomics in human disease diagnosis: Applications, limitations, and recommendations," *Metabolomics*, vol. 9, no. 5, pp. 1048–1072, 2013.

[10] X. Wang, A. Zhang, and H. Sun, "Power of metabolomics in diagnosis and biomarker discovery of hepatocellular carcinoma," *Hepatology*, vol. 57, no. 5, pp. 2072–2077, 2013.

[11] D. B. Kell and R. Goodacre, "Metabolomics and systems pharmacology: why and how to model the human metabolic network for drug discovery," *Drug Discovery Today*, vol. 19, no. 2, pp. 171–182, 2014.

[12] R. Powers, "NMR metabolomics and drug discovery.," *Magnetic resonance in chemistry : MRC*, vol. 47 Suppl 1, no. June, pp. S2–S11, 2009.

[13] E. J. Davidov, J. M. Holland, E. W. Marple, and S. Naylor, "Advancing drug discovery through systems biology," *Drug Discovery Today*, vol. 8, no. 4, pp. 175–183, 2003.

[14] D. B. Kell, "Systems biology, metabolic modelling and metabolomics in drug discovery and development," *Drug Discovery Today*, vol. 11, no. 23–24, pp. 1085–1092, 2006.

[15] A. Kullgren, F. Jutfelt, R. Fontanillas, K. Sundell, L. Samuelsson, K. Wiklander, P. Kling, W. Koppe, D. G. J. Larsson, B. T. Björnsson, and E. Jönsson, "The impact of temperature on the metabolome and endocrine metabolic signals in Atlantic salmon (Salmo salar)," *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, vol. 164, no. 1, pp. 44–53, 2013.

[16] N. J. C. Bailey, M. Oven, E. Holmes, J. K. Nicholson, and M. H. Zenk, "Metabolomic analysis of the consequences of cadmium exposure in Silene cucubalus cell cultures via 1H NMR spectroscopy and chemometrics," *Phytochemistry*, vol. 62, pp. 851–858, 2003.

[17] G. Marti, M. Erb, J. Boccard, G. Glauser, G. R. Doyen, N. Villard, C. a M. Robert, T. C. J. Turlings, S. Rudaz, and J. L. Wolfender, "Metabolomics reveals

herbivore-induced metabolites of resistance and susceptibility in maize leaves and roots," *Plant, Cell and Environment*, vol. 36, no. 3, pp. 621–639, 2013.

[18] N. Morrison, D. Bearden, J. G. Bundy, T. Collette, F. Currie, M. P. Davey, N. S. Haigh, D. Hancock, O. A. H. Jones, S. Rochfort, S.-A. Sansone, D. Štys, Q. Teng, D. Field, and M. R. Viant, "Standard reporting requirements for biological samples in metabolomics experiments: environmental context," *Metabolomics*, vol. 3, no. 3, pp. 203–210, 2007.

[19] B. P. Lankadurai, E. G. Nagato, and M. J. Simpson, "Environmental Metabolomics: an emerging approach to study organisms responses to environmental stressors.," *Environmental Review*, vol. 21, no. July, pp. 180–205, 2013.

[20] M. R. Viant, "Recent developments in environmental metabolomics," *Molecular BioSystems*, vol. 4, no. 10, p. 980, 2008.

[21] M. J. Simpson and J. R. McKelvie, "Environmental metabolomics: New insights into earthworm ecotoxicity and contaminant bioavailability in soil," *Analytical and Bioanalytical Chemistry*, vol. 394, pp. 137–149, 2009.

[22] S. a E. Brown, J. R. McKelvie, A. J. Simpson, and M. J. Simpson, "1H NMR metabolomics of earthworm exposure to sub-lethal concentrations of phenanthrene in soil," *Environmental Pollution*, vol. 158, no. 6, pp. 2117–2123, 2010.

[23] L. Zhang, X. Liu, L. You, D. Zhou, H. Wu, L. Li, J. Zhao, J. Feng, and J. Yu, "Metabolic responses in gills of Manila clam Ruditapes philippinarum exposed to copper using NMR-based metabolomics," *Marine Environmental Research*, vol. 72, no. 1–2. Elsevier Ltd, pp. 33–39, 2011.

[24] C. Ji, Q. Wang, H. Wu, Q. Tan, and W.-X. Wang, "A metabolomic investigation of the effects of metal pollution in oysters Crassostrea hongkongensis.," *Marine pollution bulletin*, vol. 90, no. 1–2. Elsevier Ltd, pp. 317–22, 2015.

[25] H. Wu, C. Ji, Q. Wang, X. Liu, J. Zhao, and J. Feng, "Manila clam Venerupis philippinarum as a biomonitor to metal pollution," *Chinese Journal of Oceanology and Limnology*, vol. 31, no. 1. pp. 65–74, 2013.

[26] C. Ji, H. Xu, Q. Wang, J. Zhao, and H. Wu, "Comparative investigations on the biological effects of As (III) and As (V) in clam Ruditapes philippinarum using multiple biomarkers," *Fish & Shellfish Immunology*, vol. 47, no. 1. Elsevier Ltd, pp. 79–84, 2015.

[27] A. Trimigno, F. C. Marincola, N. Dellarosa, G. Picone, and L. Laghi, "Definition of food quality by NMR-based foodomics," *Current Opinion in Food Science*, vol. 4, pp. 99–104, 2015.

[28] C. Fauhl, F. Reniero, and C. Guillou, "1H NMR as a tool for the analysis of mixtures of virgin olive oil with oils of different botanical origin," *Magnetic Resonance In Chemistry*, vol. 38, no. 6, pp. 436–443, 2000.

[29] D. Castejón, J. M. García-Segura, R. Escudero, A. Herrera, and M. I. Cambero, "Metabolomics of meat exudate: Its potential to evaluate beef meat conservation and aging," *Analytica Chimica Acta*, vol. 901, 2015.

[30] U. K. Sundekilde, P. D. Frederiksen, M. R. Clausen, L. B. Larsen, and H. C. Bertram, "Relationship between the metabolite profile and technological properties of bovine milk from two dairy breeds elucidated by NMR-based metabolomics," *Journal of Agricultural and Food Chemistry*, vol. 59, pp. 7360–7367, 2011.

[31] C. Piras, F. C. Marincola, F. Savorani, S. B. Engelsen, S. Cosentino, S. Viale, and M. B. Pisano, "A NMR metabolomics study of the ripening process of the Fiore Sardo cheese produced with autochthonous adjunct cultures," *Food Chemistry*, vol. 141, no. 3, pp. 2137–2147, 2013.

[32] G. Picone, S. B. Engelsen, F. Savorani, S. Testi, A. Badiani, and F. Capozzi, "Metabolomics as a powerful tool for molecular quality assessment of the fish Sparus aurata," *Nutrients*, vol. 3, no. 2, pp. 212–227, 2011.

[33] H. S. Son, M. K. Ki, F. Van Den Berg, G. S. Hwang, W. M. Park, C. H. Lee, and Y. S. Hong, "1H nuclear magnetic resonance-based metabolomic characterization of wines by grape varieties and production areas," *Journal of Agricultural and Food Chemistry*, vol. 56, no. 17, pp. 8007–8016, 2008.

[34] http://foodb.ca/

[35] W. B. Dunn and D. I. Ellis, "Metabolomics: Current analytical platforms and methodologies," *TrAC Trends in Analytical Chemistry*, vol. 24, no. 4, pp. 285–294, 2005.

[36] Z. Pan and D. Raftery, "Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics," *Analytical and Bioanalytical Chemistry*, vol. 387, no. 2, pp. 525–527, 2007.

[37] B. Worley and R. Powers, "Multivariate Analysis in Metabolomics," *Current Metabolomics*, vol. 1, no. 1. pp. 92–107, 2012.

[38] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, pp. 37–52, 1987.

[39] M. Barker and W. Rayens, "Partial least squares for discrimination," *Journal of Chemometrics*, vol. 17, no. 3, pp. 166–173, 2003.

[40] R. Steuer, J. Kurths, O. Fiehn, and W. Weckwerth, "Observing and interpreting correlations in metabolomic networks," vol. 19, no. 8, pp. 1019–1026, 2003.

[41] R. Steuer, "Review: On the analysis and interpretation of correlations in metabolomic data," *Briefings in Bioinformatics*, vol. 7, no. 2, pp. 151–158, 2006.

[42] D. Camacho, A. De Fuente, and P. Mendes, "The origin of correlations in metabolomics data," vol. 1, no. 1, pp. 53–63, 2005.

[43] E. M. Marcotte, "The path not taken", *Nature Biotechnology* vol. 19, pp. 626–627, 2001.

[44] P. R. Ribeiro, L. G. Fernandez, R. D. De Castro, and W. Ligterink, "Physiological and biochemical responses of Ricinus communis seedlings to different temperatures : a metabolomics approach," *Plant Biology* vol. 14, pp. 1–14, 2014.

[45] C. Hu, J. Shi, S. Quan, B. Cui, S. Kleessen, Z. Nikoloski, T. Tohge, D. Alexander, L. Guo, H. Lin, J. Wang, X. Cui, J. Rao, Q. Luo, X. Zhao, A. R. Fernie, and D. Zhang, "Metabolic variation between japonica and indica rice cultivars as revealed by non-targeted metabolomics" *Scientific reports* vol. 4 : 5067, 2014.

[46] S. G. Oliver, M. K. Winson, D. B. Kell, and F. Baganz, "Systematic functional analysis of the yeast genome," vol. 8, no. 1997, pp. 373–378, 1998.

# 2 MVA in metabolomics

The analytical techniques that are usually used in metabolomics allow the measurement of hundreds or even thousands of metabolites in a single experiment, generating complex data tables. Extracting relevant chemical and biological information from such a huge data set is far from straightforward without appropriate tools. Chemometrics can be defined as "the application of statistical and mathematical methods to chemical data to permit maximal collection and extraction of useful information" [1]. As a MultiVariate Analysis (MVA) discipline, it is extensively used in spectroscopic applications to extract relevant chemical information from spectral data that would otherwise be hidden through classical univariate methods. Chemometric analysis of data involves various steps such as data mining, feature extraction, feature selection, data modelling and reduction of data dimensionality.

Basically, three types of analysis can be performed in chemometric studies:

1. *Exploratory analysis* which provides an unbiased and general overview of the hidden data structure (trends, patterns, or clusters);
2. *Classification analysis and discriminant analysis*, aimed to identify predetermined sample classes;
3. *Regression analysis and prediction models* which models the quantitative relationship between two block of data.

The choice of the type of analysis is largely determined by the scientific purpose.

## 2.1 Principal Component Analysis (PCA)

The analytical methods used in metabolomic analysis often provide huge data sets, containing from several hundreds to thousands of variables. Sample projection in such a huge multidimensional space is not feasible.

*Principal Components Analysis* (PCA) is an exploratory and unsupervised chemometric method widely applied in metabolomics that allows the extraction and the visualization of the systematic variation in a data matrix $X$ consisting in $i$ rows

(samples) and $k$ columns (variables) [2]. It allows the reduction of the dimensionality of a data set consisting of a large number of interrelated variables, while retaining the variation present in the data set as much as possible [3]. This is achieved by transforming the original variables into a set of new variables, the principal components (PCs), which are uncorrelated and sorted in descending order of explained variance: the first principal component (PC1) represents the direction which describes most of the systematic variation present in the original variables; the second PC (PC2), which is orthogonal to PC1, retains the next highest degree of variance in the data set and so on. The procedure continues till all the systematic variation in the $X$ matrix is explained. Commonly, most of the relevant information is contained in the first few PCs. The dimensionality of the dataset is now reduced from $k$ to the number of the calculated PCs. When two PCs have been calculated, they constitute an orthogonal two dimensional space.

The decomposition of the original $X$ matrix by PCA results in two matrices known as *scores* and *loadings*. The scores ($t$) correspond to the coordinates of the projection of the samples onto each PC and those that have similar scores will cluster together. The loadings ($p$) describe the way in which the original variables are linearly combined to new variables and define the direction of the principal components.

The decomposition of the $X$ matrix by PCA can be written as:

$$X = TP^T + E \tag{2.1}$$

where $X$ is the data matrix, $T$ and $P$ are the scores and loadings matrices, respectively, and $E$ represents the residuals matrix. The typical structure of a PCA model is reported in Figure 2.1.


**Figure 2.1.** Typical structure of a PCA model. $X$: data matrix; $T$: scores matrix; $P$: loadings matrices; E: residuals matrix.

The graphical visualization of the scores (*scores plot*) can be very informative as it reveals the inherent clustering, trends and outliers in a data matrix. The graphical visualization of the loadings (*loadings plot*) describes the influence of the measured variables in the model plane, and the relation among them. The direction of the measured variables corresponds to the observed patterns in the scores plot. The graphical output of the scores and loadings matrices can be visualized separately, as scores plot and loadings plot, or in the same graph as biplots. The biplot is a two-dimension data visualization method that overlays samples (scores) and variables (loadings) of a data matrix, highlighting the relations existing within samples, within variables and between samples and variables. It provides a useful tool of data analysis and leads to capture the most relevant features in the multivariate data set (clustering and correlations among variables) [3]. A typical biplot is reported in Figure 2.2.

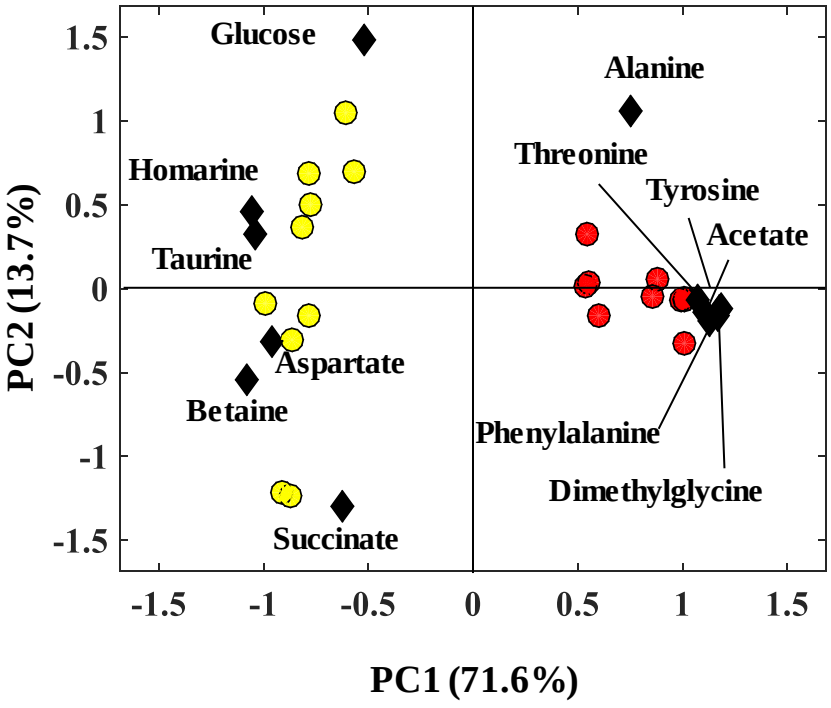



**Figure 2.2.** Representative biplot of the first two components of a PCA model. Red and yellow circles (scores) stand for different sample classes while black diamonds represent the variables (loadings) responsible for the observed clustering.

## 2.2 Partial least squares for discrimination

*Partial Least Squares (PLS) regression* is a supervised method that can be used when prior information about samples is available such as class membership and quantitative information [4]. It is a multivariate technique for assessing the relationship between a descriptor matrix $X$ (i.e. spectral intensity values) and a response matrix $Y$ (containing dependent variables). Only the $Y$-related variance in $X$ is used by the PLS method to build the models. PLS can be either used for multivariate calibration (when the responses matrix $Y$ contains quantitative information) or for class discrimination (when the responses matrix $Y$ contains qualitative information). In the latter case, the PLS method is called *PLS Discriminant Analysis* (PLS-DA). PLS-DA is one of the most well-known classification tool in chemometrics. This method is an extension of PLS1 which handles single dependent continues values, whereas PLS2 (PLS-DA) can handle multiple dependent categorical variables [5]. PLS-DA data decomposition can be written as:

$$X = TP^T + E \tag{2.2}$$

$$Y = UQ^T + F \tag{2.3}$$

where $X$ and $Y$ are the predictors and the responses matrices, respectively, $T$ and $U$ are the matrices of the projections of X (the X scores) and Y (the Y scores), respectively, $P$ and $Q$ are loading matrices, respectively, and matrices $E$ and $F$ represent the residuals. The decompositions of $X$ and $Y$ are made in such a way to maximize the covariance between the independent variable $X$ and the corresponding dependent variable $Y$ of highly multidimensional data [6]. Although, in PLS discriminant analysis, the utilization of class membership allows a better separation between classes, the variation not directly correlated with $Y$ is still present in the scores, complicating the interpretation of the model [6]. *Orthogonal PLS-DA* incorporates an Orthogonal Signal Correction (OSC) filter [7] into the PLS model, separating thus $Y$-predictive variation from the $Y$-orthogonal (uncorrelated) variation:

$$X = T_P P_P^T + T_0 P_0^T + E \tag{2.4}$$

$$Y = U_P Q_P^T + F \tag{2.5}$$

where $T_P$ and $P_P^T$ represents the scores and loadings matrices, respectively, for the *Y*-predictive variation in the dataset, $T_0$ and $P_0^T$ are the scores and loadings matrices, respectively, for the Y-orthogonal variation identified by the OSC filter, and $E$ is the residual matrix. Only the Y-predictive variation is used for the modelling of Y.

The advantage of OPLS-DA compared to PLS-DA is that a single component is used as a predictor for the class, while the other components describe the variation orthogonal to the first predictive component [8].

The identification of variables influencing samples separation in an OPLS-DA model can be performed using different graphical representations for the loadings. Among these, the S-plot and S-line plot represent informative visualization tools being thus widely applied in metabolomic studies for the identification of discriminant variables. The S-plot is a scatter plot which is used to visualize both the covariance and the correlation between the X-variables and the predictive score $t_1$ (Fig.2.3).

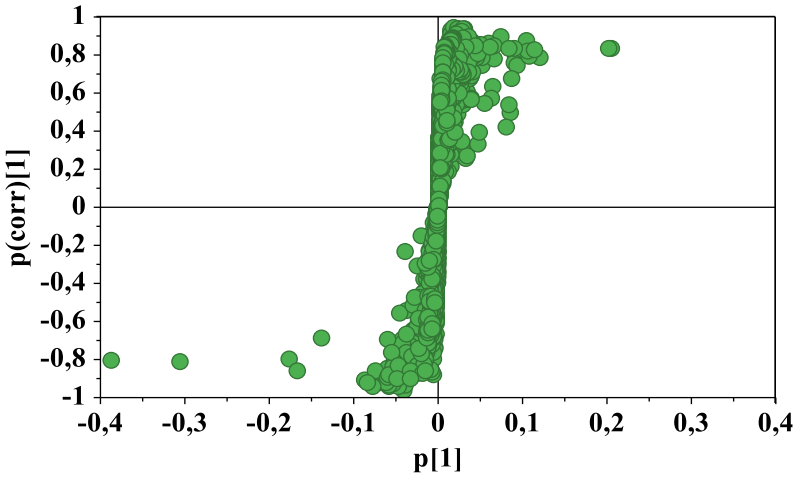



**Figure 2.3.** Representative OPLS-DA S-plot. The X axis represents the variable contribution to the clustering, while the Y axis represents variable confidence.

Each point in the plot represents a variable which is plotted in a Cartesian coordinate system: X axis represents variable contribution (covariance, p) to the observed clustering, while the Y axis represents variable confidence (correlation, p(corr)). The variable points with high correlation and covariance values are located in regions far away from the origin.

The S-line plot (Fig. 2.4) is tailor-made for spectroscopic data since it provides loadings with the same shape as that of a spectrum. This plot depicts simultaneously the outcomes of the covariance (peak height) and correlation (colour code) analyses applied to the spectral dataset. The phase of the resonance signals represents the direction in which the variables have a discriminative power.

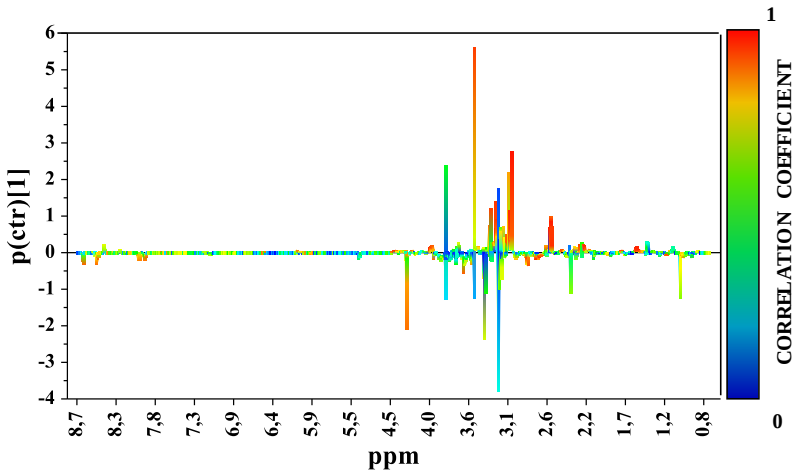



**Figure 2.4** Representative OPLS-DA S-line plot for a $^{1}$H-NMR dataset.

## 2.3 iECVA

Extended Canonical Variates Analysis (ECVA) [9] is a recent chemometric tool for samples classification. It is based on the standard Canonical Variates Analysis (CVA), which is a well-established method for classification of full rank data [10]. While data reduction is usually needed for standard discriminative methods, ECVA was developed to handle multi collinear and low rank data, such as spectroscopic datasets, where the number of variables is much larger than the number of samples [9]. Nuclear magnetic resonance, near infrared, fluorescence, Raman and mass spectrometry are all analytical methods that often provides datasets containing more variables than samples.

The ECVA method estimates the directions in space that maximize the differences between groups in the data, yielding powerful separations. It involves several steps that finally lead to the actual class membership prediction. Basically, this is achieved by calculating covariance matrices, both for the within class variation and between

class variation and by performing a PLS model between the abovementioned covariance matrices. Finally, a linear discriminant analysis is performed on the PLS results [9].

iECVA [11] is an extension of ECVA designed to find spectral regions holding the main information responsible for the separation among groups. The iECVA method performs a series of ECV analyses on the datasets: one for the whole spectrum and one for each defined interval. Finally, the performances of each interval are compared among each other and against the overall model. The interval that has the lowest prediction error is kept in the model and the procedure is repeated by selecting new intervals that contributes positively to the model. A representative iECVA plot is reported in Figure 2.5. Several intervals show a potential better discriminative power than the full spectral area.

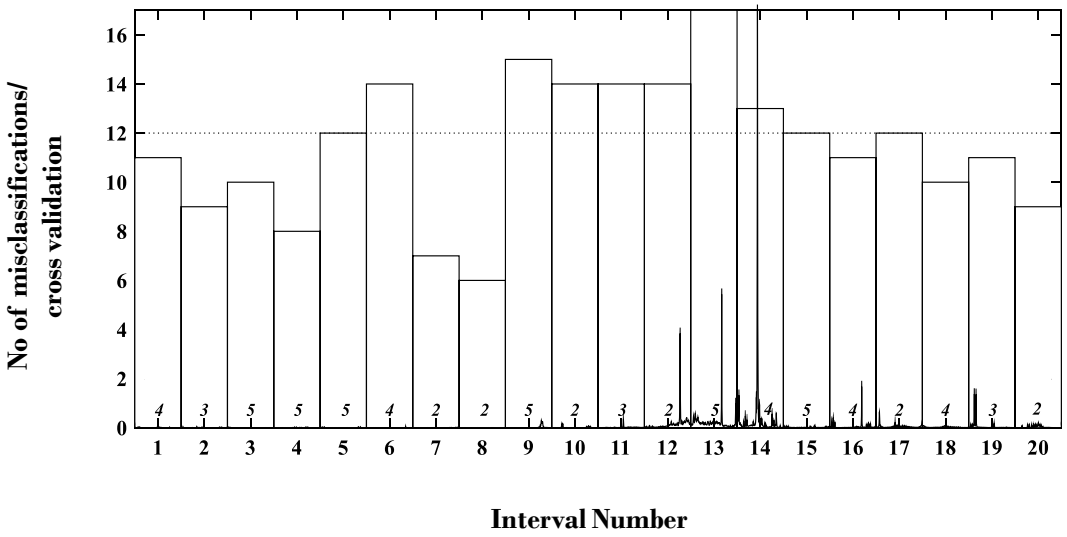



**Figure 2.5** iECVA plot describing the classification performances of the model for a $^{1}$H NMR dataset. The number of the intervals and the number of misclassifications are reported in the X and in the Y axis, respectively. Any interval presenting an error bar lower than the dotted line is potentially a better discriminant than the full spectral area. The number of components used to build each model is reported in the corresponding error bar.

## 2.4 Model performance and validation

Classifying groups of samples based on their metabolic profile is one of the main topics in metabolomics. Unfortunately, classification methods easily undergo overfitting and rigorous validation is needed. Overfitting is a common problem in multivariate statistics, particularly with spectroscopic data where the number of observation is extremely lower than the number of variables. Data overfitting provides a non-predictive model leading to good classification rates only in the training set [12]. Therefore, to provide an objective assessment of the performance and stability of a model, rigorous validation is needed. In chemometrics, the most commonly applied validation tools are:

1. *Cross-validation*
2. *Validation with a test set*
3. *Permutation test*

*Cross Validation (CV)* is used to choose the optimal model parameters as well as to test the predictability of the statistical model. It is the process of sequentially remove one or more samples, making a calibration on the remaining samples and using it to predict the removed samples [13]. It is often used for the validation of a classification model when the number of samples is dramatically low. Basically, CV is performed by dividing the dataset into several groups and then developing parallel models from the reduced data. Several systems that allow to specify the CV groups are available: venetian blinds (where the CV segments are grouped as 123-123-123), continuous subsets (where the CV segments are grouped as 111-222-333) and random subsets (where the dataset is divided in random blocks of typically 20% of data) [14]. After developing a model, the left out data are used as a test set, and differences between actual and predicted Y-values are calculated for it. Repeating the validation several times for a dataset, each time with a new randomization, will give a robust error estimate.

When an adequate number of samples is available, a rigorous way to test the predictive power of a model consists in computing predictions for an independent set of observations. The data matrix is split into two sets: a *training set* and a *validation set*

*(test set)*. The former is used for setting up a calibration model. Then, the model is tested on the validation set and the error (Root Means Square Errors of Predictions, RMSEP) is estimated by comparison with results from the reference methods. In order to perform a proper external validation, it is essential to understand the nature of the dataset so that the training set and the test set do represent the overall variation present in the X dataset [12].

Classification and prediction model performance can be evaluated also by the parameter $R^2$Y and $Q^2$Y, which represent the fraction of the variation of the Y-variable and the predicted fraction of the variation of the Y-variable, respectively. The statistical significance of $R^2$Y and $Q^2$Y values is estimated through a response permutation testing [12]. In this test, the Y-matrix is randomly re-ordered (typically 200 times), while the X-matrix is kept constant. This means that the Y-data remain numerically the same, but their positions are shifted by random shuffling. Each time a new model is fitted using X and the permuted Y matrix, providing a reference distribution of $R^2$Y and $Q^2$Y for random data. An intercept of the $Q^2$Y regression line below zero and an $R^2$Y intercept value significantly lower than the corresponding original one are indicative of a valid model [12].

# References


[1] P. K. Hopke, "The evolution of chemometrics," *Analytica Chimica Acta*, vol. 500, no. 1–2, pp. 365–377, 2003.

[2] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 2, pp. 37–52, 1987.

[3] R. Bro and A. K. Smilde, "Principal component analysis," *Analytical Methods*, vol. 6, no. 9. p. 2812, 2014.

[4] S. Wold and M. Sjostrom, "PLS-regression : a basic tool of chemometrics," pp. 109–130, 2001.

[5] M. Barker and W. Rayens, "Partial least squares for discrimination," *Journal of Chemometrics*, vol. 17, no. 3, pp. 166–173, 2003.

[6] B. Worley and R. Powers, "Multivariate Analysis in Metabolomics," *Current Metabolomics*, vol. 1, no. 1. pp. 92–107, 2012.

[7] J. Sjöblom, O. Svensson, M. Josefson, H. Kullberg, and S. Wold, "An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra," *Chemometrics and Intelligent Laboratory Systems*, vol. 44, no. 1–2, pp. 229–244, 1998.

[8] J. Trygg, E. Holmes, and T. Lundstedt, "Chemometrics in metabonomics," *Journal of Proteome Research*, vol. 6, no. 2, pp. 469–479, 2007.

[9] L. Nørgaard, R. Bro, F. Westad, S. B. Engelsen. A modification of canonical variates analysis to handle highly collinear multivariate data. *Journal of Chemometrix*, vol. 20, pp. 425–35, 2006.

[10] R. B. Darlington, S. L. Weinberg and H.t J. Walberg. Canonical Variate Analysis and Related Techniques. Review of Educational Research, vol. 43, no 4, pp. 433-454, 1973.

[11] L. Nørgaard , G. Soletormos, N. Harrit, M. Albrechtsen, O. Olsen, D. Nielsen, et al. Fluorescence spectroscopy and chemometrics for classification of breast cancer samples—a feasibility study using extended canonical variates analysis. *Journal of Chemometrix,* vol. 21, pp. 451–458, 2007.

[12] L. Eriksson, E. Johansson, N. Kettaneh-Wold, J. Trygg, C. Wikström and S. Wold. Multi- and megavariate data analysis: Principles and applications. Umetrics AB, Umeå, Sweden, 2006.

[13] J. A. Westerhuis, H. C. J. Hoefsloot, S. Smit, D. J. Vis, A. K. Smilde, E. J. J. Velzen, J. P. M. Duijnhoven, and F. A. Dorsten, "Assessment of PLSDA cross validation," Metabolomics, vol. 4, no. 1, pp. 81–89, 2008.

[14] R. Bro, K. Kjeldahl, A. K. Smilde, and H. A. L. Kiers, "Cross-validation of component models: A critical look at current methods," Analytical and Bioanalytical Chemistry, vol. 390, no. 5, pp. 1241–1251, 2008.

# 3 NMR-based metabolomics

Nuclear Magnetic Resonance (NMR) spectroscopy is a powerful analytical technique widely used in metabolomics. By virtue of its several advantages, it is remarkably suitable for the qualitative and quantitative analysis of low molecular weight metabolites in different complex biological samples.

NMR was discovered in 1946 by two independent groups of scientists: E. M. Purcell, R. V. Pound and H. C. Torrey of Harvard University and F. Bloch, W. Hansen, and M. Packard of Stanford University. The discovery first came out when it was noticed that magnetic nuclei, such as $^{1}H$, $^{13}C$ and $^{31}P$ were able to absorb radio frequency energy when placed in a magnetic field of a strength that was specific to the nucleus. Since then, NMR has been applied to solids, liquids and gasses, kinetic and structural studies, resulting in 6 Nobel prizes being awarded in the field of chemistry.

## 3.1 NMR spectroscopy

### 3.1.1 Principles

All isotopes that contain an odd number of protons and/or of neutrons have an intrinsic *magnetic moment* (μ) and *angular momentum (P)*. The two quantities are related through the expression:

$$\mu=\gamma P \tag{3.1}$$

where γ, the gyromagnetic ratio, is a constant characteristic of a particular nucleus.

According to quantum theory, angular moment and nuclear magnetic moment are quantized. The largest measurable component of the angular moment is $\hbar I$ where $I$ is the nuclear spin quantum number and ħ is the reduced Planck constant (h/2π). The spin quantum number $I$ may have integral or semi-integral values, the actual size depending on the isotope. In the absence of an external magnetic field, the individual nuclear magnetic moments are randomly oriented. When a magnetic field $B_0$ is applied, individual magnetic moments do not cancel to zero and align themselves along preferential directions (Fig. 3.1).

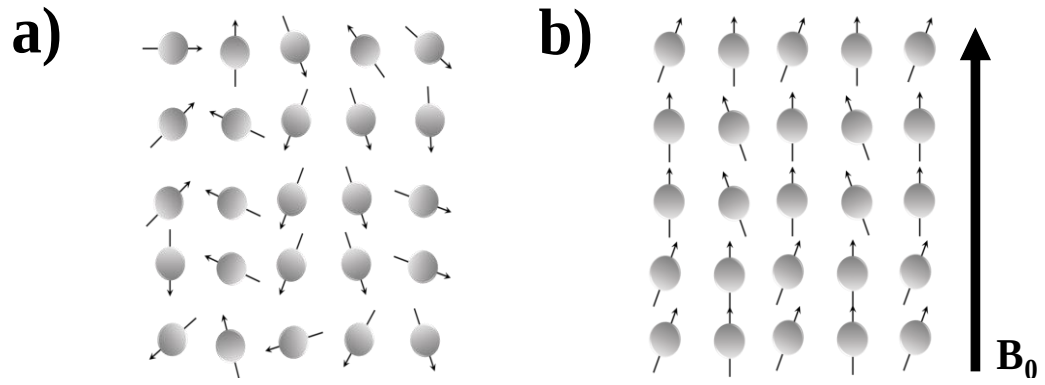



**Figure 3.1.** (a) Un-aligned spins: in the absence of a magnetic field, the net magnetization or macroscopic magnetization in a sample is zero because all the individual nuclear magnetic moments are oriented randomly. b) Aligned spins: in the presence of an applied magnetic field ($B_0$), the nuclear magnetic moment is aligned with the magnetic field. There is in this case a net nuclear magnetization.

Just as a spinning mass will precess in a gravitational field, the magnetic moment μ associated with a spinning spherical charge will precess in an external magnetic field (Fig. 3.2). The frequency of precession is called the Larmor frequency and is proportional to the strength of the applied field.

$$\nu_0 = \left(\frac{\gamma}{2\pi}\right) B_0 \qquad (3.2)$$

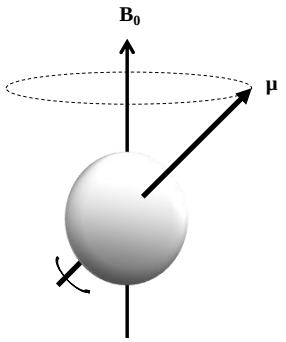



**Figure 3.2.** Precessional motion of proton nucleus in an external magnetic field $B_0$.

According to quantum mechanics, the energy level of a nucleus of nuclear spin I is thereby split in an external field $B_0$ into 2I + 1 Zeeman levels. For a nucleus with I= ½, two energy levels exist according to a parallel or antiparallel orientation of the magnetic moments with respect to the magnetic field (Fig. 3.3) [1].

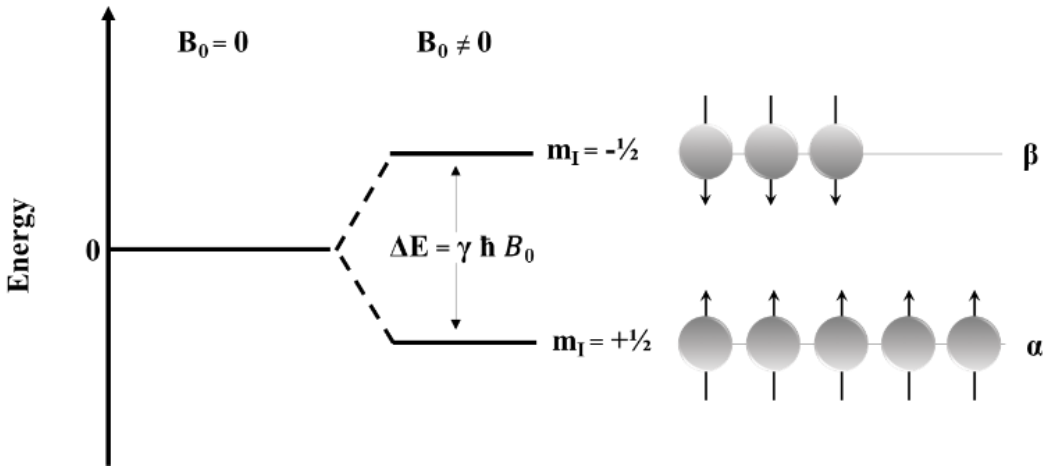



**Figure 3.3.** Nuclear energy levels of a ½ nuclear spin.

The number of nuclei adopting each orientation is not equal: the lower energy α state is more populated than β state. This gives rise to a net magnetization $M_0$, aligned with the applied magnetic field $B_0$. A representation of $M_0$ as the vector sum of all the individual nuclear magnetic moments is reported in Figure 3.4.

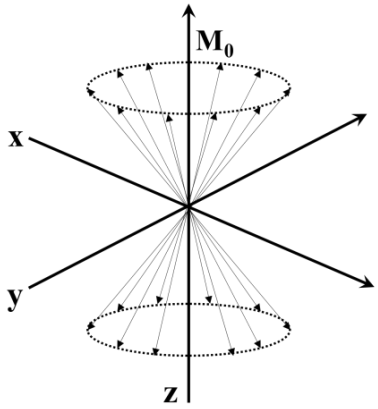



**Figure 3.4.** Net magnetization ($M_0$) originating from the sum of the individual nuclear magnetic moments.

The energy gap (ΔE) between the quantized orientations of the magnetic moments depends on the gyromagnetic ratio and the external magnetic field:

$$\Delta E = \gamma \hbar B_0 \qquad (3.3)$$

If irradiation energy having a frequency matching the Larmor frequency is introduced at a right angle to the external field, the precessing nucleus will absorb energy and nuclear spins transitions from lower to higher energy levels will be induced. The irradiation energy is in the RF range and is typically applied as a short (e.g., many microseconds) pulse. As the radiofrequency is switched off, the system will return to equilibrium. This return to equilibrium is referred to as *relaxation* and it causes the NMR signal to decay with time, producing the observed free induction decay (FID). The NMR signal is then Fourier transformed to be converted in the frequency domain.

### 3.1.2 Interpreting $^{1}$H NMR Spectra

NMR spectrum represents the actual outcome of a NMR experiment. In the case of molecular mixtures, it is constituted by a spectral line comprising a collection of several peaks arising from the molecules under investigation (Fig. 3.5). The main information provided by each spectrum is related to signals *chemical shift*, *multiplicity*, *intensity* and *linewidth*.

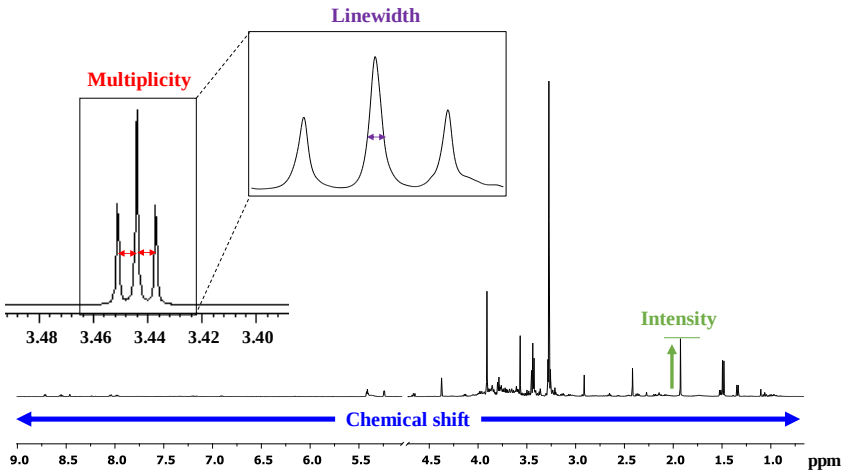



**Figure 3.5.** Representative $^{1}$H NMR spectra of hydrophilic extract. Signals multiplicity, intensities, chemical shifts and linewidth are evidenced.

The position of a peak in the NMR spectrum is strictly related to the chemical environment of the nucleus generating the signal. When an atom is placed in a magnetic field, its electrons circulate about the direction of the applied magnetic field. This circulation causes a small magnetic field at the nucleus which opposes the externally

applied field. The magnetic field at the nucleus (the effective field) is therefore generally less than the applied field by a fraction

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

where $\sigma$ is the so-called 'shielding constant' and depends on the density and distribution of the electron cloud that surrounds the nucleus. In some cases, such as the benzene molecule, the circulation of the electrons in the aromatic orbitals creates a magnetic field at the hydrogen nuclei which enhances the Bo field. This phenomenon is called deshielding. The electron density around each nucleus in a molecule varies according to the types of nuclei and bonds in the molecule, therefore the opposing field and, thus, the effective field at each nucleus vary according to their surrounding environment. This is called the *chemical shift phenomenon*. The *chemical shift* of a nucleus is the difference between the resonance frequency of the nucleus and a standard, relative to the standard:

$$\delta = \left(\frac{\nu - \nu_{ref}}{\nu_{ref}}\right) \cdot 10^6 \quad (3.5)$$

where $\nu$ is the frequency of the compound under investigation and $\nu_{ref}$ is that of a reference compound. Reference compound should be chemically inert, should have a well-resolved singlet, and its chemical shift should be independent of external variables (i.e. temperature or ionic strength). Tetramethylsilane (TMS) and 3-trimethylsililpropionic acid (TSP) are two of the most commonly used reference compounds for $^1$H NMR analysis. Chemical shift is expressed in terms of ppm and is independent of the strength of the applied magnetic field. It provides valuable information about the chemical environment of nuclei and is of great aid in structural elucidation of molecules and compounds.

Nuclei experiencing the same chemical environment (or chemical shift) are called equivalent. Those nuclei experiencing different environment (or having different chemical shifts) are nonequivalent. Nuclei which are close to one another exert an influence on each other's effective magnetic field. This effect shows up in the NMR spectrum when the nuclei are nonequivalent. If an atom under examination is perturbed or influenced by a nearby nuclear spin (or set of spins), the observed nucleus

responds to such influences, and its response is manifested in the split of its resonance signal. This spin-coupling is transmitted through the connecting bonds, and it functions in both directions. Thus, when the perturbing nucleus becomes the observed nucleus, it also exhibits signal splitting. For spin-coupling to be observed, the sets of interacting nuclei must be bonded in relatively close proximity (e.g. vicinal and geminal locations), or be oriented in certain optimal and rigid configurations. The signal splitting in proton spectra is usually small, ranging from fractions of a Hz to as much as 18 Hz, and is designated as $J$ (referred to as the *coupling constant*).

The magnitude or *intensity* of NMR resonance signals is displayed along the vertical axis of the spectrum. Peaks intensity is directly proportional to the number of nuclei giving rise to that resonance and is proportional to the molar concentration of the sample. Thus, a small or dilute sample will give a weak signal, and doubling or tripling sample concentration will increase signal intensity proportionally. If a reference compound is added to the solution under investigation in a known concentration, the concentrations of all compounds present in that sample can be precisely computed by referring the measured integrals to the integral of the reference compound. This property is very useful in metabolomic studies, where quantitative NMR (qNMR) analysis is used to determine the exact concentration of one or more chemical species in solution.

The linewidth of an NMR signal contains information on the rate of processes, including the rate of molecular motions. Usually, measured at half height of the peak, it can be influenced by several factors like: relaxation time, relaxation mechanism, chemical exchange, intramolecular rotations, temperature, presence of paramagnetic impurities, homogeneity of the sample, homogeneity of the spectrometer's magnetic field, interactions with neighbouring nuclei.

### 3.1.3 Post-acquisition processing of NMR spectra

Data processing is performed on NMR spectra after the Fourier transform of the FID. Depending on the precision and reliability of the desired parameters (i.e. chemical shift, coupling constant) that are extracted from the spectra, it allows to improve the

quality of the spectrum. This section aims at providing a short summary of the most important NMR data processing tools:

1. *Phase correction*
2. *Baseline correction*
3. *Zero filling*
4. *Apodization*

*Phase correction*. Phase correction is the process of mixing the real and imaginary signals in the complex spectrum that is obtained after Fourier transformation of the free induction decay. The phase of the spectrum is typically corrected such that a properly phased spectrum is similar to an absorption spectrum. Phasing the spectrum is a routine procedure and involves zero-order and first order phase corrections. Zero-order correction is frequency independent and is the same for all lines across the spectrum. Whereas the first-order correction is frequency dependent and applies a phase change, whose amount increases linearly with the distance to the reference signal.

*Baseline correction*. Baseline distortion can be a problem in NMR since spectra with non-flat baselines give wrong integrals, while in spectra with roll baseline small signals may not be recognized. Baseline distortion is mainly caused by the corruption of the first few data points in FID. These corrupted data points add low frequency modulations in the Fourier-transformed spectrum, and thus form the distorted baseline. Baseline correction models the baseline by fitting it to polynomial, sine or exponential functions. Horizontal baseline is a crucial point to achieve an accurate spectral signal quantification.

*Zero-filling*. In NMR spectroscopy the need for shorter acquisition times lead to the loss of digital resolution. Zero-filling is the procedure of increasing the digital resolution of the NMR spectra by increasing the FID data points. It is applied just before performing Fourier transformation, by adding zeros to the end of the FID. It improves the line shapes and the appearance of the spectrum by resolving very small couplings in multiple structures.

*Apodization* is a digital filtering manipulation of the NMR spectra which consists in multiplying the FID with different functions (such as Lorentzian or simple exponential decay, Gaussian, or Sine-bell function) which can be chosen either to enhance the sensitivity or resolution in the final resulting spectrum. For instance, exponential multiplication leads to line broadening and a reduction in noise, while other trigonometric functions produce narrowing of the spectral line while an increase in the noise.

## 3.2 Pre-processing of NMR data for chemometric analysis

In NMR-based metabolomics, NMR spectra must be transformed into a suitable format prior multivariate data analysis. The most common pre-processing techniques used to this purpose are: *binning*, *alignment*, *normalisation* and *scaling*.

One of the main complication arising from the analysis of NMR spectra derives from the strong dependence of chemical shift on temperature, pH and ionic strength. Spectral misalignment can be mitigated by dividing each spectrum into 'bins', usually sized 0.04 ppm, and integrating signals intensities within each bin to produce a smaller set of variables. The *binning* procedure can either mask chemical shift misalignments and filter noise in the spectra but it also hides potentially significant changes of low-intensity peaks nearby huge signals [2]. An alternative approach is the use of variable bin sizes, known as 'intelligent' binning. Several methods have been proposed, such as non-equidistant binning [3] and adaptive binning [4].

Spectral *alignment* is a very worthwhile procedure which retains the possibility to perform multivariate analysis with less loss of spectral information with respect to binning [5]. Spectra are initially aligned to an internal standard, typically TSP at 0.0 ppm, aiding the removal of the global shifts from the dataset. Afterwards, a fine or local alignment is needed. Various alignment algorithms have been proposed, many of which are based on a warping method. Examples of methods include, interval correlation shifting (icoshift) [6], Recursive Segment-Wise Peak Alignment (RSWPA) [7] and Correlation Optimised Warping Correlation Optimised Warping (COW) [8]. Among these, the *i*coshift algorithm, which has been applied in the present work, aligns NMR

signals in defined spectral regions by using the maximum correlation criteria. Due to the use of fast Fourier transform, it is very fast and can align all spectra in a data matrix simultaneously.

*Normalization* is a row operation that is applied to the data from each sample to make them directly comparable with each other [9]. This is achieved by the reduction of the effects of samples dilution arising from the variations in the number of cells, biofluid volume or tissue size. It can be computed either normalizing the dataset using the concentration of an internal standard as a normalization factor (normalization to a reference compound) or normalizing each spectrum in such a way that its sum is 1 (normalization to a constant sum).

Aimed to reduce the noise in the data, the *scaling* is an operation performed among the columns of the dataset allowing the enhancement of the information quality [9]. Indeed, even when all variables are expressed with identical units, metabolite levels can range over many orders of magnitude. This leads multivariate analysis to be focused on a small set of intense signals that, from a biological point of view, are not necessarily more important than those present in low concentration. Several scaling methods are used for NMR. Some of them are reported in Table 3.1.

**Table 3.1.** List of the most commonly used scaling methods in MVA

| Method | Goal | Advantage | Disadvantage |
|---|---|---|---|
| **Centering** | Focus on differences, not similarities | Removes offset from the data | Unsuitable for heteroscedastic data |
| **Unit Variance** | Compare metabolites based on correlation | All metabolites equally important | Inflation of measurements errors |
| **Range** | Compare metabolites relative to biological response range | All metabolites equally important. Biologically related scaling | Inflation of measurement errors, sensitive to outliers |
| **Pareto** | Reduces relative importance of large values; partially preserves data structure | Stay close to original measurement than UV | Sensitive to large fold changes |
| **Vast** | Focus on small fluctuations | Aims for robustness, uses prior group knowledge | Non suited for large induced variation without group structure |
| **Level** | Focus on relative response | Suited for biomarker identification | Inflation of measurement errors |

* Adapted from Worley and Powers, 2012 [9].

## 3.3 NMR spectroscopy in metabolomics

By virtue of its numerous advantages and the general ease and simplicity of the methodology, the applications of NMR spectroscopy in metabolomic studies had seen an exponential increase over the last decade (Fig. 3.6).

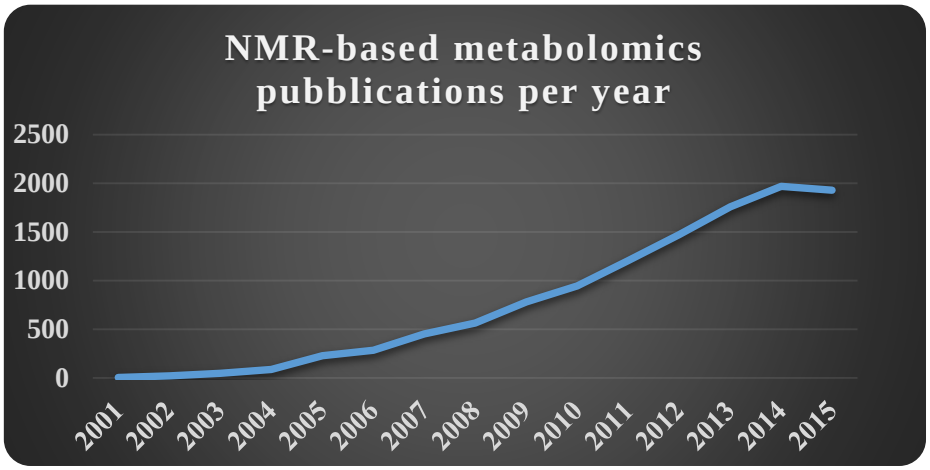



**Figure 3.6.** NMR-based metabolomics publications per year (2001-2015). Web of Science Scopus was used to search for NMR-based metabolomics publications using the terms "metabolomics" and "NMR" in "All fields". All document files were considered in the search criteria.

Beside the possibility of sample recovering, its non-destructive nature allows to perform replicate analysis on specimens and i*n vivo*/*in vitro* measurements on biological samples [10], [11]. In addition, the high reproducibility, the limited sample preparation, and the relatively short acquisition times make NMR spectroscopy a high-throughput technique capable to rapidly analyse an elevated number of samples, which represents an important requirement for subsequent multivariate statistical analysis [12]. Furthermore, the non-selective nature of this technique makes NMR a feasible analytical tool extensively applied in non-targeted analysis in which significant metabolites are, by definition, unknown.

In spite of the large number of advantages, NMR has the main drawback in its relatively low-sensitivity measurements. A list of the main advantages and drawbacks of NMR spectroscopy is given in Table 3.2.

**Table 3.2.** Advantages and disadvantages of NMR spectroscopy

| Advantages | Disadvantages |
|---|---|
| Non-destructive | Relatively low sensitivity |
| Information over a wide range of compounds | Expensive instrumentation |
| Small sample preparation required | |
| No derivatization | |
| Fast analysis ($^{1}$H NMR) | |
| Possibility to investigate different nucleus ($^{1}$H, $^{13}$C, $^{31}$P, $^{17}$O) | |
| Compatible with liquids and solids | |
| Quantitative and qualitative measurements | |

NMR-based metabolomics is nowadays playing an integral and constantly expanding role in several scientific areas form clinical medicine and pharmacology, where it has been successfully applied as advanced diagnostic tool and for drug discovery, respectively, [13], [14], to food science, for food quality/authenticity assessment [15]. NMR spectroscopy has also been widely applied in the field of environmental metabolomics where it has been demonstrated to be a viable tool in detecting metabolomics disorders arising, for example, from anthropogenic stressors [16].

Complex biological samples such as tissues and biofluids are routinely analysed in metabolomic studies. Among these, the metabolomic analysis of biofluids, such as urine, plasma, and saliva requires very small sample preparation [17]. In general, biofluids can be simply added to a deuterated aqueous buffer before transferring the solution to a sample tube to acquire a NMR spectrum. Differently, the metabolomic analysis of tissues can involve an extraction procedure. It usually requires a dual phase methanol/chloroform extraction for polar and lipophilic metabolites [18].

The development of high-resolution $^{1}$H Magic Angle Spinning (MAS) NMR spectroscopy has made feasible the acquisition of high-resolution NMR data also on small pieces of intact tissues with no pre-treatment [19]. NMR spectroscopy on a tissue sample in a MAS experiment is the same as liquid-state NMR, versatile to study metabolic changes and to perform molecular structure elucidation and molecular

dynamic studies. Furthermore, metabolic profiling successes in the characterization *in vitro* of cell systems such as microalgae and tumor cells [20], [21].

In NMR spectroscopy the quality of the samples has a profound effect on the quality of the resulting spectrum. In order to get the best possible NMR spectra in the shortest amount of time, samples should be properly prepared. Nowadays, the occurrence of large cohorts in clinical studies increases the demand of automating the whole metabolomics pipeline (sample preparation, sample injection, NMR acquisition, and data extraction and analysis). Automatic sample preparation is possible for NMR spectroscopy, and standard NMR spectra typically take only a few minutes to be acquired using robotic flow-injection methods [22].

As in the case of MS, NMR can be coupled to a variety of LC techniques. The improvements in NMR hyphenation achieved during the recent years have made NMR analysis of single components of complex mixtures achievable rapidly and without laborious preparative scale separation and purification procedures [23].

In NMR-metabolomics analysis, a crucial point is the metabolites identification. Although their inner complexity, due to the presence of crowded regions with remarkable peak overlapping, $^{1}$H NMR spectra of biological samples allow many resonances to be assigned directly based on their chemical shifts and signal multiplicities. For one-dimensional NMR experiments (1D NMR), where spectra are depicted in plots of intensity vs. frequency, extensive chemical information regarding a variety of metabolites is currently available. Several public databases, such as the Human Metabolome DataBase (HMDB) [24] as well as commercial database in the Chenomx suite (available at http://www.chenomx.com), collect reference NMR spectra of a variety of chemical compounds, allowing metabolites identification by matching the observed spectra to the reference spectra in the database. Furthermore, an accurate peak assignment can be carried out by adding the reference compounds, leading thus the direct identification of the metabolites.

Finally, two-dimensional NMR spectroscopy (2D NMR) can be used for structure elucidation of molecules. Among 2D NMR experiments, COSY (COrrelation SpectroscopY) and TOCSY (Total Correlation SpectroscopY) are commonly applied in metabolomic studies for metabolites identification [25].

# References


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

[2] R. Stoyanova, A. W. Nicholls, J. K. Nicholson, J. C. Lindon, and T. R. Brown, "Automatic alignment of individual peaks in large high-resolution spectral data sets," *Journal of Magnetic Resonance*, vol. 170, no. 2, pp. 329–335, 2004.

[3] F. Dieterle, A. Ross, G. Schlotterbeck, and H. Senn, "Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in1H NMR metabonomics," *Analytical Chemistry*, vol. 78, no. 13, pp. 4281–4290, 2006.

[4] T. De Meyer, D. Sinnaeve, B. Van Gasse, E. Tsiporkova, E. R. Rietzschel, M. L. De Buyzere, T. C. Gillebert, S. Bekaert, J. C. Martins, and W. Van Criekinge, "NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm," *Analytical Chemistry*, vol. 80, no. 10, pp. 3783–3790, 2008.

[5] S. Wu, "1D and 2D NMR Experiment Methods," *Development*, 2011.

[6] F. Savorani, G. Tomasi, and S. B. Engelsen, "icoshift: A versatile tool for the rapid alignment of 1D NMR spectra," *Journal of Magnetic Resonance*, vol. 202, no. 2, pp. 190–202, 2010.

[7] K. a. Veselkov, J. C. Lindon, T. M. D. Ebbels, D. Crockford, V. V. Volynkin, E. Holmes, D. B. Davies, and J. K. Nicholson, "Recursive segment-wise peak alignment of biological 1H NMR spectra for improved metabolic biomarker recovery," *Analytical Chemistry*, vol. 81, no. 1, pp. 56–66, 2009.

[8] N. P. V Nielsen, J. M. Carstensen, and J. Smedsgaard, "Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping," *Journal of Chromatography A*, vol. 805, no. 1–2, pp. 17–35, 1998.

[9] B. Worley and R. Powers, "Multivariate Analysis in Metabolomics," *Current Metabolomics*, vol. 1, no. 1. pp. 92–107, 2012.

[10] G. L. Chadzynski, R. Pohmann, G. Shajan, R. Kolb, S. Bisdas, U. Klose, and K. Scheffler, "In vivo proton magnetic resonance spectroscopic imaging of the healthy human brain at 9.4 T: initial experience," *Magnetic Resonance Materials in Physics, Biology and Medicine*, vol. 28, no. 3, pp. 239–249, 2014.

[11] U. Sharma, D. Upadhyay, S. Mewar, A. Mishra, P. Das, S. D. Gupta, S. N. Dwivedi, G. K. Makharia, and N. R. Jagannathan, "Metabolic abnormalities of gastrointestinal mucosa in celiac disease: An *in vitro* proton nuclear magnetic resonance spectroscopy study," *Journal of Gastroenterology and Hepatology*, vol. 30, no. 10, pp. 1492–1498, 2015.

[12] H. Winning, F. H. Larsen, R. Bro, and S. B. Engelsen, "Quantitative analysis of NMR spectra with chemometrics," *Journal of magnetic resonance (San Diego, Calif. : 1997)*, vol. 190, no. 1, pp. 26–32, 2008.

[13] A. H. M. Emwas, R. M. Salek, J. L. Griffin, and J. Merzaban, "NMR-based metabolomics in human disease diagnosis: Applications, limitations, and recommendations," Metabolomics, vol. 9, no. 5, pp. 1048–1072, 2013.

[14] R. Powers, "NMR metabolomics and drug discovery," Magnetic resonance in chemistry : MRC, vol. 47 Suppl 1, no. June, pp. S2–S11, 2009.

[15] A. Trimigno, F. C. Marincola, N. Dellarosa, G. Picone, and L. Laghi, "Definition of food quality by NMR-based foodomics," Current Opinion in Food Science, vol. 4, pp. 99–104, 2015.

[16] R. S. Viant, M.; Rosenblum, E. S.; Tjeerdema, "NMR-based metabolomics: a powerful approach for characterizing the effects of environmental stressors on organism health.," *Environ. Sci. Technol.*, vol. 37, no. 21, pp. 4982–4989, 2003.

[17] M. E. Bollard, E. G. Stanley, J. C. Lindon, J. K. Nicholson, and E. Holmes, "NMR-based metabonomic approaches for evaluating physiological influences on biofluid composition," *NMR in Biomedicine*, vol. 18, no. 3, pp. 143–162, 2005.

[18] J. Folch, M. Lees, and G. H. S. Stanley, "A simple method for the isolation and purification of total lipids from animal tissues," *J Biol Chem*, vol. 226, no. 1. pp. 497–509, 1957.

[19] L. Shintu, S. Caldarelli, and B. M. Franke, "Pre-selection of potential molecular markers for the geographic origin of dried beef by HR-MAS NMR spectroscopy," *Meat Science*, vol. 76, no. 4, pp. 700–707, 2007.

[20] M. S. Chauton, T. R. St\orseth, and G. Johnsen, "High-resolution magic angle spinning $^{1}$H NMR analysis of whole cells," *Journal of Applied Phycology*, vol. 15, pp. 533–542, 2003.

[21] B. Sitter, S. Lundgren, T. F. Bathen, J. Halgunset, H. E. Fjosne, and I. S. Gribbestad, "Comparison of HR MAS MR spectroscopic profiles of breast cancer tissue with clinical parameters," *NMR in Biomedicine*, vol. 19, no. 1, pp. 30–40, 2006.

[22] M. Spraul, "Flow injection proton nuclear magnetic resonance spectroscopy combined with pattern recognition methods: implications for rapid structural studies and high throughput biochemical screening," *Anal. Comm.*, vol. 34, no. November, pp. 339–341, 1997.

[23] C. Seger, M. Godejohann, L. H. Tseng, M. Spraul, A. Girtler, S. Sturm, and H. Stuppner, "LC-DAD-MS/SPE-NMR hyphenation. A tool for the analysis of pharmaceutically used plant extracts: Identification of isobaric iridoid glycoside regioisomers from Harpagophytum procumbens," *Analytical Chemistry*, vol. 77, no. 3, pp. 878–885, 2005.

[24] D. S. Wishart, D. Tzur, C. Knox, R. Eisner, A. C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, L. Nikolai, M. Lewis, M. A. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeroncic, P. Stothard, G. Amegbey, D. Block, D. D. Hau, J. Wagner, J. Miniaci, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang, G. E. Duggan, G. D. MacInnis, A. M. Weljie, R. Dowlatabadi, F. Bamforth, D. Clive, R. Greiner, L. Li, T. Marrie, B. D. Sykes, H. J. Vogel, and L.

Querengesser, "HMDB: The human metabolome database," Nucleic Acids Research, vol. 35, no. SUPPL. 1, pp. 521–526, 2007.

[25] W. B. Dunn and D. I. Ellis, "Metabolomics: Current analytical platforms and methodologies," *TrAC Trends in Analytical Chemistry*, vol. 24, no. 4, pp. 285–294, 2005.

# 4 Bivalves metabolomics

Invertebrates constitute the largest pool of macroscopic organisms in the marine environment with the Mollusca phylum comprising a large amount of all the known marine organisms. Among the several classes constituting the phylum, the Bivalvia class covers a prominent role in the aquaculture industry [1]. In the following subsections, a brief description of their biology and distribution is given, with a particular focus on the application of metabolomics in bivalves' research field.

## 4.1 Ecology and distribution of bivalves

Bivalve mollusks are aquatic organisms, encompassing both marine and fresh-water habitats throughout the world. The name "bivalves" refers to the two-part shell that characterizes these mollusks. Most of them adopt a sedentary or even sessile lifestyle (living under the seabed, buried in soft substrates such as sand, silt, mud, gravel, or coral fragments) which makes them relatively safe from predation. Other bivalves, such as mussels, attach themselves to hard surfaces through byssus filaments made of keratin and proteins.

Among bivalves, edible lamellibranch mollusks, such as scallops, clams, oysters, and mussels, are sessile burrowing marine organisms extensively harvested all over the world. In particular, the common European clam *Ruditapes decussatus* (Linnaeus, 1758), the Philippine clam *Ruditapes philippinarum* (Adams & Reeve, 1850) and the blue mussel *Mytilus galloprovincialis* (Lamarck, 1819) stand up for being an important fishing resource in Italy and Sardinia Island as well as in the whole Mediterranean coasts. Indeed, due to their ubiquitous distribution and easy accessibility, bivalve mollusks cover a fundamental role in the aquaculture industry where they are widely commercialised as food.

Owing to their filter feeding and sessile attitude, bivalve mollusks have been regarded as preferential species for environmental biomonitoring. Indeed, their capability of accumulating high levels of pollutants, provides an integrated overview on the degree of the environmental contamination [2], [3]. The Mussel Watch Program

(MWP) is a practical example of using bivalve mollusks as bioindicators of environmental pollution. MWP is a contaminant monitoring program, developed in the USA in 1986 by the National Oceanic and Atmospheric Administration (NOAA), aimed at monitoring the concentration of pollutants in bivalves (mussels and oysters) and sediments in lakes, coastal and estuarine waters. Toxicants to be monitored include dichlorodiphenyltrichloroethane (DDT), polycyclic aromatic hydrocarbons (PAHs), polychlorinated biphenyls (PCBs) and heavy metals.

## 4.2 Bivalves omics

Bivalve mollusks cover a fundamental role in marine ecosystems and in aquaculture industry, where they constitute important commercial resources [1].

The increasing interest of "omics" sciences toward the study of bivalves has posed new challenges for their integrative characterization [4]. So far, most of the "omics" studies have been focused towards the characterization of bivalves' genomes, transcriptomes, and proteomes. Among these applications, the development of specific methodologies in the field of genomics led to the complete sequencing of *Pinctada fucata* genome [5], while a differential gene expression in response to pollutants has been evidenced in *Perna viridis* [6]. Finally, proteome investigation is currently being used to study bivalves' responses to different environmental perturbations such acute heat stress [7]. Metabolomics applications on the study of bivalves have regarded the characterization of mussels metabolome to assess the effect of environmental contamination [8], to find sex specific metabolites [9], and to understand the mode of action of pesticides like atrazine and lindane [10]. Additionally, the metabolome of several clams species has been characterized to evaluate the impact of the exposure to heavy metals [11]–[14].

In the following subsections a general overview of bivalve metabolomics is given, with a focus on the applications in environmental and food sciences.

### 4.2.1 "Bivalves Metabolomics" in environmental science

The main topic in environmental metabolomics is the assessment of the interactions of organisms with their environment. This discipline is aimed at characterising the metabolic responses of an organism to both natural and anthropogenic stressors that can occur in the surrounding environment [15], [16]. Indeed, the effects of environmental disturbances, in terms of biochemical and physiological alterations (deemed as biomarkers of exposure), are manifested first at the molecular level in an organism [15]. Biomarkers of pollution exposure have been extensively used in the last few decades to monitor the health of organisms and hence to assess the quality of the environment.

Several metabolomic studies have been carried out as a high-throughput screening tool for monitoring the effect of environmental pollutants on bivalves' metabolic profile. Among these, Campillo et al. [14] recently have investigated the effect of environmental pollution on the clam *Ruditapes decussatus*. *Ruditapes philippinarum* metabolic response to heavy metal pollution has been characterized by several authors [11]–[13], [17], while Ji et al. [18] have assessed the effect of heavy metal exposure in the oyster *Crassostrea hongkongensis*.

As far as mussels are concerned, gender-based metabolic responses to organic pollutants have been investigated on *Mytilus galloprovincialis* species [10]. The combined effects of several physico-chemical stressors have been also investigated in several bivalves' species. Ellis et al. [19] have studied the contrasting metabolic responses by male and female mussels exposed to reduced seawater pH, increased temperature, and a pathogen, while the metabolic alterations arising from the exposure to different salinities have been investigated in both clams and mussels [20], [21].

### 4.2.2 "Bivalves Metabolomics" in food science

Metabolomics has been widely demonstrated to be a high throughput technique in several fields of food science such as food safety, food authentication, food storage and processing [22], [23]. Among these, food spoilage is one of major concerns in food industry because it leads to considerable economic losses. Furthermore, food-deterioration is a hazard to public health.

Due to several intrinsic factors, seafood products easily undergo deterioration. Among these, their poikilothermic nature, the high *post-mortem* pH in the flesh (usually > 6.0), the presence of large amounts of Non-Protein-Nitrogen (NPN) and TriMethylAmine-N-Oxide (TMAO), make them an ideal substrate for bacterial growth [24]. Seafood deterioration can be also the result of several extrinsic processes such as direct sunlight irradiation, oxygen, storage temperature, humidity or microbial contamination by bacteria, moulds or yeasts [25]. Bivalve mollusks, in particular, are among the most perishable seafood. Indeed, the shelf life of this fishery product is quite limited, primarily due to the lack of less digestive connective tissues. Moreover, the elevate water activity leads to a faster tissue degradation by microorganisms [26].

Only a few metabolomic studies have been carried on seafood to assess the biochemical modifications occurring under different storage conditions. Picone et al. [27] and Savorani et al. [28] have investigated the effect of storage conditions on *Spaurus aurata*, showing as metabolomics can represent a powerful tool for detecting alterations of the biochemical profiles in fish as a consequence of storage. Chen et al [29] have assessed the influence of the fermentation procedure on the taste of crab *Portunus trituberculatus*. Their results confirm metabolomics as a useful tool to provide important information on the crab paste quality. Finally, only two studies concerning the evaluation of the effects of storage time and temperature on mussels' metabolic profile has been published so far (Paper I and II in the present PhD thesis) [30], [31].

# References


[1] S. Pawiro, "Bivalves: Global production and trade trends," *Safe management of shellfish and harvest waters*, pp. 11–19, 2010.

[2] S. Stankovic, P. Kalaba, and A. R. Stankovic, "Biota as toxic metal indicators," *Environmental Chemistry Letters*, vol. 12, no. 1. pp. 63–84, 2014.

[3] R. Chiarelli and M. C. Roccheri, "Marine Invertebrates as Bioindicators of Heavy Metal Pollution," *Open Journal of Metal*, no. December. pp. 93–106, 2014.

[4] V. Suárez-Ulloa, J. Fernández-Tajes, C. Manfrin, M. Gerdol, P. Venier, and J. Eirín-López, "Bivalve Omics: State of the Art and Potential Applications for the Biomonitoring of Harmful Marine Compounds," *Marine Drugs*, vol. 11, no. 11, pp. 4370–4389, 2013.

[5] T. Takeuchi, T. Kawashima, R. Koyanagi, F. Gyoja, M. Tanaka, T. Ikuta, E. Shoguchi, M. Fujiwara, C. Shinzato, K. Hisata, M. Fujie, T. Usami, K. Nagai, K. Maeyama, K. Okamoto, H. Aoki, T. Ishikawa, T. Masaoka, A. Fujiwara, K. Endo, H. Endo, H. Nagasawa, S. Kinoshita, S. Asakawa, S. Watabe, and N. Satoh, "Draft genome of the pearl oyster Pinctada fucata: A platform for understanding bivalve biology," *DNA Research*, vol. 19, no. 2, pp. 117–130, 2012.

[6] P. T. Y. Leung, Y. Wang, S. S. T. Mak, W. C. Ng, and K. M. Y. Leung, "Differential proteomic responses in hepatopancreas and adductor muscles of the green-lipped mussel Perna viridis to stresses induced by cadmium and hydrogen peroxide," *Aquatic Toxicology*, vol. 105, pp. 49–61, 2011.

[7] L. Tomanek and M. J. Zuzow, "The proteomic response of the mussel congeners Mytilus galloprovincialis and M. trossulus to acute heat stress: implications for thermal tolerance limits and metabolic costs of thermal stress," *Journal of Experimental Biology*, vol. 213, no. 20, pp. 3559–3574, 2010.

[8] S. Fasulo, F. Iacono, T. Cappello, C. Corsaro, M. Maisano, A. D'Agata, A. Giannetto, E. De Domenico, V. Parrino, G. Lo Paro, and A. Mauceri, "Metabolomic investigation of Mytilus galloprovincialis (Lamarck 1819) caged in aquatic environments," *Ecotoxicology and Environmental Safety*, vol. 84. Elsevier, pp. 139–146, 2012.

[9] A. Hines, W. H. Yeung, J. Craft, M. Brown, J. Kennedy, J. Bignell, G. D. Stentiford, and M. R. Viant, "Comparison of histological, genetic, metabolomics, and lipid-based methods for sex determination in marine mussels," *Analytical Biochemistry*, vol. 369, no. 2, pp. 175–186, 2007.

[10] W. Tuffnail, G. A. Mills, P. Cary, and R. Greenwood, "An environmental 1H NMR metabolomic study of the exposure of the marine mussel Mytilus edulis to atrazine, lindane, hypoxia and starvation," *Metabolomics*, vol. 5, no. 1, pp. 33–43, 2009.

[11] H. Wu, X. Liu, J. Zhao, and J. Yu, "NMR-Based Metabolomic Investigations on the Differential Responses in Adductor Muscles from Two Pedigrees of Manila Clam Ruditapes philippinarum to Cadmium and Zinc," *Marine Drugs*, vol. 9, no. 12. pp. 1566–1579, 2011.

[12] X. L. Liu, L. B. Zhang, L. P. You, J. B. Yu, M. Cong, Q. Wang, F. Li, L. Z. Li, J. M. Zhao, C. H. Li, and H. F. Wu, "Assessment of Clam Ruditapes philippinarum as Heavy Metal Bioindicators Using NMR-Based Metabolomics," *Clean-Soil Air Water*, vol. 39, no. 8. pp. 759–766, 2011.

[13] C. Ji, H. Wu, M. Zhou, and J. Zhao, "Multiple biomarkers of biological effects induced by cadmium in clam Ruditapes philippinarum.," *Fish & shellfish immunology*, vol. 44, no. 2. Elsevier Ltd, pp. 430–5, 2015.

[14] J. a Campillo, A. Sevilla, M. Albentosa, C. Bernal, A. B. Lozano, M. Cánovas, and V. M. León, "Metabolomic responses in caged clams, Ruditapes decussatus, exposed to agricultural and urban inputs in a Mediterranean coastal lagoon (Mar Menor, SE Spain).," *The Science of the total environment*, vol. 524–525C. Elsevier B.V., pp. 136–147, 2015.

[15] B. P. Lankadurai, E. G. Nagato, and M. J. Simpson, "Environmental Metabolomics: an emerging approach to study organisms responses to environmental stressors.," *Environmental Review*, vol. 21, no. July, pp. 180–205, 2013.

[16] M. R. Viant, "Recent developments in environmental metabolomics," *Molecular BioSystems*, vol. 4, no. 10, p. 980, 2008.

[17] L. Zhang, X. Liu, L. You, D. Zhou, H. Wu, L. Li, J. Zhao, J. Feng, and J. Yu, "Metabolic responses in gills of Manila clam Ruditapes philippinarum exposed to copper using NMR-based metabolomics," *Marine Environmental Research*, vol. 72, no. 1–2. Elsevier Ltd, pp. 33–39, 2011.

[18] C. Ji, Q. Wang, H. Wu, Q. Tan, and W.-X. Wang, "A metabolomic investigation of the effects of metal pollution in oysters Crassostrea hongkongensis.," *Marine pollution bulletin*, vol. 90, no. 1–2. Elsevier Ltd, pp. 317–22, 2015.

[19] R. P. Ellis, J. I. Spicer, J. J. Byrne, U. Sommer, M. R. Viant, D. A. White, and S. Widdicombe, "1H NMR metabolomics reveals contrasting response by male and female mussels exposed to reduced seawater pH, increased temperature, and a pathogen," *Environmental Science and Technology*, vol. 48, no. 12, pp. 7044–7052, 2014.

[20] V. Carregosa, E. Figueira, A. M. Gil, S. Pereira, J. Pinto, A. M. V. M. Soares, and R. Freitas, "Tolerance of Venerupis philippinarum to salinity: Osmotic and metabolic aspects," *Comparative Biochemistry and Physiology - A Molecular and Integrative Physiology*, vol. 171, pp. 36–43, 2014.

[21] J. A. Bussell, E. A. Gidman, D. R. Causton, D. Gwynn-Jones, S. K. Malham, M. L. M. Jones, B. Reynolds, and R. Seed, "Changes in the immune response and metabolic fingerprint of the mussel, Mytilus edulis (Linnaeus) in response to lowered salinity and physical stress," *Journal of Experimental Marine Biology and Ecology*, vol. 358, no. 1, pp. 78–85, 2008.

[22] J. M. Cevallos-Cevallos, J. I. Reyes-De-Corcuera, E. Etxeberria, M. D. Danyluk, and G. E. Rodrick, "Metabolomic analysis in food science: a review," *Trends in Food Science & Technology*, vol. 20, no. 11–12, pp. 557–566, 2009.

[23] L. M. Samuelsson and D. G. J. Larsson, "Contributions from metabolomics to fish research.," *Molecular bioSystems*, vol. 4, no. 10, pp. 974–9, 2008.

[24] L. Gram and H. H. Huss, "Microbiological spoilage of fish and fish products.," *International journal of food microbiology*, vol. 33, no. 1, pp. 121–37, 1996.

[25] E. Ghaly, D. Dave, S. Budge, and M. S. Brooks, "Fish Spoilage Mechanisms and Preservation Techniques : Review Department of Process Engineering and Applied Science , Dalhousie University Halifax , Nova Scotia , Canada," *American Journal of Applied Sciences*, vol. 7, no. 7, pp. 859–877, 2010.

[26] E. Goulas, I. Chouliara, E. Nessi, M. G. Kontominas, and I. N. Savvaidis, "Microbiological, biochemical and sensory assessment of mussels (Mytilus galloprovincialis) stored under modified atmosphere packaging," *Journal of Applied Microbiology*, vol. 98, pp. 752–760, 2005.

[27] G. Picone, S. B. Engelsen, F. Savorani, S. Testi, A. Badiani, and F. Capozzi, "Metabolomics as a powerful tool for molecular quality assessment of the fish Sparus aurata," *Nutrients*, vol. 3, no. 2, pp. 212–227, 2011.

[28] F. Savorani, G. Picone, A. Badiani, P. Fagioli, F. Capozzi, and S. B. Engelsen, "Metabolic profiling and aquaculture differentiation of gilthead sea bream by 1 H NMR metabonomics," *Food Chemistry*, vol. 120, no. 3, pp. 907–914, 2010.

[29] D. Chen, Y. Ye, J. Chen, and X. Yan, "Evolution of metabolomics profile of crab paste during fermentation," *Food Chemistry*, vol. 192, pp. 886–892, 2016.

[30] V. Aru, M. B. Pisano, P. Scano, S. Cosentino, and F. C. Marincola, Changes in the $^{1}$H NMR metabolic profile of mussels (*Mytilus galloprovincialis*) with storage at 0°C. In Capozzi, F., Laghi, L., & Belton, P. S (Eds), Magnetic Resonance in food science, RSC publishing, pp. 181-198, 2015.

[31] V. Aru, M. Barbara, F. Savorani, S. Balling, S. Cosentino, and F. Cesare, "Metabolomics analysis of shucked mussels' freshness," *Food Chemistry*, vol. 205, pp. 58–65, 2016.

# 5 Summary of the results

The work presented in this Ph.D. thesis shows some examples of NMR-based metabolomics applications in environmental and food sciences for the study of complex biological samples, such as bivalves. The overall aim was to evaluate the potential of this approach for the identification of candidate biomarkers as fast and sensitive indicators of: (a) storage conditions; (b) heavy metal pollution; (c) seasonality patterns. The outcome of such studies encourages the use of NMR-based metabolomics in both environmental and food sciences.

## 5.1 Metabolomics and spoilage detection

One of the main problems of seafood marketing is the ease with which fish and shellfish undergo deterioration after death. Indeed, beside the post-mortem degradation due to autolytic processes, the high glycogen content, the lack of less digestible connective tissues and the high water activity make them an ideal substrate for bacterial growth, producing a wide variety of hydrolytic enzymes that contribute to seafood spoilage [1].

Although the role of bacteria in seafood spoilage is well recognize, data on the possible link of metabolic profiles to the microbial activity are still lacking. Understanding these relations may help the selection of the most appropriate indices for monitoring seafood freshness and quality changes during storage. Only three studies in the literature have investigated the changes in seafood composition during storage by metabolomics [2]–[4], whereas no metabolomic study, concerning the effect of storage in bivalves, has been reported.

In the present Ph.D. work, metabolomics has been combined with microbiological analysis to explore possible correlations between the biochemical alterations of mussels and the microbial loads over cold storage. Microbiological and metabolomic analysis were performed on *Mytilus galloprovincialis* stored at 0 °C over a period of 7 days (**Paper I**). Several microbial groups were quantified in mussels' samples, such as mesophilic and psychrotrophic bacteria. The results of the microbiological analysis evidenced a

significant increase ($p < 0.01$) in the total viable counts of mesophilic and psychrotrophic bacteria with a very similar slow growth rate. The identification of the post-mortem compositional changes occurring in mussels' metabolic profiles was performed by NMR-based metabolomics. In order to identify the key metabolic signatures characterizing ice stored samples, OPLS-DA models were generated in pairwise comparisons. An increased level in the relative amount of free amino acids, acetate, and TriMethylAmine (TMA) was the main outcome after 7 days of storage. Among these, TMA is the most commonly used metabolite to evaluate the freshness of fish, since its production is strictly related to the bacterial use of TriMethylAmine-*N*-Oxide (TMAO) [1]. As seafood spoils, proteins are broken down to peptides, free amino acids, amines, and volatile ammonia [5]. Consequently, the observed increase in free amino acids can be considered the result of microorganisms' development. Although glucose and lactate seem to be the substrates which are attacked first by the various groups of spoilage bacteria under aerobic and anaerobic conditions, the changes in their levels over storage at 0°C did not show statistical significance.

So as to deeper explore the utility of this approach for identifying new molecular biomarkers of spoilage, the abovementioned investigation was further extended by comparing the changes occurring in *Mytilus galloprovincialis* metabolome and microbial flora during storage at 0 °C and 4 °C for a period of 10 and 6 days, respectively (**Paper II**). Microbiological analyses concerned the quantification of different bacterial groups, such as psychrotrophs, mesophylls, *Enterobacteriaceae* and *E. coli* strains. Psychrotrophic and mesophilic bacteria increased over storage time but with differential growth rates according to storage temperature. Differently, *Enterobacteriaceae* counts did not show significant increase upon both storage conditions, while *E. coli* counts were always below the detection limit of the method.

The average distribution of the microbial species isolated in mussel samples was also determined. Psychrotrophic Gram-negative bacteria were the dominant pool of microorganisms. They were mainly represented by *Pseudomonas fluorescens*, *Aeromonas salmonicida* and *Shewanella putrefaciens* species, which are known to be the typical microbiota of many seafood, including bivalves, at refrigeration temperatures [1]. These bacteria, referred to as "Specific Spoilage Organisms" or SSOs [6], are able to produce

volatile organic compounds (VOCs) from the degradation of soluble, low molecular weight components responsible for off-flavors and taste associated with spoiled seafood. Among other species, the yeast *Rhodotorula mucilaginosa* and *Bacillus cereus* were particularly abundant in the isolates at both temperatures.

As far as NMR-metabolomics analysis is concerned, an exploratory PCA was performed on the whole NMR dataset to generate an overview of the metabolic disturbance resulting from storage time at both temperatures. The metabolic patterns at 0 °C and 4 °C were found to be similar but with faster dynamics in samples at 4 °C.

OPLS-DA models were further built at each temperature, in pairwise comparisons, where storage time class membership was defined with the discrete variable Y. The most remarkable metabolic changes at both temperatures, although with different rates, were linked to proteolysis, glycogen-lysis, glucose metabolism and the bacterial production of volatile amines which were manifested in an increase in the relative concentrations of free amino acids, organic acids (i.e. acetate, lactate, succinate), and TMA at both storage temperatures.

After identification, each discriminant metabolite was quantified from the $^{1}$H-NMR spectra and possible correlations with the abundance of specific bacterial populations (mesophilic and psychrotrophic) were investigated by calculating Pearson's correlation coefficients. Although the association between the consume or production of these metabolites to specific strains is not an easy task, the analysis of the Pearson's correlation between the above mentioned molecules and the bacterial growth suggested the presence of a worth note correlation with both mesophilic and psychrotrophic microbial counts. Thus, based on our microbiological data, some preliminary considerations can be made. TMA accumulation could derive from the action of different microorganism which have been identified in mussels' samples, such as *S. putrefaciens*-like organisms and *Vibrio* spp., which are microorganisms typical of marine environment, but also from the metabolic activity of *Aeromonas* spp. [1]. Among them, *S. putrefaciens* is considered to have a higher spoilage potential in seafood products because of its high metabolic activity even if present in low concentrations. Amino acids could arise from the proteolytic activity of *Pseudomans, Bacillus, Aeromonas* and *Vibrio*

spp [7]. It has also been demonstrated that even the yeast *Rhodotorula mucilaginosa* has proteolytic enzymes and thus a quite strong proteolytic activity [8].

It is worth noting that, besides these hypotheses, a definite interpretation of the biological significance of the above mentioned metabolic changes is not possible. The main difficulty arises from the known sensitive shift of the metabolite profiles according to specific evolutions of the microbial diversity [1]. In addition, other intrinsic factors may contribute to the measured alterations such as autolytic and chemical processes. Nevertheless, despite these limitations, this is the first work examining simultaneously the effect of two cold storage temperatures on the metabolic composition of mussels and their microbiological properties. Our initial findings may therefore serve as the baseline information for further investigations aimed at identifying new biomarkers of seafood freshness.

## 5.2 Metabolic responses to environmental stressors

### 5.2.1 Metabolic responses to heavy metal exposure

As persistent inorganic pollutants, heavy metals are one of the major harmful toxicants for the marine environment [9]. Indeed, the accumulation of such pollutants in aquatic species leads to concentrations that are several orders of magnitude higher than those of the average water [10], thus becoming of concern to public health when bioaccumulated in edible animals. In order to evaluate the impact of pollutants on the environmental quality, the use of molecular biomarkers has become highly pertinent to carry out a rapid assessment of their deleterious effects on the ecosystem [11] since their alterations reflect the pathophysiological status of an organism.

In this Ph.D. project, NMR-based metabolomics has been applied as advanced technique to analyze the effects of 48 h heavy metal exposure on *Ruditapes decussatus* and *Ruditapes philippinarum* metabolic profiles (**Paper III**). The grooved carpet shell *R. decussatus* and the Manila clam *R. philippinarum* are phenotypically similar clams' species belonging to the Veneridae family, widely reared along the Mediterranean Sea coast. In spite of their phylogenetic similarities, they have been demonstrated to differently accumulate heavy metals [12], suggesting also the occurrence of different

metabolic responses upon metal poisoning. Both species were exposed to increasing concentrations of lead nitrate (10, 40, 60, and 100 μg/L) and zinc chloride (20, 50, 100, and 150 μg/L), under laboratory conditions. A comprehensive overview on the initial pollution state of the samples and the occurrence of bioaccumulation was achieved by the direct detection of metals through Inductively Coupled Plasma Optical Emission Spectrometry (ICP-OES).

Principal component analysis of the $^{1}$H-NMR spectra of the aqueous extracts of clams was performed to seek valid metabolic signatures, potentially translatable into metal poisoning. Following 48 h metal exposure, clams' metabolic responses appeared to be both species- and metal-specific. Concerning the differences observed in terms of species, the most striking ones were the modification of organic osmolytes and Free Amino Acids (FAA) contents. All these molecules, deemed as "organic osmolytes" and "compatible solutes" (i.e. small carbohydrates, polyols, amino acids and derivatives), are known to be accumulated in aquatic invertebrates for counteracting osmotic perturbations and maintaining cell volume [13].

Metal exposure in *R. decussatus* induced an increase in the relative amount of organic osmolytes (betaine and taurine in lead-treated samples and hypotaurine and homarine in zinc-treated clams), while it decreased the content of FAA. Differently, the *R. philippinarum* metabolic response was characterized by a decrease in the organic osmolytes content and an increase of FAA (i.e. threonine, tyrosine and phenylalanine). These results suggest a differential pattern behind the restore of the cellular homeostasis in the two clams' species.

Further information on the effect of heavy metal pollution on clams' metabolic fingerprint was achieved by investigating metabolite-metabolite correlations, calculated in terms of Pearson's correlations. Lead treated samples showed a higher number of strong positive and negative correlations between amino acids and between amino acids and organic osmolytes. This result suggested a more robust metal effect in lead treated clams rather than in the zinc treated specimens.

Despite the difficulties of a strict comparison of our findings with those reported in the literature, due to dissimilar experimental conditions, we found globally a good agreement between our data and those of other studies [14]–[16], thus confirming

NMR-based metabolomics as a promising tool for the identification of putative biomarkers as fast and sensitive indicators of contaminant-induced stress. Although still much research has to be carried out to a complete characterization of the pollution-induced responses in bivalves, our findings evidenced the utility of metabolomics in identifying a biomarker contour to be used for marine quality control.

### 5.2.2 Seasonality patterns in *Ruditapes decussatus*

Bivalves are a variegated class which represent a largely exploitable resource in fishery industry. Indeed, shellfish farming is the main item of Italian aquaculture production and, within the EU, Italy is one of the main edible bivalve producer countries [17]. They comprise several species encompassing both marine and fresh-water habitats throughout the world.

Beside the remarkably economical relevance covered in the aquaculture industry, the consumption of bivalve mollusks helps to provide essential vitamins, minerals as well as polyunsaturated fatty acids with known health beneficial effects [18]. Moreover, they have been demonstrated to be feasible tools in environmental pollution monitoring programs (i.e. MWP) in which they are actually widely used as sentinel organisms.

Among bivalves, *Ruditapes decussatus* is a clam widely distributed in Sardinia island (Italy) where it represents an important seafood resource of high commercial relevance. Several nutritional studies have been carried out with the aim to characterize the fatty acid composition of these clams, demonstrating the incidence of the diet in their lipid composition [19], [20]. Although the undeniable relevance covered by this clam species in both nutrition and ecological fields, a complete characterization of the biochemical composition of *R. decussatus* has never been carried out.

In this work, $^{1}$H NMR-based metabolomics has been applied to investigate the effect of the seasonal change on *R. decussatus* metabolic profile (**Paper IV**). Principal component analysis evidenced a remarkable influence of season on clams' metabolome with sample distribution being strictly season-dependent. *i*ECVA was further applied to the spectral data in order to find the discriminant spectral regions responsible for the observed clustering. Several pre-defined intervals were used to perform iECVA. Among these, the spectral regions corresponding to the resonances of alanine and glycine were

found to be potentially better discriminants than the full spectral area. The signals arising from glycine and alanine were then integrated in each NMR spectrum; one-way ANOVA was performed to assess the significance of the observed intergroup differences. Remarkably, the observed variations in the relative amount of alanine and glycine was found to strictly mirror the seasonal change. These findings are in agreement with those found for other Mediterranean mollusks whose relative amount of alanine and glycine significantly varied over the year according to the measured cyclical salinity variations of the water [21]. Indeed, it is well known how marine invertebrate, and osmoconformers in general, can use free amino acids and osmolytes to cope with environmental osmotic stressful conditions [22] which can lead to shrinkage or swelling if the osmotic pressure of the environment dramatically changes. Although the investigation of the season-specific metabolome of *R. decussatus* is here presented at a preliminary stage, the findings of this study open the way for further investigations that would contribute to a comprehensive framework on the bio-ecology of these bivalves.

# References


[1] L. Gram and H. H. Huss, "Microbiological spoilage of fish and fish products.," *International journal of food microbiology*, vol. 33, no. 1, pp. 121–37, 1996.

[2] G. Picone, S. B. Engelsen, F. Savorani, S. Testi, A. Badiani, and F. Capozzi, "Metabolomics as a powerful tool for molecular quality assessment of the fish Sparus aurata," *Nutrients*, vol. 3, no. 2, pp. 212–227, 2011.

[3] F. Savorani, G. Picone, A. Badiani, P. Fagioli, and F. Capozzi, "Metabolic profiling and aquaculture differentiation of gilthead sea bream by 1 H NMR metabonomics," *Food Chemistry*, vol. 120, no. 3, pp. 907–914, 2010.

[4] E. Locci, C. Piras, S. Mereu, F. Cesare Marincola, and P. Scano, "1H NMR metabolite fingerprint and pattern recognition of mullet (Mugil cephalus) bottarga," *Journal of Agricultural and Food Chemistry*, vol. 59, no. 17, pp. 9497–9505, 2011.

[5] C. Delbarre-Ladrat, R. Chéret, R. Taylor, and V. Verrez-Bagnis, "Trends in postmortem aging in fish: understanding of proteolysis and disorganization of the myofibrillar structure.," *Critical reviews in food science and nutrition*, vol. 46, no. 5, pp. 409–421, 2006.

[6] P. Dalgaard, "Modelling of microbial activity and prediction of shelf life for packed fresh fish," *International Journal of Food Microbiology*, vol. 26, no. 3, pp. 305–317, 1995.

[7] M. B. Rao, A. M. Tanksale, M. S. Ghatge, and V. V Deshpande, "Molecular and biotechnological aspects of microbial proteases.," *Microbiology and molecular biology reviews : MMBR*, vol. 62, no. 3, pp. 597–635, 1998.

[8] "Proteolytic Activity of Microorganisms Isolated from Freshwater Fish," *Microbiology*, vol. 16, no. 1, pp. 128–132, 1968.

[9] A. Jakimska, P. Konieczka, K. Skóra, and J. Namieśnik, "Bioaccumulation of metals in tissues of marine animals , Part I : the role and impact of heavy metals on organisms," *Poland Journal of Environmental Studies*, vol. 20, no. 5. pp. 1117–1125, 2011.

[10] Z. P. De Bre, "Relation between metal concentration in water and metal content of marine mussels (Mytilus galloprovincilais ): impact of physiology," vol. 27, no. 7. pp. 1543–1552, 2008.

[11] A. Kaviraj, E. Unlu, A. Gupta, and A. El Nemr, "Biomarkers of Environmental Pollutants," *BioMed Research International*, vol. 2014. pp. 1–2, 2014.

[12] E. Figueira and R. Freitas, "Consumption of Ruditapes philippinarum and Ruditapes decussatus: Comparison of element accumulation and health risk," *Environmental Science and Pollution Research*, vol. 20, no. 8. pp. 5682–5691, 2013.

[13] W. Walla and P. H. Yancey, "Water Stress, Osmolytes and Proteins," *Integrative and Comparative Biology*, vol. 41, no. 4, pp. 699–709, 2001.

[14] C. Ji, H. Wu, M. Zhou, and J. Zhao, "Multiple biomarkers of biological effects induced by cadmium in clam Ruditapes philippinarum.," *Fish & shellfish immunology*, vol. 44, no. 2. Elsevier Ltd, pp. 430–5, 2015.

[15] C. Ji, Q. Wang, H. Wu, Q. Tan, and W.-X. Wang, "A metabolomic investigation of the effects of metal pollution in oysters Crassostrea hongkongensis.," *Marine pollution bulletin*, vol. 90, no. 1–2. Elsevier Ltd, pp. 317–22, 2015.

[16] H. Wu, C. Ji, Q. Wang, X. Liu, J. Zhao, and J. Feng, "Manila clam Venerupis philippinarum as a biomonitor to metal pollution," *Chinese Journal of Oceanology and Limnology*, vol. 31, no. 1. pp. 65–74, 2013.

[17] FAO, *The state of world fisheries and aquaculture*, vol. 2014. 2014.

[18] A. P. Simopoulos, "Omega-6/Omega-3 Essential Fatty Acid Ratio and Chronic Diseases," *Food Reviews International*, vol. 20, no. 1, pp. 77–90, 2004.

[19] M. Albentosa, U. Labarta, M. J. Fernández-Reiriz, and A. Pérez-Camacho, "Fatty acid composition of Ruditapes decussatus spat fed on different microalgae diets," *Comparative Biochemistry and Physiology Part A: Physiology*, vol. 113, no. 2, pp. 113–119, 1996.

[20] M. J. Fernández-Reiriz, U. Labarta, M. Albentosa, and A. Pérez-Camacho, "Lipid profile and growth of the clam spat, Ruditapes decussatus (L), fed with microalgal diets and cornstarch," *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, vol. 124, no. 3, pp. 309–318, 1999.

[21] S. Kube, A. Sokolowski, J. M. Jansen, and D. Schiedek, "Seasonal variability of free amino acids in two marine bivalves , Macoma balthica and Mytilus spp ., in relation to environmental and physiological factors," vol. 147, pp. 1015–1027, 2007.

[22] P. H. Yancey, "Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses," *Journal of Experimental Biology*, vol. 208, no. 15. pp. 2819–2830, 2005.

# 6 General conclusions

In the recent years, metabolomics has emerged as a high-throughput technique particularly feasible for the study of complex biological samples. By the study of the overall low molecular weight metabolites (metabolome), synthesized in relation to genetic alterations and/or external stimuli, it provides comprehensive insights into the pathophysiological state of an organism. The great versatility of NMR spectroscopy has enabled this analytical technique to be a useful tool for the identification and quantification of a wide range of cellular metabolites in biological samples. In particular, when combined with pattern recognition methods, it provides important contributions to the characterization of the phenotype of an organism, offering metabolomic research great possibilities in several scientific areas.

In this Ph.D. thesis, NMR-based metabolomics was applied for addressing issues related to food and environmental sciences in three bivalves' species: the clams *Ruditapes decussatus* and R*uditapes philippinarum* and the blue mussel *Mytilus galloprovincialis*. In particular, metabolomic approaches were developed to discover candidate biomarkers of mussels' freshness **(Paper I-II)**, and to study the biological impact, on clams, of several anthropogenic and natural environmental stressors, such as heavy metals **(Paper III)** and seasonal change **(Paper IV)**, respectively.

Seafood quality is a broad and complex concept embracing many factors, such as storage conditions, in terms of storage length and temperature, and microbial contamination. The need of new techniques for fast diagnosis of seafood spoilage has recently increased, representing an attractive alternative to traditional methods in terms of both costs and time. In the present Ph.D. project, the choice to characterize the freshness of edible lamellibranches mollusks stemmed from the commercial relevance covered by these organisms. Indeed, shellfish farming is the main item of Italian aquaculture production and, within the EU, Italy is one of the main edible bivalve producer countries. In addition, the consumption of bivalve mollusks provides essential vitamins, minerals as well as polyunsaturated fatty acids with well-known health beneficial effects. The combined use of NMR spectroscopy and multivariate data

analysis is here demonstrated to be an alternative and viable tool for detecting differences in the metabolic profiles of stored seafood **(Paper I-II)**. This general approach could find a wide application in the analysis of a huge variety of seafood, from spoilage detection to the establishment of predictive models capable to correlate the observed metabolic profiles to the storage time or, alternatively, to the actual microbial loads. Indeed, microbial contamination has been demonstrated to dramatically affect seafood freshness with Specific Spoilage Organisms (SSO) (i.e. *Shewanella putrefaciens*, *Pseudomonas aeruginosa* and *Vibrio parahaemolyticus*) being the main microorganisms contributing to seafood deterioration [1]. The study of the metabolome of fresh and stored seafood can provide insights into the biochemical aspects of spoilage, giving contributions to the identification of intermediate or surrogate spoilage biomarkers. This will lead to the establishment of specific approaches for the molecular assessment of seafood quality which has a remarkable impact on public health. The NMR-based metabolomic approach could be also extended to the study of bivalves' lipid composition and metabolism. Indeed, mussels' fat is rich in polyunsaturated fatty acids (PUFA, 37–48% of total fatty acids, mainly $\omega$3) [2] which are biologically important and have been associated with a decreased risk of cardiovascular disease. Shading light on the effect of different storage conditions on the bivalves' lipid profile will undeniably provide a giant step toward the characterization of the biochemical framework of these organisms, with positive implications for human health.

The importance to achieve deeper insights on bivalves' bioecology is additionally covered by the ecological and environmental relevance of these organisms. Indeed, marine bivalves' metabolic profiles have been demonstrated to narrowly mirror any change occurring in the surrounding environment (i.e. temperature and salinity changes, and pollution) [3]. This feature of marine bivalves could find its main applicability in biomonitoring programs. Understanding the biochemical consequences posed by environmental stressors is nowadays a significant challenge elevated by the large number of diverse pollutants with generally uncharacterized exposures, mechanisms, and toxicities. In the present Ph.D. project, NMR-based metabolomics has been demonstrated to be a viable tool for the identification of putative biomarkers of both anthropogenic and natural environmental stressors such as heavy metals and

seasonal change, respectively **(Paper III-IV)** in two clams' species. Although reasonable with some regards, it must remind that the biological information obtained from metabolomic studies, in term of putative biomarkers, has to be validated prior any statement. In particular, a valuable biomarker must be measurable, reproducible and linked to relevant biochemical outcomes. Therefore, to further assess the validity of the results presented in this thesis, it would be helpful to extend the present investigation to other marine bivalves' species. The detection of biomarkers of pollution in several species will become highly pertinent to carry out a rapid assessment of the deleterious effects on the ecosystem, since their alterations actually reflect the pathophysiological status of bivalves. Moreover, a targeted analysis of a specific pathway may be essential for fully clarify heavy metal-induced alterations. Finally, the possibility of integrating NMR with other analytical platforms may provide complementary information on bivalves' metabolome, offering an improved understanding of the endogenous effects induced by environmental stressors. In particular, NMR and MS have been demonstrated to be complementary and powerful analytical approaches for the complete characterization of the metabolome [4]. An integrated NMR-MS metabolomic approach will allow an improved identification of the overall metabolites and will offer the possibility to expand the range of metabolomic research.

# References

[1] L. Gram, , and H.H. Huss, "Microbiological spoilage of fish and fish products". *International Journal of Food Microbiology*, 33, 121-137, 1996.

[2] E. Orban, G. Di Lena, T. Nevigato, I. Casini, A. Marzetti, and R. Caproni, "Seasonal changes in meat content, condition index and chemical composition of mussels (*Mytilus galloprovincialis*) cultured in two different Italian sites", *Food Chemistry*, 77, 57-65, 2002.

[3] B. P. Lankadurai, E. G. Nagato, and M. J. Simpson, "Environmental Metabolomics: an emerging approach to study organisms responses to environmental stressors," *Environmental Review*, vol. 21, no. July, pp. 180–205, 2013.

[4] Z. Pan and D. Raftery, "Comparing and combining NMR spectroscopy and mass spectrometry in metabolomics," *Analytical and Bioanalytical Chemistry*, vol. 387, no. 2, pp. 525–527, 2007.

# Paper I

---

## Changes in the $^{1}$H NMR metabolic profile of mussels (*Mytilus galloprovincialis*) with storage at 0°C

**Violetta Aru**[1], Maria Barbara Pisano[2], Paola Scano[1], Sofia Cosentino[2], and Flaminia Cesare Marincola[1] (2015)

[1] Department of Chemical and Geological Sciences, University of Cagliari, S.S. 554 Bivio Sestu, 09042, Monserrato (Italy)

[2] Department of Public Health, Clinical and Molecular Medicine, University of Cagliari, Italy, S.S. 554 Bivio Sestu, 09042, Monserrato (Italy).

**In Capozzi, F., Laghi, L., & Belton, P. S (Eds), Magnetic Resonance in food science (pp. 181-198). RSC publishing.**

---

# CHANGES IN THE $^{1}$H NMR METABOLIC PROFILE OF MUSSELS (*Mytilus galloprovincialis)* WITH STORAGE AT 0°C

Violetta Aru[1], Maria Barbara Pisano[2], Paola Scano[1], Sofia Cosentino[2] and Flaminia Cesare Marincola[1]

[1] Dipartimento di Scienze Chimiche e Geologiche - Università degli Studi di Cagliari, S.S. 554 Bivio Sestu, 09042 Monserrato (Italy)

[2]Dipartimento di Sanità Pubblica, Medicina Clinica e Molecolare - Università degli Studi di Cagliari, S.S. 554 Bivio Sestu, 09042 Monserrato (Italy)

Abstract

This study investigated the chemical and microbiological changes of mussels (Mytilus galloprovincialis) during storage at 0°C for 7 days. The time-related metabolic signature of mussels was characterized by Nuclear Magnetic Resonance (NMR) spectroscopy and analyzed by orthogonal partial least squares discriminant analysis (OPLS-DA). The OPLS-DA models evidenced significant alterations in the concentration of amino acids, organic acids, and osmolytes: taurine, homarine, and betaine were reduced over time, whereas acetate, alanine, glycine, succinate, and trimethylamine were elevated with respect to fresh mussels. All of these changes reflected the presence of microorganism development under cold storage conditions.

# CHANGES IN THE $^{1}$H NMR METABOLIC PROFILE OF MUSSELS (*Mytilus galloprovincialis)* WITH STORAGE AT 0°C

Violetta Aru[1], Maria Barbara Pisano[2], Paola Scano[1], Sofia Cosentino[2] and Flaminia Cesare Marincola[1]

[1] Dipartimento di Scienze Chimiche e Geologiche - Università degli Studi di Cagliari, S.S. 554 Bivio Sestu, 09042 Monserrato (Italy)

[2]Dipartimento di Sanità Pubblica, Medicina Clinica e Molecolare - Università degli Studi di Cagliari, S.S. 554 Bivio Sestu, 09042 Monserrato (Italy)

## 1 INTRODUCTION

Mussels belonging to the *Mytilus galloprovincialis* specie are very popular shellfishes harvested in the Mediterranean Sea. These bivalves are highly appreciated globally due to their exceptional nutritional value making them ideal for the human diet. Indeed, mussels are rich in minerals and vitamins (A, B1, B2, B6, B12 and C) (1, 2). Furthermore, mussel fat is rich in polyunsaturated fatty acids (PUFA, 37–48% of total fatty acids, mainly $\omega$3) (3). These fatty acids are biologically important and have been associated with a decreased risk of cardiovascular disease (4).

Although mussels are available in the market under all the possible forms (frozen, vacuum-packed, pickled, smoked, and canned), most of the mussels are kept alive on ice or refrigerated (2-4°C) until consumed. The shelf life of mussels is limited, primarily due to a variety of microbial and biochemical degradation mechanisms depending on the duration and conditions of storage as well as the initial quality of the product (5, 6). There are many well established traditional analytical techniques and methods available to assess the quality of seafood (7), including sensory evaluation based on quality index method, microbial inspection based on total viable counts, biochemical methods related, and proteome analysis. Although chemical and microbiological methods are useful both for

research and product development, they are not practical for routine use, as they require expensive laboratory equipment and trained staff, are destructive, and can be labor intensive and time consuming. In order to surmount the aforementioned disadvantages, fast reliable methods are therefore necessary for assuring freshness specification of the starting material and making sure that the product will not become stale when distributed and displayed.

Nuclear Magnetic Resonance (NMR) spectroscopy is a technique able to provide detailed chemical information on a wide range of compounds simultaneously present in a sample. In combination with multivariate analysis, NMR is recognized to make available relevant information on the composition of food in many area of food science such as foodstuffs quality, raw material safety, and authentication (8).

The goal of this work was to investigate the NMR metabolic changes in the aqueous extracts of *Mytilus galloprovincialis* samples stored at 0° over a period of 7 days and to find putative metabolites-markers influenced by the conservation conditions. In order to achieve a more complete characterization of mussel spoilage, the most significant microbial groups were also investigated during storage.

## 2. MATERIAL AND METHODS

### 2.1 Chemicals

All solvents used, of the highest available purity, were purchased from Merck (Darmstadt, Germany), while deuterium oxide ($D_2O$, 99.9%) and sodium 3-trimethylsilyl-propionate-2,2,3,3,-d4 (TSP) were acquired from Sigma–Aldrich (Milan, Italy).

### 2.2 Samples

Live mussels (*Mytilus galloprovincialis*) were obtained from the local seafood market in Cagliari (Italy) and immediately transported to the laboratory in portable coolers at approximately 4°C. In the laboratory, mussels were washed and scrubbed under running potable water to remove debris; dead mussels or those with broken shells were discarded. The remaining mussels (70 individues) were manually shucked by cutting the adductor muscle with a sterile knife and each sample was put into insulated sterile plastic boxes without ice or water. Mussels were stored for 7 days at 0°C. Microbiological analyses were performed at 0, 2, and 6 days intervals, while the extraction of the hydrosoluble component to be analysed by NMR spectroscopy was carried out on days 0, 3, and 7.

2.3 Microbiological and statistical analyses

A 100 g sample, composed by the tissues and shell liquor from mussels, was mixed with an appropriate volume (1/2) (w⁄v) of 0.1% peptone diluent (Biolife). The mixture was homogenized in a stomacher (Stomacher Lab-Blender 400, PBI, Italy) for 1 min at normal speed. A volume of 70 ml of diluent was added to 30 ml of the homogenate to make a master dilution. After mixing, further decimal dilutions in 0.1% diluent were made and plated in specific media to enumerated microbial groups.

The analysis for *Escherichia coli* was performed according to the International Organization for Standardization, using the five-tube, three-dilution most-probable-number (MPN) method (9). Briefly, 10 ml volumes of the $10^{-1}$ dilution were added to each of five tubes of double strength Mineral Modified Glutamate Broth (MMGB, Biolife), 1 ml of the $10^{-1}$ dilution was added to each of five tubes of single strength MMGB and 1 ml of the $10^{-2}$ dilution was added to a second set of five tubes of single strength MMGB. The inoculated tubes were incubated at 37 °C for 24 h. Tubes showing any sign of a yellow coloration were sub-cultured onto Tryptone Bile X-glucuronide medium (TBX, Oxoid) and incubated at 44°C for 22 h. Tubes from which subcultures yielded blue or blue-green colonies on TBX were deemed positive. MPN tables (10) were used to calculate *E. coli* numbers per 100 g of flesh and inter-valve liquid (11). The limit of assay sensitivity was a MPN of 20 *E. coli* cells per 100 g of shellfish. Negative samples were reported as MPN < 20/100 g.

Total counts of mesophilic (MMC) and psychrotrophic (PMC) microorganisms were determined by the pour plate method, using Plate Count Agar (PCA, Microbiol, Cagliari, Italy) and incubating at 30 °C for 48 h and at 4 °C for 7 days, respectively.

Total *Enterobacteriaceae* levels were determined by pour plating on Violet Red Bile Glucose Agar (Microbiol) and incubating at 30° C for 24–48 h. Qualitative *Vibrio parahaemolyticus* analysis was carried out on 25 g of blended mussels which were added to 225 ml of phosphate buffered peptone water, homogenized and incubated for 24 h at 37 °C. A loopful of peptone water was streaked onto Thiosulphate Citrate Bile Salt Agar (TCBS, Oxoid) and incubated for 24 h at 37 °C. The suspected colony types (yellow and green) were picked out, streaked on to TCBS and incubated at 37 °C for 24 h. Isolates were examined for Gram reaction, cell morphology, moltility, oxidase and catalase reaction and for behavior in TSI (Triple Sugar Iron Agar). They were then grouped according to the criteria of Mossel et al. (12) and identified by API 20E and API 20NE (Biomérieux, Marcy l'Etoile, France). The galleries were incubated at 30 °C for 24-48 h and the API profiles were compared with the analytical Profile Index (Apilab plus version 3.3.3)

The presence of food-borne pathogens *Salmonella* spp. and *Listeria monocytogenes* was also investigated. Twenty-five grams of sample were diluted in 225 ml of buffered peptone water, homogenized as previously described and incubated for 18 h at 37 °C for the detection of *Salmonella* spp. Another 25 g of sample was diluted in Half-fraser broth, homogenized and then incubated for 24 h at 30 °C for the detection of *L. monocytogenes* according to the ISO methods (13,14).

One-way ANOVA was performed on microbiological data for each storage time. Tukey test for multiple comparisons was used to separate treatment means. All statistics were performed using GraphPad Prism Statistics software package version 3.00 (GraphPad

Prism Software Inc., San Diego, CA, USA). Statistical significance was inferred at $P < 0.01$.

## 2.4 pH measurement of the samples

The pH of the homogenized samples (100 g in 200 ml of 0.1% peptone diluent) was measured with a HI8520 pH meter (Pool Bioanalysis Italiana, Milan, Italy).

## 2.5 Sample preparation for NMR analysis

Water-soluble metabolites were extracted from mussels using a methanol/chloroform/water mixture according to the Folch method (15). Each mussel sample was dissolved in 12 mL of a mixture chloroform–methanol (2:1,v/v). After the addition of 4 mL of $H_2O$ and centrifugation at 1300 rpm for 1 h at 4 °C, the methanol/water mixture, containing the low molecular weight water-soluble components, was separated from the chloroform fraction, containing the lipid components. The $CH_3OH/H_2O$ phase was dried using a rotary evaporator. The water-soluble fraction was dissolved in 1.2 mL of a $D_2O$ solution of the internal standard (TSP) 0.80 mM. The pH of the final sample was accurately adjusted to 6.52 ± 0.03. Then, an aliquot of 650 µL was placed into a 5 mm NMR tube for NMR analysis.

## 2.6 NMR measurements

$^1H$ NMR spectra were recorded at 300 K using a Varian Unity Inova 500 MHz NMR spectrometer (Agilent Technologies, CA, USA) operating at 499.839 NMR spectra were acquired with a sweep width of 6000 Hz, a 45° pulse, an acquisition time of 1.5 s, a relaxation delay of 4 s, and 256 scans. The residual signal of $H_2O$ was suppressed by applying a presaturation technique with low power radiofrequency irradiation for 1.5 s. After Fourier transformation with 0.3 Hz line broadening and a zero-filling to 64 K, $^1H$ spectra were phased and baseline corrected. Spectral chemical shift referencing on the TSP $CH_3$ signal at 0.00 ppm was performed in all spectra.

## 2.7 NMR data preparation and analysis

For multivariate statistical analysis, the NMR spectra were manually phased and baseline corrected using MestReNova (Version 8.1, Mestrelab Research SL). Each NMR spectrum was used to construct a data matrix by subdividing it into regions having an equal bin size of 0.02 ppm over a chemical shift range of 0.5-9.5 ppm. The regions between 4.6 and 5.2 ppm and 0.5 and -0.5 ppm were excluded for multivariate data analysis because of the signals of water and TSP, respectively. Bins were normalized to the sum of total spectral area to compensate for the overall concentration differences. The final data set was automatically reduced to ASCII files, converted into an Excel file and then imported to

SIMCA version 13.0 (Umetrics, Umea, Sweden) for statistical analysis. Data were Pareto scaled and analyzed using principal component analysis (PCA) (16) and orthogonal partial least squares discriminant analysis (OPLS-DA) (17). Unsupervised PCA was used for the overview of all data set, to observe clustering or separation trends and identify outliers, while supervised OPLS-DA was performed to remove variability not relevant to class separation. OPLS-DA is a classification technique that uses class information to maximize the separation between groups of observations. With respect to PCA, this method allows improved interpretation of the variations between discriminated groups, by removing information that has no impact on discrimination. Briefly, OPLS-DA facilitates the separation of the systematic variation in X into two parts, one that is linearly related to Y (predictive information) and one that is unrelated to Y (orthogonal information). The predictive information of Y in X is concentrated in the first predictive component and is associated with the between groups variation, while the variation in X which is unrelated to Y is put in the second and orthogonal component and is linked to the within groups

variation. The quality of the model is described by the parameter $R^2Y$ and $Q^2Y$, which represent the fraction of the variation of Y-variable and the predicted fraction of the variation of Y-variable, respectively. The statistical significance of $R^2Y$ and $Q^2Y$ values is estimated through a response permutation testing (18). In this test, the Y-matrix is randomly re-ordered, while the X-matrix is kept constant. This means that the Y-data remain numerically the same, but their positions are shifted by random shuffling. Each time a new OPLS-DA model is fitted using X and the permuted Y matrix, providing a reference distribution of $R^2Y$ and $Q^2Y$ for random data. An $R^2Y$ intercept value less than 0.3-0.4 and a $Q^2Y$ intercept value less than 0.05 are indicative of a valid model.

## 3. RESULTS AND DISCUSSION

### 3.1 Microbiological analysis

Changes in microbial flora of mussels during storage in ice are shown in Table 1. We identified 58 of the 65 isolates from the samples (Figure 1). The total viable counts of mesophilic and psychrotrophic bacteria significantly increased ($P < 0.01$) at a similar rate throughout the storage period at 0°°C, both microbial groups showing a slow growth.

*Enterobacteriaceae* are active seafood spoilers that grow rapidly and become predominant in spoiling fish (19). Therefore, the contribution of *Enterobacteriaceae* to the microflora of mussels and its spoilage potential must be taken into consideration, especially in the case of polluted water. Although *Enterobacteriaceae* can grow at low temperatures, the storage at 0°C, under our experimental conditions, provided a good control of this group. Indeed, the corresponding levels decreased over time up to being below the detection limit of 100 CFU/g at 6 days of storage.

**Table 1.** Microbial changes in *Mytilus galloprovincialis* samples stored at 0 °C

| Days of storage | Microbial parameters* | | | |
|---|---|---|---|---|
| | Mesophilic bacteria*** | Psychrotrophic bacteria*** | *Enterobacteriaceae* | *E. coli*** |
| 0 | 3.08 ± 0.07 | 2.30 ± 0.05 | 2.48 ± 0.35 | <20 |
| 2 | 3.50 ± 0.17 | 2.60 ± 0.05 | 2.48 ± 0.16 | <20 |
| 6 | 3.90 ± 0.30 | 3.08 ± 0.07 | <2 | <20 |

*Values are means ± S.D of log CFU per gram

**Values are MPN per 100 per 100 g using ISO method 16649-3

*** p<0.01

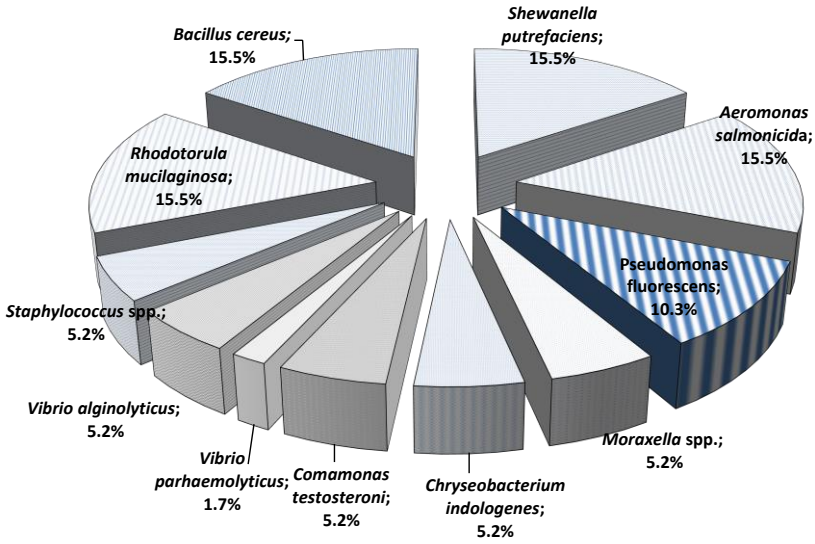



**Figure 1**. Distribution % of the species isolated from *M. galloprovincialis* samples stored at 0 °C

Pre-harvest contamination with pathogens from the animal/human reservoir (*Salmonella*, *Shigella*, *E. coli*, enteric virus) may pose a risk factor for humans and animals since in some cases a very low infective dose is required to cause illness (20). In the samples under investigation, *E. coli* concentrations were lower than the limit of 20 MPN per 100 g throughout all storage period, *Salmonella* and *L. monocytogenes* were never isolated, while *V. parahaemolyticus* was present in one sample

3.2 pH

The change in pH of seafood is usually a good index for quality assessment. Indeed, it is commonly related to the accumulation of lactic acid generated by glycogen in anoxic condition and/or the accumulation of basic substances, such as ammonia and trimethylamine (TMA), mainly derived from microbial developments. Under our experimental conditions, the pH of mussels slightly decreased from 6.56 to 6.17 during 7 days of storage at 0°C, suggesting a prevalent contribution from glycogen degradation.

3.2 NMR Spectroscopy

A representative $^{1}$H NMR spectrum of the aqueous phase extracted from mussels is shown in Figure 2. The assignment of the peaks was based on data reported in the literature (21,22). The spectra indicated the presence of numerous metabolites including amino acids, organic osmolytes, organic acids, alcohols, alkaloids, sugars and other carbohydrates.

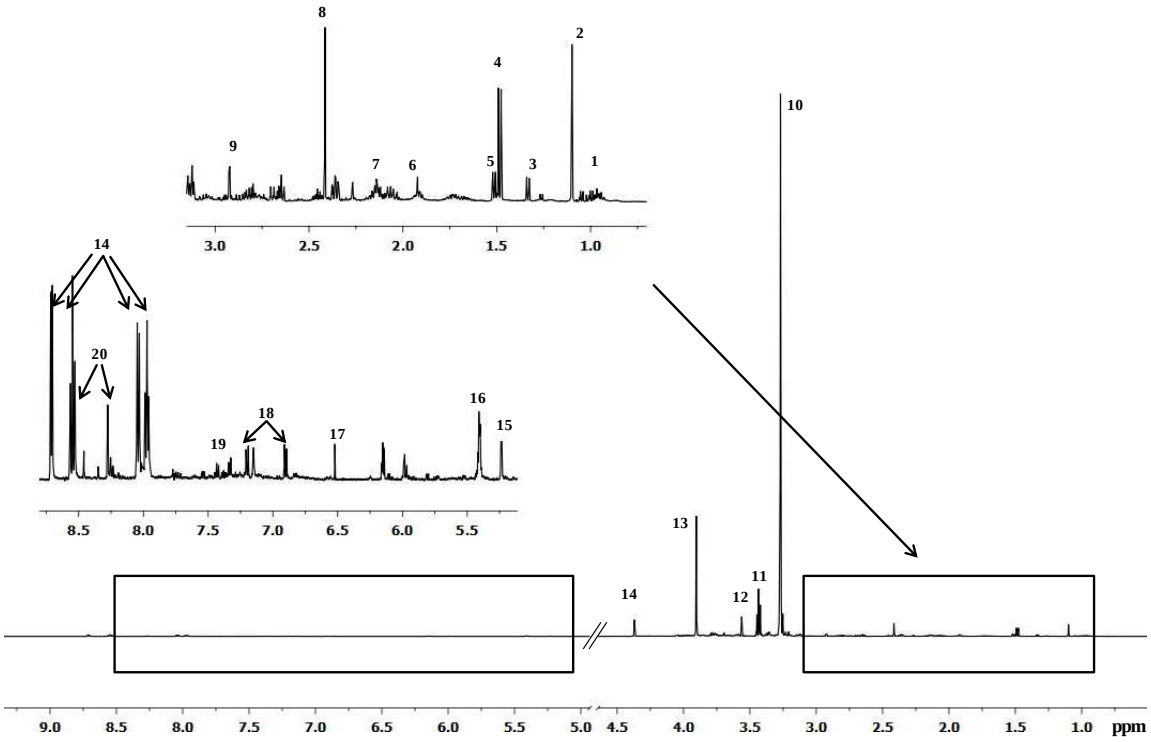



***Figure 2*** *Typical $^{1}$H NMR spectra of the aqueous extract of mussels. Peaks: 1. Branched Chain Amino acids (Isoleucine, Leucine, Valine), 2. unknown, 3. Lactate, 4. Alanine, 5. unknown, 6. Acetate, 7. Methionine, 8. Succinate, 9. Trimethylamine, 10. Betaine/Taurine/Trimethylamine N-Oxide, 11. Taurine, 12. Glycine, 13. Betaine, 14. Homarine, 15. Glucose, 16. Glycogen, 17. Fumarate, 18. Tyrosine, 19. Phenylalanine, 20. ADP.*

An exploratory unsupervised PCA analysis of the overall data set was firstly performed in order to achieve the natural interrelationship (grouping, clustering, or outlier detection) among samples. The model showed noticeable overlaps among samples, thus preventing any class differentiation in terms of storage time, possibly because of the structure variation within each sampling (data not shown). By applying OPLS-DA and using storage days as classifiers, a better separation was obtained in the scores map (Figure 3) with the following parameters resulting from internal cross validation: $R^2Y$ 0.981 and $Q^2$ 0.704. The goodness of fit and the predictability of this result were subjected to validation to test the possibility of correlation by chance. Thus, the model was subjected to ‘‘Y-scrambling” statistical validation. We randomly permutated the Y–variable, re-built the statistical model, and observed the trends of the predictive power and goodness of fit at each step. Two hundred rounds of such reshuffling gave coherent decreases in both parameters. The result of permutation test indicated that the OPLS-DA model was statistically sound and that the predictability was not due to over-fitting of the data (intercepts: $RY^2 = 0.101$; $Q^2 = -0.216$).

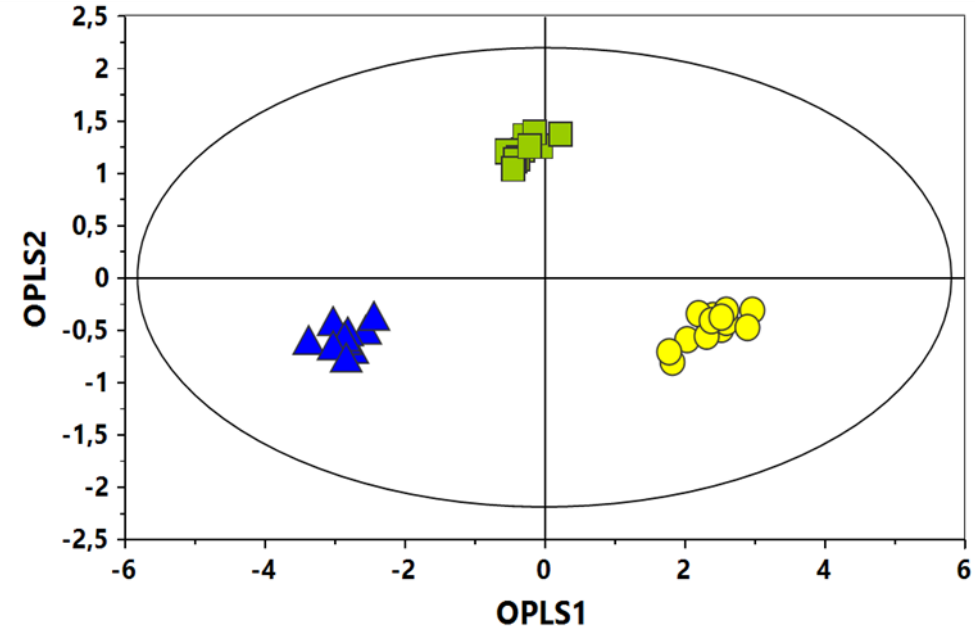



***Figure 3*** *OPLS-DA scores plot applied to the $^1H$ NMR spectra of the hydrosoluble component of mussels. Each score represents a NMR spectrum at a certain storage time at 0°C: 0 (○), 3 (■) and 7(▲) days. The quality factors for this models were $R^2X = 0.935$, $R^2Y = 0.981$, and $Q^2 = 0.704$.*

In order to further identify the key metabolic changes occurring during the storage time at 0°C, OPLS-DA models were generated in pairwise comparisons (Figure 4). All models

showed satisfactory values of the parameters that evaluate the explained variance and the prediction capability (see figure legend). The S-plots of the OPLS-DA models were used for identification of potential markers of group separation. S-plot is a scatter plot which combines the covariance and correlation loading profiles arising from the predictive component of the OPLS-DA model. In the S-plot, each point represents a variable (i.e. bin); the X axis represents variable contribution (covariance, p), where the farther the distance the variable points from zero, the more the variable contributes to the difference between two groups; the Y axis represents variable confidence (correlation, p(corr)), where the farther the distance the variable points from zero, the higher the confidence level of the variable to the difference between two groups. So, the variable points with high correlation and covariance values are located in regions far away from the origin. The variable importance coefficient (VIP) value reflecting the influence of each variable on the classification was used as an additional selection criterion using the "greater than one" rule. In the S-plot, the range of the variables selected is highlighted with a dotted rectangle.

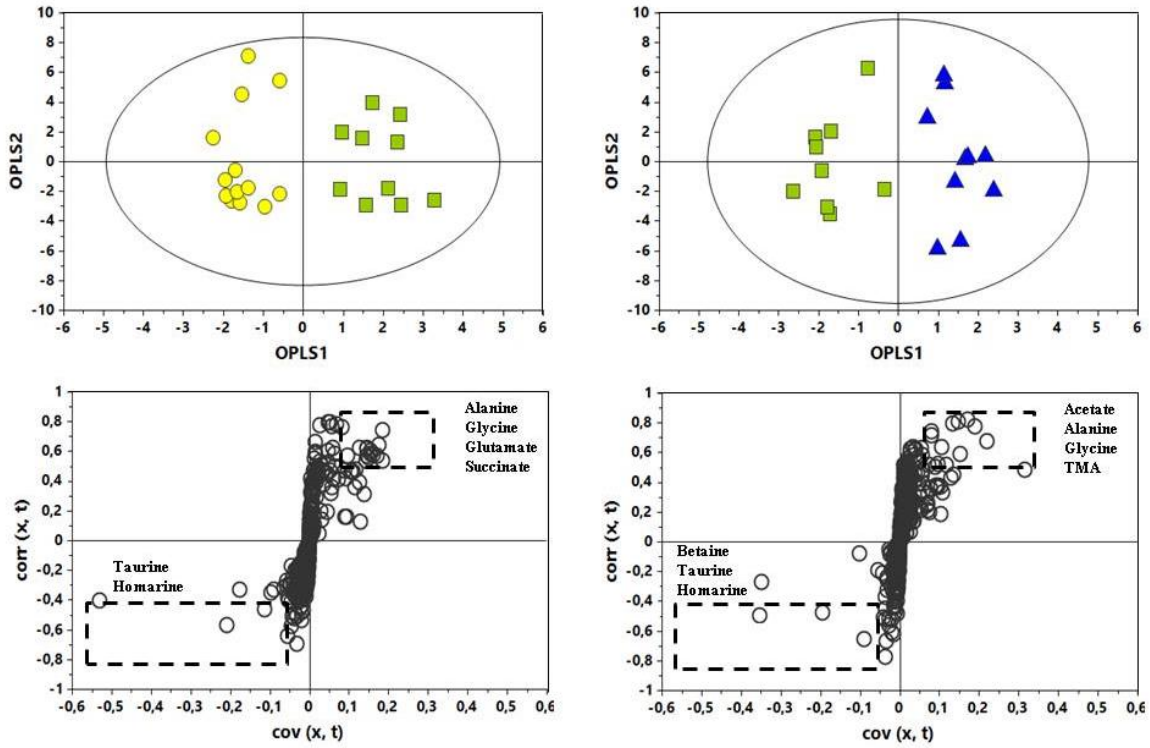



***Figure 4*** *OPLS-DA scores plots (A-B) and their respective S-plots (C-D) from the $^{1}H$ NMR spectra of the hydrosoluble component of mussels on 0 (○), 2 (■) and 7 (▲) days of storage at 0°C. The quality factors for these models were: $R^2X = 0.681$, $R^2Y = 0.893$, and $Q^2 = 0.75$ (right side); $R^2X = 0.851$, $R^2Y = 0.882$, and $Q^2 = 0.544$ (left side).*

In particular, the time-related metabolic signatures of mussels stored at 0°C were characterized mainly by alterations in the concentration of amino acids, organic acids, and osmolytes: taurine, homarine, and betaine were reduced over time, whereas acetate, alanine, glycine, succinate, and TMA were elevated with respect to fresh mussels. TMA, the most commonly used metabolite to evaluate the freshness of fish, is formed from bacterial use of Trimethylamine-*N*-oxide (TMAO), naturally occurring osmoregulating substance found in most marine fish species. This compound is found in very low levels in fresh fish, and its formation is associated with bacterial spoilage (23). Since, as seafood spoils, proteins are broken down to peptides, free amino acids, amines and volatile ammonia (24), the increase of alanine and glycine can be also involved in the microorganism development. Although glucose and lactate seem to be the substrates which are attacked first by the various groups of spoilage bacteria under aerobic and anaerobic conditions, the changes in their levels over storage at 0°C did not show statistical significance.

## 4 CONCLUSIONS

The results of the present study indicated that the NMR spectroscopy in conjunction with multivariate analysis is a feasible approach to analyze the time-related changes of the metabolic profile of mussels stored at 0°C. Although the observed modifications are in good agreement with the occurrence of microbiological development, as pointed out by the microbial counts, an accurate interpretation of the spoilage influence on the metabolic fingerprint of mussels is a difficult task. One of the reason is that the changes in the metabolic profile arise from the complex interplay of different biochemical processes, related to the concomitant actions of different microorganisms. Further integration of biological and chemical data is needed to demonstrate the utility of this analytical approach in providing rapid, accurate, and quantitative results on microbiological spoilage.

**References**

P. A. Karakoltsidis, A. Zotos and S.M. Constantinides, *J. Food Comp. Anal*., 1995, **8**, 258.

K. Vareltzis, *Fish. News*, 1996, **11**, 38.

E. Orban, G. Di Lena, T. Nevigato, I. Casini, A. Marzetti and R. Caproni, *Food Chem*., 2002, **77**, 57.

P. M. Kris-Etherton, W. S. Harris and J. A. Lawrence, *Circulation*, 2002, **106**, 2747.

A.E. Goulas, I. Chouliara, E. Nessi, M.G. Kontominas and I.N. Savvaidis, *J. Appl. Microbiol*., 2005**, 98**, 752.

M. A. Khean, C. C. Parrish and F. Shaidi, *J. Agric. Food Chem*. 2005, **53**, 7067.

G. Blafsdbttir, E. Martinsdbttir, J. Oehlenschbger, P. Dalgaard, B. Jensen, I. Undeland, I.M. Mackie, G. Henehan, J. Nielsen and H. Nilsen, 1997, Vol. 8.

L. Laghi, G. Picone and F. Capozzi, *Trend Anal. Chem.***,** 2014, **59**, 93.

ISO/TS 16649-3:2005. Microbiology of food and animal feeding stuffs - Horizontal method for the enumeration of beta-glucuronidase-positive Escherichia coli - Part 3: Most probable number technique using 5-bromo-4-chloro-3-indolyl-beta-D-glucuronide. Brussels: European Committee for Standardization.

T. J. Donovan, S. Gallacher, N. J. Andrew, M.H. Greenwood, J. Graham, J. E. Russel, D. Roberts and R. Lee, *Comm. Dis. Publ. Health*, 1998, **188**, 196.

T.L. Maugeri, M. Carbone, M.T. Fera, G.P. Irrera and C. Gugliandolo, *J. Appl. Microbiol.*, 2004, **97**, 354.

D. A. A. Mossel, J. E. Corry, C. B. Struijk and R. M. Baird in *Essentials of the microbiology of foods: a textbook for advanced studies.* John Wiley and Sons, Chichester (England), p 699.

EN ISO 6579:2004. Microbiology of Food and Animal Feeding Stuffs – Horizontal Method for the Detection of *Salmonella* spp. Brussels: European Committee for Standardization.

ISO/DIS 11290-1: 2001. Microbiology of the food chain -Horizontal method for the detection and enumeration of *Listeria monocytogenes* and other *Listeria* spp. -Part 1: Detection method. Brussels: European Committee for Standardization.

J. Folch, M. Less and G. H. S. Stanley, *J. Biol. Chem.*, 1957, **226**, 497.

H. Hotelling, *J. Educ. Psychol.*, 1933, **24**, 417.

M. Bylesjo, M. Rantalainen, O. Cloarec, J.K. Nicholson, E. Holmes and J. Trygg, *J. Chemometr.*, 2006, **20**, 341.

L. Eriksson, E. Johansson, N. Kettaneh-Wold, J. Trygg, C. Wikström and S. Wold. *Multi- and megavariate data analysis: Principles and applications*. Umetrics AB, Umeå, Sweden, 2006.

L. Gram and H.H. Huss, *Int. J. Food Microbiol*., 1996, **33**, 121.

G. Teophilo, R. dos Fernandes Vieira, D. dos Prazeres Rodrigues and F. R. Menezes *Int. Microbiol.*, 2002, **5**, 11.

S. J. Rochfort, V. Ezernieks, A. D. Maher, B. A. Ingram and L. Olsen, *Food Res. Int.*, 2013, **54,** 1302.

W. Tuffnail, G. A. Mills, P. Cary and R. Greenwood, *Metabolomics*, 2009, **5**, 33.

T. A. Gill, *Food Rev. Int.*, 1990, **6**, 681.

L. Gram, L. Ravn, M. Rasch, J. Bartholin Bruhn, A. B. Christensen and M. Givskov, *Int. J. Food Microbiol.*, 2002, **78**, 79.

# Paper II

---

# Metabolomics analysis of shucked mussels' freshness

**Violetta Aru**[1], Maria Barbara Pisano[2], Francesco Savorani[3,4], Søren Balling Engelsen[3], Sofia Cosentino[2], and Flaminia Cesare Marincola[1] (2016)

[1] Department of Chemical and Geological Sciences, University of Cagliari, S.S. 554 Bivio Sestu, 09042, Monserrato (Italy)

[2] Department of Public Health, Clinical and Molecular Medicine, University of Cagliari, Italy, S.S. 554 Bivio Sestu, 09042, Monserrato (Italy).

[3] Quality & Technology, Department of Food Science, Faculty of Science, University of Copenhagen, Rolighedsvej 26, 1958 Frederiksberg C, Denmark

[4] Department of Applied Science and Technology (DISAT), Polytechnic University of Turin – Corso Duca degli Abruzzi 24, 10129 Torino (TO), Italy

**Food Chemistry**

---

## Accepted Manuscript


Metabolomics analysis of shucked mussels' freshness

Violetta Aru, Maria Barbara Pisano, Francesco Savorani, Søren Balling Engelsen, Sofia Cosentino, Flaminia Cesare Marincola

PII: S0308-8146(16)30330-2
DOI: http://dx.doi.org/10.1016/j.foodchem.2016.02.152
Reference: FOCH 18873

To appear in: *Food Chemistry*
Received Date: 23 November 2015
Revised Date: 17 February 2016
Accepted Date: 26 February 2016

Please cite this article as: Aru, V., Pisano, M.B., Savorani, F., Engelsen, S.B., Cosentino, S., Marincola, F.C., Metabolomics analysis of shucked mussels' freshness, *Food Chemistry* (2016), doi: http://dx.doi.org/10.1016/ j.foodchem.2016.02.152

**Abstract**

In this work a NMR metabolomics approach was applied to analyse changes in the metabolic profile of the bivalve mollusk *Mytilus galloprovincialis* upon storage at 0°C and 4°C for 10 and 6 days, respectively. The most significant microbial groups involved in spoilage of mussels were also investigated. The time-related metabolic signature of mussels was analysed by Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) which revealed a clear discrimination between the fresh samples and those stored at 0°C and 4°C. The results evidenced a noticeable increase in acetate, lactate, succinate, alanine, branched chain amino acids, trimethylamine and a progressive decline of osmolytes like betaine, homarine and taurine during storage. Exploration of the correlations of these metabolites with microbial counts suggested their use as potential biomarkers of spoilage. The results support the use of NMR metabolomics as a valuable tool to provide information on seafood freshness.

**Keywords**

Metabolomics; $^{1}$H-NMR; microbiology; mussels; spoilage; biomarkers.

**Highlights**

- Storage effects on mussels were investigated by metabolomics.
- The metabolic profile of mussels was analysed by NMR spectroscopy.
- Microbial features of stored mussels were characterized.
- Correlation between metabolites and microbiological counts was investigated.
- Results confirm metabolomics as a reliable method to assess seafood freshness.

## 1. Introduction

Food quality is commonly described using terms related to nutritional, microbiological, and biochemical traits. Many well established analytical techniques and methods are used to assess the quality of food, including sensory evaluation (Hyldig & Green-Petersen, 2004), microbial inspection (Gram & Huss, 1996), biochemical methods (Dainty, 1996), and proteome analysis (Pischetsrieder & Baeuerlein, 2009). At present, the increasing demand for high standards in quality assurance and process control addresses the technological progress toward the setup of faster, more powerful, and cheaper analytical procedures (Ibañez & Cifuentes, 2001). Among these, metabolomics is a relatively new investigative tool in food science (Cevallos-Cevallos, Reyes-DeCorcuera, Etxeberria, Danyluk & Rodrick, 2009*;* Khakimov, Gürdeniz & Engelsen, 2015). Metabolomics can be defined as the comprehensive study of the low molecular weight (<1500 Da) metabolites (metabolome) in biological cell, tissue, organ or organism, which are the end products of cellular processes (Dunn & Ellis, 2005). Analytical characterization of the identities of these food metabolites may provide critical information for addressing the problems associated with food control, classification, and authenticity assessments (Cevallos-Cevallos et al., 2009; Cubero-Leon, Peñalver & Maquet, 2014). Furthermore, applications of metabolomics to food are increasing in the area of nutrition research to enhance the understanding of the complex interaction between diet and health (Jones, Park & Ziegler, 2014).

The mussel Mytilus galloprovincialis, belonging to the Mytilidae family, is a very popular shellfish, widely distributed around the globe, especially in

temperate regions and localities where there are large shipping ports (Branch & Stephani, 2004). These mussels are rich in minerals and vitamins (Karakoltsidis, Zotos & Constantinides, 1995; Vareltzis, 1996) and in polyunsaturated fatty acids (mainly ω3) (Orban, Di Lena, Nevigato, Casini, Marzetti & Caproni, 2002) and, thus, have an exceptional nutritional value. M. galloprovincialis is mainly commercialized in the market as a fresh product and can be eaten both cooked and raw. The shelf life of fresh mollusks is limited because of their high perishability caused by chemical, enzymatic, or microbial activities (Gram, 1996). Although the first two processes may cause sensory changes leading to spoilage, it is well established that microbial growth and activity are the main reasons for the development of off-odors and flavors rendering non frozen fish products unacceptable or spoiled. One of the most common preservation methods applied to extend the shelf-life of seafood is the low temperature storage. This method does not kill the microorganisms but reduces microbial metabolism which is responsible for spoilage. Recently, we have performed a NMR-based metabolomics study of the post-mortem compositional changes occurring in the metabolic profiles of M. galloprovincialis specimens stored at 0 °C over a period of 7 days (Aru, Pisano, Scano, Cosentino & Cesare Marincola, 2015). Microbiological characterization of samples allowed also the analysis of mussels' spoilage in terms of the microbial diversity occurring during storage time. Our results suggested NMR-metabolomics as a promising analytical method to investigate seafood spoilage.

In order to deeper explore the utility of this approach for identifying new molecular biomarkers of spoilage, in the present work the investigation is

extended by comparing the changes of *M. galloprovincialis* metabolome occurring during storage at 0 °C and 4 °C for a period of 10 and 6 days, respectively. It is worth noting that mussels' storage length used at both temperatures does not reflect relevant practice conditions. However, such extended time conditions were intentionally used here to enhance compositional changes and make them more easily measurable and interpretable by NMR. The levels of the most significant microbial groups were also determined at both temperatures.

## 2.Materials and methods

### *2.1 Chemicals*

All solvents used, of the highest available purity, were purchased from Merck (Darmstadt, Germany), while deuterium oxide ($D_2O$, 99.9 %), sodium deuteroxide (NaOD, 40 wt. % in D2O, 99.5 atom % D), deuterium chloride (DCl, 99 atom % D) and sodium 3-trimethylsilyl-propionate-2,2,3,3,-d4 (TSP) were acquired from Sigma–Aldrich

(Milan, Italy).

### *2.2 Samples*

Fresh live mussels (*Mytilus galloprovincialis*) were bought from the local seafood market in Cagliari (Italy) and immediately transported to the laboratory in portable coolers at approximately 4°C. In the laboratory, mussels were washed and scrubbed under running potable water to remove debris; dead mussels or those with broken shells were discarded. The remaining mussels (100 individuals)

were manually shucked with a sterile knife and each sample was put into insulated sterile plastic boxes without ice or water. Mussels were stored at 4 °C for 6 days and at 0 °C for 10 days. For NMR and microbiological analysis, sampling was done on day 0 (control), after 2 and 6 days of storage at 4 °C and after 2, 6, and 10 days of storage at 0 °C (see Figure 1 in Aru, Pisano, Savorani, Engelsen, Cosentino & Cesare Marincola, 2016).

*2.3 Microbiological and statistical analyses*

For microbiological analysis, triplicate aliquots of each sample at different storage times and temperatures were analyzed. A 100 g aliquot, composed by the tissues and shell liquor from mussels, was mixed with an appropriate volume (1/2, v/v) of 0.1 % peptone diluent (Microbiol, Cagliari, Italy). The mixture was homogenized in a stomacher (Stomacher Lab-Blender 400, PBI, Milan, Italy) for 1 min at normal speed. A volume of 70 ml of diluent was added to 30 ml of the homogenate to make a master dilution. After mixing, further decimal dilutions in 0.1 % diluent were made and plated in specific media to enumerated microbial groups.

Total Counts of Mesophilic (MMC) and Psychrotrophic (PMC) microorganisms were determined by the pour plate method, using Plate Count Agar (PCA, Microbiol, Cagliari, Italy) and incubating at 30 °C for 48 h and at 4 °C for 7 days, respectively.

The analysis for *Escherichia coli* was performed according to the International

Organization for Standardization, using the five-tube, three-dilution Most-Probable- Number (MPN) method (ISO/TS 16649-3, 2005). Briefly, 10 ml volumes of the $10^{-1}$ dilution were added to each of five tubes of double strength Mineral Modified Glutamate Broth (MMGB, Biolife, Milan, Italy), 1 ml of the $10^{-1}$ dilution was added to each of five tubes of single strength MMGB and 1 ml of the $10^{-2}$ dilution was added to a second set of five tubes of single strength MMGB. The inoculated tubes were incubated at 37°C for 24h. Tubes showing any sign of a yellow coloration were sub-cultured onto Tryptone Bile X-glucuronide medium (TBX, Microbiol) and incubated at 44 °C for 22 h. Tubes from which subcultures yielded blue or blue-green colonies on TBX were deemed positive. MPN tables (Donovan et al., 1998.) were used to calculate *E. coli* numbers per 100 g of flesh and inter-valve liquid (Maugeri, Carbone, Fera, Irrera, & Gugliandolo, 2004). The limit of assay sensitivity was a MPN of 20 *E. coli* cells per 100 g of shellfish. Negative samples were reported as MPN < 20 per 100 g.

*Enterobacteriaceae* levels were determined by pour plating on Violet Red Bile Glucose Agar (Microbiol) and incubating at 30 °C for 24–48 h. Qualitative *Vibrio parahaemolyticus* analysis was carried out on 25 g of blended mussels which were added to 225 ml of phosphate buffered peptone water, homogenized and incubated for 24 h at 37 °C. A loopful of peptone water was streaked onto Thiosulphate Citrate Bile Salt Agar (TCBS, Oxoid, England) and incubated for 24 h at 37 °C. The suspected colony types (yellow and green) were picked out, streaked onto TCBS and incubated at 37 °C for 24 h.

Isolates were examined for Gram reaction, cell morphology, motility, oxidase and catalase reaction and for behavior in TSI (Triple Sugar Iron Agar). They were then grouped according to the criteria of Mossel et al. (Mossel, Corry, Struijk & Baird, 1995) and identified by API 20E and API 20NE (Biomérieux, Marcy l'Etoile, France). The galleries were incubated at 30 °C for 24-48 h and the API profiles were compared with the analytical Profile Index (Apilab plus version 3.3.3)

The presence of food-borne pathogens *Salmonella* spp. and *Listeria monocytogenes* was also investigated. Twenty-five grams of sample were diluted in 225 ml of buffered peptone water, homogenized as previously described and incubated for 18 h at 37 °C for the detection of *Salmonella* spp. Another 25 g of sample was diluted in Halffraser broth, homogenized and then incubated for 24 h at 30 °C for the detection of *Listeria monocytogenes* according to the ISO methods (ISO 6579:2004; ISO 11290-1: 2001).

Two-way ANOVA was performed on microbiological data, with temperature and time as factors, using GraphPad Prism Statistics software package version 3.00 (GraphPad Prism Software Inc., San Diego, CA, USA). Statistical significance was inferred at $p < 0.01$.

### *2.4 pH measurement of the samples*

The pH of the homogenized samples (100 g in 200 ml of 0.1 % peptone diluent) was measured with a HI8520 pH meter (PBI).

### *2.5 Sample preparation for NMR analysis*

Water-soluble metabolites were extracted according to the Folch method (Folch, Less & Stanley, 1957). Each mussel sample was dissolved in 12 mL of a mixture chloroform– methanol (2:1, v/v). After the addition of 4 ml of $H_2O$ and centrifugation at 1700 g for 1 h at 4 °C, the methanol/water mixture was separated by the chloroform fraction. The former contained the low molecular weight water-soluble components, while the latter the lipid components. Water/methanol solvent mixture was removed with an Eppendorf concentrator 5301 (Eppendorf AG, Hamburg, Germany), and the remaining aqueous extract was freeze-dried. The freeze-dried sample was dissolved in 1.2 ml of a $D_2O$ solution of the internal standard (TSP) 0.80 mM and centrifuged at 13000 g for 5 min at 4 °C to remove particulate matter. The pH of the final sample was accurately adjusted to 6.52 ± 0.03 by adding a small amount of NaOD or DCl to minimize pH-based peak movements. Then, an aliquot of 650 µl was placed into a 5 mm NMR tube for NMR analysis.

*2.6 NMR measurements*

$^1$H-NMR spectra were acquired with a Varian Unity Inova 500 spectrometer (Agilent Technologies, CA, USA) operating at 499.84 MHz at 300 K. For each spectrum, 256 scans were collected into 32k data points over a spectra width of 6000 Hz, with a 45° pulse, an acquisition time of 1.5 s, and a relaxation delay of 4 s. The solvent (water) residual signal was suppressed by applying a presaturation technique with low power radiofrequency irradiation for 1.5 s. An exponential function corresponding to 0.3 Hz was applied to each free induction

decay (FID) before Fourier transformation and a zerofilling to 64 K. Spectral chemical shift referencing on the TSP $CH_3$ signal at 0.00 ppm was performed in all spectra.

*2.7 NMR data pre-treatment and chemometric analysis*

NMR spectra were manually phased and baseline corrected using MestReNova (Version 8.1, Mestrelab Research SL). Each NMR spectrum was integrated between 0.5 and 9.5 ppm over a series of 0.005 ppm integral regions (bins). The regions between 4.6 and 5.0 ppm and 0.5 and -0.5 ppm were excluded for multivariate data analysis because of the signals of water and TSP, respectively. The noisy region between 9.5 and 10.5 ppm was also removed. Bins were normalized to the sum of total spectral area to compensate for the overall concentration differences. The final data set, sized 62 samples and 1665 variables, was converted into an Excel file and then imported to SIMCA version 13.0 (Umetrics, Umeå, Sweden) for statistical analysis. Data sets were pre-processed using Pareto scaling by weighting each integral region or variable by $(1/S_k)^{½,}$ where $S_k$ represents the standard deviation of the variable. This increased the representation of lower concentration metabolites in the resultant data models, while minimizing the contribution from noise.

Principal components analysis (PCA) was performed to sample overview and to show trends, groupings and outliers in the data (Wold, Esbensen & Geladi 1987). PCA model performance was evaluated using the correlation coefficient $R^2$ and the default method of 7-fold internal cross-validation of SIMCA to extract the coefficient $Q^2$. $R^2$ is defined as the proportion of variance in the data explained

by the model and indicates goodness of fit. $Q^2$ is defined as the proportion of variance in the data predictable by the model and indicates the goodness of prediction. A good prediction model is achieved when $Q^2$>0.5 and, if $Q^2$>0.9, it is regarded as an excellent predictive ability (Hotelling, 1933).

Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA), a supervised classification technique, was used to extract the systematic differences between known classes of samples in predictive variation (systematic variation correlated to the class separation) and orthogonal variation (systematic variation unrelated to the class separation) (Bylesjo, Rantalainen, Cloarec, Nicholson, Holmes & Trygg, 2006). Model validity was verified using permutation tests (Y-scrambling) and CV-ANOVA (Eriksson, Trygg & Wold, 2008). The metabolites that most contributed to class separation were selected by analyzing the correlation coefficient loading plots. This plot depicts simultaneously the outcomes of the covariance (peak height) and correlation (color code) analyses applied to the $^1$H-NMR dataset. The observed phase of the resonance signals represents the relative changes in the concentration of metabolites: negative (positive) peak phase reflects metabolites with decreased (increased) concentration over storage time. The hot colored metabolites (e.g. red) show more significant contribution than the cold colored (e.g. blue) ones for the intergroup discrimination.

## 3. Results

### *3.1 Microbiological analysis*

Microbial culture techniques were used to estimate viable counts of the various microbial floras in mussel samples during cold storage at 0° C and 4°C for 10 and 6 days, respectively. As can be seen in Table 1, at time 0 mesophilic bacteria showed the highest mean count (3.08 Log CFU/g), while psychrotrophs and *Enterobacteriaceae* displayed very similar loads (2.30 and 2.48 Log CFU/g). At 4 °C, both mesophilic and psychrotrophic viable bacterial counts significantly increased throughout the storage period up to 4.90 and 4.70 Log CFU /g, respectively. Significant lower counts were observed at refrigerated storage at 0 °C. A difference of 1 and 1.9 Log CFU/g was recorded in mesophilic and psychrotrophic mean counts, respectively, between storage at 4°C and 0 °C for 6 days.

*Enterobacteriaceae* counts starting from a level of 2.48 Log CFU/g remained constant during 6 days of storage at 4 °C, while they were lower than 2 Log CFU/g in samples stored at 0 °C. In all samples the concentration of *E. coli* was below the level of sensitivity of the assay (20 MPN per 100 g) throughout the whole storage period, being thus within the legal limits of 230 MPN/100 g (EC Reg. n. 1441/2007). *Salmonella* spp. and *Listeria monocytogenes* were never isolated. Regarding the presence of *Vibrio parhaemolyticus*, only one aliquot of the sample stored at 0°C and 4 °C at 2 days proved to be positive.

The average distribution of the microbial species isolated in mussel samples was also determined (see Figure 2 in Aru et. al., 2016). Psychrotrophic Gram-negative bacteria were the dominant microorganisms. They were mainly represented by *Pseudomonas fluorescens*, *Aeromonas salmonicida* and *Shewanella putrefaciens* species, which are known to be the typical microbiota of

many seafood, including bivalves, at refrigeration temperatures (Gram, 1996). These bacteria, referred to as "specific spoilage organisms" or SSOs (Dalgaard, 1995) are able to produce volatile organic compounds (VOCs) from the degradation of soluble, low molecular weight components responsible for off-flavors and taste associated with spoiled seafood.

Among other species, the yeast *Rhodotorula mucilaginosa* and *Bacillus cereus* were particularly abundant in the isolates at both temperatures.

### *3.2 pH changes during storage*

The change in pH of seafood is usually a good index for quality assessment. Indeed, it is commonly related to the accumulation of lactic acid generated by glycogen in anoxic condition and/or the accumulation of basic substances, such as ammonia and TMA, mainly derived from the action of alkalinizing bacteria. As reported in literature (Khan, Parrish & Shahidi 2005), fresh and good-quality bivalves exhibit pH values varying from 6 to 7. In our study, the pH of mussels significantly decreased from 6.56 (fresh samples) to 6.17 after 6 days of storage at 0°C and to 5.9 after 10 days (Table 1). A more remarkable decrease in the pH was recorded in samples stored at 4°C. Indeed, after 6 days of storage the pH reached a value of 5.77, suggesting a more pronounced glycogen degradation at this storage temperature.

### *3.3 NMR Spectroscopy and chemometrics analysis*

A total of 62 water-soluble extracts from mussels were analysed by $^{1}$H-NMR spectroscopy. A representative spectrum is shown in Fig. 1. The spectrum

displays several hundred peaks that arise from the different functional groups of a large number of metabolites including amino acids, organic acids, organic osmolytes, and carbohydrates. Among these, 29 metabolites were identified (see Table 1 in Aru et. al. 2016). PCA analysis was performed to generate an overview of the metabolic disturbance resulting from storage time at both temperatures. Because of a marked PC scores overlapping, the model did not evidence clear dynamic trends of the metabolic profiles over the experimental period (see Figure 3 in Aru et. al., 2016). Therefore, to better monitor time-related modifications, the metabolic trajectories for each temperature were drawn by taking the mean position in the PC scores plot for samples having the same storage time and temperature and connecting the coordinates chronologically. As it can be observed in Fig. 2A, the trajectories at 0 °C and 4 °C were similar in that a progressive change along the first principal component (PC1) occurred at both temperatures from the left to the right part of the plot as the storage period increased. This result suggests comparable metabolic changes for the two temperatures. However, it is worth noting that the point of mean score on day 6 at 4 °C moved further away from the control with respect to the score representing day 6 at 0 °C, suggesting the occurrence of faster dynamic changes at 4 °C. Besides, temperature-related modifications were visible along the second principal component (PC2). From the PC1 loadings plot (Fig. 2B), it can be noted that proteolysis, glycogen-lysis, glucose metabolism and the bacterial production of volatile amines are the main processes involved in mussels' spoilage at refrigerated temperatures. Indeed, branched chain amino acids (BCAA), alanine,

glucose, lactate, acetate, succinate and TMA increased over time. These changes were, in particular, more pronounced at 4 °C rather than at 0 °C (Fig. 2C).

In order to maximize the separation between experimental groups and focus on metabolic variations at the different time points, OPLS-DA was performed, at each temperature, in pairwise comparisons, where storage time class membership was defined with the discrete variable Y. Since no differences could be modelled between samples stored at 0 °C for 6 and 10 days, a combined class including both storage time-points was used. All models exhibited a reasonable separation between groups and acceptable goodness-of-fit ($R^2$Y) and goodness-of-prediction ($Q^2$Y). Furthermore, to validate the goodness of fit and the predictability of these results, permutation tests calculated by 200fold randomization of sample class were performed on the data sets. All the permuted models showed lower $R^2$Y values if compared with the original model's $R^2$Y value and all the $Q^2$ regression lines showed negative intercepts (see Table 2 in Aru et. al., 2016). These results confirm that the cluster separations in the OPLS-DA scores plots (Figs. 3A-B and 4A-B) are statistically significant. The metabolic information regarding the class separation was extracted from the correlation coefficient loadings plots (Figs. 3C-D and 4C-D). The major changes in the metabolic profile of water-soluble extracts pointed toward the same compounds at both storage temperatures: acetate, lactate, succinate, glucose, BCAA, alanine, and TMA increased over storage time, while the organic osmolytes betaine, homarine and taurine decreased with increasing of storage days. The discriminatory metabolites were quantified from the $^1$H-NMR spectra (see Table 3 in Aru et al., 2016) and possible correlations with the abundance of specific bacterial populations (mesophilic and

psychrotrophic) were investigated by calculating Pearson's correlation coefficients. The correlation matrices are reported in Table 2. Interestingly, all metabolites exhibited a quite strong correlation ($r \geq 0.6$) with microbial counts at both temperatures.

## 4. Discussion

One of the main problems of seafood marketing is the ease with which fish and shellfish undergo deterioration after death. Indeed, beside the post-mortem degradation due to autolysis and chemical processes, the high glycogen and free amino acids content, the lack of less digestible connective tissues and the high water activity make them an ideal substrate for bacterial growth, producing a wide variety of hydrolytic enzymes that contribute to seafood spoilage (Lougovois & Kyrana, 2005). The rate of deterioration during storage depends on multiple factors such as microbial contamination, biochemical compositions of the substrates and metabolites present in the tissue, and the conditions of storage. The changes in quality can be monitored by chemical, microbiological, or physical methods. Although all of these methods are useful for research or product development, they are not practical for routine use and not always applicable to all commercial species.

Even though the role of bacteria in the development of spoilage-related molecules in fish is recognized, data on the possible link of metabolic profiles to the microbial activity are still lacking. Understanding these relations may help to the selection of the most appropriate indices for monitoring seafood freshness and quality changes during storage. Only four studies in the literature have

investigated the changes in seafood composition during storage by metabolomics (Savorani, Picone, Badiani, Fagioli, Capozzi & Engelsen, 2010; Picone, Engelsen, Savorani, Testi, Badiani & Capozzi, 2011; Piras, Scano, Locci, Sanna & Cesare Marincola, 2014; Aru et al., 2015). Differently from these investigations, in the present work, we combined metabolomics with microbiological analysis to explore possible correlations between the chemical modifications of mussels and the microbial loads over cold storage. The most remarkable metabolic changes at 0 °C and 4 °C temperatures, although with different rates, were an increase in the relative concentrations of acetate, lactate, succinate, BCAA, alanine, glucose, TMA, and, conversely, a time-related decrease of betaine, homarine, and taurine. The modifications in the content of these metabolites led back to some of the components mainly involved in fish spoilage: proteins, carbohydrates, and other non-protein nitrogen compounds. Indeed, the observed increase in alanine and BCAA levels can be considered as an indication of protein degradation (Ababouch, Afilal, Benabdeljelil & Busta, 1991). TMA is one of the volatile bases largely responsible for the pungent, fishy, ammonia-like odor experienced in fish after the initial phase of freshness. It can be the result of the bacterial or enzymatic degradation of TMAO and betaine (King, 1988) which are naturally occurring osmoregulating compounds found in most marine fish species. Marine bivalve mollusks accumulate large amounts of polysaccharides, like glycogen, as energetic reserves (De Zwaan & Zandee, 1972). Many microorganisms, including yeast and bacteria, have the capacity to utilize glycogen to cope with the continuously changing environmental conditions (Wilson et al., 2010), thus leading to carbohydrates release which represent an

ideal substrate for several bacterial species (Adams & Moss, 2008). It is therefore likely that the observed accumulation of lactate, acetate and succinate in mussels at both temperatures over time is related to microbial glucose metabolism.

Although the association between the consume or production of the above-mentioned metabolites to specific strains is not an easy task, the analysis of the Pearson's correlation between these molecules and the bacterial growth suggested a strong correlation with both mesophilic and psychrotrophic microbial counts. Based on our microbiological data, some preliminary considerations can be made. TMA accumulation could derive from the action of *Shewanella putrefaciens*-like organisms and *Vibrio* spp., which are microorganisms typical of marine environment, but also from the metabolic activity of *Aeromonas* spp. (Gram & Dalgaard, 2002). Among them, *S. putrefaciens* is considered to have a higher spoilage potential in seafood products because of its high metabolic activity even if present in low concentrations. Amino acids could arise from the proteolytic activity of *Pseudomonas, Bacillus, Aeromonas* and *Vibrio* spp. (Rao, Tanksale, Ghatge, & Deshpande, 1998). It has also been demonstrated that even the yeast *Rhodotorula mucilaginosa* has proteolytic enzymes and thus a quite strong proteolytic activity (Kazanas, 1968).

It is worth noting that, besides these hypotheses, a definite interpretation of the biological significance of the above mentioned metabolic changes is not possible. The main difficulty arises from the known sensitive shift of the metabolite profiles according to specific evolutions of the microbial diversity (Gram, 2002). In addition, other intrinsic factors may contribute to the measured alterations such as autolytic and chemical processes. Nevertheless, despite these limitations, this

is the first work examining simultaneously the effect of two cold storage temperatures on the metabolic composition of mussels and their microbiological properties. Our initial findings may therefore serve as the baseline information for further investigations aimed at identifying new biomarkers of seafood freshness.

## 5. Conclusions

In recent years, the need of new techniques for fast diagnosis of seafood spoilage has increased, representing an attractive alternative to traditional methods in terms of both costs and time. In this work, we applied the NMR-based metabolomics approach to analyze cold storage effects on mussels. The data revealed substantial time-related changes in the metabolic profile of samples stored at 0 °C and 4 °C. These chemical modifications were in good agreement with the microbiological quality of samples, reflecting changing in their microbial loads. These results confirm the potential use of metabolomics as a reliable method to assess seafood freshness. Further studies are needed to advance understanding on the microbiological significant of spoilage related molecules.

## Acknowledgments

Violetta Aru gratefully acknowledges Sardinia Regional Government for the financial support of her PhD scholarship (P.O.R. Sardegna F.S.E. Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2007-2013 - Axis IV Human Resources, Objective l.3, Line of Activity l.3.1.).

**References**


Ababouch, L., Afilal, M. E., Benabdeljelil, H., & Busta, F. F. (1991). Quantitative changes in bacteria, amino acids and biogenic amines in sardine (*Sardina pilchardus*) stored at ambient temperature (25–28 °C) and in ice. *International Journal of Food Science & Technology*, 26, 297–306.

Adams, M. R., & Moss, M. O. (2008). Food Microbiology. (3rd Ed.). RSC publishing.

Aru, V., Pisano, M. B., Scano, P., Cosentino, S., & Cesare Marincola, F. (2015). Changes in the $^{1}$H NMR metabolic profile of mussels (*Mytilus galloprovincialis*) with storage at 0°C. In Capozzi, F., Laghi, L., & Belton, P. S (Eds), Magnetic Resonance in food science (pp. 181-198). RSC publishing.

Aru, V., Pisano, M. B., Savorani, F., Engelsen, S. B., Cosentino, S., Cesare Marincola, F. (2016). Data supporting the metabolomics analysis of shucked mussels' freshness. *Data in Brief*¸ submitted.

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

Branch, G. M., & Stephani, C. N. (2004). Can we predict the effects of alien species? A case history of the invasion of South Africa by *Mytilus galloprovincialis* (Lamarck). *Journal of Experimental Marine Biology and Ecology*, 300, 189–215.

Cevallos-Cevallos, J. M., Reyes-De-Corcuera, J. I., Etxeberria, E., Danyluk, M. D., & Rodrick, G. E. (2009). Metabolomic analysis in food science: a review. *Trends in Food Science & Technology*, 20, 557–566.

Cubero-Leon, E., Peñalver, R., & Maquet, A. (2014). Review on metabolomics for food authentication. *Food Research International*, 60, 95–107.

Dainty, R. H. (1996). Chemical/biochemical detection of spoilage. *International Journal of Food Microbiology*, 33, 19-33.

Dalgaard, P. (1995). Qualitative and quantitative characterization of spoilage bacteria from packed fish. *International Journal of Food Microbiology*, 26, 319-333.

De Zwaan, A. & Zandee, D. I., (1972). Body distribution and seasonal changes in the glycogen content of the common sea mussel *Mytilus edulis*. *Comparative biochemistry and phisiology*, 43, 53-58.

Donovan, T. J., Gallacher, S., Andrews, N. J., Greenwood, M. H., Graham, J., Russell, J. E., Roberts, D., & Lee, R. (1998). Modification of the standard method used in the United Kingdom for counting *Escherichia coli* in live bivalve mollusks. *Communicable Disease and Public Health*, 1, 188-196.

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

Eriksson, L., Trygg, J., & Wold, S. (2008). CV-ANOVA for significance testing of PLS and OPLS1 models. *Journal of Chemometrics*, 22, 594–60.

Folch, J., Less, M., & Stanley, G. H. S. (1957). A simple method for the isolation and purification of total lipides from animal tissues. *Journal of Biological Chemistry*, 226, 497-509.

Gram, L., & Huss, H.H. (1996). Microbiological spoilage of fish and fish products. *International Journal of Food Microbiology*, 33, 121-137.

Gram, L., & Dalgaard, P. (2002). Fish spoilage bacteria – problems and solutions. *Current Opinion in Biotechnology*, 13, 262–266.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417-441.

Hyldig, G. & Green-Petersen, D. M. B. (2004). Quality Index Method-An Objective Tool for Determination of Sensory Quality. *Journal of Aquatic Food Product Technology*, 13, 71-80.

Ibañez, E., & Cifuentes, A. (2001). New Analytical Techniques in Food Science. *Critical Reviews in Food Science and Nutrition*, 6, 413-450.

ISO/DIS 11290-1:2001. Microbiology of the food chain -Horizontal method for the detection and enumeration of *Listeria monocytogenes* and other *Listeria* spp. -Part 1: Detection method. Brussels: European Committee for Standardization.

ISO 6579:2004. Microbiology of Food and Animal Feeding Stuffs – Horizontal Method for the Detection of *Salmonella* spp. Brussels: European Committee for Standardization.

ISO/TS 16649-3:2005. Microbiology of food and animal feeding stuffs - Horizontal method for the enumeration of beta-glucuronidase-positive Escherichia coli - Part 3: Most probable number technique using 5-bromo-4-chloro-3-indolyl-beta-D-glucuronide. Brussels: European Committee for Standardization.

Jones, D. P., Park, Y., & Ziegler, T. R. (2014). Nutritional metabolomics: Progress in addressing complexity in diet and health. *Annual Review of Nutrition*, 32: 183–202.

Khakimov, B., Gürdeniz, G., & Engelsen, S. B. (2015). Trends in the application of chemometrics to foodomics studies. *Acta Alimentaria*, 44, 4-31.

Khan, M. A., Parrish, C. C., & Shahidi, F. (2005). Quality Indicators of Cultured Newfoundland Blue Mussels (*Mytilus edulis*) during Storage on Ice: Microbial Growth, pH, Lipid Oxidation, Chemical Composition Characteristics, and Microbial Fatty Acid Contents. *Journal of Agricultural and Food Chemistry*, 53, 7067–7073.

Karakoltsidis, P. A., Zotos, A., & Constantinides, S. M. (1995). Composition of the commercially important Mediterranean finfish, crustaceans and mollusks. *Journal of Food Composition and Analysis*, 8, 258-273.

Kazanas, N. (1968). Proteolytic Activity of Microorganisms Isolated from Freshwater Fish. *Applied Microbiology*, 128-132.

King, G.M. (1988). Distribution and metabolism of quaternary amines in marine sediments. In Blackburn, T. H. and Sørensen, J., (Eds), *Nitrogen Cycling in Coastal Marine Environments* (pp. 143-173). John Wiley & Sons Chichester publishing.

Lougovois, V. P., & Kyrana, V. R. (2005). Freshness quality and spoilage of chill-stored fish. In Arthur P. Riley (Eds.), *Food policy control and research* (pp. 35-86). Nova science

publishing.

Maugeri, T. L., Carbone, M., Fera, M.T., Irrera, G. P., & Gugliandolo, C. (2004). Distribution of potentially pathogenic bacteria as free living and plankton associated in a marine coastal zone. *Journal of Applied Microbiology*, 97, 354-361.

Mossel, D. A. A., Corry, J. O. E. L., Struijk, C. B., & Baird, R. M. (1995). Essentials of the microbiology of foods: a textbook for advanced studies, John Wiley & Sons, West Sussex.

Orban, E., Di Lena, G., Nevigato, T., Casini, I., Marzetti, A. & Caproni, R. (2002). Seasonal changes in meat content, condition index and chemical composition of mussels (*Mytilus galloprovincialis*) cultured in two different Italian sites. *Food Chemistry*, 77, 57-65.

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

Piras, C., Scano, P., Locci, E., Sanna, R. & Cesare Marincola, F. (2014). Analysing the effects of frozen storage and processing on the metabolite profile of raw mullet roes using $^{1}$H NMR spectroscopy. *Food Chemistry*, 159, 71-79.

Pischetsrieder, M. & Baeuerlein, R. (2009). Proteome research in food science. *Chemical Society Reviews*, 38, 2600–2608.

Rao, M. B., Tanksale, A. M., Ghatge, M. S., & Deshpande, V. V. (1998). Molecular and Biotechnological Aspects of Microbial Proteases. *Microbiology and Molecular Biology Reviews*, 3, 597-635.

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, 907-914.

Vareltzis, K. (1996). Mussels as Food. *Fishing News*, 11, 38-47.

Wilson, W. A., Roach, P. J., Montero, M., Baroja-Fernandez, E., Muñoz, F. J., Eydallin, G., Viale A. M. & Pozueta-Romero, J. (2010). Regulation of glycogen metabolism in yeast and bacteria, *FEMS Microbiology Reviews*, 34, 952–985.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 2, 37-52.

## Tables and figures

**Table 1** - Viable microbial counts and pH changes in *Mytilus galloprovincialis* samples during storage at 0°C and 4 °C.

| Microbial parameters | Storage temperature | Days of storage (means ± S.D of Log CFU per gram) | | | | |
|---|---|---|---|---|---|---|
| | | **0** | **2** | **6** | **10** | ***p*[b]** |
| Mesophilic bacteria | 0 °C | 3.08 ± 0.07 | 3.58 ± 0.23 | 3.86 ± 0.19 | 4.18 ± 0.16 | ≤ 0.01[c, d] |
| | 4 °C | 3.08 ± 0.07 | 3.50 ±0.04 | 4.90 ± 0.07 | / | |
| Psychrotrophic bacteria | 0 °C | 2.30 ± 0.05 | 2.83 ± 0.09 | 3.08 ± 0.05 | 3.20 ± 0.20 | ≤ 0.01[c, d] |
| | 4 °C | 2.30 ± 0.05 | 3.90 ± 0.07 | 4.70 ± 0.20 | / | |
| *Enterobacteriaceae* | 0 °C | 2.48 ± 0.35 | 2.48 ± 0.16 | <2 | <2 | / |
| | 4 °C | 2.48 ± 0.35 | 2.48 ± 0.35 | 2.60 ± 0.05 | / | / |
| *E. coli*[a] | 0 °C | <20 | <20 | <20 | <20 | / |
| | 4 °C | <20 | <20 | <20 | / | / |
| pH | 0 °C | 6.56 ± 0.04 | 6.23 ± 0.01 | 6.17 ± 0.05 | 5.90 ± 0.02 | ≤ 0.01[c, d] |
| | 4 °C | 6.56 ± 0.04 | 5.98 ± 0.01 | 5.77 ± 0.02 | / | |

[a] Values are MPN per 100 per 100 g using ISO method 16649-3.
[b] Statistical differences were assessed by two-way ANOVA.
[c] Effect of storage temperature.
[d] Effect of storage time.

**Table 2.** Pearson's correlation coefficients for discriminant metabolites associated with mesophilic and psychrotrophic total counts at 4 °C and 0 °C.

| | **4°C** | | **0°C** | |
|---|---|---|---|---|
| ***Metabolites*** | ***Mesophilic bacteria*** | ***Psychrotrophic bacteria*** | ***Mesophilic bacteria*** | ***Psychrotrophic bacteria*** |
| **Ace** | 0,826 | 0,804 | 0,871 | 0,763 |
| **Ala** | 0,849 | 0,828 | 0,754 | 0,619 |
| **BCAA** | 0,831 | 0,809 | 0,861 | 0,809 |
| **Bet** | -0,856 | -0,744 | -0,910 | -0,953 |
| **Glc** | 0,896 | 0,878 | 0,980* | 0,982* |
| **Hom** | -0,998** | -0,999* | -0,943 | -0,867 |
| **Lac** | 0,842 | 0,821 | 0,977* | 0,992** |
| **Suc** | 0,946 | 0,933 | 0,969* | 0,989* |
| **Tau** | -0,984 | -0,990 | -0,922 | -0,832 |
| **TMA** | 0,788 | 0,763 | 0,900 | 0,870 |

Keys: Ace=acetate; Ala=alanine; BCAA=branched chain amino acids; Bet=betaine; Glc=glucose; Hom=homarine; Lac=lactate; Suc=succinate; Tau=taurine; TMA=trimethylamine.
*p≤0.05
**p≤0.01

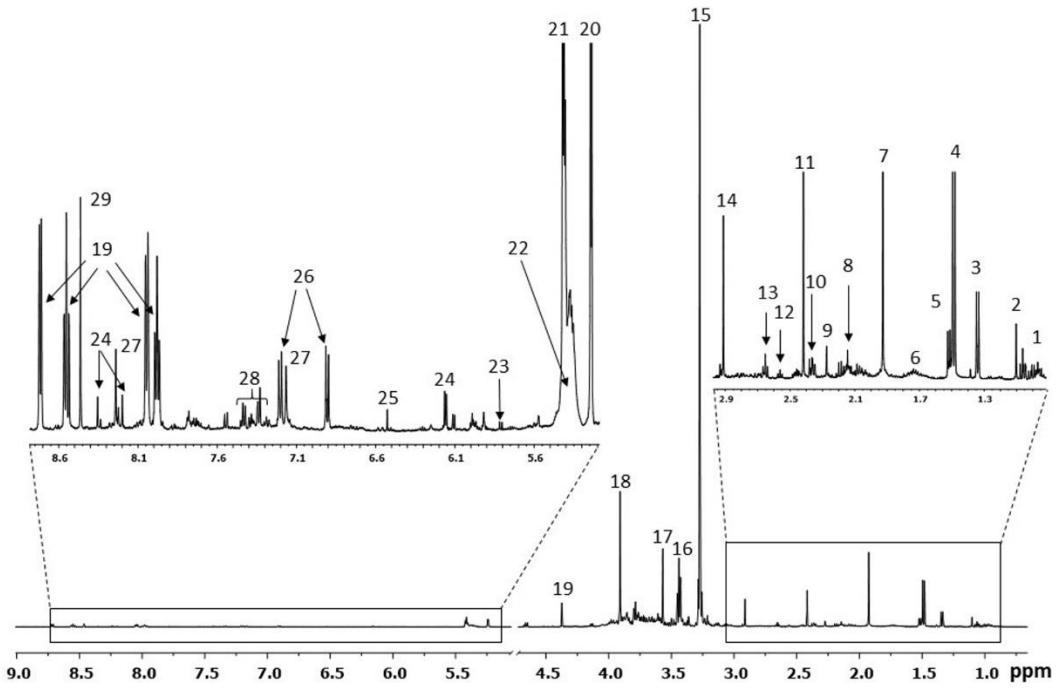



**Figure 1**. Representative $^{1}$H-NMR spectrum of the hydrosoluble fraction of mussel. 1. Isoleucine *(t, d)*, 0.93, Leucine *(t)*, 0.96 - Valine *(d, d)*, 0.98, 1.03; 2. Unknown *(s)*, 1.10; 3. Lactate *(d)*, 1.31; 4. Alanine *(d)*, 1.48; 5. Unknown *(d)*, 1.50; 6. Arginine *(m, m)*, 1.65, 1.72; 7. Acetate *(s)*, 1.91; 8. Methionine *(m)*, 2.16; 9. Acetoacetate *(s)*, 2.24; 10. Glutamate *(dd)*, 2.35; 11. Succinate *(s)*, 2.40; 12. β-Alanine *(t)*, 2.55; 13. Hypotaurine *(t)*, 2.64; 14. TMA *(s)*, 2.90; 15. Betaine *(s)*, Taurine *(t)*, TMAO *(s)*, 3.26; 16. Taurine *(t)*, 3.41; 17. Glycine *(s)*, 3.54; 18. Betaine *(s)*, 3.91; 19. Homarine *(s)*, 4.35; 20. α-glucose *(d)*, 5.25; 21. Maltose *(d)*, 5.40; 22. Glycogen *(bb)*, 5.41; 23. Uracil *(d)*, 5.80; 24. Inosine *(d)*, 6.11; 25. Fumarate *(s)*, 6.50; 26. Tyrosine *(d, d)*, 6.91, 7.20; 27. Histidine *(s, s)*, 7.18, 8.08; 28. Phenylalanine *(d, m, m)*, 7.31, 7.36, 7.42; 29. Formate *(s)*, 8.45. Metabolites, chemical shifts, and multiplicity are listed in Table 1 of Aru et al., 2016.

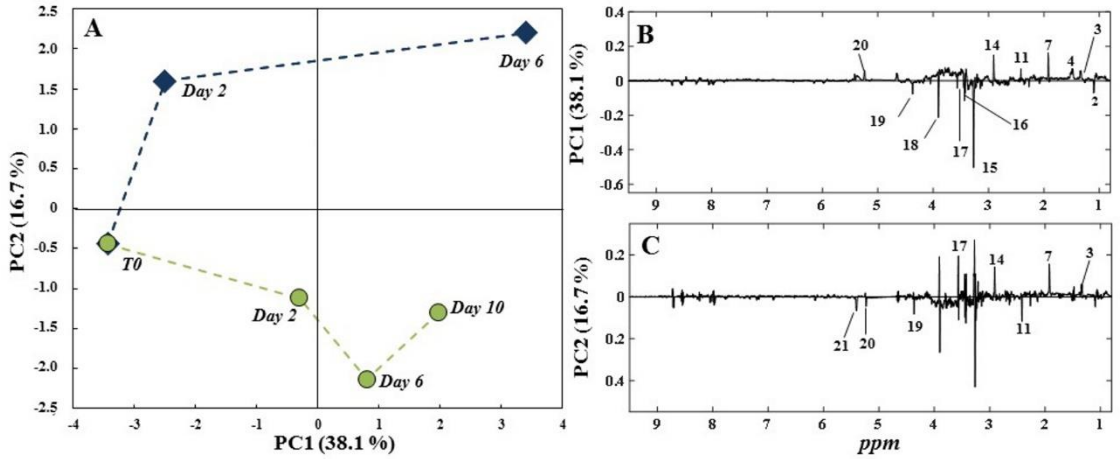



**Figure 2.** PCA metabolic trajectory plot derived from the 1H-NMR spectra of mussel aqueous extracts stored at 0 °C (○) and 4 °C (◊) (A) and corresponding loadings plots for PC1 (B) and PC2 (C) (R2= 0.878, Q2= 0.679). Each dot in the scores plot represents mean values of the scores from the first and second principal components at a time point.

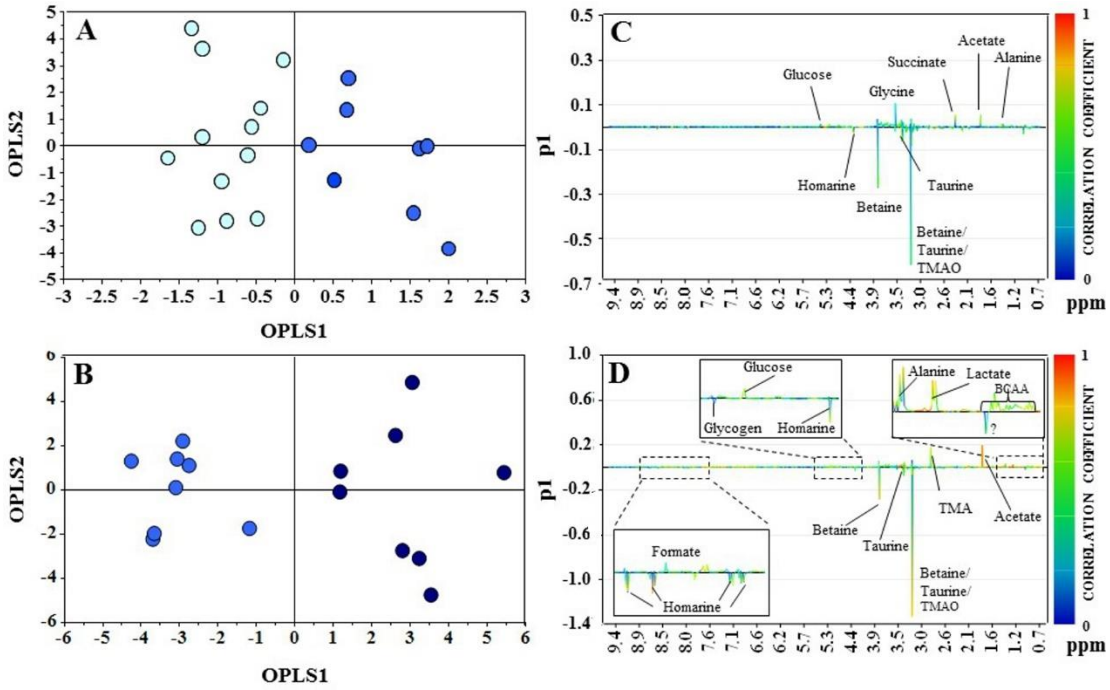



**Figure 3.** OPLS-DA cross-validated scores plots (A-B) and correlation coefficient loadings plots (C-D) from the $^{1}$H-NMR spectra of the water-soluble component of mussels stored at 4 °C: control (light blue); 2 days of storage (blue); 6 days of storage (dark blue). The quality of fit and predictability for these models are reported in Table 2 of Aru et al., 2016. For a statistically significant biomarker, cutoff values for the covariance of p≥|0.1| and for the correlation of p(corr)≥ |0.5| were used.

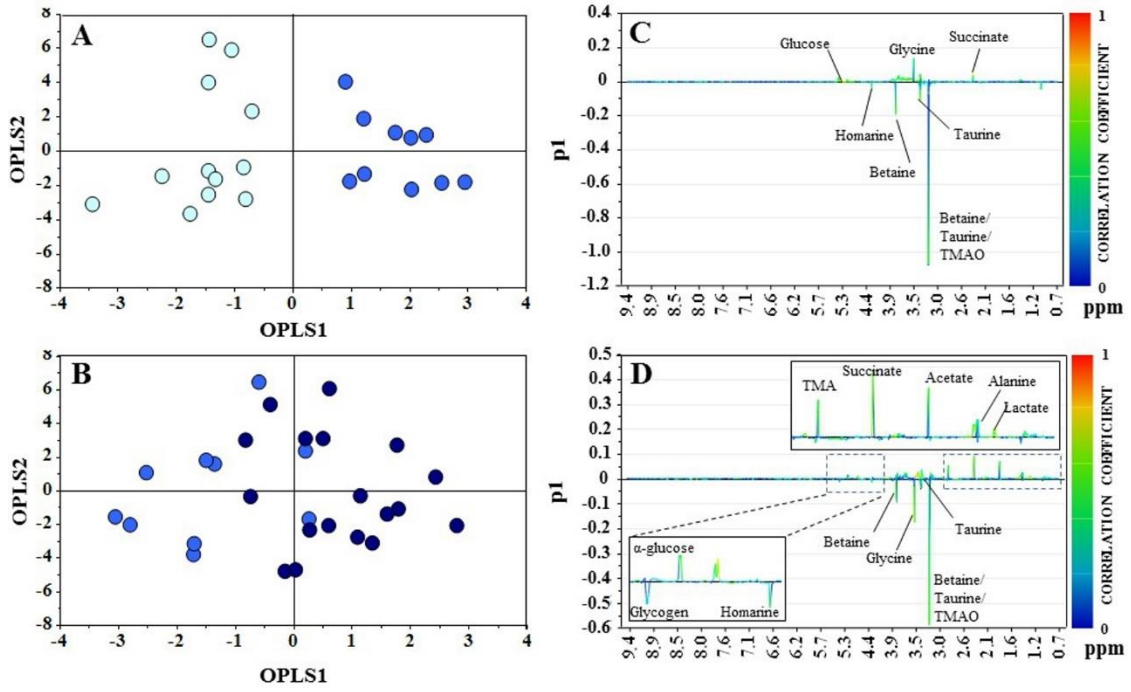



**Figure 4.** OPLS-DA cross-validated scores plots (A-B) and correlation coefficient loadings plots (C-D) from the $^{1}$H-NMR spectra of the water-soluble component of mussels stored at 0 °C: control (light blue); 2 days of storage (blue); 6-10 days of storage (dark blue). The quality of fit and predictability for these models are reported in Table 2 of Aru et al., 2016. For a statistically significant biomarker, cutoff values for the covariance of p≥|0.1| and for the correlation of p(corr)≥ |0.5| were used

# *Data on the changes of the mussels' metabolic profile under different cold storage conditions*

[2]

Violetta Aru[1], Maria Barbara Pisano[2], Francesco Savorani[3, 4,] Søren Balling Engelsen[3], Sofia Cosentino[2] and Flaminia Cesare Marincola[1]*

[1] Department of Chemical and Geological Sciences, University of Cagliari, S.S. 554 Bivio Sestu, 09042, Monserrato (Italy)

[2] Department of Public Health, Clinical and Molecular Medicine, University of Cagliari, Italy, S.S. 554 Bivio Sestu, 09042, Monserrato (Italy).

[3] Spectroscopy & Chemometrics, Department of Food Science, University of Copenhagen, Rolighedsvej 26, 1958 Frederiksberg C, Denmark

[4] Department of Applied Science and Technology (DISAT), Polytechnic University of Turin – Corso Duca degli Abruzzi 24, 10129 Torino (TO), Italy.

**Abstract**

One of the main problems of seafood marketing is the ease with which fish and shellfish undergo deterioration after death. $^{1}$H NMR spectroscopy and microbiological analysis were applied to get deeper insights into the effects of cold storage (4 °C and 0 °C) on the spoilage of the mussel *Mytilus galloprovincialis*. This data article provides information on the average distribution of the microbial loads in mussels' specimens and on the acquisition, processing, and multivariate analysis of the $^{1}$H NMR spectra from the hydrosoluble phase of stored mussels. This data article is referred to the research article entitled "Metabolomics analysis of shucked mussels' freshness" [1].

[2] Food Chemistry, Submitted

**Specifications Table**

| | |
|---|---|
| Subject area | *Chemistry, microbiology* |
| More specific subject area | *NMR-based Metabolomics* |
| Type of data | *Tables; Figures* |
| How data was acquired | *$^{1}$H NMR (Varian Unity Inova 500 spectrometer, Agilent Technologies, CA, USA);*<br>*MestReNova (Version 8.1, Mestrelab Research SL);*<br>*SIMCA 13 software package (Umetrics, Umeå, Sweden* |
| Data format | *Analysed data* |
| Experimental factors | *Mussels were stored at different cold temperatures (0 °C and 4 °C) and sampled at different storage days* |
| Experimental features | *Integrated 1H NMR-based metabolomics and microbiological analysis* |
| Data source location | *Cagliari, Sardinia (Italy)* |
| Data accessibility | *Data available within this article* |

**Value of the data**

- The multivariate analysis of the NMR data highlight significant differences between the metabolic profile of fresh mussels (Mytilus galloprovincialis) and those stored at 4 °C and 0 °C.
- The microbiological analysis provides indication of the average distribution of the microbial species isolated from mussels' samples.
- These data would serve as an important reference for developing targeted analysis for applications in food research.

**Data**

Datasets provided in this article represent the results of a combined microbiological and NMR-based metabolomics investigation of mussels under different cold storage conditions (0 °C, 4 °C and different storage days). With respect to the analysis reported in the published manuscript [1], here we provide additional information on the mussels' microbiological characterization and on the Principal Component Analysis (PCA) and Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) performed.

**1. Experimental Design, Materials and Methods**

1.1 Sample collection and experimental design

Fresh live mussels (*Mytilus galloprovincialis*) were bought from a local seafood market in Cagliari (Italy) and immediately transported to the laboratory in portable coolers at approximately 4°C. Mussels were subsequently inspected and dead ones or those with broken shells were discarded. The remaining mussels (100 individuals) were manually shucked with a sterile knife and each sample was put into insulated sterile plastic boxes without ice or water. Mussels were stored at 4 °C and 0 °C for 6 and 10 days, respectively. NMR and microbiological analysis were performed on fresh mussels (0 day) and after 2 and 6 days of storage at 4 °C and after 2, 6, and 10 days of storage at 0 °C (Fig. 1).

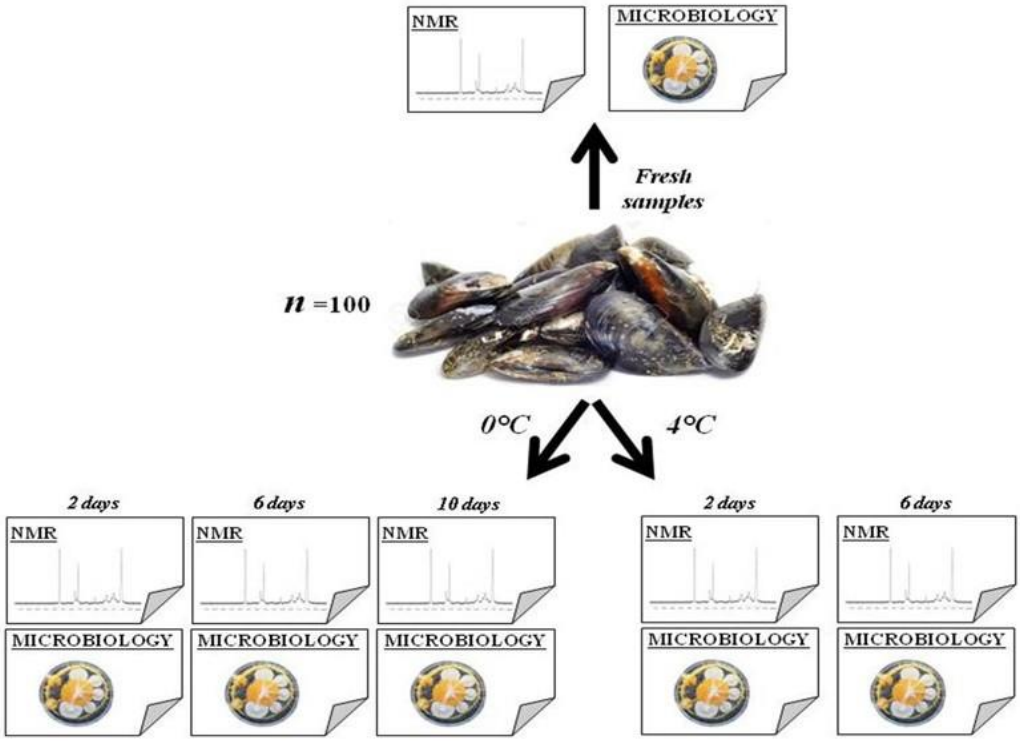



**Figure 1.** Experimental design for microbiological and NMR analysis on fresh and cold stored mussels.

1.2 Microbiological analysis

Microbiological analysis was performed as previously described [1]. Total Counts of Mesophilic (MMC) and Psychrotrophic (PMC) microorganisms were determined by the pour plate method, using Plate Count Agar (PCA, Microbiol, Cagliari, Italy) and incubating at 30 °C for 48 h and at 4 °C for 7 days, respectively. Two-way ANOVA was performed on microbiological data, with temperature and time as factors, using GraphPad Prism Statistics software package version 3.00 (GraphPad Prism Software Inc., San Diego, CA, USA). Statistical significance was inferred at $p < 0.01$. The average distribution of the microbial species isolates from mussels' samples is shown in Fig. 2.

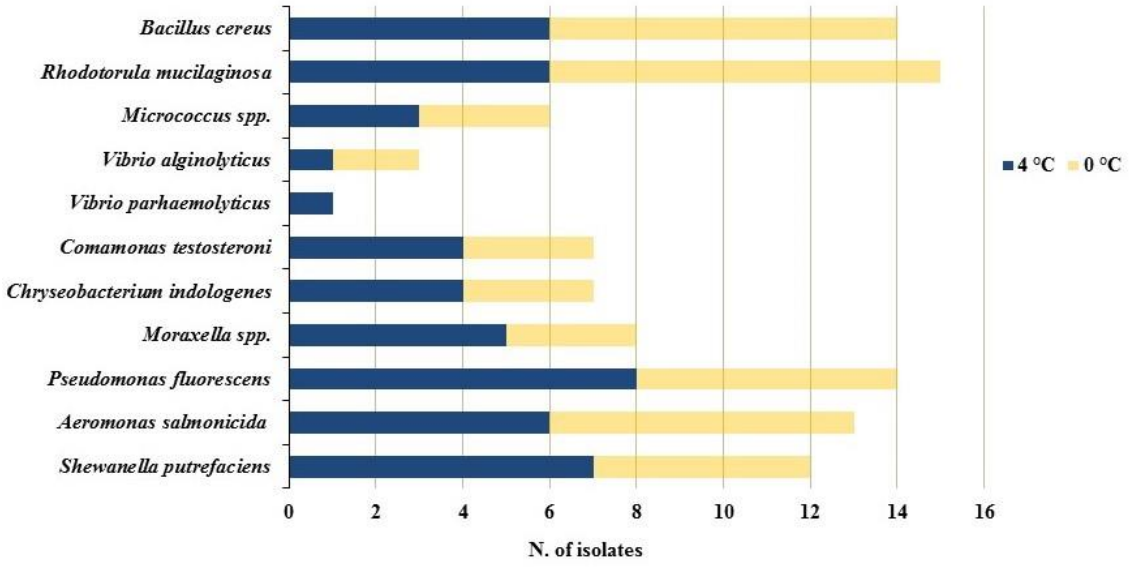



**Figure 2.** Average distribution of the microbial species isolated from mussels' samples (*Mytilus galloprovincialis*) stored at 4°C and 0 °C.

1.3 Sample preparation and NMR analysis

Water-soluble metabolites were extracted according to the Folch method [2]. $^{1}$H-NMR experiments were performed at 300 K on a Varian Unity Inova 500 spectrometer (Agilent Technologies, CA, USA) operating at the frequency of 499.84 MHz. One-dimensional (1D) $^{1}$H NMR spectra were obtained by applying a presaturation technique with low power radiofrequency irradiation for 1.5 s to suppress solvent (water) residual signal. For each spectrum, a total of 256 scans were collected into 64k points over a spectral width of 6000 Hz, with a 45° pulse, an acquisition time of 1.5 s, and a relaxation delay of 4 s. After Fourier transformation with 0.3 Hz line broadening, spectra were phased and baseline corrected, and the chemical shift scale was set by assigning a value of $\delta = 0.00$ ppm to the signal for the internal standard sodium 3-trimethylsilyl-propionate-2,2,3,3,-d4 (TSP). NMR spectra displayed several hundred peaks that arise from the different functional groups of a large number of metabolites including amino acids, organic acids, organic osmolytes, and carbohydrates. A total of 29 metabolites were identified whose assignments, multiplicity and chemical shifts are reported in Table 1.

**Table 1.** List of the metabolites identified in the $^{1}$H NMR spectra of mussels' hydrosoluble extract.

| Reference | Metabolites | Multiplicity [a] | ppm |
|---|---|---|---|
| **1** | Isoleucine | *t* | 0.93 |
| | | *d* | 1.00 |
| **1** | Leucine | *t* | 0.96 |
| **1** | Valine | *d* | 0.98 |
| | | *d* | 1.03 |
| **2** | Unknown | *s* | 1.10 |
| **3** | Lactate | *d* | 1.31 |
| **4** | Alanine | *d* | 1.48 |
| **5** | Unknown | *d* | 1.50 |
| **6** | Arginine | *m* | 1.65 |
| | | *m* | 1.72 |
| **7** | Acetate | *s* | 1.91 |
| **8** | Methionine | *m* | 2.16 |
| **9** | Acetoacetate | *s* | 2.24 |
| **10** | Glutamate | *dd* | 2.35 |
| **11** | Succinate | *s* | 2.40 |
| **12** | β-alanine | *t* | 2.55 |
| **13** | Hypotaurine | *t* | 2.64 |
| **14** | TMA | *s* | 2.90 |
| **15** | Betaine (s), Taurine (t), TMAO (s) | - | 3.26 |
| **16** | Taurine | *t* | 3.41 |
| **17** | Glycine | *s* | 3.54 |
| **18** | Betaine | *s* | 3.91 |
| **19** | Homarine | *s* | 4.35 |
| **20** | α-Glucose | *d* | 5.24 |
| **21** | Maltose | *d* | 5.40 |
| **22** | Glycogen | *bb* | 5.41 |
| **23** | Uracil | *d* | 5.80 |
| **24** | Inosine | *d* | 6.11 |
| **25** | Fumarate | *s* | 6.50 |
| **26** | Tyrosine | *d* | 6.91 |
| | | *d* | 7.20 |
| **27** | Histidine | *s* | 7.18 |
| | | *s* | 8.08 |
| **28** | Phenylalanine | *d* | 7.31 |
| | | *m* | 7.36 |
| | | *m* | 7.42 |
| **29** | Formate | *s* | 8.45 |

[a] *s*: singlet, *d*: doublet, *t*: triplet, *dd*: doublet of doublets, *m*: multiplet, *bb*: broad band.

2.4 Chemometrics analysis

NMR spectra were manually phased and baseline corrected using MestReNova (Version 8.1, Mestrelab Research SL). Each NMR spectrum was integrated between 0.5 and 9.5 ppm over a series of 0.005 ppm integral regions (bins). The regions between 4.6 and 5.0 ppm and 0.5 and 0.5 ppm were excluded because of the signals of water and TSP, respectively. The noisy region between 9.5 and 10.5 ppm was also removed. NMR data set, sized 62 samples and 1665 variables, was converted into an Excel file and then imported to SIMCA version 13.0 (Umetrics, Umeå, Sweden) for statistical analysis.

Principal Components Analysis (PCA) [3] of the $^{1}$H NMR spectra of the mussels' water-soluble extracts was performed to sample overview and to show trends, groupings and outliers in the data (Fig. 3).

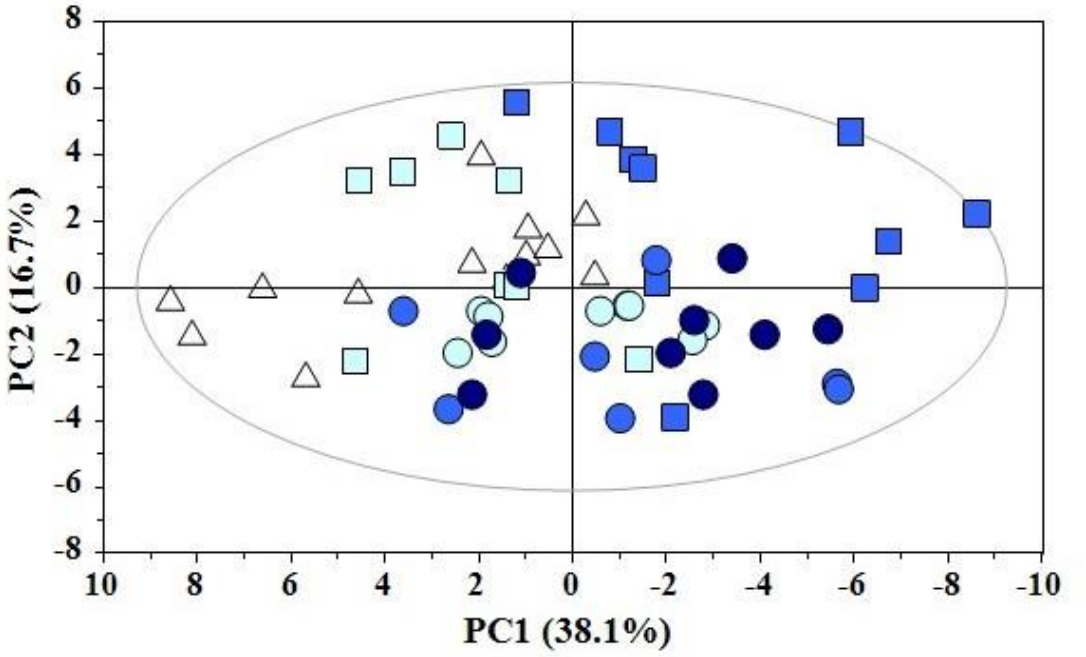



**Figure 3**. PCA scores plot of NMR spectra of mussels water-soluble extracts (R2X=0.878, Q2=0.679). Scores were labelled according to temperature (fresh samples, Δ; 0°C, ●; 4°C, ■) and colored according to storage days (2d, light blue; 6d, blue; 10d, dark blue).

Orthogonal Partial Least Squares Discriminant Analysis (OPLS-DA) [4] was applied at each temperature, in pairwise comparisons, to extract and maximize the systematic differences between experimental groups and to focus on the metabolic variations at the different time points. Model validity was assessed by using permutation tests (Y-scrambling) and CV-ANOVA [5]. The performance parameters of the models are reported in Table 2.

**Table 2.** Performance parameters of the OPLS-DA models.

| | | OPLS-DA parameters | | | | | |
|---|---|---|---|---|---|---|---|
| | | | | | *Permutation test* (n=200) | | |
| **Storage temperature (°C)** | **Comparison Group (days *vs* days)** | $R^2X$ | $R^2Y$ | $Q^2Y$ | $R^2Y$ intercept | $Q^2Y$ intercept | [a] *p-value* |
| 0 | 0 *vs* 2 | 0.875 | 0.988 | 0.833 | 0.294 | -0.284 | 0.0047 |
| | 2 *vs* 6/10 | 0.715 | 0.743 | 0.512 | 0.305 | -0.185 | 0.011 |
| 4 | 0 *vs* 2 | 0.639 | 0.917 | 0.648 | 0.410 | -0.2 | 0.017 |
| | 2 *vs* 6 | 0.618 | 0.938 | 0.864 | 0.401 | -0.294 | 0.00001 |

The OPLS discriminant analysis allowed the identification of several putative biomarkers of mussels' freshness. The identified discriminant metabolites were quantified from the $^1$H-NMR spectra and their relative amount is reported in Table 3

**Table 3.** Relative amounts of discriminant metabolites. Values are reported as mean ± SD.

| | | ***Metabolites*** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **days** | **Ace** | **Ala** | **BCAA** | **Bet** | **Glc** | **Hom** | **Lac** | **Suc** | **TMA** | **Tau** |
| | / | 0.30±0.12 | 1.35±0.71 | 1.30±0.41 | 1.59±0.71 | 1.01±0.49 | 1.69±0.50 | 0.50±0.20 | 1.12±0.43 | 0.05±0.01 | 1.64±0.83 |
| **4°C** | **2** | 0.73±0.22** | 1.45±0.32* | 1.37±0.28** | 1.59±0.30* | 1.18±0.36** | 1.50±0.51 | 0.91±0.34** | 1.39±0.50* | 0.14±0.06** | 1.32±0.57 |
| | **6** | 5.40±0.82** | 2.11±0.74* | 2.00±0.70** | 1.35±0.42 | 1.78±0.32** | 1.39±0.47 | 4.01±0.79** | 1.89±0.72* | 7.29±0.90** | 1.24±0.22 |
| **0°C** | **2** | 0.35±0.15** | 1.37±0.66 | 1.30±0.43 | 1.49±0.41 | 1.12±0.29 | 1.48±0.40** | 0.91±0.30* | 1.58±0.37* | 0.14±0.03** | 1.49±0.30 |
| | **6** | 0.53±0.15** | 1.38±0.38 | 1.38±0.35 | 1.39±0.48* | 1.24±0.53 | 1.41±0.36** | 0.10±0.39* | 1.99±0.75* | 1.07±0.29** | 1.39±0.77 |
| | **10** | 1.03±0.38** | 1.82±0.21 | 1.40±0.45 | 0.96±0.24* | 1.27±0.36 | 1.00±0.38** | 1.13±0.79* | 2.03±0.28* | 1.20±0.47** | 0.99±0.27 |

Keys: Ace=acetate; Ala=alanine; BCAA=branched chain amino acids; Bet=betaine; Glc=glucose; Hom=homarine; Lac=lactate; Suc=succinate; TMA=trimethylamine; Tau=taurine.
*p≤0.05
**p≤0.01

**Acknowledgements**
Violetta Aru gratefully acknowledges Sardinia Regional Government for the financial support of her PhD scholarship (P.O.R. Sardegna F.S.E. Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2007-2013 - Axis IV Human Resources, Objective l.3, Line of Activity l.3.1.).

**References**


[1] V. Aru, M. B. Pisano, F. Savorani, S. B. Engelsen, S. Cosentino, F. Cesare Marincola, Metabolomics analysis of shucked mussels' freshness, Food Chem. 205 (2016) 58–65.

[2] J. Folch, M. Less, G. H. S. Stanley, A simple method for the isolation and purification of total lipides from animal tissues, J. Biol. Chem. 226 (1957) 497-509.

[3] S. Wold, K. Esbensen, P. Geladi, Principal component analysis, Chemometr. Intell. Lab. Syst. 2 (1987) 37-52.

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

[5] L. Eriksson, J. Trygg, S. Wold, CV-ANOVA for significance testing of PLS and OPLS1 models, J. Chemometr. 22 (2008) 594–660.

# Paper III

---

## Metabolic responses of clams, *Ruditapes decussatus* and *Ruditapes philippinarum*, to short-term exposure to lead and zinc

Violetta Aru[1], Giorgia Sarais[2], Francesco Savorani[3,4], Søren Balling Engelsen[3], and Flaminia Cesare Marincola[1*]

[1]Departement of Chemical and Geological Sciences – University of Cagliari, S.S. 554 Bivio Sestu, 09042 Monserrato, Italy

[2]Departement of Life and Environmental Sciences, University of Cagliari, via Ospedale 72, 09124 Cagliari, Italy

[3]Quality & Technology, Department of Food Science, University of Copenhagen, Rolighedsvej 26, 1958 Frederiksberg C, Denmark

[4]Department of Applied Science and Technology (DISAT), Polytechnic University of Turin – Corso Duca degli Abruzzi 24, 10129 Torino (TO), Italy

**Marine Pollution Bulletin, Submitted**

---

**Abstract**

This study investigates the effects of 48 h heavy metal exposure upon the metabolic profiles of *Ruditapes decussatus* and *Ruditapes philippinarum* using $^{1}$H NMR metabolomics. Both species were exposed to increasing concentrations of lead nitrate (10, 40, 60, and 100 µg/L) and zinc chloride (20, 50, 100, and 150 µg/L), under laboratory conditions. ICP-OES analysis was further performed on the clams' samples in order to verify the occurrence of heavy metal bioaccumulation. With respect to the controls, the metabolic profiles of treated *R. decussatus* exhibited higher levels of organic osmolytes and lower contents of free amino acids. An opposite behavior was shown by *R. philippinarum*. In terms of heavy metal, the exposure effects were more evident in the case of Pb rather than Zn. These findings show that NMR metabolomics has the required sensitivity and specificity for the identification of metabolites that can act as sensitive indicators of contaminant-induced stress.

**Keywords**

NMR metabolomics, clams, bioindicators, pollution, lead, zinc

**Highlights**

- Clams exposure to Pb and Zn was investigated by metabolomics.
- The metabolic profile of clams was analyzed by NMR spectroscopy.
- Metabolite-metabolite and metal-metabolite correlations were investigated.
- Results confirm metabolomics as a powerful tool to study contaminant-induced stress.

## 1. Introduction

Over the past few decades, an increased worldwide attention has been paid to the issue of heavy metals contamination in marine environments as the result of anthropogenic activity (Rainbow, 1995). Trace amounts of metals have been demonstrated to affect, at different levels, marine food chains causing adverse effects on aquatic organisms (Wang and Rainbow, 2005). Indeed, heavy metals such as arsenic (As), mercury (Hg), cadmium (Cd), chromium (Cr), lead (Pb) and zinc (Zn) are not easily expelled from marine aqua systems and animals and thus tend to be accumulated in the environment and in the tissues of aquatic organisms (Jakimska et al., 2011). The accumulation of such pollutants in aquatic species leads to concentrations that are several orders of magnitude higher than those of the average water (Casas, 2008), thus becoming of concern to public health when accumulated in edible animals.

In order to evaluate the impact of pollutants on the environmental quality, it has become highly pertinent to carry out a rapid assessment of their deleterious effects on the ecosystem. Mollusks and, in particular, bivalves are among the most commonly-used sentinel organisms in pollution studies due to their sedentary nature, their wide-spread geographical distribution, their great capacity for accumulating contaminants and their easy sampling (Stankovic et al., 2014). Bivalve biological response (biomarkers) has been proposed as a sensitive tool to assess the environmental quality of coastal areas (Viarengo and Nott, 1993). In this context, molecular biomarkers are defined as early-warning signals that reflect alterations occurring in cells, tissues or body fluids, and which can be related to exposure to or toxic effects of environmental pollutants (Kaviraj et al., 2014). Therefore, biomarkers play a valuable role in assessing whether or not the organism has been exposed to pollutants, and they have the potential to provide an early diagnosis of disorders caused

by anthropogenic contaminants (Lyons et al., 2010). The “biomarker approach” has been incorporated into several pollution monitoring programs in Europe and the USA (OSPAR, 2012; ICES, 2013).

The search of candidate biomarkers as fast and sensitive indicators of contaminant-induced stress within organism has gained much attention over the last decades (Bebianno et al., 2004). Among the analytical approaches used to this aim, the “omics” techniques offer added value compared with the classical ones, providing, at different levels, information on the molecular basis of exposure. In particular, metabolomics provides a more immediate physiological measure of organisms than other omics approaches (Weckwerth, 2003). It can be defined as the comprehensive study of the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological spectroscopic data (Nicholson et al., 1999). By the overall characterization of the low molecular-weight metabolites (<1500 Da), regarded as metabolic fingerprints, metabolomics aims at studying the regulation of the metabolic pathways and metabolic networks of a biological system (Dunn et al., 2005). The metabolomics studies on marine organisms cover a wide range of application areas. With regard to pollutant exposure, a special interest has been turned on bivalve mollusks to detect and elucidate mechanisms of environmental stress (Campillo et al., 2015; Chiarelli and Roccheri, 2014; Fasulo et al., 2012; Ji et al., 2015a; Jin and Wang, 2015).

The grooved carpet shell *Ruditapes decussatus* and the Manila clam *Ruditapes philippinarum* are phenotypically similar species belonging to the Veneridae family, widely reared along the Mediterranean Sea coast. In spite of their phylogenetic similarities, they have been demonstrated to accumulate heavy metals differently (Figueira and Freitas, 2013), suggesting also the occurrence of different metabolic responses to metal exposure.

In this study, $^{1}H$ NMR metabolomics is used to investigate and compare the biochemical responses of *R. decussatus* and *R. philippinarum* to 48 hours lead nitrate ($Pb(NO_3)_2$) and zinc chloride ($ZnCl_2$) exposure. $^{1}H$-NMR spectra of the aqueous extracts of clams were analyzed by Principal Component Analysis (PCA) to search for valid metabolic signatures for metal exposure. A comprehensive overview on the initial pollution state of the samples and the occurrence of bioaccumulation was achieved by complementing NMR data with a direct heavy metal detection in clams by Inductively Coupled Plasma Optical Emission Spectrometry (ICP-OES).

## 2. Materials and methods

### 2.1 Chemicals

Deuterium oxide ($D_2O$, 99.9 %), sodium deuteroxide (NaOD, 40 wt. % in $D_2O$, 99.5 atom % D), deuterium chloride (DCl, 99 atom % D) and sodium 3-trimethylsilyl-propionate-2,2,3,3-d4 (TSP) were purchased from Sigma–Aldrich (Milan, Italy).

Lead nitrate ($Pb(NO_3)_2$) and zinc chloride ($ZnCl_2$) were used as a source of $Pb^{2+}$ and $Zn^{2+}$ for laboratory exposure experiments. The salt aqueous solutions were prepared from analytical-grades $Pb(NO_3)_2$ and $ZnCl_2$ standards (Sigma-Aldrich, Milan, Italy). The concentrations of standard stock solutions were 130 and 230 mg/L, respectively.

High-purity water, with electrical resistivity of 18 MΩ cm, was produced by a Milli-Q system (Millipore, Bedford, MA). Nitric acid at 65% (Carlo Erba, Milan, Italy), hydrochloric acid at 37% (Carlo Erba, Milan, Italy) and hydrogen peroxide at 30 wt. % in $H_2O$ (Aldrich, Milan, Italy) were of reagent-grade. Stock solutions (1mg/ml) of Pb, and Zn were RS-Standard solution for AAS (Analyticals Carlo Erba, Milan, Italy). Mix working standard solutions for ICP-OES calibration were prepared diluting opportunely stock

solutions using high-purity water as diluent. All glass and plastic ware were soaked overnight in 5% v/v nitric acid, and rinsed three times with deionized water before use.

2.2 Experimental protocol

A total of four hundred individuals of adult *Ruditapes decussatus* and *Ruditapes philippinarum* (shell length: 32–37 mm) were purchased from a local seafood market in Cagliari (Italy) and immediately transported to the laboratory. For each metal exposure experiment, almost one hundred clams were randomly divided into five tanks: one for controls and four for heavy metals exposure experiments. A schematic representation of the experimental design is reported in Figure 1. Briefly, individuals were allowed to acclimatize in aerated seawater (25°C and 33 psu) for 1 week and daily fed with *Chlorella vulgaris Beij* at a ratio of 2% tissue/dry weight. After the acclimatization period, samples in 4 tanks were exposed to dissolved lead nitrate (nominal concentrations: 10, 40, 60 and 100 µg/L) and zinc chloride (nominal concentrations: 20, 50, 100 and 150 µg/L) for 48 h with continuous suppling of air and food. After 48 hours of exposure, all clams were immediately shucked, deprived of the digestive gland, and stored at -80°C before metabolite extraction.

2.3 ICP-OES analysis

2.3.1 Microwave-assisted acid decomposition (or digestion procedure)

In order to dissolve clams samples for elemental analysis, microwave-assisted acid decomposition was performed at high pressure and temperature according to ICRAM method (ICRAM, 2003). Approximately 0.3 g of dried sample, weighed accurately, was transferred into microwave Teflon vessels; 3 ml of a mixture of nitric acid (65%) and

hydrochloric acid (37%) (1:3, v/v) and 1 ml of hydrogen peroxide at 30% were added. All samples were placed in the microwave carousel together with a blank prepared with a solvent of analytical-reagent grade. A CEM model MarsX Microwave-Assisted system was used. The microwave operation parameters were: power 600 W (50%) for 3 min, 600 W (60%) for 2 min, 600 W (70%) for 3 min, and 600 W (80%) for 10 min. After cooling, the residues were taken up and brought to a final volume of 25 ml with bi-distilled water. Sample digests were then filtered with a 25mm syringe filter RC, pore size 0, 45 µm. Quantitative determinations were made by the external standard method. The digests and the blank were analyzed with an ICP-OES spectrometer.

2.3.2 Apparatus

A simultaneous atomic emission spectrometer ICP-OES Varian Vista-MPX (Varian Inc., Palo Alto, CA) with axial viewed plasma, was used. The plasma was operated with a concentric nebulizer coupled to a cyclone type nebulization chamber. The operating conditions were set as follows: power 1.2 kW; plasma gas flow 15 l/min; auxiliary gas flow 1.5 l/min; nebulizer gas flow 0.9 l/min. The wavelengths used for the quantification were: Pb 220.353 nm, and Zn 206.200 nm. The sensibility of the method was evaluated by determining the limit of detection (LOD) and limit of quantification (LOQ) for each element. The LOD is the smallest quantity of analyte that can be shown to be significantly greater than the measurement (random) error of the blank at the prescribed level of confidence (usually 95%). It is the lowest concentration of analyte in a sample that can be detected, but not necessarily quantified, and it was calculated as 3 times the SD of the background noise. The LOQ is the smallest amount of analyte in a test sample that can be quantitatively determined with suitable precision and accuracy under previously

established methods and conditions. It was taken as 10 times the SD of the background noise (Magnusson, 2014).

## 2.4 $^{1}$H NMR spectroscopy

### 2.4.1 Sample preparation

A modified Folch (Folch et al., 1957) method was applied to extract polar metabolites from clams. Each clam sample was dissolved in 12 ml of a mixture chloroform–methanol (2:1, v/v). After the addition of 4 ml of $H_2O$ and centrifugation at 1700 g for 1 h at 4 °C, the methanol/water mixture was separated from the chloroform fraction. The $CH_3OH/H_2O$ phase was dried using a rotating evaporator (Eppendorf Concentrator Plus). The freeze-dried sample was dissolved in 1.2 ml of a $D_2O$ solution of the internal standard (TSP) 0.80 mM and centrifuged at 13000 g for 5 min at 4 °C to remove particulate matter. The pH of the final sample was accurately adjusted to 6.52 ± 0.03 by adding a small amount of NaOD or DCl to minimize pH-based peak misalignments. Then, an aliquot of 650 µl was placed into a 5 mm NMR tube for NMR analysis.

### 2.4.2 NMR measurements

$^{1}$H-NMR spectra were recorded at 300 K using a Varian Unity Inova 500 MHz NMR spectrometer (Agilent Technologies, CA, USA) operating at a proton Larmor's frequency of 499.839 MHz. NMR spectra were acquired with a sweep width of 6000 Hz, a 45° pulse, an acquisition time of 1.5 s, a relaxation delay of 4 s and 256 scans. The residual signal of $H_2O$ was suppressed by applying a presaturation technique with low power radiofrequency irradiation for 1.5 s. After Fourier transformation with 0.3 Hz line broadening and a zero-filling to 64K, $^{1}$H-NMR spectra were manually phased and baseline corrected using

MestReNova (Version 8.1, Mestrelab Research SL). Spectral chemical shift referencing on the TSP $CH_3$ signal at 0.00 ppm was performed on all spectra.

2.5 Chemometric analysis of $^1$H NMR data

$^1$H-NMR spectra were converted to ASCII files and then imported into Matlab software (Mathworks, MA) and a data matrix sized 300 × 65536 (samples × variables) was built. Spectra were aligned using the *i*coshift algorithm, which is based on correlation optimized shifting of spectral intervals and aligns all spectra simultaneously (Savorani et al., 2010). The regions between 4.6 and 5.2 ppm and 0.5 and -0.5 ppm were excluded for the following multivariate data analysis because of the signals of (residual) water and TSP, respectively. The noisy region between 9 and 11 ppm was also excluded making the final data matrix sized 300 × 39891. NMR spectra were normalized using total area normalization. This normalization facilitates a direct comparison among the samples and minimizes the effects of small variations in the dilution of the samples (Craig et al., 2006). Data sets were then Pareto scaled prior to chemometric analysis. This scaling pre-treatment procedure is commonly used for the multivariate analysis of NMR data: it allows the reduction of the influence of intense signals while emphasizing weaker ones that may have more biological relevance (Worley and Powers, 2012). Data set were then imported in LatentiX 2.12 (Latent5, Copenhagen, Denmark) and analyzed by Principal Component Analysis (PCA). PCA is an unsupervised chemometric technique that allows the reduction of the dimensionality of a data set consisting of a large number of interrelated variables, while retaining the variation present in the data set as much as possible (Bro and Smilde, 2014). This is achieved by transforming the original variables into a smaller set of new variables, the principal components (PCs), which are uncorrelated and ordered so that the

first one retains most of the systematic variation present in all of the original variables: in this way the first component explains the most variation, the second component the second most, etc. Commonly, most of the relevant systematic information is contained in the first few PCs, while the following ones are modelled just by chance variation and noise. Samples are assigned scores that correspond to the variation along the principal components and those that have similar scores will cluster together. The loadings are about the relation between the original variables and define the direction of the principal components. The results of this exploratory tool can be displayed via scores plots and loadings plots. The biplot is a two-dimension data visualization method that overlays samples (scores) and variables (loadings) of a data matrix, highlighting the relations existing within samples, within variables and between samples and variables. It provides a useful tool to capture the most relevant features in the multivariate data set (clustering and correlations among variables) (Gabriel, 1971). Although the biplot can be an effective visualization tool for small systems with few samples and variables, its interpretation can be very complex for large megavariate data sets such as the ones coming from $^{1}$H NMR spectroscopy. In order to reduce model complexity, the PCA was performed only on the quantitative relative amounts of 21 metabolites. Baseline-resolved *signals* of these metabolites were accurately integrated from the NMR spectra and their relative amounts were used as variables. Finally, a metabolic correlation analysis was conducted in order to highlight the interactions among metabolites, and the relationship between metabolites and the added pollutant. Pearson's correlation analysis was further performed on those variables, and between each metabolite and the metal treatment. 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 around 0 (zero) indicates no relationship between two variables.

## 3. Results

### 3.1 ICP-OES

A total of five samples were collected from each tank and analyzed by ICP-OES to measure Pb and Zn levels. However, Pb could not be detected in any clam before or after metal exposure, and no significant changes in Zn levels could be observed between control and exposed specimens. These findings showed the lack of metal exposure effects on the heavy metal accumulation by clams, under the experimental conditions. A possible explanation of this result could be the short time of exposure which was not long enough to provide a significant bioaccumulation above the detection limits. Additionally, it is not excluded that possible mechanisms of heavy metal release from these organisms may contribute to the present results (Rainbow, 2002).

### 3.3 Chemometric and statistical analysis of NMR data

A total of 300 hydrophilic extracts of clams were analyzed by $^{1}$H-NMR spectroscopy. Representative spectra of *R. decussatus* and *R. philippinarum* are reported in Figures 2A and 2B, respectively. The spectra display several hundred peaks deriving from a large number of metabolites including amino acids, organic acids, organic osmolytes, and carbohydrates whose assignments, multiplicity and chemical shifts are reported in Table 1.

Firstly, an exploratory PCA was performed on the dataset consisting of the complete NMR spectra. Figure 3A shows the PCA scores plot, while Figures 3B-C display the corresponding loadings plots with the loadings contributions on PC1 and PC2, respectively.

As can be observed, the two species are clearly separated along PC1. The analysis of the PC1 loadings plot (Fig. 3B) revealed that *R. decussatus* spectra were mainly characterized by an abundance of organic osmolytes (hypotaurine, betaine and homarine), as well as of carbohydrates such as glucose, maltose, and glycogen. In contrast, *R. philippinarum* spectra exhibited a higher amount of amino acids (BCAA, alanine, glycine, tyrosine and phenylalanine and DMG), organic acids (acetic, succinic, malonic and formic acids) and the organic osmolyte taurine.

The effect of metal exposure was found to be smaller than the species variation but the score distribution along PC2 was visibly metal-sensitive (Fig. 3A). Indeed, lead-treated clams were mainly distributed in the upper side of the plot (along positive PC2 values), while zinc-treated samples were mainly located in the lower side of the plot (along negative PC2 values). The analysis of the loadings plot of PC2 (Fig. 3C) revealed that the metabolic changes that mostly characterized lead nitrate exposure were increased levels of amino acids such as glycine, tyrosine and phenylalanine, the monosaccharide glucose, the disaccharide maltose and the polysaccharide glycogen, as well as acetate, DMG, hypotaurine and ATP. In contrast, the spectra of zinc-treated clams showed metabolic profiles characterized by an increase in betaine, taurine, alanine, succinic and formic acid.

In order to obtain deeper insights on the biochemical consequences occurring after metal exposure, a PCA analysis was performed separately for each clam species by comparing controls with samples exposed at each nominal metal concentration. Within each series of experiments (species), the PCA models revealed that the metabolites affected by metal exposure were basically the same at all pollution levels (data not shown). Figure 4 shows the biplots that compare the controls and the samples exposed to the highest metal concentration. Only metabolites that contribute strongly to sample clustering are shown in

the biplots. Figure 4A shows the PCA models built for lead-treated *R. decussatus.* When compared to the non-treated controls, the lead-treated clams' metabolic profiles displayed higher levels of organic osmolytes (betaine and taurine), formate and aspartate and lower levels of glucose and free amino acids such as BCAA, alanine, phenylalanine, tyrosine and threonine. Compared to control, the zinc-treated *R. decussatus* displayed higher levels of homarine, hypotaurine, phosphocholine and aspartate (Fig. 4B) and lower levels of free amino acids such as BCAA, phenylalanine and tyrosine.

For the *R. philippinarum* the situation is quite different. The metabolic response of lead-treated *R. philippinarum* clams (Fig. 4C) involved an increase in the levels of free amino acids (BCAA, threonine, alanine, tyrosine and phenylalanine and DMG) and acetate and a concomitant decrease in the content of glucose, aspartate, taurine, betaine, homarine and the Krebs's cycle intermediate succinate. Finally, the metabolome of the zinc-treated *R. philippinarum* (Fig. 4D) displayed a higher amount of acetate, succinate, threonine, tyrosine, phenylalanine, and DMG and corresponding lower levels of aspartate, alanine, taurine, homarine, glucose and malonate.

Further information on the effect of heavy metals exposure on clams' metabolism was achieved by the analysis of the metabolite-metabolite and metal-metabolite Pearson's correlations. Only strong correlations, which may reveal robust associations between variables, have been considered for further remarks. At a threshold of $r \geq \left| 0.8 \right|$, 98 and 91 positive correlations and 104 and 103 negative correlations were found in Pb-treated *R. decussatus* and *R. philippinarum*, respectively. A much lower number of highly correlated metabolite correlations were encountered in Zn-treated *R. decussatus* and *R. philippinarum*, 54 and 51 positive correlations and 42 and 39 negative correlations, respectively. As far as the Pb-treatment is concerned, globally, the highest numbers of correlations were either

amongst amino acids or between amino acids and osmolytes. Furthermore, a high degree of strong positive and negative metal-metabolite correlations was evidenced. Differently, Zn-treated clams' metabolome exhibited a lower amount of both metabolite-metabolite and metal-metabolite correlations.

## 4. Discussion

Most of coastal marine ecosystems, including lagoons, coastal lakes and estuaries, are regularly affected by anthropogenic pollution deriving from industrial activity and sewage discharges that undermine the ecological quality of the environment. Due to their persistent nature, trace metals have a noticeable deleterious effect on biota, thus being considered to be the most toxic inorganic environmental pollutants (Stankovic et al., 2014). Indeed, they can be accumulated in the trophic chain at different levels, becoming a concern for ecosystems and humans' health, especially if the accumulation takes place in edible seafood.

Due to the capacity of accumulating chemical contaminants well above environmental levels, bivalve mollusks have gained a global importance as bioindicators of marine and estuarine pollution (Gupta and Singh, 2011). Toxicity, kinetics of bioaccumulation and elimination, and biomarker responses of these marine organisms to metal exposure are well documented (Doyen et al., 2015; Ji et al., 2015b; Velez et al., 2015). In this work, the metabolic responses of the clams, *R. decussatus* and *R. philippinarum*, to short-term lead and zinc exposure have been investigated by using NMR-based metabolomics. The significance of testing the effect of lead contamination in seafood is covered by its persistent and noxious nature. Indeed, according to its remarkable neurotoxicity (Lidsky and Schneider, 2003), it is considered one of the most dangerous trace element amongst the

heavy metals. Concerning the biological role played by zinc, although this trace metal is present as cation in numerous proteins, enzymes and cofactors in several phyla, it may also cause mortality or sublethal stress, depending on the sensitivity of the organism (Solomons, 2001).

Following 48 h metal exposure, clams' metabolic responses appeared to be both species- and metal-specific. Concerning the differences observed in terms of species, the most striking ones were the modification of organic osmolytes and free amino acid contents. Specifically, the metal exposure of *R. decussatus* induced a significant increment in organic osmolytes production (betaine and taurine in lead-treated samples and hypotaurine and homarine in zinc-treated clams), while it decreased the content of free amino acids. In contrast, *R. philippinarum* metabolic response was characterized by a decrement in the organic osmolytes content and an increment in the content of the free amino acids. A similar behavior by *R. philippinarum* was observed also in the case of Cd exposure in previous metabolomics research work (Ji et al., 2015b). In particular, an increased level of branched chain amino acids (isoleucine, leucine and valine), following heavy metal exposures, has been proposed to be a biomarker of immune stress.

Organic osmolytes and 'compatible' solutes (i.e. small carbohydrates, polyols, amino acids and derivatives) are known to be accumulated in aquatic invertebrates for counteracting osmotic perturbations and maintaining cell volume (Walla and Yancey, 2001). Some of them are also involved in chemical reactions that protect cells and tissues other than osmotically (i.e. antioxidant activity of taurine and hypotaurine) (Yancey, 2005). For instance, free amino acids are also substrates for energy metabolism and protein synthesis and their levels can be altered by a variety of environmental and developmental processes, such as anaerobic energy metabolism (Powell et al., 1982), seasonal patterns

(Kube et al., 2007), reproduction (Kasschau, 1982) and pollutants exposure (Ji et al., 2015b). Due to the interplay of several concomitant acting bioecological and anthropogenic stressors (i.e. environment, pollution, physiology), the interpretation of the metabolic profiles is not straightforward. In order to minimize the interaction of these effects, factors such as temperature, salinity, dissolved oxygen, food availability and seasonality were taken, as much as possible, under control in our laboratory conditions. Owing to the precise control of all these important environmental factors, the data obtained in the present study suggest that the metal exposure had the major influence over the metabolic responses of bivalves.

Despite the difficulties of a strict comparison of our findings with those reported in the literature due to dissimilar experimental conditions (such as pollutant nature and concentration, exposure time, physiological conditions, etc.), there is a good agreement between present data and other studies (Ji et al., 2015a, 2015b; Wu et al., 2013). Indeed, the effect of contaminants on the biological response of bivalves tend to alter the levels of organic osmolytes and free amino acids, depending on the species and contaminants used, providing a reasonable indication of the occurrence of biochemical disturbance under 48 h metal exposure. It is interesting to note that no significant metal accumulation was monitored in the same specimens by using ICP-OES analysis. The comparison of this finding with the biological response of clams, as monitored by NMR spectroscopy, points out the sensitivity of the metabolomics approach in revealing short-time metal-induced stress.

Pearson's correlations analysis was also performed to investigate metabolite-metabolite and metal-metabolites correlations (heat maps). Correlation analysis is widely applied in the study of systems biology and has been demonstrated to be a valuable tool also in the

field of metabolomics where it is commonly applied to explore the relationships among metabolites (Steuer, 2006). In this study, a higher number of strong metabolite-metabolite and metal-metabolite correlations was observed upon lead exposure with respect to the zinc one. These correlations can be viewed as a direct consequence of the interplay of metabolic networks characterizing clams' biological response to metal exposure. Although a definite interpretation of these results is not yet possible, due to the complex dynamics of the biochemical networks involved in the restore of clams' homeostasis, it is reasonable to interpret the observed patterns of correlations as a "fingerprint" of the specific physiological state of clams upon metal exposure. Thus, our findings suggest a more concerted action of metabolites in the presence of lead rather than zinc salt. In particular, lead seems to exert a detrimental effect of wider extension on clams' metabolome probably due to its noxious nature.

## 5. Conclusions

Heavy metals are considered to be among the most harmful pollutants that can contaminate marine environments. Heavy metals toxicity rises from their persistent nature which leads to environmental accumulation. Given the high accumulation rates associated with their filter feeding attitude, bivalves mollusks are considered to be feasible monitoring organisms and are widely used in biomonitoring programs. *R. decussatus* and *R. philippinarum*, two bivalves' species widely distributed along the Italian coasts, were selected in the present study for assessing lead and zinc contamination effects on their metabolic profiles. The obtained data point out a sensible metabolic response of bivalves to short-time metal exposure. A major variability in the content of amino acids and organic osmolytes was observed in clams, even though to a different extent depending on the metal

nature and bivalve species. Although still much research work has to be carried out to achieve a complete characterization of the pollution-induced responses in bivalve, our findings suggest a potentially important associations between short-time metals exposure and clams' metabolism and stress the capability of metabolomics in identifying a biomarker contour (Bro et al., 2015) rather than relying on a single biomarker, to be used for marine quality control. Future detailed investigations into the biochemical role covered by these metabolites with respect to heavy metal pollution will allow elucidation of the network of key metabolic regulators in both clams' species.

**Acknowledgments**

Violetta Aru gratefully acknowledges Sardinia Regional Government for the financial support of her PhD scholarship (P.O.R. Sardegna F.S.E. Operational Programme of the Autonomous Region of Sardinia, European Social Fund 2007-2013 - Axis IV Human Resources, Objective l.3, Line of Activity l.3.1.).

## References

Bebianno, M.J., Géret, F., Hoarau, P., Serafim, M.A., Coelho, M.R., Gnassia-barelli, M., Roméo, M., 2004. Biomarkers in *Ruditapes decussatus* : a potential bioindicator species. Biomarkers 9, 305–330.

Bro, R., Kamstrup-Nielsen, M.H., Engelsen, S.B., Savorani, F., Rasmussen, M.A., Hansen, L., Olsen, A., Tjønneland, A., Dragsted, L.O., 2015. Forecasting individual breast cancer risk using plasma metabolomics and biocontours. Metabolomics 11, 1376–1380.

Bro, R., Smilde, A.K., 2014. Principal component analysis. Analytical Methods 6, 2812–2831.

Campillo, J. a, Sevilla, A., Albentosa, M., Bernal, C., Lozano, A.B., Cánovas, M., León, V.M., 2015. Metabolomic responses in caged clams, Ruditapes decussatus, exposed to agricultural and urban inputs in a Mediterranean coastal lagoon (Mar Menor, SE Spain). The Science of the total environment 524-525, 136–147.

Casas, S, Gonzalez, J.L, Andral, B., and Cossa, D.,2008. Relation between metal concentration in water and metal content of marine mussels *(Mytilus alloprovincialis):*impact of physiology. Environmental Toxicology and Chemistry, 27, 1543–1552.

Chiarelli, R., Roccheri, M.C., 2014. Marine Invertebrates as Bioindicators of Heavy Metal Pollution. Open Journal of Metal 1, 597-616.

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, 2262–2267.

Doyen, P., Morhain, E., Rodius, F., 2015. Modulation of metallothionein, pi-GST and Se-GPx mRNA expression in the freshwater bivalve *Dreissena polymorpha* transplanted into polluted areas. AIMS Environmental Science 2, 333-344.

Dunn, W.B., Bailey, J.C., Johnson, H.E., Dunn, W.B., 2005. Measuring the metabolome : current analytical technologies 130, 606–625.

Fasulo, S., Iacono, F., Cappello, T., Corsaro, C., Maisano, M., D'Agata, A., Giannetto, A., De Domenico, E., Parrino, V., Lo Paro, G., Mauceri, A., 2012. Metabolomic investigation of Mytilus galloprovincialis (Lamarck 1819) caged in aquatic environments. Ecotoxicology and Environmental Safety 84, 139–146.

Figueira, E., Freitas, R., 2013. Consumption of Ruditapes philippinarum and Ruditapes decussatus: Comparison of element accumulation and health risk. Environmental Science and Pollution Research 20, 5682-5691.

Folch, J., Lees, M., Stanley, G.H.S., 1957. A simple method for the isolation and purification of total lipids from animal tissues. The Journal of Biological Chemistry 226, 497-509.

Gabriel, K.R., 1971. The biplot-graphical display of matrices with applications to principal components analysis. Biometrika 58, 453-467.

Gupta, S.K., Singh, J., 2011. Evaluation of mollusc as sensitive indicatior of heavy metal pollution in aquatic system: A review. IIOAB Journal 2, 49-57.

Istituto Centrale per la Ricerca scientifica e tecnologica Applicata al Mare (ICRAM), 2001. Metodologie analitiche di riferimento. Lo Studio Editoriale s.r.l., Roma – Giorgio

Vitale.

Jakimska, A., Konieczka, P., Skóra, K., Namieśnik, J., 2011. Bioaccumulation of metals in tissues of marine animals , Part I : the role and impact of heavy metals on organisms. Poland Journal of Environmental Studies 20, 1127-1146.

Ji, C., Wang, Q., Wu, H., Tan, Q., Wang, W.-X., 2015a. A metabolomic investigation of the effects of metal pollution in oysters Crassostrea hongkongensis. Marine pollution bulletin 90, 317-322.

Ji, C., Wu, H., Zhou, M., Zhao, J., 2015b. Multiple biomarkers of biological effects induced by cadmium in clam Ruditapes philippinarum. Fish & shellfish immunology 44, 430-435.

Jin, Y., Wang, W.-X., 2015. Biokinetics and metallothionein-like proteins response in oysters facing metal challenges in an estuary. Environmental Toxicology and Chemistry 34, 1818–1825.

Kasschau, R., 1982. Glycine concentration as a biochemical indicator of sex and maturation in the Sea anemone Bundusoma cavernata. Comparative Biochemistry and Physiology Part A Physiology 72, 595–597.

Kaviraj, A., Unlu, E., Gupta, A., Nemr, A. El, 2014. Biomarkers of Environmental Pollutants. BioMed Research International 2014, 1-2.

Kube, S., Sokolowski, A., Jansen, J.M., Schiedek, D., 2007. Seasonal variability of free amino acids in two marine bivalves , Macoma balthica and Mytilus spp ., in relation to environmental and physiological factors 147, 1015–1027.

Lidsky, T.I., Schneider, J.S., 2003. Lead neurotoxicity in children: basic mechanisms and

clinical correlates. Brain 126, 5-19.

Lyons, B.P., Thain, J.E., Stentiford, G.D., Hylland, K., Davies, I.M., Vethaak, A.D., 2010. Using biological effects tools to define Good Environmental Status under the European Union Marine Strategy Framework Directive. Marine Pollution Bulletin 60, 1647–1651.

Magnusson, O., 2014. Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics, Eurachem Guide.

Nicholson, J.K., Lindon, J.C., Holmes, E., 1999. "Metabonomics": understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica 29, 1181–1189.

Powell, E. N., Chen, E., Koenig, M., Pecon, J., 1982. Changes in the free amino acid pool during environmental stress in the gill tissue of the oyster, Crassostrea virginica. Comparative Biochemistry and Physiology 71A, 591–598.

Rainbow, P.S., 1995. Biomonitoring of heavy metal availability in the marine environment. Marine Pollution Bulletin 31, 183–192.

Rainbow, P.S, 2002. Trace metal concentrations in aquatic invertebrates: why and so what? Environmental Pollution 120, 497–507.

Savorani, F., Tomasi, G., Engelsen, S.B., 2010. i coshift : A versatile tool for the rapid alignment of 1D NMR spectra. Journal of Magnetic Resonance 202, 190–202.

Solomons, N.W., 2001. Dietary sources of zinc and factors affecting its bioavailability. Food and Nutrition Bulletin 22, 138-154.

Stankovic, S., Kalaba, P., Stankovic, A.R., 2014. Biota as toxic metal indicators.

Environmental Chemistry Letters 12, 63–84.

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

Velez, C., Galvão, P., Longo, R., Malm, O., Soares, A.M.V.M., Figueira, E., Freitas, R., 2015. Ruditapes philippinarum and Ruditapes decussatus under Hg environmental contamination. Environmental science and pollution research international 22, 11890–11904.

Viarengo, a., Nott, J. a., 1993. Mechanisms of heavy metal cation homeostasis in marine invertebrates. Comparative Biochemistry and Physiology - C Pharmacology Toxicology and Endocrinology 104, 355–372.

Walla, W., Yancey, P.H., 2001. Water Stress, Osmolytes and Proteins. Integrative and Comparative Biology 41, 699–709.

Wang, W.-X., Rainbow, P.S., 2005. Influence of metal exposure history on trace metal uptake and accumulation by marine invertebrates. Ecotoxicology and Environmental Safety 61, 145–159.

Weckwerth, W., 2003. Metabolomics in systems biology. Annual Review of Plant Biology 54, 669–689.

Worley, B., Powers, R., 2012. Multivariate Analysis in Metabolomics. Current Metabolomics 1, 92-107.

Wu, H., Ji, C., Wang, Q., Liu, X., Zhao, J., Feng, J., 2013. Manila clam Venerupis philippinarum as a biomonitor to metal pollution. Chinese Journal of Oceanology and Limnology 31, 65-74.

Yancey, P.H., 2005. Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses. Journal of Experimental Biology 208, 2819-2830.

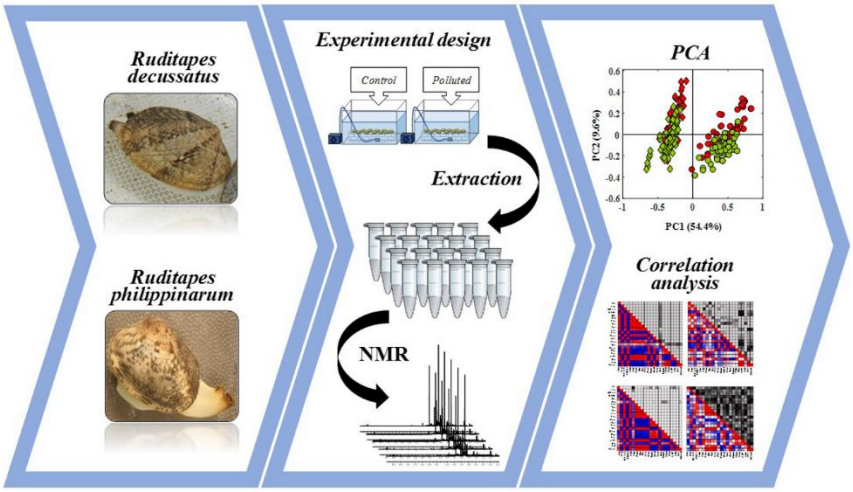
## Graphical abstract, Tables and Figures

**Table 1.** List of metabolites identified in the $^{1}$H-NMR spectra of clams' hydrophilic extract.

| Ref. n° | Metabolites | Multiplicity[a] | ppm |
|---|---|---|---|
| **1** | Isoleucine | *t* | 0.93 |
| | | *d* | 1.00 |
| **1** | Leucine | *t* | 0.96 |
| **1** | Valine | *d* | 0.98 |
| | | *d* | 1.03 |
| **2** | Unknown | *s* | 1.10 |
| **3** | Threonine | *d* | 1.31 |
| **4** | Alanine | *d* | 1.48 |
| **5** | Arginine | *m* | 1.65 |
| | | *m* | 1.72 |
| **6** | Acetate | *s* | 1.91 |
| **7** | Methionine | *m* | 2.16 |
| **8** | 2- Aminoadipate | *t* | 2.24 |
| **9** | Glutamate | *dd* | 2.35 |
| **10** | Succinate | *s* | 2.40 |
| **11** | Hypotaurine | *t* | 2.64 |
| **12** | Aspartate | *dd* | 2.66 |
| | | *dd* | 2.79 |
| **13** | Dimethylglycine | *s* | 2.90 |
| **14** | Malonate | *s* | 3.12 |
| **15** | Unknown | *s* | 3.14 |
| **16** | Phosphocholine | *s* | 3.21 |
| **17** | Betaine (s)- Taurine (t)-TMAO (s) | - | 3.26 |
| **18** | Taurine | *t* | 3.41 |
| **19** | Glycine | *s* | 3.54 |
| **20** | Betaine | *s* | 3.91 |
| **21** | Homarine | *s* | 4.35 |
| **22** | Glucose | *d* | 5.24 |
| **23** | Maltose | *d* | 5.40 |
| **24** | Glycogen | *bb* | 5.41 |
| **25** | ATP | *s* | 6.11 |
| **26** | Fumarate | *s* | 6.50 |
| **27** | Tyrosine | *d* | 6.90 |
| | | *d* | 7.21 |
| **28** | Hystidine | *s* | 7.18 |
| | | *s* | 8,08 |
| **29** | Phenylalanine | *d* | 7.31 |
| | | *m* | 7.36 |
| | | *m* | 7.42 |
| **30** | Formate | *s* | 8.45 |

[a] *s*: singlet, *d*: doublet, *t*: triplet, *dd*: doublet of doublets, *m*: multiplet, *bb*: broad band.

## Graphical abstract



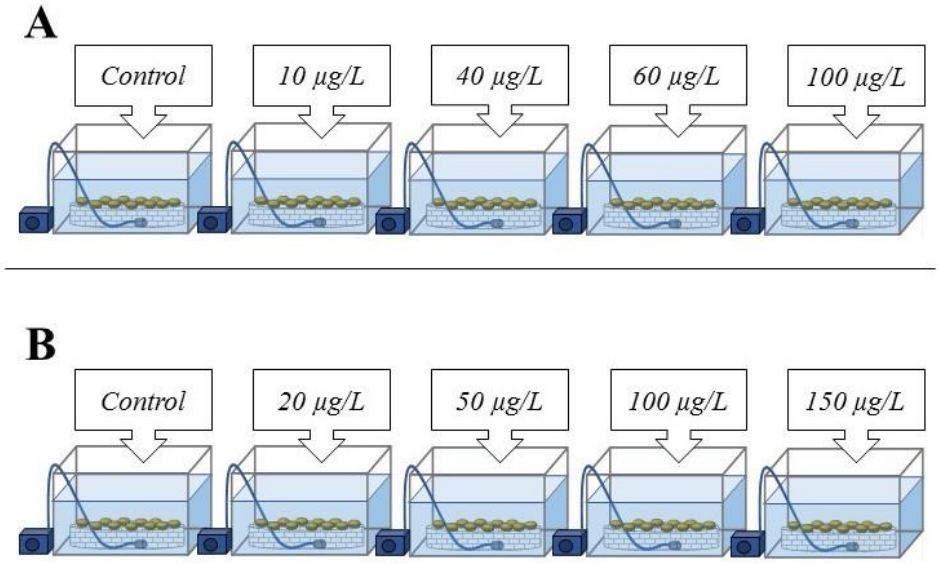




**Figure 1. Experimental design for short-term lead nitrate (A) and zinc chloride (B) exposure of clams.** After acclimatization, samples were exposed to four increasing concentrations of the above-mentioned pollutants.

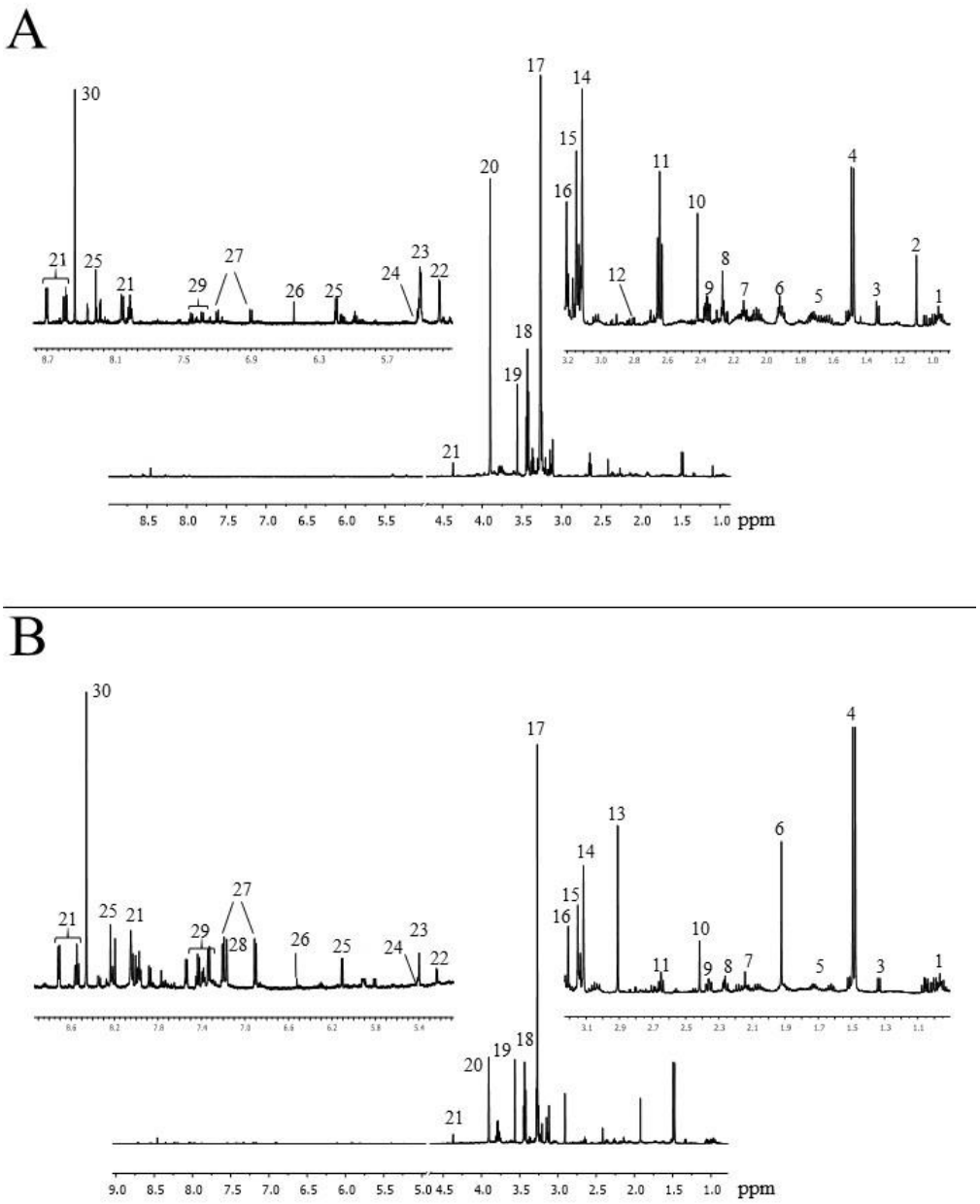



**Figure 2. Representative $^{1}$H NMR spectra of the hydrosoluble fraction of metal exposed Ruditapes decussatus (A) and Ruditapes philippinarum (B).** Expansions of the aliphatic and aromatic zones are reported on the left and right side of each figure. The assigned peaks are labelled with numbers sorted from up-field to down-field. Metabolites assignments, chemical shifts and multiplicity are reported in Table 1.

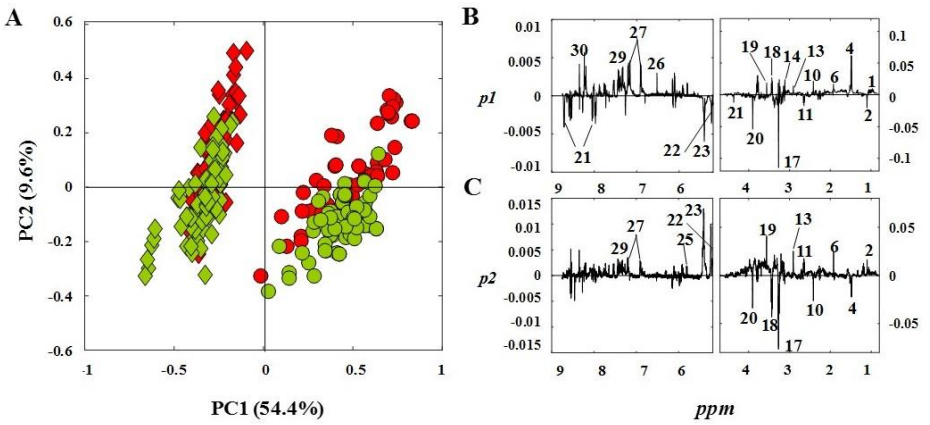



**Figure 3. PCA scores (A) and corresponding PC1 (B) and PC2 (C) loadings plots derived from the $^{1}$H-NMR spectra of clams' aqueous extracts.** Symbols indicate bivalves' species: diamonds, *R. decussatus;* circles, *R. philippinarum*. Colors denote the metal: red, Pb; green, Zn. Metabolites assignments are reported in Table 1.

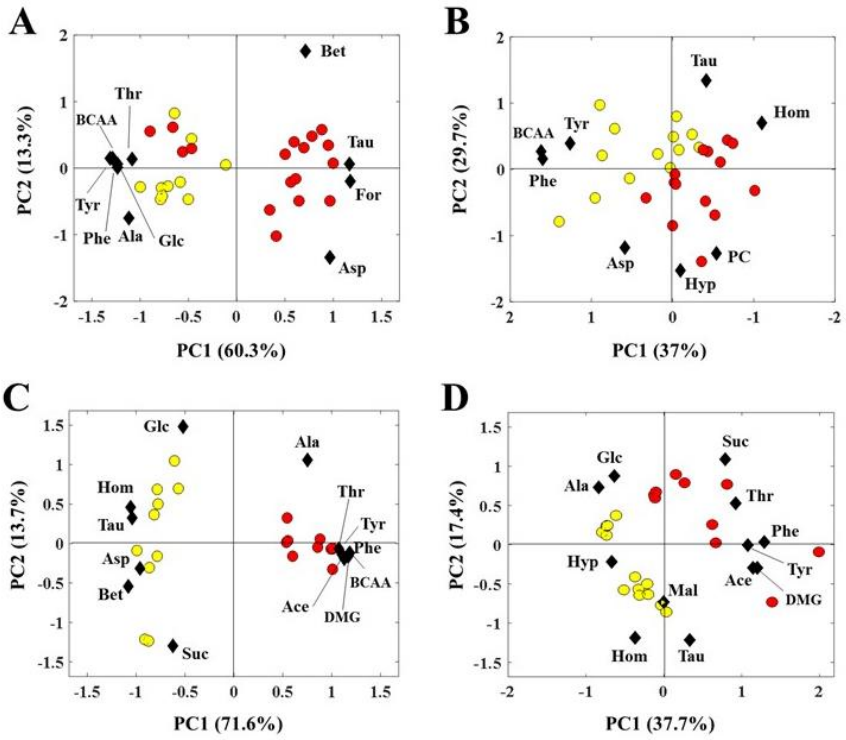



**Figure 4. PCA bi-plots showing the pairwise comparison between the controls (yellow circles) and the most heavily polluted clams (red circles).** Lead- (A) and zinc-treated (B) *R. decussatus*; lead- (C) and zinc-treated (D) *R. philippinarum*. Keys: Ace=acetate; Ala=alanine; Asp=aspartate; BCAA=branched chain amino acids; Bet=betaine; For=formate; Glc=glucose; Gly=glycine; Hom=homarine; Hyp=hypotaurine; Phe=phenylalanine; Thr=threonine; Tyr=tyrosine; Suc=succinate; DMG=dimethylglycine; Tau=taurine.

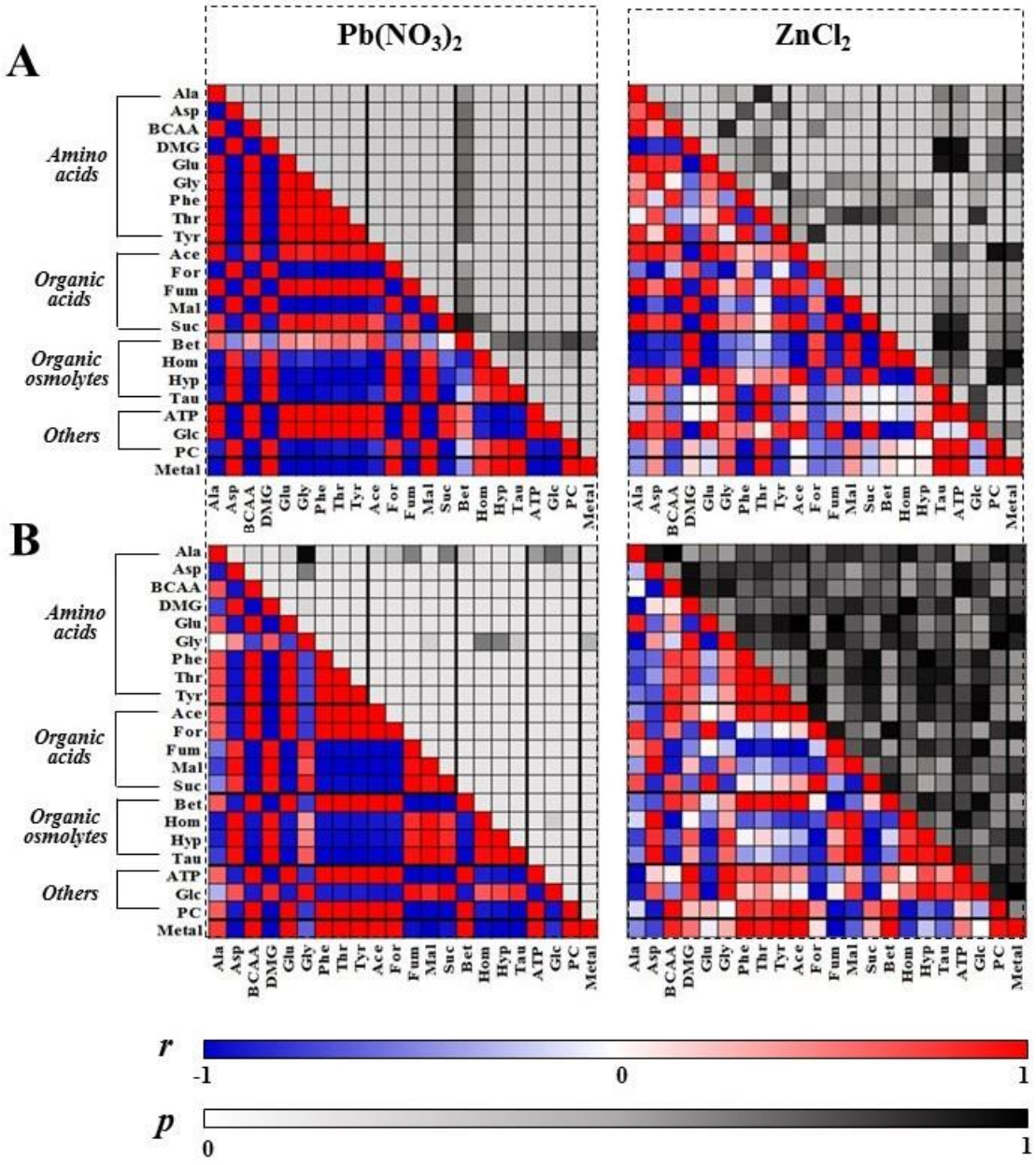



**Figure 5. Metabolite-metabolite correlation matrices (heat maps) for *R. decussatus* (A) and *R. philippinarum* following Pb and Zn exposure.** Metabolites are grouped by compound class. The heat maps are colored according to Pearson's correlation coefficient (r) and the corresponding p-values (p). The red and blue color of each cell depicts the Pearson's correlation coefficient value, with deeper colors indicating higher positive (red) or negative (blue) correlation coefficients. Light gray color stands for significant correlations ($p<0.05$).

# Paper IV

---

# Seasonal variations in the metabolic profile of *Ruditapes decussatus* (Linnaeus, 1758) from Santa Gilla lagoon

**Violetta Aru**[1], Søren Balling Engelsen[2], Francesco Savorani[2,3], Jacopo Culurgioni[5], Giorgia Sarais[4], Giulia Atzori[5], Serenella Cabiddu[5] and Flaminia Cesare Marincola[1]

[1]Departement of Chemical and Geological Sciences – University of Cagliari, S.S. 554 Bivio Sestu, 09042 Monserrato, Italy

[2]Quality & Technology, Department of Food Science, University of Copenhagen, Rolighedsvej 26, 1958 Frederiksberg C, Denmark

[3]Department of Applied Science and Technology (DISAT), Polytechnic University of Turin – Corso Duca degli Abruzzi 24, 10129 Torino (TO), Italy

[4]Departement of Life and Environmental Sciences, University of Cagliari, via Ospedale 72, 09124 Cagliari, Italy

[5]Departement of Life and Environmental Sciences, University of Cagliari, via T. Fiorelli 1, 09126 Cagliari, Italy

**Manuscript Draft**

---

## Abstract

The impact of the seasonal change on *Ruditapes decussatus* was investigated through global metabolite profiling by $^{1}H$ Nuclear Magnetic Resonance (NMR) spectroscopy. Principal Component Analysis (PCA) of the NMR data set evidenced a clear seasonal trend in samples grouping, while Interval Extended Canonical Variate Analysis (iECVA) allowed the identification of the discriminant metabolites responsible for the observed distribution pattern. In particular, the relative amount of alanine and glycine underwent to significant variations according to sampling seasons. These results could be explained in terms of an adaptive behavior of clams to better cope with different climatic and environmental conditions. Although in a preliminary stage, the findings reported in this study highlight that NMR spectroscopy coupled with multivariate statistical analysis can shed light on the metabolic features of bivalves in relation with season variability. This will open the way for further investigations that would contribute to achieve deeper insights on bivalves' bio-ecological framework.

## 1. Introduction

Shellfish farming covers a prominent role in the Italian aquaculture production and, within the EU, Italy is one of the main edible bivalves producer countries [1]. Bivalves are a variegated class of mollusks which comprise several species encompassing both marine and fresh-water habitats throughout the world. These seafood products represent a largely exploitable resource in the fishery industry being widely raised for human consumption. Moreover, marine bivalve mollusks are rich in essential vitamins and minerals as well as polyunsaturated fatty acids with known health beneficial effects [2]. Beside the remarkably economical relevance covered in the aquaculture industry, they have been demonstrated to be feasible tools in environmental pollution monitoring programs in which they are widely employed as sentinel organism [3].

*Ruditapes decussatus* is a burrowing clam widely distributed along the Italian coasts where it represents a seafood resource of prominent commercial relevance. This clam species represents a traditional resource of fishing activity in Sardinia Island where it is not cultured, but fishermen commonly pick them up from the bottom of brackish areas. Several nutritional studies have been carried out with the aim of characterizing the fatty acid composition of these clams [4], [5]. Although the undeniable relevance covered by this clam species in both nutrition and ecological fields, a complete characterization of the biochemical composition of *R. decussatus* has never been carried out.

Metabolomics is a scientific discipline mainly focused on the high-throughput characterization of small metabolites in biological matrices [6]. It has been established to be a feasible analytical approach in the fields of both food science, from food component analysis to food quality/authenticity assessment [7], and environmental science for the detection of biomarkers of pollution [8]. Metabolomics has been also deemed to be a reliable tool in the study of bivalves since it may reveal the constraints they are subjected to, and may help in clarifying the most affected biochemical pathways and the homeostatic mechanisms triggered [9].

In this work, $^{1}$H NMR-based metabolomics has been applied to study the influence of seasonal changes on the *R. decussatus'* metabolic profile. The main objectives of the

present study were to: (i) use NMR to metabolically characterize *R. decussatus*, (ii) ascertain how the metabolic profiles differ according to the sampling season and (iii) provide new insights on the potential of NMR-based metabolomics as a rapid and informative screening tool in marine biology research.

The potential of metabolomics in the discrimination among clams sampled in different seasons was explored by Principal Component Analysis (PCA) and assessed by Interval Extended Canonical Variate Analysis (*i*ECVA). Each discriminant metabolite was quantified in the $^{1}$H NMR spectra leading thus to gain further information on the effect of seasonal changes in clams' biochemical composition. Additional microscopic inspection of clams' gonadal tissues was carried out on samples.

## 2.Materials and methods

### *2.1 Chemicals*

All solvents used, of the highest available purity, were purchased from Merck (Darmstadt, Germany), while deuterium oxide ($D_2O$, 99.9 %), sodium deuteroxide (NaOD, 40 wt. % in D2O, 99.5 atom % D), deuterium chloride (DCl, 99 atom % D) and sodium 3-trimethylsilyl-propionate-2,2,3,3,-d4 (TSP) were acquired from Sigma–Aldrich (Milan, Italy).

### *2.2 Samples collection*

A total of one hundred adult specimens of *R. decussatus* (shell length 36.9 ± 1.0 mm) were randomly collected in Santa Gilla lagoon, in the south of Sardinia (Italy) (Fig. 1) by using a shellfish rake. Clams sampling was performed in different months of the year, starting from May 2013 to July 2014. After sample collection, clams were transported to the laboratory in portable coolers, at approximately 4°C. In the laboratory, clams were immediately shucked and deprived of the digestive gland and the intestine. The soft tissue of the sample was then put into insulated sterile plastic boxes and stored at -80°C before metabolites extraction.

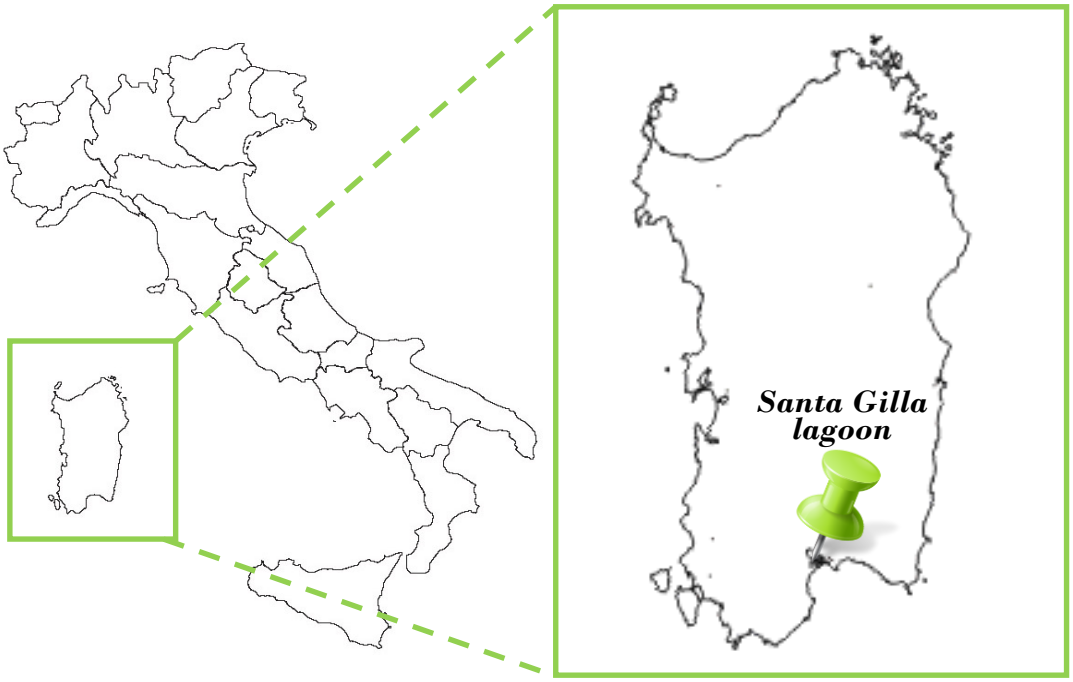



**Figure 1.** Clams sampling was carried out in Santa Gilla lagoon (Sardinia, Italy). The sampling location is evidenced by the green push pin.

*2.3 Sex and parasitosis determination*

Gender determination was performed according to the squashing method [10]. A drop of water was put on the glass slide and then a smear of the gonads and/or a small piece of the gonadal tissue into the water. The specimens were located in the centre of the glass slide, and the cover glass was placed on top of it. After squashing, samples were examined microscopically using a Zeiss Primo Star Halogen/LED Microscope. Clams gender was determined by the presence of ovary or sperm follicles within the mantle. Parasitological analysis was also performed in the same way as describe before.

*2.5 Sample preparation for NMR analysis*

Water-soluble metabolites were extracted according to the Folch method [11]. A mixture of 12 ml chloroform–methanol (2:1, v/v) was added to the clams' samples. After the addiction of 4 ml of $H_2O$ and centrifugation at 1700 rcf for 1 h at 4 °C, the methanol/water mixture was separated by the chloroform fraction and transferred into several Eppendorf vials. Water/methanol solvent mixture was evaporated with an

Eppendorf concentrator 5301 (Eppendorf AG, Hamburg, Germany). Samples were subsequently re-dissolved in 1.2 ml of a $D_2O$ solution of the internal standard (TSP) 0.80 mM. and centrifuged at 13000 rcf for 5 min at 4 °C to remove particulate matter. The pH of the final sample was accurately adjusted to 6.52 ± 0.03 by adding a small amount of NaOD or DCl to minimize pH-based peak movements. Then, an aliquot of 650 µl was placed into a 5 mm NMR tube for NMR analysis.

*2.6 NMR measurements*

$^1$H-NMR spectra were acquired with a Varian Unity Inova 500 spectrometer (Agilent Technologies, CA, USA) operating at 499.84 MHz at 300 K. For each spectrum, 256 scans were collected into 32k data points over a spectra width of 6000 Hz, with a 45° pulse, an acquisition time of 1.5 s, and a relaxation delay of 4 s. The solvent (water) residual signal was suppressed by applying a presaturation technique with low power radiofrequency irradiation for 1.5 s. An exponential function corresponding to 0.3 Hz was applied to each free induction decay (FID) before Fourier transformation and a zero-filling to 64 K. NMR spectra were manually phased and baseline corrected using MestReNova (Version 8.1, Mestrelab Research SL). Spectral chemical shift referencing on the TSP $CH_3$ signal at 0.00 ppm was performed in all spectra.

*2.7 Preprocessing and chemometric analysis of NMR data*

$^1$H NMR spectra were converted into ASCII files and imported in Matlab software (Mathworks, MA) and an overall data matrix, sized 60 samples × 65536 variables, was built. Spectra were then aligned using the *ic*oshift algorithm, which is based on correlation optimized shifting of spectral intervals and aligns all spectra simultaneously [12]. Beside the removal of the noisy region between 9.5 and 10.5 ppm, also the regions comprised between 0.5 and -0.5 ppm and 4.6 and 5.0 ppm were excluded for multivariate data analysis because of the signals of TSP and water, respectively. The final data matrix sized 60 × 41499.

In order to minimize the effect of sample dilution, NMR spectra were normalized using total area normalization. This procedure allowed the direct comparison among samples. The data set was then Pareto-scaled. This pre-processing scaling increases the

representation of lower concentration metabolites in the resultant data models while minimizing the contribution from noise [13]. The pre-processed data matrix was than imported on LatentiX 2.12 (www.latentix.com, Latent5, Copenhagen, Denmark).

Principal Component Analysis (PCA) was carried out on the dataset with the aim to identify inherent group clustering and trends. PCA is a chemometric tool for exploratory data analysis that displays the intrinsic structure of the dataset in a low-dimensional orthogonal projection. It highlights sample grouping and trends by transforming the original variables into few latent factors, the Principal Components (PCs), sorted by explained variance [14]. In order to identify the discriminant metabolites whose relative amount significantly changed according to season, Interval Extended Canonical Variable Analysis (*i*ECVA) was further performed in Matlab using the ECVA toolbox (http://www.models.life.ku.dk/ source/). *i*ECVA [15] is a recent classification chemometric tool designed to find spectral regions holding the main information responsible for the separation among groups. It is an extension of ECVA [16] which represents a new approach for grouping samples based on the standard Canonical Variates Analysis (CVA), but with an underlying PLS engine. It is able to cope with several different classes yielding powerful separations.

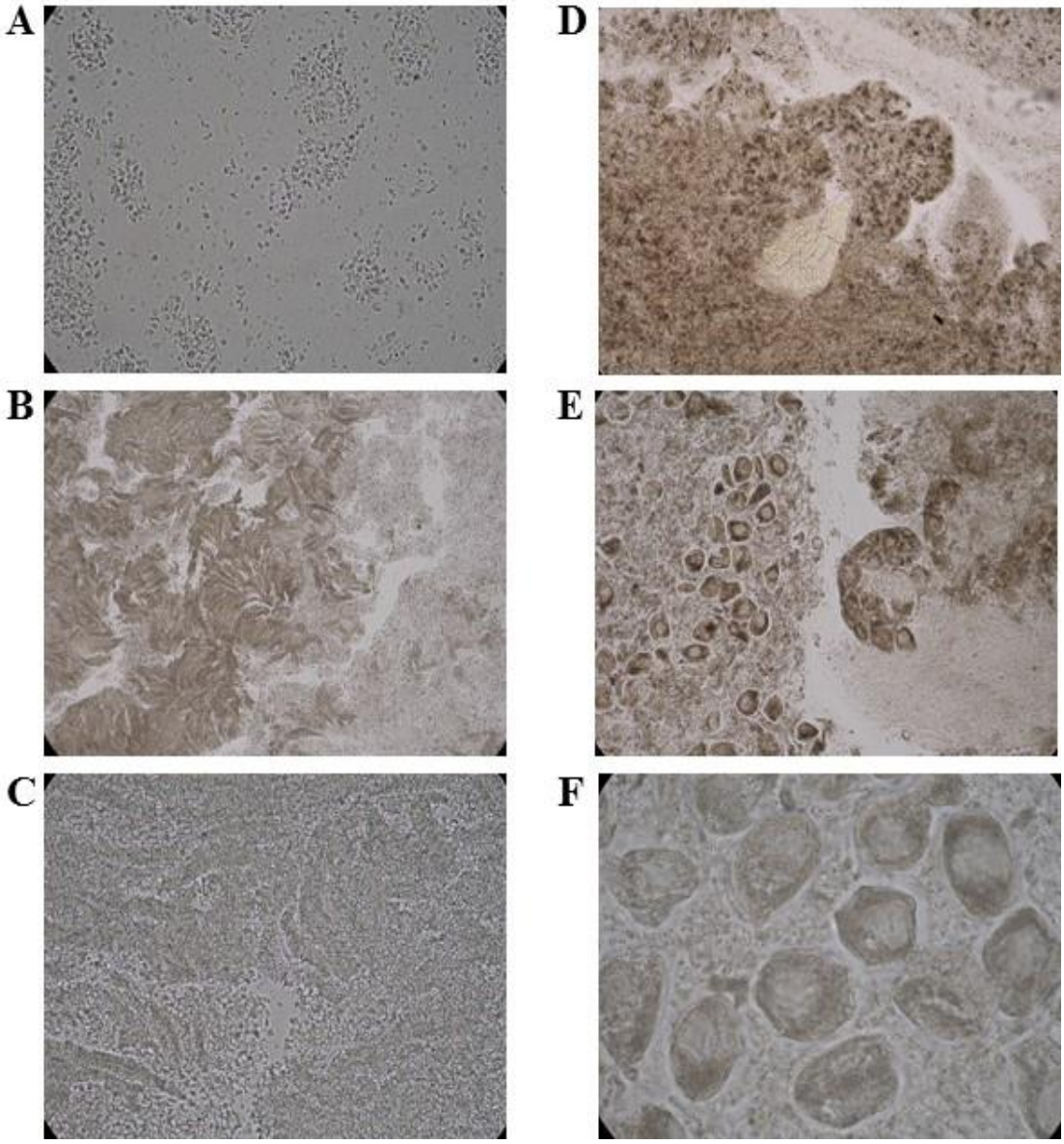
In the present study, *i*ECVA was applied to several pre-defined intervals with the aim to find those holding the higher discriminative power. Careful validation was performed using a "venetian blind" cross-validation.

## 3. Results and discussion

### *3.1 Sex determination*

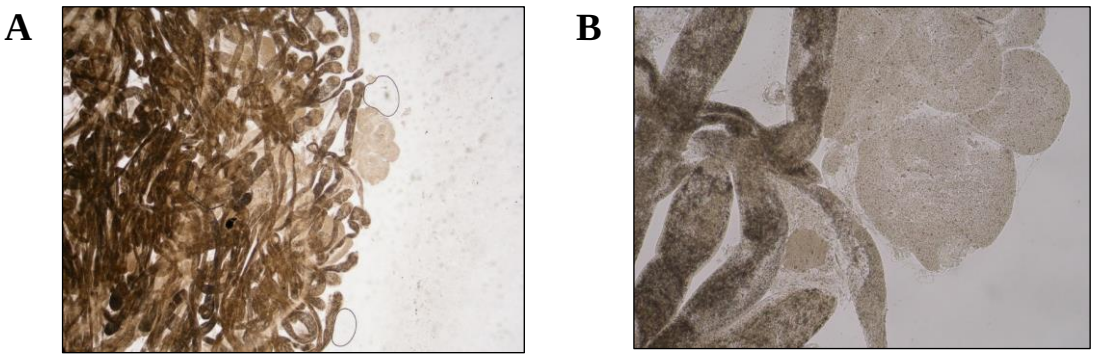
The macroscopic inspection of the visceral mass of the clam specimens allowed the identification of the gonadal tissue. It was composed of a series of compact granular follicles representing the structural units of the gonadal tissue. The follicles were surrounded by connective tissues and contained numerous gametes. Ova and spermatozoa were detected in the gonadal tissue of female and male ripe clams, respectively (Fig. 2). This enabled all 60 clams to be classified as males or females. In Figure 2, representative microscopic views at different magnifications of clams' gonadal

tissues, showing ripe female gonads with ova (right-side) and ripe male gonad with spermatozoa (left-side), are shown.




**Figures 2.** Gonadal tissues of male (left side) and female (right side) *R. decussatus*. Ovary and sperm follicles are shown. Pictures were taken under a light microscope. Magnifications 20 X(A/D), 100 X (B/E) and 400 X (C/F) are reported.

The occurrence of parasite infections was also investigated in clam samples. Sporocyst and cercariae of *Bacciger bacciger* (Platyhelminthes, Trematoda, Digenea) were found in the specimens, with an infection rate of 10.0%.




**Figures 3.** Gonadal tissues of *R. decussatus* with mother sporocysts of *Bacciger bacciger* containing daughter sporocysts. Pictures were taken under a light microscope. Magnifications 20 X (A) and 100 X (B) are reported.

*B. bacciger* is a foodborne pathogen commonly infecting bivalve mollusks. The infection by this parasite has already been reported for several commercial native bivalves from St. Gilla lagoon, among which *R. decussatus* [17]. Specimens showing the highest levels of parasites, were excluded from further analysis. A total of sixty clam samples were selected for the following analysis.

*3.2 NMR Spectroscopy and chemometrics analysis*

A total of 60 water-soluble extracts from clams were analysed by $^{1}$H NMR spectroscopy. A representative spectrum is shown in Fig. 4. The spectrum displays several hundred peaks arising from the different functional groups of a large number of metabolites including amino acids, organic acids, organic osmolytes, and carbohydrates. Although several metabolite classes were observed, the spectra were mainly dominated by the presence of free amino acids and osmolytes, all of which are recognized to be involved in invertebrate osmotic regulation [18]. Thirty-one compounds were identified on the basis of data published in the literature [19] and by adding standard compounds directly to the sample solution and recording again the NMR spectrum under the same conditions. Their assignments, chemical shifts, and multiplicity are reported in Table 1.

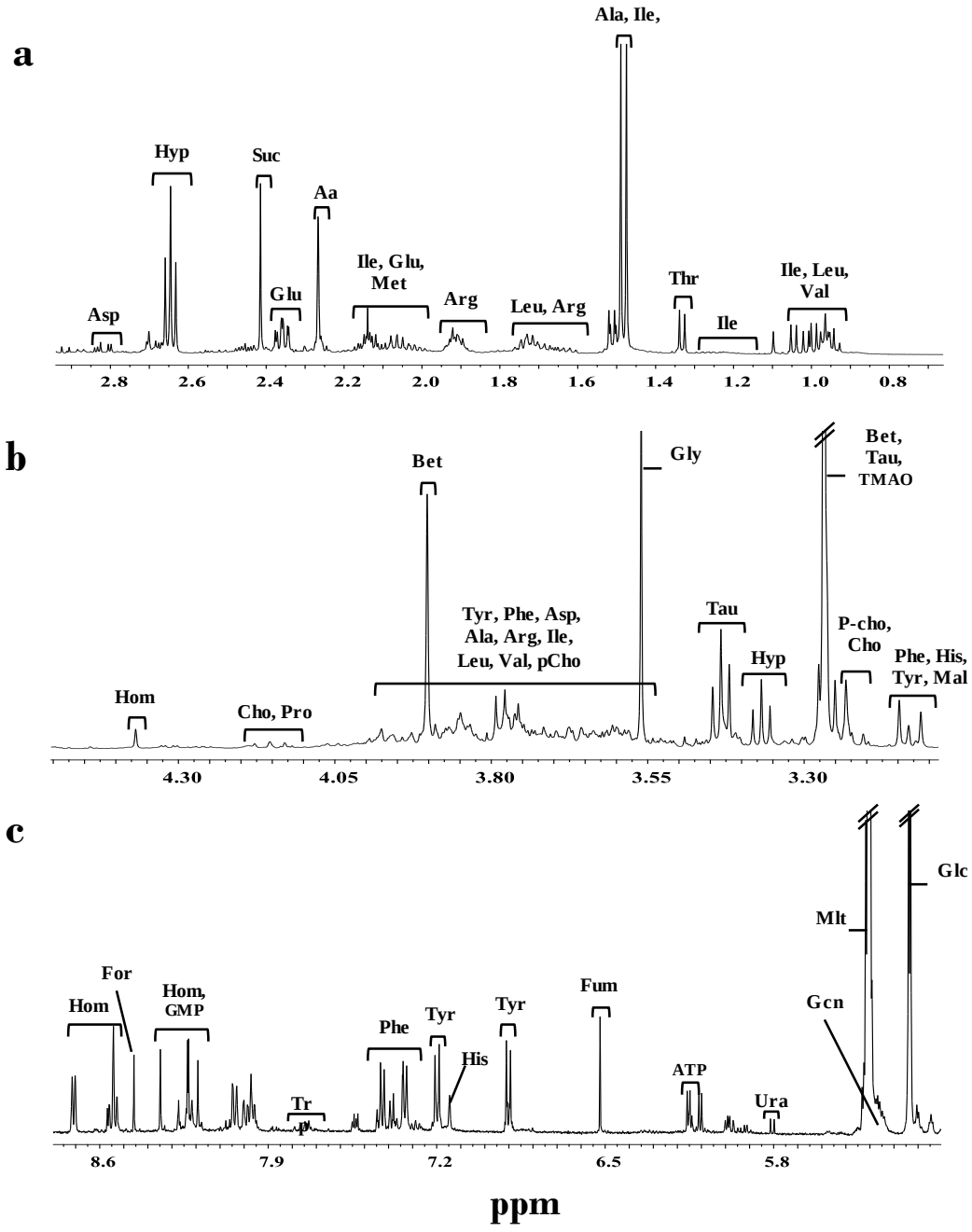



**Figure 4.** Representative $^{1}$H-NMR spectrum of the aqueous extract of *R. decussatus*.

**Table 1.** List of compounds identified in the $^{1}$H NMR spectrum of the aqueous extract of *R. decussatus*, and their respective chemical shifts and multiplicity.

| Compound | ppm | Multiplicity* |
|---|---|---|
| Acetoacetate (Aa) | 2.27 | *s* |
| Alanine (Ala) | 1.49 | *d* |
| Arginine (Arg) | 1.70/1.92 | *m*/*m* |
| Aspartate (Asp) | 2.68/2.80 | *dd*/*dd* |
| Adenosine triphosphate (ATP) | 8.52 | *s* |
| Betaine (Bet) | 3.26/3.39 | *s*/*s* |
| Choline (Cho) | 3.21 | *s* |
| Formate (Form) | 8.46 | *s* |
| Fumarate (Fum) | 6.50 | *s* |
| Glutamate (Glu) | 2.38 | *t* |
| Glycine (Gly) | 3.57 | *s* |
| Glycogen (Gcn) | 5.40 | *bb* |
| Glucose (Glc) | 5.24 | *d* |
| Histidine (His) | 7.15/7.79 | *s*/*s* |
| Homarine (Hom) | 4.31 | *s* |
| Hypotaurine (Hyp) | 2.64/ 3.36 | *t*/*t* |
| Isoleucine (Ile) | 0.94/1.02 | *d*/*t* |
| Leucine (Leu) | 0.96 | *d* |
| Malonato (Mal) | 3.10 | *s* |
| Maltose (Mlt) | 5.40 | *d* |
| Methionine (Met) | 2.14 | *s* |
| Phenylalanine (Phe) | 7.33/7.42/7.42 | *m*/*m*/*m* |
| Phosphorylcholine (P-Cho) | 3.24 | *s* |
| Succinate (Suc) | 2.41 | *s* |
| Taurine (Tau) | 3.24/3.43 | *t*/*t* |
| Trimethylamine oxide (TMAO) | 3.24 | *s* |
| Threonine (Thr) | 1.34 | *d* |
| Tryptophan (Trp) | 7.72 | *d* |
| Tyrosine (Tyr) | 6.89/7.19 | *d*/ *d* |
| Uracil (Ura) | 5.81/7.54 | *d*/*d* |
| Valine (Val) | 0.99/1.04 | *d*/ *d* |

**s*: singlet; *d*: doublet; *t*: triplet; *dd*: doublet of doublets; *m*: multiplet; *bb*: broad band.

An exploratory PCA analysis was performed on the whole dataset to seek for sample grouping and trends which would be hard-to identify through the simple visual inspection of each spectrum. The application of PCA allowed the reduction of the large $^1$H NMR data set to four principal components with 64% of total variance explained. As it can be observed in Figure 3, the samples distribution in the scores plot is remarkably season-sensitive while, no gender-related clustering could be detected. As can be seen, a noteworthy seasonal trend of the metabolic profiles lined samples distribution from spring (green circles) to autumn (blue circles).

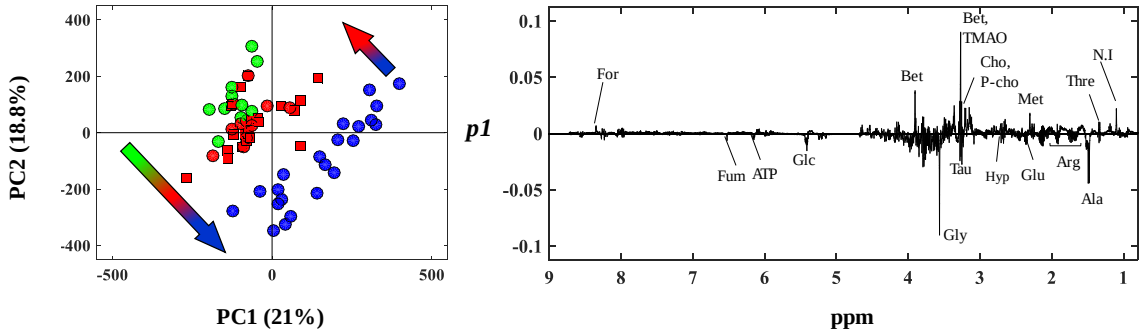



**Figure 3.** PCA score plot of PC1 versus PC2 (left side) and PC1 loadings plot (right side) of the $^1$H NMR spectral data of clams aqueous extracts. The most significant metabolites are highlighted. Spring 2013 (green circles), Summer 2013 (red circles), Autumn 2013 (blue circles), Summer 2014 (red squares). Arrows are coloured according to the observed seasonal trend.

Examination of the loadings plot enabled us to determine the variables with the highest impact on the variance and, thus, to identify those better contributing to the observed trend. They included, among the others, free amino acids (i. e. alanine, glutamate, glycine), choline and phosphocholine, and the osmolytes taurine and hypotaurine. In order to investigate in more details the actual discriminant nature of these metabolites, *i*ECVA was performed on these pre-defined intervals. Basically, these spectral intervals were utilized to calculate the corresponding ECVA models whose performances (in terms of no. of misclassifications) can be compared to each other and with respect to that of the ECVA model built using the whole spectrum. Thus, for each of them, the number of misclassified samples, resulting from "venetian blind" cross-validation, was calculated. Any interval presenting an error bar lower than the

dotted line is potentially a better discriminant than the full spectral area. Interestingly, the intervals showing the lower number of misclassifications were the spectral regions of alanine and glycine. In Figure 5 the scores plot of the iECVA model built by using alanine and glycine intervals is reported. The discriminant nature of these intervals led to sample clustering according to seasons: an outstanding seasonal trend is evidenced along PC1 from the left to the right side of the scores plots. An enlargement of the superimposed NMR signals belonging to the two discriminant metabolites is reported.

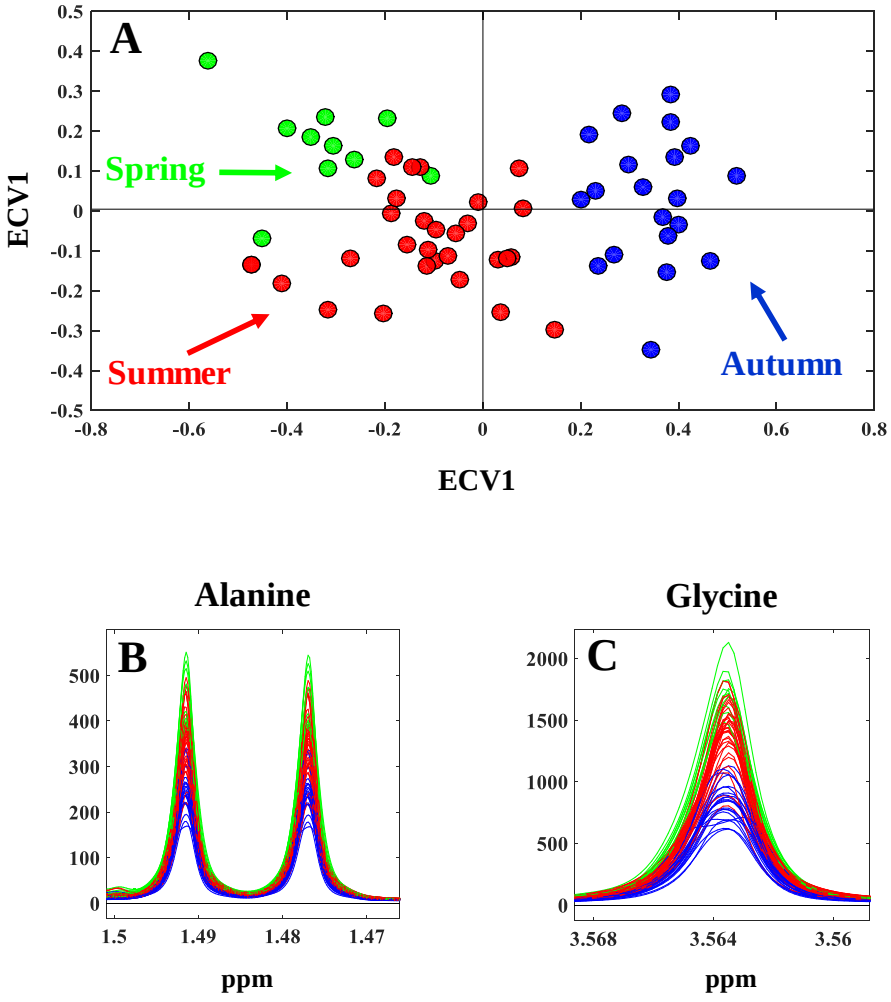



**Figure 5.** (A) iECVA model calculated by using alanine and glycine intervals. The spectral features of the two discriminative intervals are reported in Figs. B and C.

The base-line resolved signals of these metabolites were accurately integrated in the NMR spectra and their relative amount was calculated. One-way ANOVA was used to

further assess the statistical significance of the results: different letters (a, b, c) stand for significant variations in the relative amount of these discriminant metabolites. In Figure 6 the changes in the content of alanine and glycine are represented in a box-plot format. As it can be observed, both amino acids exhibited a significant decrease until autumn, then the level of alanine and glycine started to increase again.

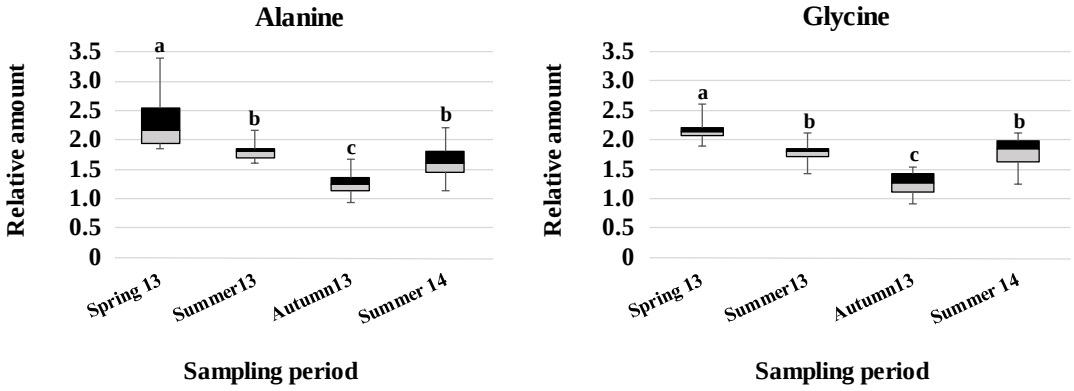



**Figure 6.** Variations in the relative amount of alanine and glycine. Both metabolites showed a remarkable seasonal pattern.

These findings are in agreement with those found for other Mediterranean mollusks whose relative amount of alanine and glycine significantly varied according to the measured cyclical salinity variations of the water [20] due to seasonal change. Indeed, it is well known how marine invertebrates, and osmoconformers in general, can use free amino acids and osmolytes to cope with osmotic stressful conditions [18] which can lead to shrinkage or swelling if the osmotic pressure of the environment dramatically changes.

## 4. Preliminary conclusions

The results of the present study demonstrated that the use of advanced multivariate statistical techniques, such as *i*ECVA, in conjunction with $^{1}$H NMR spectroscopy is a feasible approach that allows the discrimination between specimens of *R. decussatus* sampled in different periods of the year. In agreement with what already found in the literature for other Mediterranean species, the relative amounts of alanine and glycine

exhibited a significant variation according to seasonal change. Indeed, free amino acids in marine invertebrates are known to be involved in a variety of environmental processes, such as osmotic regulation, suggesting an adaptive behavior of the clams to better cope with different climatic conditions.

# References


[1] FAO, *The state of world fisheries and aquaculture*, vol. 2014. 2014.

[2] A. P. Simopoulos, "Omega-6/Omega-3 Essential Fatty Acid Ratio and Chronic Diseases," *Food Reviews International*, vol. 20, no. 1, pp. 77–90, 2004.

[3] S. Stankovic, P. Kalaba, and A. R. Stankovic, "Biota as toxic metal indicators," *Environmental Chemistry Letters*, vol. 12, no. 1. pp. 63–84, 2014.

[4] M. Albentosa, U. Labarta, M. J. Fernández-Reiriz, and A. Pérez-Camacho, "Fatty acid composition of Ruditapes decussatus spat fed on different microalgae diets," *Comparative Biochemistry and Physiology Part A: Physiology*, vol. 113, no. 2, pp. 113–119, 1996.

[5] M. J. Fernández-Reiriz, U. Labarta, M. Albentosa, and A. Pérez-Camacho, "Lipid profile and growth of the clam spat, Ruditapes decussatus (L), fed with microalgal diets and cornstarch," *Comparative Biochemistry and Physiology - B Biochemistry and Molecular Biology*, vol. 124, no. 3, pp. 309–318, 1999.

[6] U. Roessner and J. Bowne, "What is metabolomics all about?," *BioTechniques*, vol. 46, no. 5, pp. 363–365, 2009.

[7] A. Trimigno, F. C. Marincola, N. Dellarosa, G. Picone, and L. Laghi, "Definition of food quality by NMR-based foodomics," *Current Opinion in Food Science*, vol. 4, pp. 99–104, 2015.

[8] B. P. Lankadurai, E. G. Nagato, and M. J. Simpson, "Environmental Metabolomics: an emerging approach to study organisms responses to environmental stressors.," *Environmental Review*, vol. 21, no. July, pp. 180–205, 2013.

[9] M. R. Viant, "Metabolomics of aquatic organisms: The new 'omics' on the block," *Marine Ecology Progress Series*, vol. 332, pp. 301–306, 2007.

[10] R. D. Guerrero, W. L. Shelton, An aceto-carmine squash method of sexing

juvenile fishes. *The Progressive Fish-Culturist*, vol. 36, pp. 36-56, 1974.

[11] J. Folch, M. Lees, and G. H. S. Stanley, "A simple method for the isolation and purification of total lipids from animal tissues," *J Biol Chem*, vol. 226, no. 1. pp. 497–509, 1957.

[12] F. Savorani, G. Tomasi, and S. B. Engelsen, "icoshift: A versatile tool for the rapid alignment of 1D NMR spectra," *Journal of Magnetic Resonance*, vol. 202, no. 2, pp. 190–202, 2010.

[13] B. Worley and R. Powers, "Multivariate Analysis in Metabolomics," *Current Metabolomics*, vol. 1, no. 1. pp. 92–107, 2012.

[14] R. Bro and A. K. Smilde, "Principal component analysis," *Analytical Methods*, vol. 6, no. 9. p. 2812, 2014.

[15] L. Nørgaard, R. Bro, F. Westad, S. B. Engelsen. A modification of canonical variates analysis to handle highly collinear multivariate data. *Journal of Chemometrix*, vol. 20, pp. 425–35, 2006.

[16] L. Nørgaard, G. Soletormos, N. Harrit, M. Albrechtsen, O. Olsen, D. Nielsen, et al. Fluorescence spectroscopy and chemometrics for classification of breast cancer samples—a feasibility study using extended canonical variates analysis. *Journal of Chemometrix*, vol. 21, pp. 451–458, 2007.

[17] J. Culurgioni, V. D' Amico, R. De Murtas, G. Canestri Trotti, and V. Figus, "Parasitological monitoring of commercial native bivalves from St. Gilla lagoon (Sardinia, South Western Mediterranean)," *Ittiopatologia*, vol. 3, 243-252, 2006.

[18] P. H. Yancey, "Organic osmolytes as compatible, metabolic and counteracting cytoprotectants in high osmolarity and other stresses," *Journal of Experimental Biology*, vol. 208, no. 15. pp. 2819–2830, 2005.

[19] A. P. Tikunov, C. B. Johnson, H. Lee, M. K. Stoskopf, and J. M. Macdonald, "Metabolomic investigations of American oysters using 1H-NMR spectroscopy," *Marine Drugs*, vol. 8, no. 10, pp. 2578–2596, 2010.

[20] S. Kube, A. Sokolowski, J. M. Jansen, and D. Schiedek, “Seasonal variability of free amino acids in two marine bivalves , Macoma balthica and Mytilus spp ., in relation to environmental and physiological factors,” vol. 147, pp. 1015–1027, 2007.