Chapter 11

NMR-Based Metabolomics in Food Quality Control

Alberta Tomassini, Giorgio Capuani, Maurizio Delfini and Alfredo Miccheli*

Department of Chemistry, Sapienza University of Rome, Rome, Italy

Corresponding author: alfredo.miccheli@uniroma1.it

Chapter Outline

1 INTRODUCTION

Metabolomics has been defined as the identification and quantification of all metabolites in a biological system. This approach has emerged as an important tool in many different fields such as human disease, drug discovery, and finally plant physiology. Over the past few years, metabolomics has also emerged as a tool of increasing interest of food and nutrition scientists [1]. Metabolomics gives the opportunity to gain deeper insights into a food to obtain a better definition of the biochemical composition and to evaluate the quality, the processing, and the safety of raw materials and final products. Traditionally, food component analysis is focused on identifying and classifying food constituents into very broad categories such as carbohydrates, proteins, fats, fibre, vitamins, trace elements, solids, and/or ash. Otherwise a metabolomic analysis, through a detailed characterization of foods and

Data Handling in Science and Technology, Vol. 28. http://dx.doi.org/10.1016/B978-0-444-59528-7.00011-9

beverages into their biochemical constituents, offers to food chemists the opportunity not only to understand what gives to foods and beverages their aroma, taste, texture, or colour, but also to evaluate the nutritional properties and their impact on human health.

The analytical tools employed in metabolomic analysis are techniques that allow obtaining a high throughput of data, the two most employed being mass spectrometry (MS) or nuclear magnetic resonance (NMR) spectroscopy. By applying multivariate analysis to NMR spectroscopy or MS data, it is possible to interpret the wide range of data and to provide comprehensive information about biochemical constituents of a biological system [2,3]. The field of food analysis metabolomics, together with transcriptomics and proteomics, is involved in "Foodomics", that is defined "as a new discipline that studies the food and nutrition domains through the application of advanced omics technologies to improve consumer's well-being, health, and confidence" [4,5]. Consequently, the control of quality, the safety of food, and the monitoring of the food processing by metabolomics techniques acquire a greater importance both in the field of manufacturing output and the well-being of the consumers. Although still at their beginning, metabolomics-based strategies are already being applied to food processing and quality control, as well as plant breeding for improved crop varieties and in the development of novel foodstuffs. The importance of this technology has also recently been recognized by the EU Research division, which funds many multi-national projects like, for example, METAPHOR (metabolomics for Plants, Health and OutReach) 2007–2012 [6]. Successful development of this approach requires the collaboration between a multitude of disciplines, such as analytical chemistry and spectroscopy, food chemistry, plant biology, nutrition, computational science, bioinformatics, and statistics. Overall, it requires the management of different databases, acquired in standard conditions, and the application of appropriate models for the classification, the acceptability of quality, and the compliance with food regulations: this represents the actual challenge for the chemometricians in metabolomics of foods.

In this chapter, we report the NMR-based metabolomic approach in food analysis and display its more instructive applications in quality control in order to illustrate the set of problems related to peculiar data source, the potentiality, and the development features of main interest for chemometricians in this field. Therefore, to overview the different MS-based strategies applied to food analysis, we address the reader to the most recent comprehensive reviews [5,7].

2 METHODOLOGY

Foodstuff can be considered as complex mixtures containing a wide range of metabolites having different concentrations and chemical properties. The biochemical components of a foodstuff can already be present in the crude products, or arise from the storage conditions and processing, or be added for preservation or peculiar nutritional purposes.

In this context, NMR spectroscopy has shown some important advantages in respect to the other analytical techniques, which can be summarized by ease of quantification, straightforward metabolite identification, and ability to determine unexpected metabolites [8]. Indeed, NMR analysis of a sample allows the simultaneous detection of low molecular weight metabolites present in a solution at a concentration beyond 10^{-6} M.

However, the sensitivity of NMR spectroscopy is the main limitation in its application in food analysis, when compared to MS techniques. The presence of toxins or pesticides in a foodstuff cannot be investigated by NMR spectroscopy, even if the increasing developments of NMR hardware are greatly improving its sensitivity. In spite of this, NMR spectroscopy turns out to be one of the most suitable techniques for food quality control and assurance, mainly because of its reproducibility and of its ability to analyse a foodstuff sample without or with minimal chemical treatment.

Due to these characteristics, the possibility to build databases of NMR-based metabolic profiles of foodstuffs, in order to assess their quality, is becoming more and more actual [9]. However, when we apply multivariate analysis to our data fields or wish to compare databases obtained by NMR spectroscopy by other researchers, it is very important to consider the source of extrinsic variability originated by the employed experimental conditions, in terms of sample preparation, NMR acquisition parameters, signal-to-noise ratio, NMR spectral pre-treatment, NMR data pre-processing, as well as the strategy for NMR-based metabolomics (for an exhaustive introduction to the reported critical concepts see the manual by Axelson [10]).

2.1 NMR Sample Preparation

The numbers forming the data fields handled by chemometricians are entities embedding the whole history of the experimental protocol performed from the raw material to the data obtained by spectroscopy: therefore it is of utmost importance that each step of the experimental protocol is well understood, in order to have the correct idea of what the degrees of confidence and the uncertainties associated with that number are.

In the case of NMR spectroscopy, the more sensitive steps of the protocols can be roughly located in the extraction methods, the choice of the appropriate NMR parameters and the post-processing of the obtained spectra (that is the first step of data pre-processing).

The preparation of samples to be analysed by high-resolution NMR spectroscopy must accomplish two main goals: yielding homogeneous samples (mainly homogeneous solutions in liquid phase), avoiding as much as possible the presence of high- and medium-molecular weight metabolites. Indeed, the presence of different phases in the sample—either two or more separated phases or particulate—interferes with the homogeneity of the magnetic field resulting in broader NMR signals, with a loss of both resolution and reproducibility, while

relatively high molecular weight molecules would appear as broad signals in the spectra superimposed to the relatively narrow resonances from the low molecular weight molecules: this would affect the quantification of the latter in an unreproducible way, introducing a further source of uncertainty, and therefore large molecules should be removed as much as possible at this stage, or they will need to be managed at a later stage by applying NMR acquisition or post-processing techniques that can have a more detrimental effect on the remaining signals.

Most of the foodstuffs that will be tested by QC analyses originate from living systems, which undergo time evolution by their nature. This is a factor that should be taken into account for the samples to be actually representative: apart of all the issues related to the gathering of raw materials (like times and conditions of harvest or collection, storage, transport, and processing) that are upstream to the laboratory, the metabolites content of the food must be kept as steady as possible by blocking the biological reactions that are still occurring, that is, by quenching the enzymatic activity.

Generally, in liquid food, for instance fruit juice or tomato paste, the enzymes are inactivated by exposition to high temperature through manufacturing processes for sterilization (e.g. pasteurization or UHT). Consequently, these foodstuffs can be analysed without any metabolic quenching. The other way round, solid or fresh liquid foods require different steps of sample handling in order to achieve a liquid sample containing the soluble components of the foodstuff (Figure 1).

Quenching can be performed by low temperature (liquid nitrogen bath, blast chilling) or by solvent addition: some of these steps are also preparatory to the following procedures—indeed, after liquid nitrogen bath the sample is suitable for powdering, while solvent addition can be both a quenching and an extraction step.

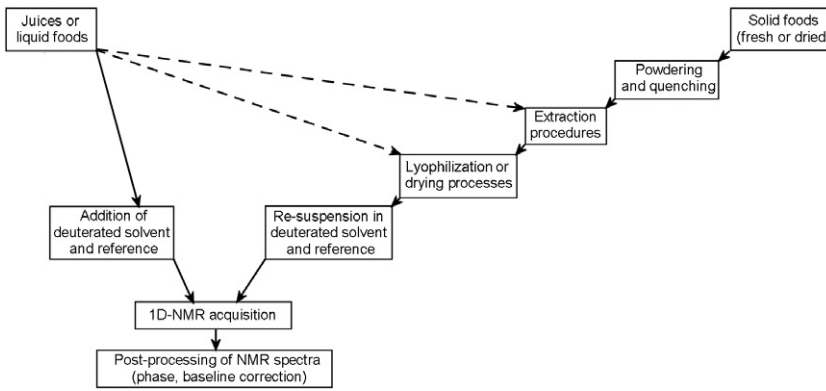

FIGURE 1 Procedures to obtain NMR data from raw foodstuffs.

Currently, there is not any univocal procedure used as a standard: different procedures of extraction have been used depending on the characteristic of foodstuffs (for a comprehensive list of the extraction procedures used for food NMR analysis, we address the reader to a recent review of Mannina *et al*. [8]).

In the case of liquid foodstuffs, like fruit juices [9], or olive oils [11], or vinegars [12], the samples can be analysed without any further treatment or, at most, by procedures involving either to concentrate diluted samples (wine, for example) [13] or to dilute by adding suitable solvents for viscous liquid foodstuffs, like in the case of honey [14].

Sometimes it is worth to perform lyophilization or extraction procedures on liquid foods too, not only to yield more homogeneous solutions or to remove higher molecular weight metabolites as stated above, but also to have a strict control on concentrations: as a matter of fact, liquid foods can be regarded generally as solutions (or suspensions) whose concentration depend on a number of factors that are out of the examiner's control; furthermore, the presence of the intrinsic solvent (water in almost any case) is often an issue for ^{1}H 1D high-resolution NMR spectroscopy, that is by far the most common applied NMR technique, and its removal is generally a positive factor in the design of the experimental protocol. The choice of the eventual extraction procedure should be undertaken by considering the nature of the sample and the effects that extraction will have on it. For example, the co-presence of fatty and hydrosoluble metabolites could suggest a multisolvent extraction procedure, like chloroform/methanol/water, which gives rise to two distinct hydrophilic and lipophilic phases and, in turn, two separate samples that will be dried and redissolved in different solvents (typically deuterated water and deuterated chloroform, respectively) [15]. The use of aggressive chemical reactants as extraction agents might lead to the degradation of some metabolite, and the extraction conditions must be carefully controlled to avoid the underestimation of that metabolite and the overestimation—or the unpredicted presence—of its degradation products.

2.2 NMR Acquisition and Processing Parameters

As we stated earlier in this chapter, data yielded by any analytical technique depends on the experimental protocol and the characteristics of the specific detection method employed. NMR spectroscopy is a multiparametric technique, and the choice of the experimental set-up is critical to assure the accuracy, reliability, and representativeness of data.

NMR spectra are constituted by a series of resonance signals, each of them originated by the magnetically equivalent nuclei that undergo to the NMR experiment (i.e. hydrogen nuclei for ^{1}H NMR spectroscopy). Their position in the spectrum (chemical shift) is peculiar of the chemical group to which the nucleus is bound and of its chemical environment (chemical neighbourhood—both directly linked and spatially displaced, stereochemistry), their intensities

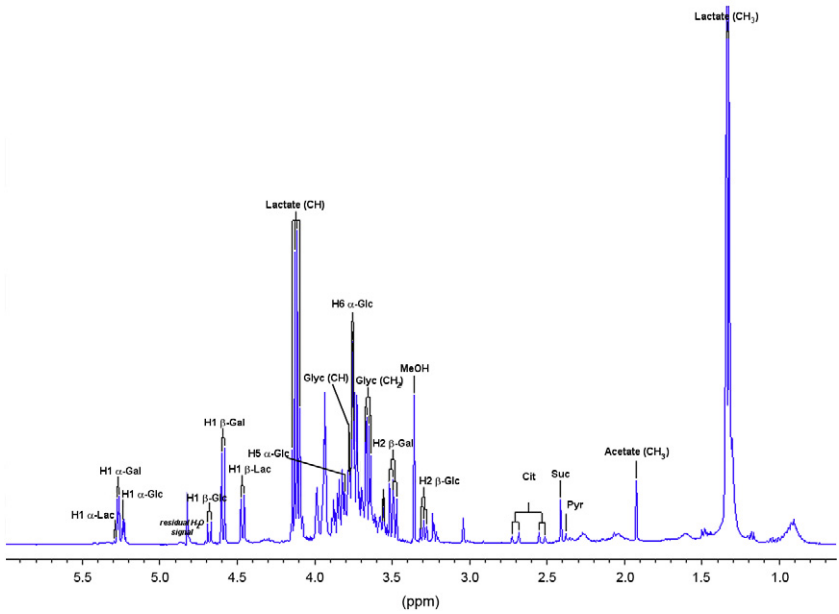

FIGURE 2 NMR spectrum (0.5–6.0 ppm region) of the aqueous extract of mozzarella cheese: signals relative to the most relevant metabolites are tagged.

(areas) are proportional to the concentration of the chemical group, and thence of the molecule. Since any metabolite is composed of more than one chemical group having hydrogen atoms, it comes out that each metabolite is represented by more than one NMR signal—indeed, it usually gives rise to an even more complicated pattern due to spin coupling. The ^{1}H 1D NMR of a mixture (Figure 2), like the solution derived from foodstuffs is, produces therefore an overcrowded ensemble of resonance signals, that is a snapshot of all the metabolites contained therein: it is often described as its fingerprint. Such a situation has some disadvantages and some advantages: many NMR peaks from very similar groups bound to different chemical species can be partially or totally overlapped, and thence difficult to be detected, assigned to a specific metabolite and measured; on the other hand, this redundancy can be helpful because each molecule has its own spectral pattern and can be recognized on this basis, and its quantification can be performed anyway by measuring a non-overlapped signal even if the remaining are entangled in intricate regions. Assignment of the overlapping signals can be made much easier by performing 2D NMR experiments that develop the 1D spectrum complexity into a second dimension.

It is straightforward that any possible experimental factor influencing the intensity of the signals must be carefully avoided or minimized, while the physicochemical environment of the specimens must be kept as much

constant as possible to ensure the reproducibility of chemical shifts throughout the whole experimental set.

Signal intensities can be affected by either intrinsic sample-related issues, like very broad signals spanning a wide spectral range due to the presence of macromolecules or a huge signal covering some part of the spectrum due to a very high ratio of concentrations between the solvent and the metabolites to be measured; or the choice of inappropriate spectral conditions (e.g. saturation of all or some signal due to short relaxation delays or inadequate solvent suppression parameters); or inadequate spectrum post-processing (e.g. phase and baseline correction). Each of these issues can be addressed by applying a correct experimental protocol: in the case of ^{1}H NMR spectroscopy, solvent signal can be greatly reduced by drying the sample and re-suspending it in deuterated solvents, or by using solvent suppression sequences; signals from macromolecules can be managed removing most of the high- or medium-molecular weight molecules by microfiltration or extraction procedures, or else by using NMR sequences that are able to minimize signals from heavy, slow-moving molecules, like spin-echo-based techniques.

The stability of chemical shifts along the experimental set is another crucial issue, since some signals are more prone to change their position within the spectrum than others: for instance, in ^{1}H NMR spectroscopy, the signals from polar groups can move upon the influence of charged ions, whose concentration determines the entity of the shift.

Depending on the different polarity of the compounds presents in an extract from plants or vegetables, the dried samples yielded by a solid foodstuff extract can be re-suspended in solvent mixtures, at different v/v ratios depending on the sample characteristic (commonly deuterated water and methanol). This can result in slight chemical shift differences of the resonances arising from the same metabolite among different samples originated from the same foodstuff, making them difficult to be compared. Another important step in the NMR sample preparation, which influences the chemical shifts of signals, is the final value of pH of the liquid sample, either reconstituted from dried extracts or already in liquid state. Therefore, it is important to control the pH of the samples.

As it can be seen, the experimental procedure can have qualitative and quantitative influences on the comparability among data obtained by NMR spectroscopy. In any case, each step of the laboratory protocol implies some losses and introduces a degree of uncertainty associated to the data that will be obtained: preliminary experiments should always be planned to acquire the necessary knowledge of the system under analysis and of the experimental pitfalls that come with it, and finally to define a scheme that is the best compromise.

Post-processing of NMR spectra is a multistage procedure basically comprising Fourier transform, phase correction, and baseline correction. Each of these steps can be done automatically but, since they can greatly influence

the following quantification of the NMR signals, they should never be applied uncritically. In particular, the results of the application of phase and baseline correction to a series of spectra should be at least inspected and, in case, corrected by the operator. Phase correction follows from the phase uncertainty given by the Fourier transform algorithm, and a bad correction can lead to frequency-dependent, that is, chemical shift-dependent, errors. Baseline correction is maybe even more critical, since it directly affects the absolute values of the NMR signal integral (area). The very basic, mandatory baseline correction is DC drift correction (baseline zeroing), but many more algorithms are available (polynomial interpolation, fid reconstruction, linear prediction) to cope with the specific baseline patterns associated to the spectra [16–18].

A standardization of the experimental protocols to obtain standard sample is desirable in order to compare database acquired in different laboratories. In this context, several initiatives, carried out by the Standard Metabolic Reporting Structures group, were focused to produce draft policy documents covering all aspects of a metabolomic study that are recommended for recording, from the raw biological sample, the analysis of material from that sample, and chemometric and statistical approaches to retrieve information from the sample data [19–21].

The use of 1D NMR spectroscopy as a high-throughput analytical technique for metabolomic studies has been very enhanced by the introduction of automated procedures: autosamplers allow to handle more than a hundred samples per day with no human attendance, while auto-tuning and auto-shimming routines assure an average good quality of the spectra, although the best performance of the spectrometer is not reached unless each step of the analysis is optimized by the operator. If this is the case, anyway, it is worth noting that the optimal spectral features, like signal-to-noise ratio and resolution, must be achieved for every single spectrum in order to have comparable data: the homogeneity of those spectral features is a priority since they affect the signal quantification, and should be pursued rather than having some perfect spectra and some average ones.

2.3 Targeted Analysis and Pattern Recognition

Quantification is the passage that transforms a set of NMR spectra into a data matrix. This task can be performed following two complementary routes (Figure 3): either by resolving the signal forest assigning each (or most) of them to the proper metabolite—this implies that we *a priori* assume to know which molecule can be present in the mixture—and measuring the intensities of the specific peaks of interest (targeted analysis) or by processing the spectra "as they are" regardless to any other consideration and relying on the correlation structures embedded in the ^{1}H 1D spectrum due to its intrinsic redundancy (pattern recognition).

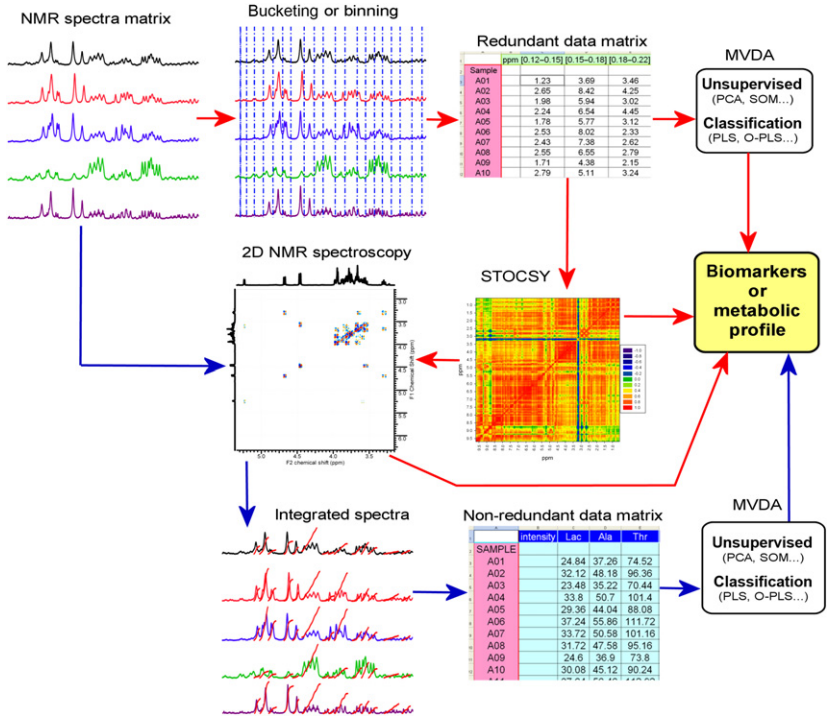

FIGURE 3 Strategy for NMR-based metabolomic analysis: pattern recognition (red route) and targeted analysis (blue route).

This latter strategy is performed on the whole set of spectra at once and requires a series of automatic, or semi-automatic steps that can include peak alignment [22,23] and then discretization by subdividing the spectral range in small fractions (bucketing or binning). The first, more obvious binning procedure divided the spectra in regular intervals of equal size: this can lead to unreliable results if peaks are not perfectly aligned and do not have similar linewidths in each spectrum. More sophisticated algorithms have been developed [24,25] to address this issue and are included in the most popular NMR analysis software packages. The result is a data matrix that embeds all the information included in the spectra, but whose variables (spectral intervals) do not have a direct metabolic meaning: multivariate data analysis is then applied for variable reduction to unravel the structural and metabolic relationships subtended by the spectra.

Both targeted analysis and pattern recognition method require the assignment of the resonance signals to their source metabolites in work flow (Figure 3). This procedure needs the operator's skill, which can be based on a large set of 2D NMR techniques (bidimensional homonuclear spectroscopy

like correlation spectroscopy and total correlation spectroscopy experiments or bidimensional heteronuclear spectroscopy like heteronuclear single quantum coherence or heteronuclear multiple bond correlation), on a number of either publicly available [26–28] or proprietary databases, on reference spectra, and on literature data. Some effort has been done to automate this process [29,30]. One more available tool is given by Statistical Total Correlation Spectroscopy [31]: variables from the unsupervised matrix are correlated pairwise and the correlation coefficients are arranged in a correlation matrix. The strongest correlations are due to signals belonging to the same molecules.

The variable of the data matrix yielded by supervised analysis are the real concentrations of the detected metabolites, and the application of multivariate methods can directly lead to the understanding of the metabolic system under study.

Both the "targeted analysis" and pattern recognition method contribute to provide the metabolic profiles of the foodstuff and to detect the biomarkers that characterize the biological processes occurring in it (for detailed methodological reviews, see Refs. [32,33]).

Pre-processing of data matrices from NMR spectra for multivariate analysis should be applied by keeping in mind the peculiar characteristics of their origin. Centring and scaling allow to equalize the importance of all the considered metabolites, and this is important since many of them are intermediates of the metabolic pathways, their concentrations are regulated by enzymatic patterns and are therefore generally low in respect of the dominant signals from end products. On the other hand, NMR data are always positive quantities and centring introduces a factitious threshold that should be taken into account in some cases, while scaling can overestimate the variations of small signals because of stochastic fluctuations: in this respect, Pareto scaling can represent a good compromise. After these steps, multivariate techniques are applied for detection—in an unsupervised way to pick out intrinsic correlations in the data field and/or the presence of outliers (commonly by PCA, as first) or else supervised against known external variables (like with PLS, OPLS, MCA) or for classification (PLS, SIMCA, multiway analysis [34–36]).

3 NMR-BASE METABOLOMICS APPLICATIONS

3.1 Food Quality Control

The term "quality control" (QC) in foods can indicate different concepts at two different levels of complexity. In a broader sense of acceptation, QC may signify the adherence to the ever stricter standards controlling the food chain promulgated by official authority as, for example, the European Union. In this regard, food safety control represents the major challenge and pertains principally to finished products. The EU integrated approach to food safety aims to assure a high level of food safety, animal health, animal welfare,

and plant health within the European Union through coherent "From Farm to Fork" measures and adequate monitoring, tracing the different stages of the food chain system and examines the practices and procedures that ensure the safety of our food. As a consequence, the QC at these levels is mainly represented by the detection of toxic/harmful or adulteration substances, either voluntarily or not, present in the finished product or formed during the preparation process.

Otherwise, the intra-factory QC may also concern a control of the starting raw materials and the stability of products during processing, ideally leading to a "golden standard" that can differ from factory to factory.

Foods can be grossly divided into vegetable- or animal-derived products. They are characterized by an intrinsic complexity and heterogeneity of the matrix that is reflected in the presence of several classes of metabolites with different chemical properties and wide concentration ranges, which must be preserved as much as possible during the analysis. In this respect, NMR-based metabolomics is more suitable for the application on liquid products (i.e. juices, beverages, etc.) where the samples are practically ready for the analysis requiring a dilution step at most. On the contrary, solid foodstuffs have to be extracted to obtain a liquid representative sample.

The complexity of the chemical analysis is well described when, for example, we have to assess the quality of plant-derived raw materials or products. In fact, natural products are characterized by a wide range of metabolite composition linked to factors like structural variation in plant organs, different cultivars, and/or species, geographic origin, growing sites and seasons, agronomic treatments, developmental stages, climatic differences in production sites, and many more. The same complexity can be found for animal-derived products, either liquid (i.e. milk) or solid (i.e. meat, cheese), where the metabolite composition reflects species difference, of course, but also different livestock farming and feeding.

However, it must be underlined that an agreement on "omics"-derived food standards is presently still lacking and needs, as a first step, the standardization of experimental designs, technical protocols, and results presentation for a comparative analysis and toward the preparation of common databases.

Furthermore, when evaluating metabolomics data, it should be advisable to consider that statistically significant differences between different foods do not necessarily imply better or worse quality or altered nutritional values, representing biologically meaningful effects. For example, one can find large differences in metabolite levels but in minor food constituents or statistically significant differences due to relatively small differences in metabolite levels.

Strict standardization protocols will be a prerequisite when considering the relative new world of "nutraceuticals" and "functional foods" that, from a legal point of view, fall in a grey area between the food and the drug categories. The term "nutraceutical" has been defined as "a food or parts of food that provide medical or health benefits, including the prevention and/or

treatment of disease". In a broader acceptation, the nutraceuticals include food supplements, food ingredients, fortified foods that are subjected to strict ever-changing regulations particularly when health claims are involved [37]. Here the quality and safety control, besides the efficacy one, is of paramount importance and prefigures the identification of bioactive components, their quantification, and their stability in complex mixtures as for any plant-derived food. Furthermore, in the substantiation of health claims by the European Union, two key considerations are specified, namely "the extent to which the food/constituent is defined and characterized, and the extent to which the claim is defined and is a beneficial physiological effect" (Regulation (EC) No. 1924/2006 of the European Parliament and of the Council of December 20, 2006 on nutrition and health claims made on foods) [38].

From an historical point of view, the first studies that applied a metabolomic approach mainly regarded the discrimination of botanical species, cultivars, and clones. Since then, several papers appeared in the literature both on raw materials (e.g. grape, tomato, lettuce, melon, etc.) and on finished or semi-finished products (juices, oils, cheeses, wines, beers, etc.). It is out of the scope of this chapter to discuss all the single evidences that can be found well reported in many recent reviews [1,4,5,7,8,39].

Here, two studies will be reported as an exemplification of the potentiality of the metabolomic-based approach in food analysis; the first one is discussed because it led to a patented method (Bruker JuiceScreener™ for SGF Profiling™) for juice authentication, and the second one because it showed the potentiality to discriminate different cultivars of the same botanical species (kiwi fruits) and to follow the whole ripening process through metabolic fingerprinting.

SGF Profiling™ represents an NMR-based screening method for the quality control of fruit juices [9,40]. Through this system where sample transfer, measurement, data analysis, and reporting are fully automated, a huge spectral reference database from more than 6000 samples of 50 different types of fruit juices from more than 50 countries has been built. Then the data has been analysed through both a targeted and non-targeted approach. Through the first approach the identification and quantification of single compounds for each type of fruit juice were provided. The fruit juice industry has established a Code of Practice in Europe, which contains concentration ranges of several indicative compounds for different types of juices. SGF profiling compares the values found by NMR and indicates if a concentration is within the range defined by the Code of Practice. Furthermore, NMR allows additional information to be collected for new indicative compounds not listed in the Code of Practice. In fact, any deviation in the spectral profile could imply quality problems and/or adulteration like, for example, the addition of sugars or citric acid to the juices or the usage of unripe fruits.

An exhaustive statistical analysis was applied to the same data in the untargeted approach. First of all, the amount of data of the spectra was

reduced via bucketing, and a series of statistical tests and classification/discrimination steps were done leading to the differentiation of 20 types of fruits (namely apple, orange/mandarin, sour cherry, pineapple, black currant, passion fruit, lemon, grapefruit, banana, peach, raspberry, strawberry, pear, apricot, mango, guava, carrot, elderberry, pomegranate, and grape). The applied statistical method was a combination of PCA and discrimination analysis, and the accuracy was checked via cross-validation and Monte Carlo analyses.

This type of classification could allow the detection of frauds as, for example, the dilution of orange juices with the cheaper mandarin ones that is difficult to be discovered with conventional analysis. On the contrary, the proposed NMR-based method can detect mandarin juice at a level of 10% or more and therefore is suggested as a valuable pre-screening method prior to DNA- or other conventional analyses. Furthermore, more specialized models could distinguish between direct juice and juice from concentrate and could detect the origin of the fruit. As an example, the results for estimation of origin for a particular orange juice sample are showed in Figure 4. In this case, the possible sources of orange juices included Spain, Greece, Brazil, Belize/Mexico/Costa Rica, Cuba, and USA. A 3D projection of the discrimination model space represents the ellipsoids of probability for each source, and the sample of interest is marked by a star. Similarity factors were calculated in the complete discrimination space and provided the probabilities for the estimation of source. It has to be underlined again that the success in the applicability of these type of approaches is strictly linked to the collection of a huge reference dataset; the fact that the majority of samples in the presented study were from orange and apple juices has allowed the separation of these juices in additional subgroups related to their geographical origin and their production from direct or

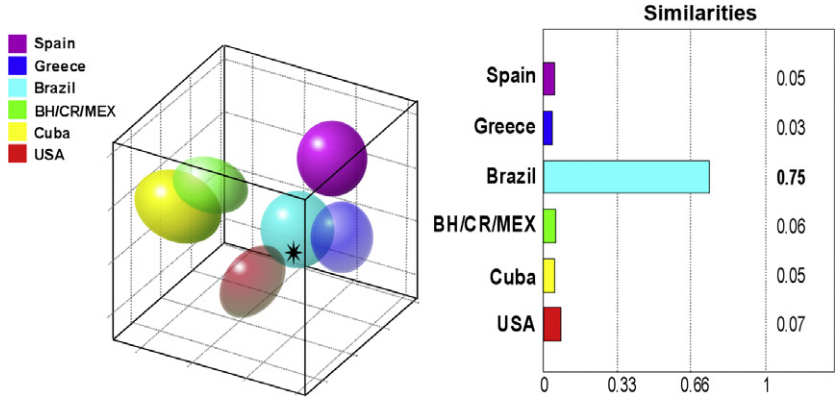

FIGURE 4 Estimation of the exact variety of origin of an orange juice. The left plot shows a 3D projection of the model space with the ellipsoids of possible groups. The star symbolizes the actual sample. *Modified from Ref. [9].*

concentrated semi-finished material. It has been calculated that at least 30–40 samples of each subclass are needed to start building models for more rare juice types [9,40].

In the second type of approach, the metabolic profiling of aqueous extracts of one cultivar (Hayward) of kiwifruits (*Actinidia deliciosa*) was used to investigate the kiwifruit composition at different harvesting times over a 7-month period [41]. The qualitative and quantitative analyses of metabolites in kiwifruits are very important for commercial success as flavour, fragrance, and healthful properties are obviously linked to the fruit chemical composition. In particular, the balance between soluble sugars and non-volatile organic acids influence the flavour of the fruit flesh, while the different sugars are responsible of different sweetness levels and organic acids give a different perception of the acidity. Furthermore, the industries that extract specific compounds from the fruit heavily rely on the knowledge of nutritional profile of the fruits. The biochemical composition of fruits determines the shelf life too. In fact, a fruit harvested too early has a low level of carbohydrates, whereas a fruit harvested too late may be subjected to mould, with a resulting reduced shelf life.

The NMR analysis clearly showed the presence of some metabolites, although in different concentrations, during the whole growth period while other metabolites were characteristic of particular months. A data matrix was built from the intensities of each selected NMR signal where samples were represented by a row and each NMR signal was represented by a column. The harvesting months were added as the "external" variable, against which the multivariate regression was performed. ^{1}H resonances from 29 metabolites always present during the investigated period were selected and the obtained PLS model was validated through a full cross-validation test procedure. The significance of the correlation between the original variables and the PLS models was assessed by the uncertainty test. Two LVs account for 95% of the variability within the data. A clear grouping of the samples according to the harvesting month was shown with early samples (June, July, and August) separated mainly along LV2, while the later samples appeared separated mainly along LV1. The findings suggested August as the key month for the kiwifruit development, showed by a drastic change in the metabolic trend.

Then two further PLS analyses were singly performed on the early (collected in June, July, and August) and late (collected in September, October, November, and December) data in order to investigate in more detail the growth and the ripening processes. In these analyses, the metabolites present only in the first or in the last harvesting period were also taken in account. In the first PLS, the first two LVs account for the 97% of the variability within the data. In the second PLS, the first two LVs account for the 96% of the variability within the data. A detailed definition of the processes occurring during the fruit growth was obtained from the PLS analysis over the early

stage where, while LV1 fully confirmed the findings from the previous analysis on the whole time set, LV2 also involved metabolites present only in the early period like, for example, those involved in the shikimic acid pathway. In a following paper, three *Actinidia* varieties characterized by different developmental times were investigated [42]. Beside the Hayward cultivar, Zespri™ Gold kiwifruits and a crossbreed from different species of *A. deliciosa* (CI.GI.) were used.

Zespri fruits are usually fully mature about 1 month earlier than Hayward fruits, whereas the developmental time of CI.GI. fruit is intermediate between Zespri and Hayward.

In Figure 5, the means and standard deviations of the PCA scores at each stage of development relative to the first two principal components obtained on the whole data matrix of the common metabolites are shown. The earlier ripening of the Zespri variety was observable by the evolution of the scores along PC1 that always preceded their Hayward and CI.GI. counterparts. These last two varieties were characterized by metabolic profiles substantially coupled along the first as well as the second latent variables, in agreement with the common origin of CI.GI. and Hayward that are both varieties of *A. deliciosa*. In this study, the stage of kiwifruit development and kiwifruit cultivar was used as external variables along which the multivariate regression was performed through PLS2, an extension of the standard partial least-squares (PLS)

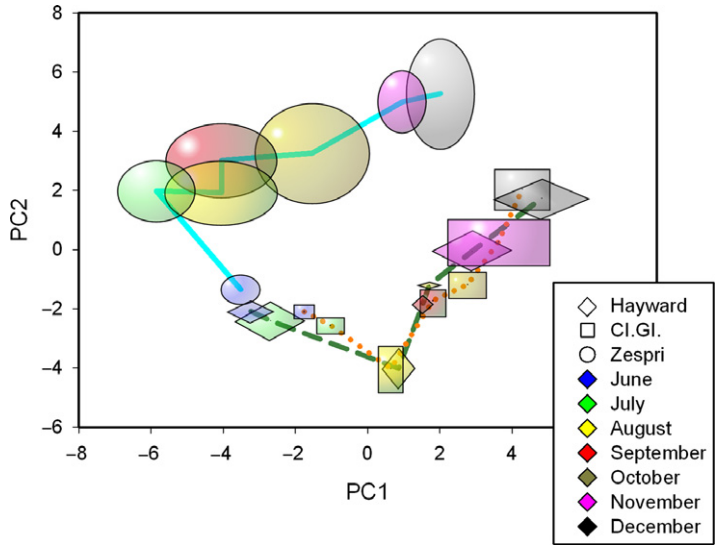

FIGURE 5 Scores from PCA on the metabolites common to Zespri, CI.GI., and Hayward kiwifruits (modified from Ref. [42]). Data are expressed as means and standard deviations of the scores for each cultivar and month: the centre of each ellipsoid represents the mean values, the widths of the ellipsoid axes are the standard deviations along PC1 and PC2.

analysis. This analysis gives a single set of scores, a single set of loadings, and separate sets of correlation coefficients for each external variables. Again, the effect of the developmental stage was visible as a trend along LV1 with the samples collected from June to August grouped according to the harvesting time along LV1 and LV2, in agreement with the results reported in the case of Hayward kiwifruits.

In the following chapters, some application examples on quality control issues in geographical origin discrimination, safety, and adulteration are discussed. However, the proposed division in subchapters is somehow artificial. In fact, if a food can be characterized depending on its geographical origin through the acquisition of a collection of metabolic profiles, it follows that this food can obtain an authentication. By fixing a range of acceptable natural variations through, for example, different year crops, one can readily be warned about adulterations that will be revealed as outliers in my statistical cloud. If such a process would be applied in the future starting from raw materials to the finished products and further to the food shelf life, a real "From farm to fork" quality control will be obtained.

3.2 Quality Control: Geographical Origin and Authentication

Several metabolomics studies have been focused on quality control and geographical origins of wines reflecting the commercial interest in the exclusive use of particular brands and vinegars leading to a novel debut of a genuine "wine science". Protected Denomination of Origin (PDO) and Protect Geographical Indication (PGI) are legal terms recognized by the European Union to protect high-quality foods.

In this regard, metabolomics can provide valuable information ranging from chemical composition, production area, climate influences, and grape varieties to processing and vintage contribution to the overall quality (for a comprehensive review, see Ref. [39]). The NMR analysis of wines represents a good paradigm of problems to be faced. Wine samples are usually analysed as such, with just the addition of buffered solutions, or after lyophilization: the first used to minimize changes in NMR chemical shifts induced by variation in pH, the second used to avoid the overlapping of metabolites with the ethanol peak. However, volatile metabolites are lost during lyophilization, or multiple solvent suppression sequences have to be used to get rid of the ethanol peak in buffered samples.

Geographical dependences of wines in samples from Italy, France, Australia, California, and Korea have been reported, and the application of NMR-based multivariate statistical analysis has successfully allowed the association of wine metabolites with wine quality not only on the basis of geography, but also depending on climate, soil, sun exposition, or rainfalls ("terroir" effect) [13,43–46].

In a study of Son *et al*., pattern recognition methods applied to ^{1}H NMR spectra were used to investigate the influence of different regions in relation to metabolic differences in grape pulp, skin, and seeds and in wines from South Korea [13]. The large dominance of sugar resonances in ^{1}H NMR spectra of pulp and skin extracts prevented the development of PLS-DA models with high predictability; after the exclusion of the involved regions the Q^2 value increased to 0.84 resulting in a significant separation. Then, PLS-DA models were applied to ^{1}H NMR spectra data sets of wines aged for 3 and 6 months, leading to a good region-dependent separation. Interestingly, the authors could correlate the metabolic clustering of grapes in relation to microclimate parameters (sun exposure, high/low rainfalls) showing different levels of sugar, proline, malate, citrate, alanine, threonine, and trigonelline in grapes. Similar environmental effects were also observed in the complementary wines, suggesting that the proposed approach can be used to determine appropriate harvest times for grapes according to individual circumstances.

A metabolomic approach was also shown as a very useful one in the field of beer characterization, allowing a classification on the basis of different raw materials (malt), beer types, brewing sites, or even production times of the same brewery [47–49]. Beer is obtained from malted grains, hops, and yeast through a fermentation process; other components (fruits, herbs, and spices) can then be added to produce a particular blend. Different fermentation processes distinguish the two principal types, namely ales and lagers, and alcohol-free beers are produced too.

In an exhaustive study of Lachenmeier *et al*., metabolic fingerprinting and PCA allowed the classification of beers in relation to types and origin; furthermore, beers made from barley or wheat malt could be distinguished as well as brewing sites [48].

As observed with NMR analysis of wine, sugar and organic acids signals showed small 0.002–0.003 or 0.06 ppm shifts, respectively, due to pH sensitivity of the aliphatic region, in particular. This problem causes a decreased ability of the multivariate analysis to classify the data and was overcome through buffering all the samples. A bucket table was then generated as an input for PCA, scaling to total intensity was applied and the number of principal components needed was determined applying a cross-validation technique. By plotting the data defined by the two or three largest PCs, it was possible to highlight the similarities and differences among the spectra of 80 beers. The plot of the first two PCs in Figure 6 (95.5% of total variability expressed) displays a large group of 76 samples with 4 outliers; the bucket 4.20–4.24 ppm was identified as the cause of clusterization in the four samples and was ascribed to lactic acid. A microbiological analysis revealed anomalous numbers of beer-spoilage lactic acid bacteria in these samples that, furthermore, originated from home breweries. Thus, the metabolomics analysis could discriminate beer origin in terms of different breweries and, more

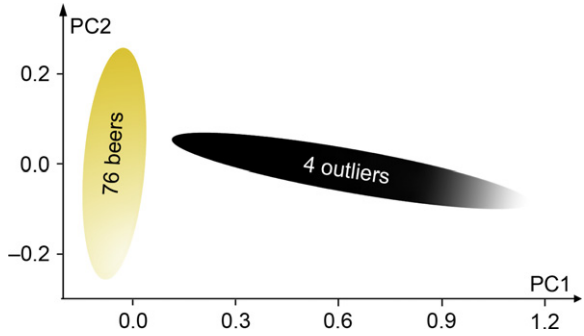

FIGURE 6 Scores grouping of the PCA from beer samples (0.4–9 ppm region of NMR spectra). *Modified from Ref. [48].*

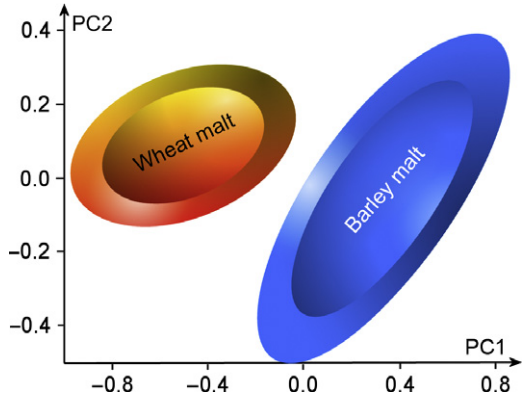

FIGURE 7 Scores grouping of the PCA from barley and wheat beer samples for the aromatic region (6.7–7.6 ppm) of the NMR spectra. *Modified from Ref. [48].*

important, could evidence deficiencies in hygiene during processing, probably due to the recontamination after wort boiling. Interestingly, if the beers from the same brewery are tagged in the scatter plot, the corresponding points were close together. The possible causes of this clustering were ascribed to differences in raw materials, namely the particular water used, the origin of malt, the strain of the yeast [48].

By performing PCA on buckets from specific NMR aromatic regions (7.6–6.7 ppm; no bucket exclusion; 0.01 ppm range), a differentiation between barley and wheat malt types was observed (Figure 7). Wheat beers cluster in a region of negative PC1, whereas the beers made with barley malt (Pilsener, Export, and Alt beers) give scores of positive PC1 or PC1 around zero. Interestingly, this PCA model could be used to verify the compliance with German beer purity laws that prescribe the declaration of the malt type [48].

In the German beer legislation, the beer categories also depend on quality parameters as the original gravity, the ethanol, and lactic acid contents. The authors correlated the NMR spectra with results from other analytical data for calculating these parameters by using the PLS analysis. The PLS regression model used two data tables where the *X* corresponded to the bucket table calculated from NMR spectra, and the *Y* contained extra information on the sample (extra variables).

For example, in order to determine original gravity, the bucket table of the entire NMR range (10.0–0.3 ppm), including ethanol, was correlated to the wet chemical analysis. The PLS model, verified through cross-validation, showed an excellent correlation ($R^2 = 0.998$) between the standard laboratory values and the NMR prediction. A very good correlation was observed for ethanol and lactic acid contents ($R^2 = 0.985$ and 0.954, respectively) showing that the proposed approach is suitable as a fast and cost-efficient screening procedure since quantitative determination of essential compounds through NMR spectroscopy as well as chemometric classification are simultaneously possible.

As underlined above, some problems can arise when a metabolomics approach is applied to solid samples that need an extraction step before NMR analysis (see above). This is the case for tea analysis starting from leaves extracts. Teas can be grossly divided in three types on the basis of the fermentation process, namely green (unfermented), oolong (semi-fermented), and black teas (fermented), and their quality is assessed through appearance, flavour, and aroma. As the quality, and then the price, of teas depends on the judgment of professional tasters, there is a growing interest in a quality assessment based on chemical measurement considering, in particular, the health-related properties claimed, especially of green teas.

In this regard, a relatively recent but exhaustive work of Le Gall *et al*. can represent a good example of the problems to be faced [50]. A very comprehensive database of 168 green teas originating from different countries (145 from China, 7 from Japan, 6 from Indonesia, 5 from India, 3 from Vietnam, 2 from Bangladesh) were collected to study if the teas can be discriminated according to the country of origin by chemometric analysis of their ^{1}H NMR spectra. When analysing NMR data by PCA, it appeared that the pattern of covariance was seriously affected by some problem in that it described an "arabesque" style figure in both PC1/PC2 and PC2/PC3 score plots. The problem was due to a shift in some individual peaks (i.e. caffeine) from sample to sample depending on the association of caffeine molecules both with each other and with polyphenols in the used relatively polar solvents. Therefore an in-house Matlab macroprogram was written based on the principles of partial linear fit to allow the local alignment of the peaks within the data matrix. The improvement in the PCA was significant and showed the need of the preparation of NMR data and the importance of peak alignment before chemometric analysis, in order to obtain reliable results in particular situations and, above all, evidences again the need of an international

scientific agreement toward the issuing of standardized protocols in the field of NMR sample preparation and data analysis.

Notwithstanding the applied correction, and although a clustering of the scores of non-Chinese teas was observable in the PC1/PC2 plot (41% of explained variance), the Chinese tea scores were scattered evenly within the plot with no further obvious clustering according to the country of origin, probably due to the small number of samples: this fact underlines the need to balance the amplitude of the samples in metabolomics studies.

For this reason, in the same work, a hierarchical cluster analysis was applied to a geographically balanced set of teas (10 Chinese and 23 non-Chinese). An agglomerative clustering procedure based on standardized Euclidean distances and the nearest neighbour (single-link) method was used, with the PC scores of the samples as the inputs for the cluster analysis. An improved effect is achieved by using more than 3 PCs, as suggested by the figures for cumulative variance. A number of samples remained clustered regardless the number of PCs used. The outputs (dendrograms) differed from each other highlighting the influence of the geographical origin on green tea classes [50].

In this study, both PCA and cluster analysis indicated that the green teas could only partially be classified according to their country of origin, because other factors, namely the genetic strain, climate, altitude of growth, type of soil, use of fertilizer, or even tea leaves processing, can influence the chemical composition, simultaneously affecting the multivariate analysis results. Again, this underlines the need to carefully plan the experimental data field in metabolomic studies, beside the amplitude and balance of the sample size.

In a similar study, non-targeted ^{1}H NMR-based metabolomics and a pattern recognition technique has been applied for the quality prediction of Japanese green tea [51]. Fifty-three green tea samples from dried leaves were ranked in high and low quality upon the judgment of professional tea tasters before analysis. The PCA score plots showed clustering of green tea samples related to the high and low qualities in the second and third PCs only when considering a middle-frequency region of the ^{1}H NMR spectra. Ignoring the overlapped sugar peaks, the loading plots from caffeine, theanine, and cathechins contributed most to the cluster separation; these same compounds corresponded to the major taste-related metabolites that are usually used to evaluate the quality of tea. Furthermore, a good quality model to be employed to evaluate unknown tea samples was obtained when the tea quality ranking was used as a dependent variable in a PLS regression model ($R^2 = 0.987$; $Q^2 = 0.671$).

The metabolomics approach is also suitable to investigate micro-geographical variations as shown in a recent study of Ohno *et al*. [52]. In this study, black tea leaves cultivated in four regions with different elevations in Sri Lanka were harvested and aqueous solutions of tea, extracted from the dried tea leaves with boiling water for 4 min, were used for 1D and 2D

NMR analyses. Each ^{1}H NMR spectrum was subdivided into regions having an equal bin size of 0.04 ppm over a chemical shift range of 0.5–9.0 ppm and, excluding the region around the water signal, each spectral bin was integrated. The PLS-DA modelling indicated a 42% of variance contribution rate and 29% predictability in the multidimensional space for LV1 ($R_x^2 = 0.42$, $R_y^2 = 0.33$, and $Q^2 = 0.29$). On the other hand, LV2 gave a low contribution rate of 14% ($R_x^2 = 0.14$). The results suggested that the spectra collected from the highest cultivation site (>1800 m) were separated from the other ones and likely reflected differences in chemical composition. Potential components separated by difference of elevation were subsequently identified using *S*-plots, which were represented with covariance (p) against correlation (p_{corr}). The separation between the highest and three other cultivation sites by LV1 in the PLS-DA score plot was mainly attributable to two compounds, TF (theaflavin) and TFG (theaflavin 3,3′-digallate), and the separation among the lower cultivation sites by LV2 was due to three compounds, TRG (thearubigin 3,3′-digallate), caffeine, and theanine. Interestingly, these differences could be ascribed to microclimatic geographical differences. In fact, although the fermentation process in each factory was almost identical, some elevation-linked factors, such as oxygen levels, humidity, and temperature, might affect the fermentation condition, resulting in differentiation of the composition in black tea. In manufacturing at very high altitude, the insufficient oxidation process may result in the formation of TF and TFG, which are formed from polymerization of catechins at this stage, concomitantly with the low amount of TRG. On the contrary, higher levels of TRG that result from the enzyme-catalysed oxidation of catechin via the formation of TF or TFG, characterized the lower altitude teas [52].

Variations of theanine, isoleucine, leucine, valine, alanine, threonine, glutamine, quinic acid, glucose, epicatechin, epigallocatechin, epigallocatechin-3-gallate, and caffeine levels were responsible for the discriminations of green teas cultivated in three different growing areas of South Korea [53]. The results were correlated with climatic conditions like high/low temperature, long/short sun exposure time, and high/low rainfall and, interestingly, demonstrated a differently stimulated synthesis of theine and caffeine during the spring season. It should be underlined that, although some sample models between two different areas were overlapped in the PCA, all samples were clearly separated in the OPLS-DA model indicating that uncorrelated variation in X variables (metabolites) to Y variables (growing areas) could be removed, resulting in maximum separation in the OPLS-DA model.

Many studies have demonstrated how NMR-based metabolomics constitutes a powerful tool for the geographical characterization of olive oils on Mediterranean, national, and regional scales (for an exhaustive review, refer to Ref. [54]). These studies will be applied to comply to an important act of legislation, the PDO [55] that allows the identification of some Mediterranean extra-virgin olive oils with the names of the geographical areas where they

are produced, assuring, in the meanwhile, a high-quality product. In this regard ^{1}H NMR spectroscopy data, analysed through PLS-DA and SIMCA modelling, could confirm the declared geographical origin of oil in samples from Liguria (Italy) chosen as a case study among 896 samples from the Mediterranean area (other Italian regions, Spain, Greece, France, and Turkey) during three harvesting years [56]. In order to build a model able to discriminate the Ligurian from all other oils (global model), samples from all harvesting years were included in the analysis to represent the maximum variability, and samples used as training or test set were randomly chosen. The balanced training set and the test set were built using 140 Ligurian and 132 non-Ligurian, and all the other samples (624) not included in the training set, respectively. A very good classification accuracy (>85%) was obtained in both sets through PLS-DA, corresponding to 15 and 18 Ligurian samples being misclassified in modelling and cross-validation, respectively, with comparable values when this model was used to predict the test samples (prediction accuracy). When the data were analysed by the SIMCA class-modelling technique, a 84.3% sensitivity (118 of 140 Ligurian samples accepted by the model) and a 70.6% specificity (92 of the 132 samples from the other regions rejected by the model of Liguria) were obtained.

A good discrimination of virgin oils from various Mediterranean areas in terms of national, regional, and PDO levels was also obtained by Alonso-Salces *et al*. [57] through ^{1}H NMR fingerprinting, isotopic ratios from ^{13}C and ^{2}H data, combined with PCA, LDA, PLS-DA, and Probabilistic Neural Networks with this last showing the best results.

Interestingly, a metabolomics approach was able to discriminate the geographical origin also directly by using a semi-finished product ready for reprocessing as the starting material [58]. In the study of Consonni *et al*., triple concentrated tomato paste samples coming from Italy and China were analysed by ^{1}H NMR spectroscopy and a significant discrimination was obtained by combining NMR data from buckets and PCA and OPLS-DA. A clear sample separation was already present by using the unsupervised PCA model. Chinese samples were characterized by much higher levels of citrate (bucket 2.43 and 2.47 ppm) than in Italian ones, while Italian samples appeared to be strongly characterized by the glucose and fructose content (buckets 3.75, 3.67, 3.59, 3.35, 3.27, and 3.15 ppm). However, a geographical discrimination based only on citrate content could be misleading, as this compound can be added for pH correction and to avoid bacteria growth.

Therefore, a new training and test sets were extracted by excluding citrate buckets from the original data set: the new model revealed more buckets responsible for sample discrimination, in particular, those assigned to aspartate and glutamine characterized the Chinese samples, while Italian samples were still characterized by higher levels of both glucose and fructose.

Since the study was done considering samples from a single harvesting year, in a following work the influence of different production years as well

as different production stages (double–triple concentrated tomato paste) were investigated [59]. A sample differentiation, independent from both the concentration of samples (double- or triple-) and the tomato production year, was demonstrated by applying the bidirectional orthogonal projection to latent structures-discriminant analysis (O2PLS-DA). This technique is well suited for noisy and correlated variables and has been adopted to obtain robust classification models, having a clear interpretation of the systematic variation useful to characterize each class. The results evidenced that the sample variation related to the production year was negligible and dwelled mostly in the orthogonal space of the geographical discrimination model [59].

The preceding examples focused on plant-derived foods because still scarce information is available on animal-derived ones, with the exception of dairy products. In this field, a recent study of Jung *et al.* [60] aimed at assessing the geographical origin of beef obtained from four countries (Australia, Korea, New Zealand, and the United States) by using ^{1}H NMR spectroscopy, followed by multivariate analysis techniques. Up to now the determination of the geographical origin of beef has been mainly studied either using stable isotope ratio analyses [61] or an NMR-based metabolomic approach by HR-MAS solid state NMR spectroscopy [62]. Beef dried samples were extracted in methanol/chloroform and the polar phase was used for the analysis. The NMR spectra were phased, baseline corrected, aligned, and the spectral regions were divided into 0.01 ppm buckets after removal of the regions corresponding to the solvent signal. All imported data were Pareto scaled for multivariate analysis. A PCA model able to distinguish beef samples from different origins was established using seven components and an OPLS model was established using three predictive and one orthogonal component ($R_x^2 = 0.609$, $R_y^2 = 0.848$, and $Q^2 = 0.757$). The OPLS-DA score plots provided a paired comparison between origin countries and showed clear separation according to the first component. Some metabolites, namely succinate and various amino acids (isoleucine, leucine, methionine, tyrosine, and valine), were suggested as potentially markers to distinguish between various raw beef samples. However, the authors themselves underlined the need to be cautious in the establishment of reliable biomarkers as, for example, succinate is largely used as a substitute for salt in foods and, therefore, external factors such as feed type should be carefully considered. Furthermore, the metabolite profile of raw beef can be correlated with other environmental parameters beside its geographical origin like breed, feeding regimen, production system and even age, and weight at slaughter are factors that can influence the biochemical composition.

3.3 Quality Control, Adulteration, and Safety

Countless are the reports regarding food adulterations as many are the different kinds of compounds added to realize an economic advantage in the food market. However, such adulterations can be grossly divided into intentional

ones (the addition of beetle sugars or ethylene glycol to increase the ethanol content and the body in wines, respectively, the dilution with cheaper grapefruit or with pulp wash in juices, the addition of seed or ester oils in extra-virgin olive oils, among the others) or into unintentional ones. In this latter case, the contribution of the NMR-based metabolomics approach could represent a first choice approach to catch unintended adulterations allowing the simultaneous evaluation of hundreds of compounds without an *a priori* selection of the involved molecular class. It is worthwhile to underlie again how the already discussed assignment of geographical origin can be considered as an adulteration QC particularly when laws protecting national brands are provided.

A few selected examples will be discussed, starting from the work of Charlton *et al*. [63] where commercial carbonated soft drinks were intentionally contaminated with different amounts of paraquat, *p*-cresol, and glyphosate with the aim of verifying the usefulness of the methodology itself and establishing detection limits for forensic applications. A database of the spectra of uncontaminated samples was built to describe the normal range of product variation and contaminated samples were tested against this database through SIMCA modelling (8 PCs; 99.8% of total variance). Furthermore, this approach was applied both to the full spectral data set and to a reduced set of frequencies identified by the application of a variable selection approach. Both paraquat and glyphosate spiked samples could be successfully separated from the unadulterated samples using the first two principal components (strong outliers), while *p*-cresol spiked beverages were identified only when the PCs number were largely increased. As the discrimination was obtained only at very high concentrations (>10 mM), a variable selection algorithm was used to detect regions of the spectrum containing peaks that are characteristic of contamination, resulting in lowered detection limits (0.075, 0.2, and 0.06 mM for *p*-cresol, paraquat, and glyphosate, respectively). The authors suggested that a larger database covering different production sites and sample ages could lead to a more accurate database avoiding false positive results due to natural fluctuations in product composition.

An intentional adulteration has also been examined in honeys as the addition of sugar syrups is becoming an increasing type of falsification due to the scarcity of production. In the study of Bertelli *et al*. [64], 63 authentic honeys of different botanical origin and 63 intentionally adulterated samples, obtained adding 10%, 20%, and 40% of seven commercial syrups, were analysed. This represents a challenging falsification procedure to be detected as the raw material (honeys) and the added compounds (sugar syrups) are very similar in composition. Another strong point to be underlined is the very simple and quick preparation step of the samples that were only weighted and diluted in methyl sulfoxide-d_6. In this case, due to their high complexity, ^{1}H NMR spectra were not integrated but pre-processed through phasing and alignment and used as a direct data source by considering the intensity (quote)

of each data point: this can be thought as an extreme form of binning, where each bin comprises a single data point. The resulting matrix had 16K spectroscopic variables, 126 samples.

Then, the spectroscopic regions devoid of signals were deleted to reduce the number of data points. Factor analysis and general discriminant analysis were applied to the data set, and a leave-one-out method was used as a cross-validation procedure. 73.1% and 24.9% of total variance were explained by the first and second canonical functions, respectively, with the signals related to glucose, maltose, fructose, and sucrose representing the correlated variables. The model was able to correctly classify all of the authentic honeys, and all of the adulterated honeys were correctly misclassified. Furthermore, the method was so accurate to classify correctly the adulterated honeys in accordance with the syrup addition levels, with a prediction capacity of 90.5%.

The issues due to a similar chemical composition also represented a challenge in the adulteration of orange juices with pulp wash investigated by Le Gall *et al*. [65]. Pulp wash is obtained by washing the separated pulp with water and its "in-line" or "off-line" addition is forbidden at least in the European Union. Furthermore, citric acid, amino acids, or sugars can be added to mimic the chemical composition of the authentic orange juices. This particularly hard problem was handled by acquiring a very large database of the two "pure" products (263 pure orange juices samples; 50 pulp washes samples; 16 different country origins; 25 varieties; 7 years harvestings; hand squeezed or concentrated). To minimize the natural variation and in order to highlight the differences due to the pulp-wash addition, a stepwise LDA procedure was applied to the NMR spectra. In fact, in the classical LDA procedure the first PCs may account for the majority of variance, but the obtained information may not be sufficiently relevant to distinguish this kind of adulteration. On the contrary, in the stepwise LDA, the chosen first PC picked out by the model is the one with the highest classification rate and the other PCs are progressively added regardless of the proportion of variance associated with them. A six-PC model correctly classified the samples with 13/208 and 8/105 wrongly assigned samples in the training and validation sets, respectively. 84/88 authentic samples and 13/17 pulp washes were correctly predicted. Interestingly, the analysis of the loadings suggested a compound not previously recognized, namely dimethylproline, as a marker of pulp-wash adulteration of orange juices.

Another practice, that does not produce negative effects on health but is considered as a food adulteration, is represented by the blending of wines with a wine very rich in anthocyanins ("Rossissimo") or by adding anthocyanins extracted by black rice, in order to naturally increase the colour of red wines. However, the study by Ferrari *et al*. [66] deserves to be mentioned here because it not only compared two different techniques (Fourier transform near-infrared and ^{1}H NMR spectroscopies), but also applied a series of multivariate classification methods. In particular, PLS-DA was used as

classification method, then a variable selection/classification methods were applied, that is, iPLS-DA (interval version of PLS-DA) and WILMA-D (a novel algorithm based on the fast wavelet transform). The chemometric analysis was applied to aligned spectra intervals of the aromatic region that include the chemical shifts of phenolic compounds, in particular, the aglycone moiety of anthocyanins. Six iPLS-DA models were calculated (obtained on 20, 50, 80 variables intervals with forward or backward procedures), the best one reaching the 100% of efficiency although the prediction on the test set was surprisingly poorer than that obtained without any selection of variables. On the contrary, a classification efficiency greater than 95% in validation was obtained through WILMA-D modelling giving the most satisfactory results in authors' opinion.

NMR fingerprinting and multivariate analysis techniques have been also applied for the identification, discrimination, and quality analysis of herbal-derived nutraceuticals, products that were found to be counterfeited, causing ever-increasing concerns about their safety. The term "nutraceutical" was originally defined by Defelice in 1995 as "A food or parts of food that provide medical or health benefits, including the prevention and/or treatment of disease" [67]. As an example, Zaho *et al*. [68] assessed the chemical composition and quality of 4 *Hoodia* species and 10 *Hoodia* products sold in the market as herbal weight-loss supplements by NMR fingerprinting analysis, obtaining characteristic spectral signals. Then, the "spectral-feature-signatures" of two species, *H. gordonii* and *H. parviflora*, that represent the authentic and contaminating species, respectively, were subjected to OPLS-DA. While the NMR spectra from the two species presented a close similarity in the glycosides region, a different "spectral-feature-signature" was observed in sugar regions. When this differentiation was analysed by OPLS-DA, a model showing a clear separation between the two species along the predictive component dimension was obtained ($R_y^2 = 0.984$ and $Q_y^2 = 0.980$) with significant variables corresponding to sucrose and glucose that had higher values in *H. gordonii*. Having acquired the "spectral-feature-signature" from *H. gordonii* as a reference, the authors investigated ten commercial *Hoodia* products, sold on the market and claimed to consist solely of 100% pure *H. gordonii*. Six products showed significantly different spectral features and could be easily sorted as counterfeits according to their NMR fingerprints.

High-throughput sample screening techniques have been also applied to investigate their applicability in the control of illicit growth-promoting hormones to cattle. Up to now, the monitoring of illegal practices uses relatively simple and rapid techniques (i.e. immunoassays, chromatography, etc.) to identify potential adulterated samples, and subsequent confirmatory testing based on hyphenated gas chromatography or liquid chromatography (LC)-based MS analysis. However, both the drug types and the administration manners are continuously changing. In this regard, Graham *et al*. [69] have recently tested a metabolomic approach investigating plasma samples of beef

cattle intentionally treated with dexamethasone, prednisolone, and oestradiol. Two alternative NMR methodologies of plasma sample analysis were investigated, the first one requiring an extraction step and second one using ^{1}H Carr–Purcell–Meiboom–Gill (CPMG) pulse sequences without sample pre-treatment. From an OPLS-DA analysis, a good separation of the four animal groups (control and three different treatments) was obtained for both extracted and non-extracted samples, with LC1 explaining 11% and 17.6%, LC2 explaining 8.2% and 10.6%, LC3 explaining 68.5% and 5% of the variation, and a R^2 of 0.78 and 0.93, and Q^2 0.33 and 0.66, respectively. It was evident that the information obtained from the extracted samples provided a poorer clustering in comparison with the CPMG sequence data. However, cleaner NMR spectra can be obtained from the extracted samples, also requiring less processing times and allowing the identification of a wider range of metabolites. Interestingly, the metabolites altered in oestrogen-treated animals were quite different from those observed in dexamethasone and prednisone-treated ones, these last two being both corticosteroids and showing similar high and low levels of selected metabolites in respect to control animals.

3.4 Quality Control and Processing

In the field of food processing, the fermentation represents one of the most important steps considering the large number of fermented foods produced all over the world. Time-dependent metabolic changes during fermentation and/or ageing have been subjected to NMR analysis in wines, beers, vinegars, cheeses, and soy sauces [43,47,49,70–79], so only few selected studies will be presented here, each facing some particular problem.

One of these is the dominance of some compounds (e.g. ethanol, acetic acids) in the NMR spectra from fermented foods. In this regard, as we already stated, selective excitation is a useful tool to remove strong signals, thus allowing the detection of minor components. In a recent study, Koda *et al.* [76] obtained F2-selective TOCSY spectra of the peptide region of three rice wines types from Japan, China, and Korea. Rice wines are fermented alcohol beverages, in which brewing process rice starch is converted to glucose by a grain (e.g., rice, wheat) cultivated with mould (e.g., *Aspergillus oryzae*), and glucose is converted to ethanol by yeast. The combination of the obtained spectra with an unsupervised PCA provided a good model of classification of the different rice wines and was also capable to discriminate the type of grain used as the raw starting material (Figure 8). The authors suggested that the presented approach could be applied to other foodstuffs in order to obtain new "fingerprint" for quality control and food authentication.

A step by step metabolomic analysis of the fermentation process was investigated by Choi *et al.* [75] on a traditional Korean food, Cheonggukjang, that is the product of the fermentation of boiled soybeans with rice straw, and which is claimed to have antioxidative and antimicrobial bioactivities.

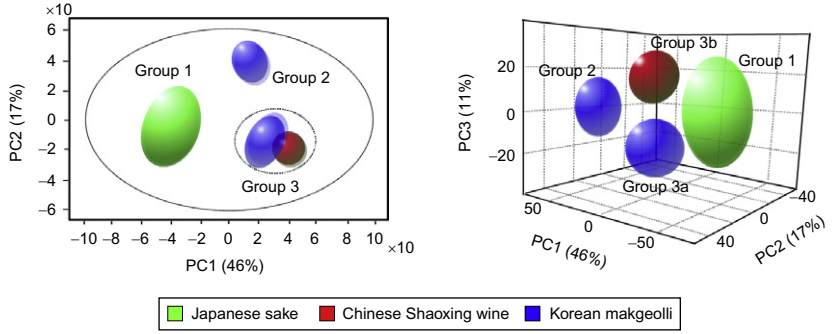

FIGURE 8 Grouping of PCA scores derived from the F2-selective TOCSY spectra of rice wines for PC1–PC2 (left) and PC1–PC2–PC3 (right). *Modified from Ref. [76].*

The first two principal components (PC1 and PC2) from PCA of the ^{1}H NMR spectra of the aqueous fraction allowed discrimination of Cheonggukjang extracts obtained at 0, 5, 10, 20, and 40 h after the start of the fermentation process and account for 98.5% of the total variation. Isoleucine/leucine, lactate, acetic acid, citric acid, choline, fructose, glucose, and sucrose were the main metabolites that contributed to discrimination of the different samples. Furthermore, because the claimed bioactivities of Cheonggukjang might vary according to the period of fermentation, the proposed approach could be used to determine the optimal period of fermentation in relation to different applications.

The metabolomic characterization of soy sauces presented in the paper of Ko *et al.* [78] deserves a particular attention because it compared both traditional and commercial, boiled and unboiled products up to 12 years of ageing, discussing the results also on the basis of the fermentative microflora composition. Indeed, the manufacturing processes differ for the production of Japanese soy sauce and that of traditional Korean soy sauce. The microflora involved in the fermentation varies between Japanese and Korean soy sauces too; therefore different kind and levels of metabolites produced from the enzymatic degradation of source material during fermentation or ageing can be expected. Commercial Korean and traditional Korean soy sauces aged for 1 year were selected for the investigation because they differ for high and low ethanol levels, respectively. The NMR spectra were either normalized to total spectra area or by dividing each spectrum by the median spectrum, and each normalization was carried out in the absence and presence of ethanol peaks. The results by PCA showed that the division by median spectrum was a better normalization method in metabolomic studies, particularly when dealing with samples that produce an abnormal spectral peak with a large area or high intensity in their NMR spectrum, like ethanol in the commercial soy sauces.

Unboiled soy sauces were clearly differentiated from boiled ones by PCA score plots with high statistical values of R_x^2 (0.85) and Q^2 (0.77).

Interestingly, scattered clusters of unboiled samples were evidenced in the score plots. This intravariation was related to the large number of microflora and suggested that unpasteurized samples could be of less verifiable quality.

Finally, metabolite levels varied as the ageing period increased; in particular, the traditional Korean soy sauces were clearly still undergoing ageing or fermentation until 12 years with a decrease in carbohydrate content and an increase in other metabolites. Among the discriminating metabolites, formate, betaine, glycine, glucose, and oligosaccharides have been suggested as markers for soy sauces that were aged or fermented for a prolonged period, being the main metabolites related to osmoprotection and growth/maintenance characteristics of the halophilic microflora involved in ageing or fermentation of soy sauce.

A similar approach has been applied in the study of Wu *et al.* [74] on a Chinese liquor typically obtained from cereals by complex fermentation processes using natural mixed culture starters (i.e. "Daqu") followed by distillation. Several types of Daqu can be distinguished, such as light-flavour, strong-flavour, and sauce flavour Daqu that differ in the formulation of ingredients and, particularly, in the incubation conditions during their processing, leading to different microflora compositions and formations of metabolites and reaction products. In this work, a different problem was dealt, namely the reliability and accuracy of the non-targeted analysis in relation to the extraction procedure of the samples. To validate the method, the disruption of microbial cells with or without glass beads, the possible effect of enzyme modifications during the extraction step, and the need of ultrafiltration for the removal of macromolecules were investigated and a standardized method was chosen. The PCA result verified that the standardized procedure was reproducible and reliable. The discussion of all the results is outside the scope of this chapter; just as an example, acetate was found as a biomarker of light-flavour Daqu by PCA, in agreement with the fact that ethylacetate is the representative aroma compound in this type of Daqu and was ascribed to its production by microorganisms. As ethyl lactate was present at high levels but lactate was not detected, the authors suggested that this compound is produced during the alcoholic fermentation stage, later in the process of liquor making.

A more accurate statistical approach has been applied in a similar study on the ageing process of traditional balsamic vinegar of Modena (Italy) [73]. Seventy two samples were divided into young (<12 years), old (>12 and <25 years), and very old (>25 years) subgroups. A partitioning into three classes was already obtained by applying unsupervised PCA on the bucketed spectra with centring pre-treatment; the separation was essentially due to the bucket corresponding to the signal of acetate, whose level decreased during ageing as expected. However, the separation in subgroups was still present after removal of the acetate bucket, and the second PCA model seemed to provide a better model as no samples of known age behaved as outliers upon reprojection (83.40% of total variance explained). Then, the *X*-matrix was

divided in two blocks, A and B, the first containing strong signal intensities (region from 1.5 to 5.1 ppm) and the second weak signal intensities (region from 5.1 to 10 ppm). PCA on block A and gave a model with eight components selected by cross-validation ($R^2 = 0.99$ and $Q^2 = 0.90$) with the first two PCs explaining the 98.1% of the total variance and PCA on block B resulted in a five components model ($R^2 = 0.94$ and $Q^2 = 0.76$) with the first two principal components explaining the 75.7% of the total variance. Thus, a clear differentiation among young, old, and extra old samples was obtained from either block. The authors remarked that the discrimination in both blocks was due to the application of a hierarchical approach and that, otherwise, the block B resulted irrelevant for vinegar characterization. The two sets of score vectors obtained by applying PCA on each block (13 orthogonal variables) were subjected to PLS-DA or Naïve Bayes classification. To prove the accuracy of the representation on both training and test sets, the robust classification model built by PLS-DA with the two classes approach was compared with three classes Naïve Bayes classifier and the models were also compared on the basis of the predictions obtained for the "unknown" vinegar samples, showing more than 80% agreement among them.

Another work deserves a brief discussion because it addressed a practical request of the industry, namely the evaluation of proprietary products in terms of different sites and times of production of the same brand [49]. The authors showed that ^{1}H NMR spectroscopy in tandem with PCA can be a promising method, offering not only a full characterization of the products in term of composition, but also representing a direct, short-term monitoring of the samples. The method was tested in the evaluation of the same type of beer (lager) originating from three different brewing sites in three different countries, and produced on different dates. The NMR spectra were divided in three spectral regions (aliphatic, aromatic, and sugar containing signals) and the different data sets were subjected to PCA. The aliphatic region could distinguish the production sites, and while two sites showed a good reproducibility in time, in one site lactic and pyruvic acid contents resulted altered, probably reflecting yeast quality and/or yeast generation number. Linear dextrins from the sugar region predominated in one site and branched dextrins in the other two while a series of compounds distinguished all the production sites. These evidences were correlated with differences in the malting stage of processing and to the degradation of amino acids, respectively. It is evident that the information achievable by the metabolomic approach is of great value for the industry to monitor and control the product process, although the correlation with quality and/or nutritional aspects is still lacking.

A time-course global analysis was also applied by Son *et al*. [79] during alcoholic fermentation and ageing in must and wine. Interestingly, the obtained data were correlated with three different used yeast strains giving information on the fermentative performances of these strains. Furthermore, a series of pattern recognition methods (PCA, PLS-DA, and OPLS-DA) were

employed and their potentiality was compared. For example, the statistics for differentiating musts during fermentation (day-to-day comparison) gave high goodness of fit and predictability (R^2 values from 0.77 to 0.96 and Q^2 values from 0.69 to 0.92). However, the R^2 values decreased to 0.66 and the Q^2 to 0.38 in the analysis at days 2 and 3, indicating larger metabolic changes at later fermentation times in respect to those at earlier times. OPLS-DA resulted as the best analytical approach in this case showing how the levels of 2,3-butanediol, succinate, and glycerol, simultaneously detected by NMR spectroscopy, can be representative of the different fermentation speeds and activities of the three yeast strains used.

This last study highlights the possibility to integrate the global analysis of metabolites (metabolomics) with other "omics" data to build comprehensive system biology networks, especially where genomics and proteomics data are affordable to be obtained as in the field of fermentation processes that utilize simple microorganisms. The same approach can be equally valuable in other industrial processes, as will be discussed in the following paragraphs for coffee and meat.

Different degrees of coffee beans roasting have been metabolically characterized through ^{1}H and ^{13}C NMR spectroscopy, detailed signal assignment, and multivariate data analysis [77]. Compounds variations during the roasting processes have already been reported using both compound-targeted techniques as chromatography and MS [80] and untargeted NMR spectroscopy on instant [81] or espresso coffee [82]. On the contrary, in this study, ^{13}C NMR spectra were subjected to the multivariate analysis, instead of the more commonly used ^{1}H NMR ones: this different approach was chosen to minimize the effect of chemical shifts of caffeine and chlorogenic acids linked to the production of caffeine–chlorogenate complexes during the roasting process. Noticeably, chlorogenic acids were degraded during the process while caffeine was relatively thermally stable as inferred by ^{1}H NMR data. The PCA indicated that some compounds, namely sucrose, chlorogenic acids, quinic acids, and polysaccharides could serve as chemical markers during coffee bean roasting [77].

Finally, a different method has been recently proposed for the rapid detection of irradiated meat in official control of food [83]. The method differs from the preceding ones both for the used NMR spectroscopic technique (^{1}H NMR lipid profiling) and for the applied multivariate statistical analysis, stepwise linear discriminant analysis and artificial neural networks (ANNs). Irradiation is a process used to improve the safety of the product, simultaneously extending its shelf life by inactivating pathogenic bacteria. However, this process could affect the nutritional quality of the product itself. In this regard, a protocol to differentiate between irradiated and non-irradiated food containing fat has been adopted by the European Committee for Standardisation as EN 1785 in 2001. The used technique, although representing a targeted approach, envisages several time-consuming steps, namely lipid extraction,

absorption chromatography, and detection by gas chromatography coupled to MS, allowing the analysis of a few samples at a time. ^{1}H NMR spectra from muscle lipids fraction of 72 beef samples (29 non-irradiated and 17, 18, and 8 irradiated at 2.5, 4.5, and 8 kGy, respectively) were acquired. Hundred percent of the beef samples were classified correctly into irradiated or non-irradiated groups, respectively, even in cross-validation and 88.9% of the beef samples were correctly classified according to their irradiation dose. Similar to the classification ability of LDA, 100% recognition and prediction ability were obtained by ANNs. Thus, the proposed metabolomic approach is suitable to be recognized by regulatory official organisms if a greater number of samples and different meat species will be used to standardize the methodology and acquire large databases.

A last, worth mentioning, application of a metabolomic approach is represented by the efforts to provide an objective "measure" of quality where quality ranking is based on subject evaluations as, for example, when professional tasters are enrolled to evaluate taste quality. In a study of Tarachiwin *et al.* [51], untargeted ^{1}H NMR-based metabolomics has been applied for the quality evaluation of Japanese green tea. ^{1}H NMR spectra were acquired from green tea quality ranked by a professional taster panel. This ranking was used as a dependent variable in the PLS regression model that showed a R^2 of 0.987 and a Q^2 of 0.671. Then, by applying the orthogonal signal correction (OSC) approach, which is reducing the number of variables from the spectra matrix by removing those linearly unrelated (orthogonal), an improved regression was obtained in which Q^2 increased from 0.671 to 0.982. OSC is normally used to remove the uncorrelated variables or orthogonal to Y from X using the non-linear iterative partial least-squares algorithm [84,85]. This approach will cope also with moderate amounts of missing data. The residual data from this orthogonal model are obscured by the variation in the data set. Furthermore, a combination between OSC and wavelet transform named OSCW, which tended to compress and de-noise complicated signals was applied [86]. In conclusion, all of the presented multivariate calibrations could be used to predict the quality of green tea with a very good fitting and excellent predictability, although the best quality predictive model with highest prediction accuracy was obtained from the PLS-OSCW regression. This kind of approach could be beneficial for the industry as, besides being instrumentally based, it could save the time-consuming and expensive method based on taster panels.

For the sake of completeness, we must cite another similar work aiming to investigate the relationship between the NMR metabolite profile and the sensory profile of sour cherry juices from different cultivars and clones. A sensory analysis was carried out as a quantitative descriptive analysis and PCA was applied to the obtained data, showing a discrimination in two groups along PC1, explaining 63% of the variation. A clustering in two groups was also obtained when PCA was applied to the NMR data although with very poor predictive ability. A PLS-DA model related to these two groups had a

high predictive ability, with *X* variables described 82.0% of the variation in the *Y*-matrix and were closely related to the genetic background of the cherry clones. In terms of metabolites, malic acid alone or the ratio between the glucose and malic acid contents were important determinants of the perception of sourness and sweetness. However, the limited number of samples evaluated by the sensory panel reduced the validity of the models and made it difficult to generalize the findings [87].

REFERENCES

[1] Wishart DS. Metabolomics: applications to food science and nutrition research. Trends Food Sci Technol 2008;19:482–93.

[2] Nicholson JK, Lindon JC, Holmes E. 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica 1999;29:1181–9.

[3] Lindon JC, Holmes E, Nicholson JK. Pattern recognition methods and applications in biomedical magnetic resonance. Prog NMR Spectrosc 2001;39:1–40.

[4] Cifuentes A. Food analysis and foodomics. J Chromatogr A 2009;1216:7109–10.

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

[6] Hall RD Food metabolomics: META-PHOR. A new European research initiative. Agro-FOOD industry hi-tech. January/February 2007, Anno 18—No. 1; 2007.

[7] Ellis DI, Brewster VL, Dunn WB, Allwood JW, Golovanov AP, Goodacre R. Fingerprinting food: current technologies for the detection of food adulteration and contamination. Chem Soc Rev 2012;41:5706–27.

[8] Mannina L, Sobolev AP, Viel S. Liquid state ^{1}H high field NMR in food analysis. Prog NMR Spectrosc 2012;66:1–39.

[9] Spraul M, Schutz B, Humpfer E, Mortter M, Schafer H, Koswig S, et al. Mixture analysis by NMR as applied to fruit juice quality control. Magn Reson Chem 2009;47:S130–S137.

[10] Axelson DE. Data pre-processing for chemometric and metabonomic analysis; MRi_Consulting. Kingston: Ontario; 2010.

[11] Sacchi R, Addeo F, Paolillo L. ^{1}H and ^{13}C NMR of virgin olive oil. An overview. Magn Reson Chem 1997;35:S133–S145.

[12] Caligiani A, Acquotti D, Palla G, Bocchi V. Identification and quantification of the main organic components of vinegars by high resolution ^{1}H NMR spectroscopy. Anal Chim Acta 2007;585:110–9.

[13] Son HS, Hwang GS, Kim KM, Ahn HJ, Park WM, Van Den Berg F, et al. Metabolomic studies on geographical grapes and their wines using ^{1}H NMR analysis coupled with multivariate statistics. J Agric Food Chem 2009;57:1481–90.

[14] Sandusky P, Raftery D. Use of selective TOCSY NMR experiments for quantifying minor components in complex mixtures: application to the metabonomics of amino acids in honey. Anal Chem 2005;77:2455–63.

[15] Manetti C, Bianchetti C, Bizzarri M, Casciani L, Castro C, D'Ascenzo G, et al. NMR-based metabonomic study of transgenic maize. Phytochemistry 2004;65:3187–98.

[16] Gan F, Ruan GH, Mo JY. Baseline correction by improved iterative polynomial fitting with automatic threshold. Chemometr Intell Lab Syst 2006;82:59–65.

[17] Chang D, Banack CD, Shah SL. Robust baseline correction algorithm for signal dense NMR spectra. J Magn Reson 2007;187:288–92.

[18] Xi Y, Rocke DM. Baseline correction for NMR spectroscopic metabolomics data analysis. BMC Bioinformatics 2008;9:324.
[19] van der Werf MJ, Takors R, Smedsgaard J, Nielsen J, Ferenci T, Portais JC, et al. Standard reporting requirements for biological samples in metabolomics experiments: microbial and in vitro biology experiments. Metabolomics 2007;3:189–94.
[20] Fiehn O, Sumner LW, Rhee SY, Ward J, Dickerson J, Lange BM, et al. Minimum reporting standards for plant biology context information in metabolomic studies. Metabolomics 2007;3:195–201.
[21] Goodacre R, Broadhurst D, Smilde AK, Kristal BS, Baker JD, Beger R, et al. Proposed minimum reporting standards for data analysis in metabolomics. Metabolomics 2007;3:231–41.
[22] Savorani F, Tomasi G, Engelsen SB. icoshift: a versatile tool for the rapid alignment of 1D NMR spectra. J Magn Reson 2010;202:190–202.
[23] Vu TN, Valkenborg D, Smets K, Verwaest KA, Dommisse R, Lemière F, et al. An integrated workflow for robust alignment and simplified quantitative analysis of NMR spectrometry data. BMC Bioinformatics 2011;12:405.
[24] De Meyer T, Sinnaeve D, Van Gasse B, Tsiporkova E, Rietzschel ER, De Buyzere ML, et al. NMR-based characterization of metabolic alterations in hypertension using an adaptive, intelligent binning algorithm. Anal Chem 2008;80:3783–90.
[25] Anderson PE, Mahle DA, Doom TE, Reo NV, DelRaso NJ, Raymer ML. Dynamic adaptive binning: an improved quantification technique for NMR spectroscopic data. Metabolomics 2010;7:179–90.
[26] Human Metabolome Data Base. www.Hmdb.ca.
[27] Cui Q, Lewis IA, Hegeman AD, Anderson ME, Li J, Schulte CF, et al. Metabolite identification via the Madison Metabolomics Consortium Database. Nat Biotechnol 2008;26:162.
[28] NMRShiftDb. http://nmrshiftdb.nmr.uni-koeln.de/.
[29] Weljie AM, Newton J, Mercier P, Carlson E, Slupsky CM. Targeted profiling: quantitative analysis of ^{1}H NMR metabolomics data. Anal Chem 2006;78:4430–42.
[30] Xia J, Bjorndahl TC, Tang P, Wishart DS. MetaboMiner—semi-automated identification of metabolites from 2D NMR spectra of complex biofluids. BMC Bioinformatics 2008;9:507.
[31] Holmes E, Cloarec O, Nicholson JK. Probing latent biomarker signatures and in vivo pathway activity in experimental disease states via statistical total correlation spectroscopy (STOCSY) of biofluids: application to HgCl2 toxicity. J Proteome Res 2006;5:1313–20.
[32] Ross A, Schlotterbeck G, Dieterle F, Senn H. Chapter 3: NMR spectroscopy techniques for application to metabonomics. In: Lindon J, Nicholson J, Holmes E, editors. Handbook of metabonomics and metabolomics. The Netherlands: Elsevier; 2007. p. 55–112.
[33] Smolinska A, Blanchet L, Buydens LMC, Wijmenga SS. NMR and pattern recognition methods in metabolomics: from data acquisition to biomarker discovery. Anal Chim Acta 2012;750:82–97.
[34] Bro R. Multi-way analysis in the food industry. Models, algorithms, and applications. PhD thesis. University of Amsterdam, Netherlands; 1998.
[35] Bro R, van den Berg F, Thybo A, Andersen CM, Jørgensen BM, Andersen H. Multivariate data analysis as a tool in advanced quality monitoring in the food production chain. Trend Food Sci Technol 2002;13:235–44.
[36] Liland KH. Multivariate methods in metabolomics—from pre-processing to dimension reduction and statistical analysis. Trend Anal Chem 2011;30:827–41.
[37] Gulati OP, Berry Ottaway P. Legislation relating to nutraceuticals in the European Union with a particular focus on botanical-sourced products. Toxicology 2006;221:75–87.

[38] Flynn A. Scientific substantiation of health claims in the EU. Proc Nutr Soc 2012;71:120–6.
[39] Hong S-Y. NMR-based metabolomics in wine science. Magn Res Chem 2011;49:S13–S21.
[40] Spraul M, Schutz B, Rinke P, Koswig S, Humpfer E, Schafer H, et al. NMR-based multiparametric quality control of fruit juices: SGF profiling. Nutrients 2009;1:148–55.
[41] Capitani D, Mannina L, Proietti N, Sobolev AP, Tomassini A, Miccheli A, et al. Monitoring of metabolic profiling and water status of Hayward kiwifruits by nuclear magnetic resonance. Talanta 2010;82:1826–38.
[42] Capitani D, Mannina L, Proietti N, Sobolev AP, Tomassini A, Miccheli A, et al. Metabolic profiling and outer pericarp water state in Zespri, CI.GI, and Hayward kiwifruits. J Agric Food Chem 2013;61:1727–40.
[43] Son H-S, Kim KM, Van der Berg F, Hwang G-S, Park W-M, Lee C-H, et al. ^{1}H nuclear magnetic resonance-based metabolomic characterization of wine by grape varieties and production areas. J Agric Food Chem 2008;56:8007–16.
[44] Viggiani L, Castiglione Morelli A. Characterization of wines by nuclear magnetic resonance: a work study on wines from the Basilicata region in Italy. J Agric Food Chem 2008;56:8273–9.
[45] Imparato G, Di Paolo E, Braca A, Lamanna R. Nuclear magnetic resonance profiling of wine blends. J Agric Food Chem 2011;59:4429–34.
[46] Papotti G, Bertelli D, Graziosi R, Silvestri M, Bertacchini L, Durante C, et al. One- and two-dimensional NMR spectroscopy or the characterization of protected designation of origin Lambrusco wines of Modena. J Agric Food Chem 2013;61(8):1741–6.
[47] Duarte IF, Barros A, Almeida C, Spraul M, Gil AM. Multivariate analysis of NMR and FTIR data as a potential tool for the quality control of beer. J Agric Food Chem 2004;52:1031–8.
[48] Lachenmeier DW, Frank W, Humpfer E, Schafer H, Keller S, Mortter M, et al. Quality control of beer using high-resolution nuclear magnetic resonance spectroscopy and multivariate analysis. Eur Food Res Technol 2005;220:215–21.
[49] Almeida C, Duarte IF, Barros A, Rodrigues J, Spraul M, Gil AM. Composition of beer by ^{1}H NMR spectroscopy: effects of brewing sites and date of production. J Agric Food Chem 2006;54:700–6.
[50] Le Gall G, Colqhuoun IJ, Defernez M. Metabolic profiling using ^{1}H NMR spectroscopy for quality assessment of green tea, *Camelia sinensis* (L.). J Agric Food Chem 2004;52:692–700.
[51] Tarachiwin L, Ute K, Kobayashi A, Fukusaki E. ^{1}H NMR based metabolic profiling in the evaluation of Japanese green tea quality. J Agric Food Chem 2007;55:9330–6.
[52] Ohno A, Oka K, Sakuma C, Okuda H, Fukuhara K. Characterization of tea cultivated at four different altitudes using ^{1}H NMR analysis coupled with multivariate statistics. J Agric Food Chem 2011;59:5181–7.
[53] Lee J-E, Lee B-J, Chung J-O, Hwang J-A, Lee S-J, Lee C-H, et al. Geographical and climatic dependencies of green tea (Camelia sinensis) metabolites: a ^{1}H NMR-based metabolomic study. J Agric Food Chem 2010;58:10582–9.
[54] Mannina L, Sobolev AP. High resolution NMR characterization of olive oils in terms of quality, authenticity and geographical origin. Magn Res Chem 2011;49:S3–S11.
[55] Council Regulation (EEC) No. 2081/92 on the protection of geographical indications and designations of origin for agricultural products and foodstuffs. Off J Eur Union 1992;208:1–15.

[56] Mannina L, Marini F, Gobbino M, Sobolev AP, Capitani D. NMR and chemometrics in tracing European olive oils: the case study of Ligurian samples. Talanta 2010;80:2141–8.
[57] Alonso-Salces RM, Moreno-Rojas JM, Holland MV, Reniero F, Guillou C, Heberger K. Virgin olive oil authentication by multivariate analyses of ^{1}H NMR fingerprints and δ^{13}C and δ^{2}H data. J Agric Food Chem 2010;58:5586–96.
[58] Consonni R, Cagliani LR, Stocchero M, Porretta S. Triple concentrated tomato paste: discrimination between Italian and Chinese products. J Agric Food Chem 2009;57:4506–13.
[59] Consonni R, Cagliani LR, Stocchero M, Porretta S. Evaluation of the production year in Italian and Chinese tomato paste for geographical determination using O2PLS models. J Agric Food Chem 2010;58:7520–5.
[60] Jung Y, Lee J, Kwon J, Lee K-S, Ryu DH, Hwang G-S. Discrimination of the geographical origin of beef by ^{1}H NMR-based metabolomics. J Agric Food Chem 2010;58:10458–66.
[61] Bong YS, Shin WJ, Lee AR, Kim YS, Kim K, Lee KS. Tracing the geographical origin of beefs being circulated in Korean markets based on stable isotopes. Rapid Commun Mass Spectrom 2010;24:155–9.
[62] Shintu L, Caldarelli S, Franke BM. Pre-selection of potential molecular markers for the geographic origin of dried beef by HR-MAS NMR spectroscopy. Meat Sci 2007;76:700–7.
[63] Charlton AJ, Robb P, Donarski JA, Godward J. Non-targeted detection of chemical contamination in carbonated soft drinks using NMR spectroscopy, variable selection and chemometrics. Anal Chim Acta 2008;618:196–203.
[64] Bertelli D, Lolli M, Capotti G, Bortolotti L, Serra G, Plessi M. Detection of honey adulteration by sugar syrups using one-dimensional and two-dimensional high-resolution nuclear magnetic resonance. J Agric Food Chem 2010;58:8495–501.
[65] Le Gall G, Puaud M, Colquhoun IJ. Discrimination between orange juice and pulp wash by ^{1}H nuclear magnetic resonance spectroscopy: identification of marker compounds. J Agric Food Chem 2001;49:580–8.
[66] Ferrari E, Foca G, Vignali M, Tassi L, Ulrici A. Adulteration of the anthocyanin content of red wines: perspectives for authentication by Fourier transform-near infrared and ^{1}H NMR spectroscopies. Anal Chim Acta 2011;701:139–51.
[67] Defelice SL. The nutraceutical revolution, its impact on food industry research and development. Trends Food Sci Technol 1995;6:59–61.
[68] Zhao J, Avula B, Joshi VC, Techen N, Wang Y-H, Smillie TJ, et al. NMR fingerprinting for analysis of Hoodia species and Hoodia dietary products. Planta Med 2011;77:851–7.
[69] Graham SF, Ruiz-Aracama A, Lommen A, Canonizzo FT, Biolatti B, Elliott CT, et al. Use of NMR metabolomic plasma profiling methodologies to identify illicit growth-promoting administrations. Anal Bioanal Chem 2012;403:573–82.
[70] Son H-S, Hwang G-S, Kim KM, Kim E-H, van der Berg F, Park W-M, et al. ^{1}H NMR-based metabolomic approach for understanding the fermentation behaviors of wine yeast strains. Anal Chem 2009;81:1137–45.
[71] Lee J-E, Hwang G-S, van der Berg F, Lee C-H, Hong Y-S. Evidence of vintage effects on grape wines using ^{1}H NMR-based metabolomic study. Anal Chim Acta 2009;648:71–6.
[72] Consonni R, Cagliani LR. Ripening and geographical characterization of Parmigiano Reggiano cheese by ^{1}H NMR spectroscopy. Talanta 2008;76:200–5.
[73] Consonni R, Cagliani LR, Benevelli F, Spraul M, Humpfer E, Stocchero M. NMR and chemometric methods: a powerful combination for characterization of Balsamic and Traditional Balsamic Vinegar of Modena. Anal Chim Acta 2008;611:31–40.

[74] Wu X-H, Zheng X-W, Han B-Z, Vervoort J, Nout MJR. Characterization of Chinese liquor starter, "Daqu", by flavor type with ^{1}H NMR-based nontargeted analysis. J Agric Food Chem 2009;57:11354–9.
[75] Choi H-K, Yoon J-H, Kim Y-S, Kwon DY. Metabolomic profiling of Cheonggukjang during fermentation by ^{1}H NMR spectrometry and principal components analysis. Process Biochem 2007;42:263–6.
[76] Koda M, Furihata K, Wei F, Miyakawa T, Tanokura M. NMR-based metabolic profiling of rice wines by F2-selective total correlation spectra. J Agric Food Chem 2012;60:4818–25.
[77] Wei F, Furihata K, Koda M, Hu F, Miyakawa T, Tanokura M. Roasting process of coffee beans as studied by nuclear magnetic resonance: time course of changes in composition. J Agric Food Chem 2012;60:1005–12.
[78] Ko B-K, Ahn H-J, van der Berg F, Lee C-H, Hong Y-S. Metabolomic insight into soy sauce through ^{1}H NMR spectroscopy. J Agric Food Chem 2009;57:6862–70.
[79] Son H-S, Hwang G-S, Kim KM, Ahn H-J, Park W-M, van der Berg F, et al. Metabolomic studies on geographical grapes and their wines using ^{1}H NMR analysis coupled with multivariate statistics. J Agric Food Chem 2009;57:1481–90.
[80] Nunes FM, Coimbra MA. Chemical characterization of the high-molecular-weight material extracted with hot water from green and roasted robusta coffees as affected by the degree of roast. J Agric Food Chem 2002;50:7046–52.
[81] Charlton AJ, Farrington WHH, Brereton P. Application of ^{1}H NMR and multivariate statistics for screening complex mixtures: quality control and authenticity of instant coffee. J Agric Food Chem 2002;50:3098–103.
[82] Bosco M, Toffanin R, de Palo D, Zatti L, Segre A. High resolution ^{1}H NMR investigation of coffee. J Sci Food Agric 1999;79:869–78.
[83] Zanardi E, Caligiani A, Padovani E, Mariani M, Ghiaini S, Palla G, et al. Detection of irradiated beef by nuclear magnetic resonance lipid profiling combined with chemometric techniques. Meat Sci 2013;93:171–7.
[84] Eriksson L, Johansson E, Kettaneh-Wold N, Wold S. In: Multi and mega variate data analysis principles and applications. Umeå, Sweden: Umetrics AB; 2001. p. 229–50.
[85] Eriksson L, Trygg J, Johansson E, Bro R, Wold S. Orthogonal signal correction, wavelet analysis, and multivariate calibration of complicated process fluorescence data. Anal Chim Acta 2000;420:181–95.
[86] Barclay VJ, Bonner RF, Hamilton IP. Application of wavelet transform to experimental spectra: smoothing, denoising, and data set compression. Anal Chem 1997;69:78–90.
[87] Clausen MR, Pedersen BH, Bertram HC, Kidmose U. Quality of sour cherry juice of different clones and cultivars (Prunus cerasus L.) determined by a combined sensory and NMR spectroscopic approach. J Agric Food Chem 2011;59:12124–30.