Send Orders of Reprints at reprints@benthamscience.net

Plant Metabolomics: From Holistic Data to Relevant Biomarkers

Jean-Luc Wolfender[1,*], Serge Rudaz[1], Young Hae Choi[2] and Hye Kyong Kim[2]

[1]*School of Pharmaceutical Sciences, EPGL, University of Geneva, University of Lausanne, 30, Quai Ernest Ansermet, CH-1211 Geneva 4, Switzerland;* [2]*Natural Product Laboratory, Institute of Biology, Leiden University, Einsteinweg 55, P.O.Box 9502, 2300RA Leiden, The Netherlands*

Abstract: Metabolomics is playing an increasingly important role in plant science. It aims at the comprehensive analysis of the plant metabolome which consists both of primary and secondary metabolites. The goal of metabolomics is ultimately to identify and quantify this wide array of small molecules in biological samples. This new science is included in several systems biology approaches and is based primarily on the unbiased acquisition of mass spectrometric (MS) or nuclear magnetic resonance (NMR) data from carefully selected samples. This approach provides the most ''functional'' information of the 'omics' technologies of a given organism since metabolites are the end products of the cellular regulatory processes. The application of state-of-the-art data mining, that includes various untargeted and targeted multivariate data analysis methods, to the vast amount of data generated by this data-driven approach leads to sample classification and the identification of relevant biomarkers. The biological areas that have been successfully studied by this holistic approach include global metabolite composition assessment, mutant and phenotype characterisation, taxonomy, developmental processes, stress response, interaction with the environment, quality control assessment, lead finding and mode of action of botanicals.

This review summarises the main MS- and NMR-based approaches that are used to perform these studies and discusses the potential and current limitations of the various methods. The intent is not to provide an exhaustive overview of the field, which has grown considerably over the past decade, but to summarise the main strategies that are used and to discuss the potential and limitations of the different approaches as well as future trends.

Keywords: Metabolomics, profiling, plants, natural products, lead finding, stress response, mass spectrometry (MS), nuclear magnetic resonance spectroscopy (NMR), multivariate data analysis.

1. INTRODUCTION

Plants produce a large array of metabolites that are either essential to sustain plant life via normal metabolic processes such as respiration and photosynthesis (primary metabolites) or non-essential but necessary for survival in a given environment (secondary metabolite). Primary metabolites include proteins, lipids, carbohydrates and chlorophyll. Secondary metabolites represent compounds that are not directly involved in the normal growth, development, or reproduction of an organism [1-3]. These compounds are mainly involved in the survivability and fecundity of a plant and are often restricted to a narrow set of species within a phylogenetic group. Secondary metabolites play a particularly important role in signalling and plant defence [4]. Many of these compounds are also referred to as plant natural products (NPs) and have historically been an important source of lead molecules in drug discovery due to their action on many pharmacological targets [5].

Until recently, most of these plant metabolites have been analysed and detected with targeted methods for very specific purposes, including the evaluation of metabolite diversity in certain plant species, the quantitation of bioactive metabolites from medicinal plants and the evaluation of chemical variations in certain environments and conditions. With the advent of powerful modern analytical methods and the development of multivariate data analysis approaches, these metabolites can be comprehensively analysed in complex natural extracts by metabolite profiling and fingerprinting. The purpose of metabolite fingerprinting is not to identify each observed metabolite but to compare patterns or ''fingerprints'' of metabolites that change in a given biological system. Conversely, metabolite profiling focuses on the analysis of a group of metabolites that is either related to a specific metabolic pathway or a class of compounds [6]. In most cases, metabolite profiling follows a hypothesis-driven approach rather than a hypothesis-generating one. Based on the questions asked, metabolites are selected for analysis, and specific analytical methods are developed for their determination. Both strategies can be used in the search for new biomarkers and can therefore be considered as complementary [7]. When these analyses are performed on a sufficient number of biological replicates, they enable researchers to discriminate and classify samples into groups and to monitor changes in metabolome composition that are related to a given physiological status, the influence of stress or a stimuli, genetic modification, interaction with other organisms, and so on. This type of holistic approach has been named metabolomics, and it has considerably modified the way in which biological issues are investigated. In the related field of phytochemistry, metabolomics is considered as the large-scale analysis of metabolites in a given organism in different physiological states [8].

The goal of metabolomics is the comprehensive and quantitative analysis of the largest possible array of low MW (<1,000 Da) metabolites in biological samples. It is playing an increasingly important role in various aspects of life science. Metabolomics is indeed a component of systems biology, which comprises a number of other 'omics' technologies such as transcriptomics (gene expression) and proteomics (protein expression). Metabolomics provides the most ''functional'' information among the "omics" technologies [9] by offering a broad view of the biochemical status of an organism that can be used to monitor significant metabolite variations. Indeed, because metabolites are the end products of cellular regulatory processes, their levels can be regarded as the ultimate response of biological systems to genetic or environmental changes. This information can be used with other systems biology approaches to assess gene function (functional genomics) and to provide a holistic view of a living system for its in-depth study [10].

Thus, metabolomics represents a promising tool for the post-genomic study of plant models with respect to variations induced by perturbations including environmental changes and physical, biotic, abiotic or nutritional stress [11]. From the beginning of the 21st century, dozens of plant genomes have been sequenced, including model plants such as *Arabidopsis thaliana* (thale cress) and crop plants such as *Oryza sativa* (rice) [12]. In this respect, metabolomics constitutes a powerful tool for plant molecular biotechnology and may be used to study the effects of plant breeding, mutation, and other manipulations, such as the expression of engineered genes [8].

*Address correspondence to this author at the School of Pharmaceutical Sciences, University of Geneva, University of Lausanne, 30 Quai Ernest Ansermet, 1211 Geneva 4, Switzerland; Tel: +41 22 379 33 85; Fax: +41 22 379 33 99;
E-mail: jean-luc.wolfender@unige.ch

1875-533X/13 $58.00+.00 © 2013 Bentham Science Publishers

Metabolomics is extensively used for studying biologic fluids in humans, where it is recognised as a key approach to monitor the state of biological organisms and is a widely used diagnostic tool for disease [13]. In plant science, the number of studies related to plant metabolomics has increased significantly in recent years [14, 15]. As an example, this approach was found to be very effective for investigating plant system issues such as plant stress responses [16] and plant-host interactions [17].

General problems that are encountered when characterising the plant metabolome are the complex nature and the large chemical diversity of the analysed compounds. Estimates of the sizes of metabolomes are highly dependent on the studied organism. Some authors have suggested a comparable number of metabolites and genes [18]; however, it should be noted that there is no direct link between every gene involved in metabolism and a given metabolite. Other estimates propose up to 15,000 distinct compounds within a given plant species [19, 20]. Overall, the plant kingdom is known to produce a wide diversity of chemical molecules that is estimated to include 200,000 primary and secondary metabolites [21]. With the exception of the primary metabolites that are directly involved in growth, development, and reproduction, most of these metabolites remain unknown. The chemical properties of this array of compounds are greatly variable, and this diversity is a critical aspect to consider. Some compounds, such as sugars and lipids, have important nutritive value and can occur in very large amounts, while secondary metabolites involved in signalling defence and resistance mechanisms may be present in very restricted amounts. The variability in molecular weight, polarity and solubility and the wide dynamic range of concentrations (from fmol to mmol) constitute an additional issue when analysing plant metabolites, as no amplification process is available.

Despite the impressive development of analytical methods in plant science, the comprehensive characterisation of a plant metabolome remains challenging. Even with the most advanced methods, a complete survey of all metabolites that are present in a crude plant extract is not possible. As will be discussed in this review, the best possible metabolome coverage will be achieved not by a single but by multiple analytical approaches [22] to provide a complete metabolome profile.

Finding relevant answers to biological questions using a metabolomic approach is the result of a series of challenges, from the choice of a specific sample preparation method and an adopted analytical platform to the data processing and modelling procedures. Each step of the process has to be carefully planned.

From a metabolomic perspective, one would ideally expect to be able to monitor, identify and quantify all of the compounds in very complex biological matrices in a single analysis with minimal sample and in an unbiased manner. Furthermore, data mining should provide a comprehensive way to interrogate datasets from various viewpoints according to the biological question at hand. Metabolomics has not yet reached this level of advancement, mainly because of the limitations of the analytical methods that are used. In most of the developed approaches, compromises have to be made, but nevertheless, a broad survey of metabolome variations can be recorded that will ultimately provide new biological knowledge.

This review gives a broad overview of the challenges that are related to plant metabolomics using either mass spectrometric (MS)- or nuclear magnetic resonance (NMR)-based approaches or a combination thereof. Various aspects will be discussed, including experimental setup, sample preparation, analytical issues and the downstream handling of data to extract biological knowledge. Since the various terms used in metabolomics have been already summarised in other reviews we invite the reader to refer to them for useful working definitions (eg.[23], [22]).

2. HARVESTING, COLLECTION AND SAMPLE PREPARATION

The different steps that are involved in a typical metabolomic study begin with the correct harvesting and sample preparation of the plant material of interest. The next steps include the analysis of a sufficient number of replicates for each set of conditions to study, multivariate analysis of the data, identification of the biomarkers of interest and the extraction of biological knowledge. These different steps are summarised in (Fig. **1**) for MS- or NMR-based approaches.

2.1. Plant Biological Variability

The levels of plant metabolites vary depending on environmental factors such as light, temperature or soil, and interaction with other organisms. Even within the same plant, metabolites can differ depending on their developmental stage and the time of harvesting [24-26]. Therefore, an appropriate experimental design is important to minimise the biological variations of the plant metabolome. When the influence of a specific stimuli, stress or physiological modification needs to be assessed, the biological variability should be minimised or the sample number will consequently need to be increased to find biomarkers that can be statistically significantly correlated to a given effect in a defined biological system.

Care should be taken to note all the information from the samples that are used for the experiments. Depending on the variety of cultivars (different accession), metabolites from the same plant can differ considerably, even if they were grown in the same conditions. A good example is *Arabidopsis thaliana*. Seeds with six different accessions were grown in an identical experimental design, but they displayed different metabolite profiles (unpublished data) (Fig. **2**). Any accession can be selected for an experiment, but the accession needs to be described.

2.2. Plant Harvesting

To reduce biological variability, plants should be cultivated in a controlled environment (light, temperature, and humidity controlled chamber). Under controlled conditions, the biological variability of samples tends to be markedly reduced. All types of treatment (insecticides or fungicides) or unwanted types of stress (e.g., insects) should be avoided because they may induce uncorrelated variability of the extract composition. When plants are grown in a field or in the wild, all external factors may influence the metabolome composition. Thus, for studies that relate to plant physiology, the comparison of independent specimens under controlled conditions is usually performed. Experiments that are performed to compare a group of specimens in their natural environment may be considered as preliminary studies because of the inherent variability (e.g., for chemotaxonomy or origin studies).

The method of harvesting varies according to the type of study (plant physiology, phenotyping, chemotaxonomy, stress response, etc.). When detailed metabolome variations that can be related to enzymatic activity have to be monitored, rapid harvesting and quenching is always recommended, e.g., by freeze-clamping or using liquid nitrogen to quench plant metabolism. These aspects are extremely important when studies are performed on plant hormones or defence signalling. For example, in the case of plant wounding, which mimics herbivore attack, harvesting can be perceived by the plant as a wound, and the plant reacts within seconds by producing signalling molecules [27]. In this case, a delay in quenching may alter the level of jasmonates and may lead to a biased interpretation of the wound induction of such metabolites.

When samples are flash-frozen they are usually kept at low temperature (-80°C to -196 °C) before use. Samples are generally dried before extraction, but metabolic processes may be re-activated during this step [28], especially if the drying process is air-, oven-

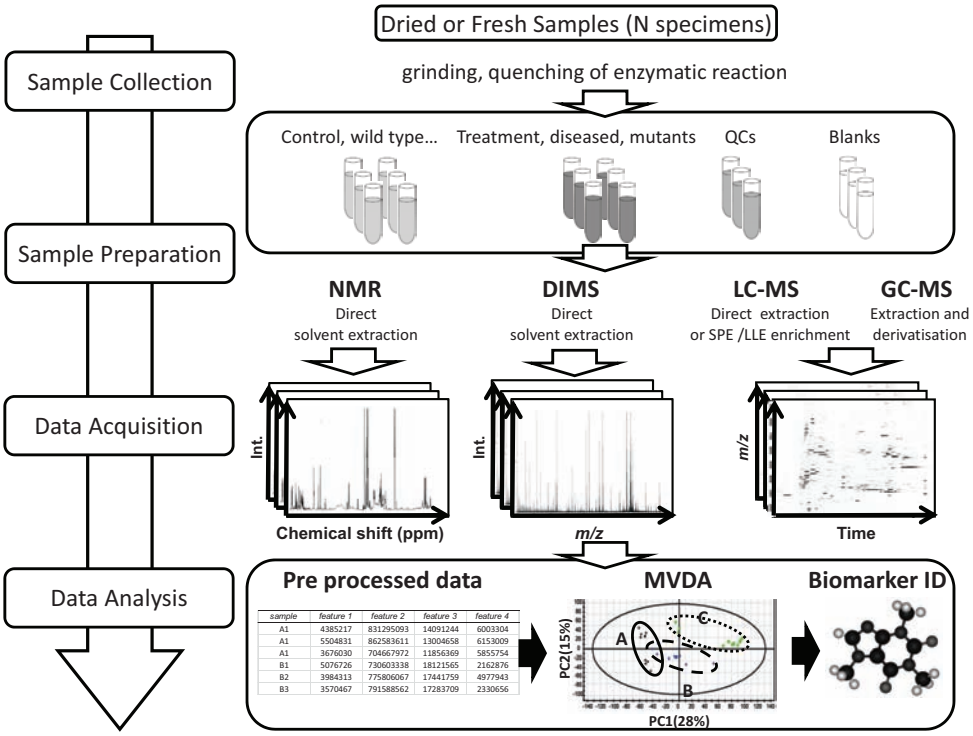

Fig. (1). Generic workflows used either for MS-based or NMR-based plant metabolomic studies.

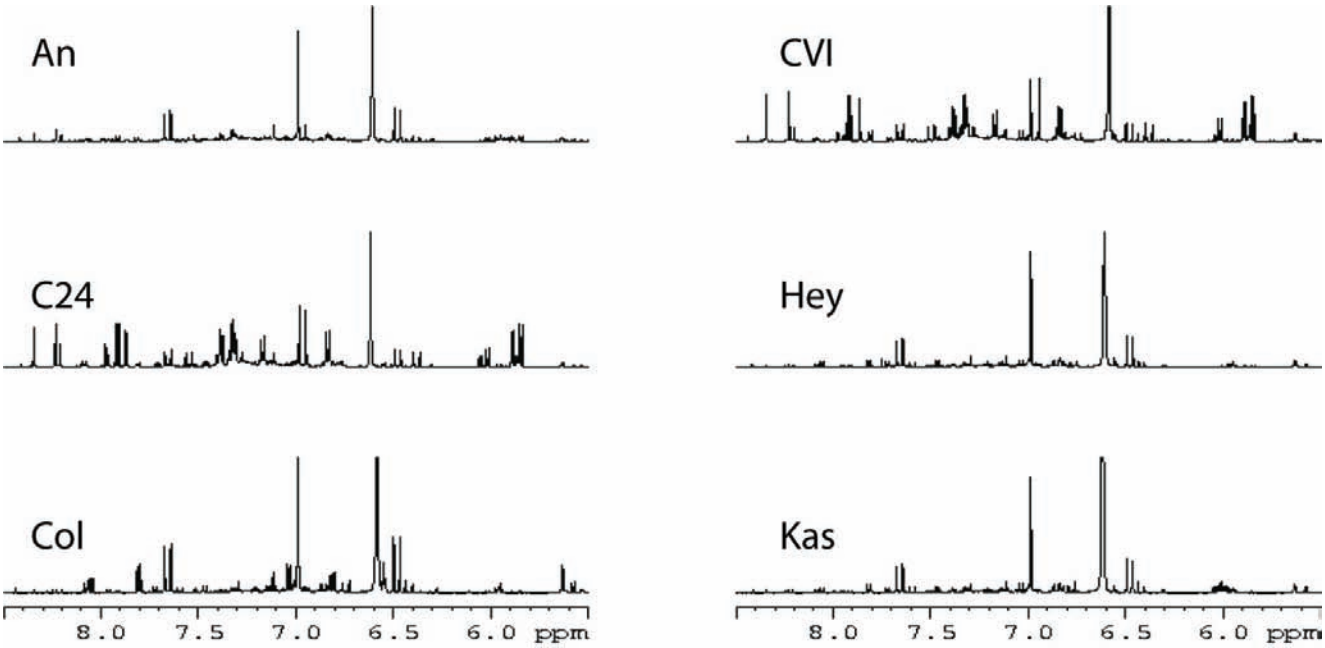

Fig. (2). ^{1}H NMR (δ 8.5-5.5) spectra of different accessions of *Arabidopsis thaliana*. Plants were grown in the same conditions and extracted with MeOH-Water (1:1) as described in the reference (Kim *et al.*, 2011). An: Antwerpen, Col: Columbia, CVI:Cape Verde Islands, Hey; Heythuysen, Kas: Kashimir.

or silica-based. A more suitable technique is freeze-drying, which prevents sample thawing because the absence of water reduces enzymatic activity [29]. However, the freeze-drying of tissue may lead to the irreversible adsorption of metabolites on cell walls and membranes [30]. Freezing also destroys the cells and several types of biochemical conversions may occur after thawing. These undesirable biochemical conversions should be avoided during thawing by extracting the frozen material using a denaturing solvent, or by a brief treatment with microwaves to avoid enzyme activation [18]. Another solution is to extract fresh frozen tissues directly using a solvent or a technique that will inhibit enzyme reactivation. For example, solvents may either be heated to 70-80°C, or acids may be added (formic or hydrochloric acid).

Accurate timing is an important issue to consider for proper sample collection. If comparisons have to be made, ideally all plants should be harvested at the same growth stage and at the same period of the day [24]. Plants may exhibit circadian variations in their metabolism [26, 31]. When the consequence of a given stimulus needs to be monitored, the harvesting should be performed at different time points after stimulation. Depending on the biological

problematic that is being studied, the harvesting can be conducted on a time scale to permit the comparison of metabolomic profiles at different time points after stimulation. The time requirements of the experiments will vary significantly. In the case of wound response, the timing of the observation of the induction onset of jasmonate will take a few hundred seconds and will reach a maximum at approximately 90 min [27]. Contrarily, the production of glucosinolates as a result of treatment with methyl jasmonate in *Brassica rapa* will not lead to the significant production of phenylpropanoids and indole derivatives until 14 days after treatment [32].

In plants, metabolite production occurs in different cell types. Ideally, it would be important to compare the metabolite profiles of single cells/cell types, but this is still extremely challenging to achieve with the current technology. However, localisation studies are important to consider, since the metabolomic results obtained will only correspond to a mean result of the production of all cells at a given time. The metabolome study of a whole organ corresponds to a snapshot picture that will not provide specific information about the localisation of individual metabolites and their dynamics in the organisms studied. A striking example of localised metabolite production is the biosynthesis of phenylphenalenones in *Dilatris* spp. (Haemodoraceae), which was demonstrated by the laser micro dissection harvesting of secretory cavities (SC) from leaves and the subsequent cryogenic ^{1}H NMR spectroscopy and high performance liquid chromatography (HPLC) analysis of the micro-dissected samples [33]. Such a level of dissection is not appropriate for most metabolomics studies, but it should be taken into consideration that the metabolite content varies between plant organs. It is therefore advisable to separate the different organs prior to extraction. In *Arabidopsis,* for example, it has been demonstrated that the concentration of glucosinolates changes spatially within a single leaf [34]. Significant differences in organ composition have been reported between primary and crown roots of maize [35]. Young and old leaves of the same plant can also exhibit differences [36].

Careful selection of the harvesting conditions and the timing are critical considerations at the beginning of plant metabolomics studies. Because metabolomics is a data-driven approach, several iterative cycles with different methods of sampling the plant may need to be performed before generating datasets that will yield significant information for a given biological issue.

2.3. Extraction

In addition to sampling, one key aspect of all metabolomic studies is the selection of an adapted extraction protocol that would be suitable for the detection of the biomarkers of interest. Extraction and sample preparation in plant metabolomics are fundamental and critical steps with important consequences for the accuracy of the results [24]. To increase extraction yields, plants are typically ground to a fine powder. When fresh material is used, great care should be taken to always keep the tissues frozen during grinding. This can be done with a mortar and a pestle using liquid nitrogen, or using a mixer mill system [37]. There are various parameters that must be considered when performing extractions.

The aim of metabolomics is to comprehensively profile the largest possible array of low MW metabolites in a given biological sample. To achieve this goal, generic extraction procedures with solvents that can extract compounds over a broad polarity range need to be devised. Because there is no single solvent that is able to extract and dissolve all plant metabolites to obtain the most exhaustive view of the metabolome, several solvents of increasing polarity can be used in succession, e.g., hexane > chloroform > methanol > water. However, this type of approach, which is used for classical phytochemical investigation and bioactivity-guided fractionation in particular, is difficult to implement in metabolomics studies because it generates many different extracts per sample that have to be analysed separately. This, however, is usually not compatible with the large number of samples that need to be analysed in metabolomics studies. Alternatively, combinations of various solvents can also be used, for example a multiple-phase solvent system composed of a mixture of chloroform-methanol-water has been proposed for plant metabolomics studies [24]. Most studies are performed either with pure MeOH or a mixture containing MeOH [24, 38, 39]. Another important point is that the selected solvents should be compatible with the analytical technique used. For example, one aspect that should be considered is that very lipophilic compounds do not elute from C18 reversed phase columns and certain MeOH-water mixtures do not extract chlorophylls from plant aerial parts, which might be an advantage for incorporating LC analysis in MS-based metabolomics. The use of water should be avoided in GC-MS analysis, whereas in LC-MS, solvents with elution strengths that are preferably lower than or equal to that of the mobile phase should be injected. This has consequences on the polarity of the solvent used for extraction. In the case of NMR-based approaches, the extracts obtained need to be fully soluble in the deuterated solvent used for NMR measurements. In NMR-based metabolomics, it is also possible to directly extract plants with deuterated solvents, and generic protocols using 750 μL CD_3OD + 750 μL KH_2PO_4 buffer in D_2O are available [40]. The sample preparation in this type of approach is minimised, facilitating high throughput analysis and repeatability.

Conventional techniques such as maceration and smooth agitation are slow and poorly adaptable to the preparation of numerous micro-samples. To accelerate the extraction process, additional energy can be applied by directly heating the system or by other techniques such as microwave extraction [41], ultrasonic extraction [42], ball mill extraction [43], pressurised liquid extraction [44] and supercritical fluid extraction [45]. Using such methods, extraction times of less than 5 minutes can be attained [37].

During the extraction of fresh plant material, enzymes can be reactivated, even in mildly polar solvents such as methanol or isopropanol, and the use of denaturing agents, heating or the use of acids could avoid such problems (see Plant Harvesting section, above). This aspect can strongly influence the final results. For example, in the case of an MS-based metabolomic study of the plant wound response [37], unwounded specimens were harvested 3 minutes after wounding and were extracted in isopropanol either at room temperature or at 75°C. Using boiling isopropanol, a strong increase in oxylipin-containing galactolipid levels was measured after wounding [46]. This increase was not recorded at room temperature because in the control samples, the enzymes were already active during extraction, which was similar to what occurred in the wounded plants, and these effects could not be observed [37].

2.4. Sample Preparation

Good sample preparation methods are important to ensure that the quality of metabolite profiling and fingerprinting will generate reliable datasets and instrument repeatability. Both parameters are critical for generating high-quality datasets that will be subjected to data mining. For example, sample preparation must avoid sample carryover in chromatography-based metabolomics that would otherwise bias a complete data set. In this respect, NMR-based approaches require minimal sample preparation and there is no interaction between the sample and the instrument. Sample preparation typically only involves the identification of a deuterated solvent that will minimise the chemical shift due to the pH of the sample solution and result in good quality spectra. In this case, the use of a buffer is advised. Among the many available buffers, phosphate buffer is commonly used because of its applicability [47]. ^{1}H NMR spectra of crude plant extracts are very complex, and sometimes simple pre-preparation of samples is required to obtain highly resolved signals of the desired metabolites. For example, the signal-to-noise ratio of the spectra of phenolic compounds in grape extracts increased after the removal of abundant primary metabolites using solid-phase extraction (SPE) [48]. Alternatively, liquid-liquid

partitioning can be performed to improve the detection of minor secondary metabolites. *n*-BuOH partitioning was used to concentrate the secondary metabolites of a beer sample. As a result, aromatic signals were considerably concentrated in the *n*-BuOH fraction with considerably higher NMR resolution [49].

Because most LC-MS-based metabolomics applications use reversed phase HPLC, liquid-liquid extraction (LLE) or SPE procedures can be used to remove interfering or undesired compounds and to concentrate samples. When using C18 columns, very lipophilic compounds such as pigments of lipids must be removed. An easy way to achieve this is to filter the samples using SPE with a solvent system that contains slightly less organic modifier than the solvent that is used for the analysis itself. This will solve the problem of peak tailing at the end of the HPLC gradient and will prevent the HPLC column from being contaminated [50].

In GC-MS-based metabolomics, volatile samples such as essential oils can be analysed without sample preparation. In this case, they can be collected by steam distillation or headspace. However, in the large majority of studies, an additional preparation step is required to increase the volatility of the samples. This is especially true for the non-volatile (polar) primary metabolites that require an extensive chemical derivatisation procedure to increase their volatility and thermal stability. Derivatisation generally involves oxime formation (with O-alkylhydroxylamines) followed by trimethylsilylation [N-acetyl-N-(trimethylsilyl)-trifluoroacetamide] to replace active hydrogens on polar functional groups with less polar trimethylsilyl groups, which consequently increases the volatility by reducing the dipole–dipole interactions [51]. While more elaborate protocols exist for amino acids [52], trimethylsilylation can be used as a universal, mild and quick process for metabolite profiling, thus avoiding the need to double or triple the number of analytical runs per plant sample, provided that adequate quality control measures are taken [53].

The biological problem that is being addressed defines the metabolomic tool that is required to obtain meaningful analytical data. It should be noted that the choice of the analytical method imposes certain constraints for the extraction method [24]. Because of the wide array of metabolites to be analysed, there is no generally applicable extraction protocol that is suitable for every biological question and for the various analytical methods. As will be discussed later, the standardisation of methods would be an absolute requirement for building a public metabolomics database that can be used globally for data-mining and biomarker identification [24]. Several protocols are now widely accepted in GC-MS based metabolomics [54], LC-MS-based metabolomics [55, 56] and NMR-based metabolomics [40, 57]. The strengths and limitations of each method have been discussed in a previous review [40].

It should be mentioned however that, even if the establishment of common protocols is of great importance, their characteristics are strongly related to the foreseen applications. When comparing plant samples, for example, the variability observed within different specimens of the same species (model plant) might be important to understand if a given biomarker is relevant or not to a given physiological status or stimuli. In control quality issues sampling might differs and, in this case, variations among batches that represent pool of plant specimens will be assessed. In this case often the observation of a restricted set of specific biomarkers previously identified as being related to a specific property will be monitored. Depending on the aim of the study (e.g. sample classification or biomarker discovery) the protocols might also have to be adapted.

3. ANALYTICAL METHODS FOR METABOLITE PROFILING AND FINGERPRINTING

With the aim of profiling all the metabolites in plant extracts, the analytical method of choice should be able to objectively detect a broad range of metabolites over a wide dynamic range, with high sensitivity, resolution and repeatability. Furthermore such a method should provide qualitative spectroscopic data for biomarker identification and quantitative information on the observed metabolome modifications. Because there is currently no single analytical method that can provide all the information needed, two main approaches are used: NMR and/or MS [22, 58]. Chromatography-based approaches are employed and most of them rely on MS detection (e.g., GC-MS or LC-MS) [6]. LC-NMR is powerful for rapidly providing structural information on single constituents of complex mixtures [59]. However, to our knowledge, metabolite profiling data obtained by LC-NMR have not been directly used for differential metabolomics, mainly because of the lack of sensitivity, cost, and low throughput of the method. The use of micro NMR methods in direct relation to HPLC, mainly by at-line hyphenation, is a however powerful way to identify *de novo* biomarkers that are highlighted by metabolomics and this will be discussed later. The types of fingerprints that are generated by these different analytical methods, together with spectra that can be used for metabolite identification, are illustrated in (Fig. **3**).

Each of these analytical methods has its strengths and limitations that are summarised in (Table **1**).

3.1. NMR-Based Metabolomics

NMR-based metabolite fingerprinting is used to directly analyse crude plant extracts without the need for prior chromatographic separation [60]. The method is simple, reproducible and has a high throughput capacity that does not require specific sample preparation and enables the detection of all protons [22, 61]. Another advantage of NMR analysis is the possibility to obtain quantitative information because the proton (^{1}H) signal intensity is proportional to the molar concentration [62]. In addition, NMR is a very useful tool for the structural elucidation of metabolites. As will be discussed below, MS-based techniques may help the identification of metabolite structure by providing molecular formula assignment or MS/MS characteristic fragments, but in most cases, NMR is required for the final identification. In any case, if new metabolites are found, diverse one- and multi-dimensional NMR techniques are needed, as they provide information for all H and C atoms in molecules with highly reproducible chemical shifts, coupling constants, and correlation between atoms [63]. The advantage of NMR spectroscopy for structure elucidation is a strong point for plant metabolomics when identifying unknown metabolites. The quality of metabolomic studies is dependent not only on the number of detected signals but also on the unambiguous identification of biomarkers. Depending on the type of study (group classification vs. biomarker investigation), a single identified metabolite is more meaningful than several hundred unidentified signals. Although NMR spectroscopy is a major technique to elucidate the chemical structures of small organic molecules, the identification procedure is still mostly based on the visual inspection of NMR spectra. In contrast to MS, GC-MS databases (e.g., NIST) in particular, the lack of generally available databases of NMR spectra of metabolites (particularly plant secondary metabolites), is an obstacle for NMR-based plant metabolomics. However, the high reproducibility and robustness of NMR data make it an ideal candidate for the establishment of a public database for metabolomics.

3.1.1. Sensitivity and Resolution

The low sensitivity of NMR is considered the major limitation of this application. The signal in an NMR experiment arises from the transition between low and high nuclear spin energy states. The low sensitivity is because the NMR signal depends on small differences in the populations of the Zeeman energy levels. For example, a 14.1 T magnet produces a 600 MHz frequency for ^{1}H. At room temperature, thermal energy is nearly 5 orders of magnitude larger than the energy associated with 600 MHz; thus, the two nuclear energy levels are almost equally populated, with the low energy

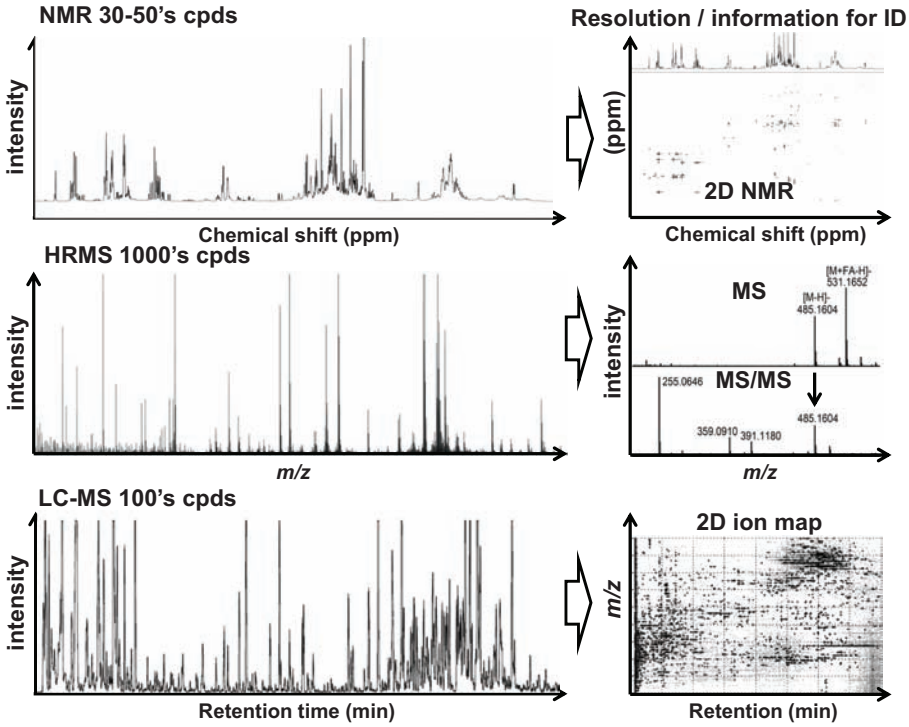

Fig. (3). Various methods of metabolite profiling or fingerprinting of crude extracts. Lefts panel: NMR (^{1}H-NMR), DIMS (TOFMS) and LC-MS (UHPLC-TOFMS) profiling. For each method a rough estimates of the number of compounds that can detected in a plant extract are indicated, this range from a few tens to about one thousand metabolites according to the technique used. Right panel: 2D NMR of the extract profiled by ^{1}H-NMR for enhanced resolution of the features, HR-MS and corresponding MS/MS spectra of s single compounds for metabolite identification, 2D ion map of the HR UHPLC-TOF-MS showing all features detected with UHPLC-TOFMS hyphenation.

Table 1. Possibilities and Limitations of the Analytical Methods Used for Metabolite Fingerprinting in Metabolomics.

	Sample Preparation	Reproducibility Short / long[1]	Absolute Quant.	Relative Quant.	Resolution LR / HR[2]	Sensi-tivity	Univer-sality	Cpds Nb[3]	Features type	Data Mining	ID
MS	**+[4]**	**++/-**	**-**	**+[5]**		**++/+++[6]**	**+/++[7]**				**+/++[8]**
DIMS	++	++ / -	-	+	+ / +++[9]	++	++	1000s	*m/z* x int	1D	+/++[10]
LC-MS	+	++ / -	-	+	++ / +++	++	+/++	100s	RT x *m/z* x int	2D	+/++
GC-MS	-	++ / +	-	++	++ / +++	++	+	1000s	RT x *m/z* x int	2D	++/+++ [11]
MS/MS	++	++ / +	-	++	NA[12]	+++[13]	-	100s	*m/z* x *m/z* x int	1D	++[14]
NMR	**+++**	**+++ / +++**	**+++**	**+++**	**- / +**	**-/+[15]**	**+++[16]**				**++/+++**
1D NMR	+++	+++ / +++	+++	+++	- / +	-	+++	10s	ppm x int	1D	++
J resolv	+++	+++ / +++	+++	+++	+ / ++	+	+++	10s	ppm x int	1D	+
2D NMR	+++	+++ / +++	+++	+++	+/ ++[17]	-	+++	10s	ppm x ppm x int	2D	++/+++

[1]Reproducibility is different if series of samples are compared on a long term perspective (long) or within a series of samples analysed at a given moment (short).
[2]Low (LR) or High (HR) resolution refers to the resolution of MS or NMR respectively.
[3]The number of compounds is different from the number of features detected. The numbers represent a rough estimation. Methods like GC/MS or NMR will generate many features for each compound detected contrary to LC-MS or DIMS.
[4]In general in MS sample preparation is necessary to maintain instrumental repeatability; this can be very limited as for DIMS for very elaborated as for GC-MS.
[5]In MS relative quantification usually expressed as fold changes might be prone to ion suppression. Ion suppression decrease as follows: (ESI>APCI-APPI) , (DIMS>LC-MS>GC-MS).
[6]In MS sensitivity is strongly dependant on the ionisation methods and the type of instrument used.
[7]For untargeted analyses, the universality of MS will be improved if both PI and NI detection are performed and if multiple ionisation methods are used. DIMS may ionise all metabolites if no ion suppression occurs. LC-MS is more suited for polar to medium polar compounds in RP mode. GC-MS is more suited for volatile or polar small metabolites after derivatisation.
[8]The biomarker identification potential of MS is strongly dependant on available MS or MS/MS databases which are generally instrument specific.
[9]Depending on the MS analyser, resolution isobaric compounds maybe resolved or not. Isomers cannot be seprated even on HR FT-ICR-MS.
[10]Depending on the resolution molecular formulae can be determined unambiguously.
[11]EI ionisation in GC-MS provides spectra that are searchable in widespread databases.
[12]Not applicable
[13]MS/MS acquisition can be used in the MRM mode for semi-targeted metabolomics approaches that give higher sensitivity but are more selective
[14]MS/MS alone or in combination with LC (LC-MS/MS) provide complementary structural information to MS. The identification is however strongly dependant on the MS/MS library at disposal.
[15]Intrinsic sensitivity of NMR is at least one order of magnitude less than MS but the use of new probe technology and high fiel NMR magnet cope in part this problem
[16]NMR enables the detection of all metabolites bearing protons provided that they are soluble in the solvent used.
[17]2D NMR (eg ^{1}H-^{13}C correlations) give and additional dimension for resolving the features

nuclear state in excess of the upper (i.e. the polarisation) by only 1 in 10,000 spins. The net result is that most of the potential NMR signal is cancelled, and only the small excess is detected [64]. Due to this low sensitivity, the required amount of samples is often higher than those of other analytical platforms. One way to improve sensitivity would be to increase the magnetic field because the S/N ratio roughly increases between $B_0^{1.5}$ and $B_0^{1.75}$ (nuclear relaxation and detection efficiency complicates this relationship) [65]. However, this is an extremely costly option because the price of NMR spectrometers roughly doubles for each increase of 100 MHz in ^{1}H operation frequencies. A recent advance in the constant efforts to improve sensitivity was the development of microcoil probe heads where the sample volume can be reduced by a factor of 20–100 compared to conventional probe heads [66, 67]. A further milestone in sensitivity improvement is the use of low temperature probes [68]. In these probes, the sample remains at room temperature, but the electronics, one of the main sources of noise in NMR spectroscopic measurements, are cooled down to 20 K (−253 °C) by bathing them in a cryogen such as liquid helium. These typically enable a 3–4-fold enhancement of the detection sensitivity in high-resolution NMR compared to the corresponding conventional probes [68]. Consequently, it is possible to measure considerably smaller amounts of substances or to decrease the measurement time for a fixed sample concentration.

Another aspect of NMR to be considered, particularly in metabolomics, is the resolution of NMR spectroscopy. ^{1}H-NMR detects all molecules containing protons; therefore, large numbers of signals are measured in a single ^{1}H-NMR spectrum. Hundreds of signals are detected in a ^{1}H-NMR spectrum, and a considerable number of signals overlap, which hampers identification. There are several ways to overcome this problem, and one of the most useful methods is the use of two-dimensional (2D) NMR spectroscopy. 2D NMR is useful for increasing signal dispersion and for elucidating the connectivity between signals, thus helping to identify metabolites. Many 2D NMR experiments are useful for signal assignment and further structure identification, but the spectra can also be used for multivariate data analysis with increased resolution [60, 69].

Considering their sensitivity and resolution, the use of 11.7 T (500 MHz proton frequency) or 14.1 T (600 MHz) magnets is generally accepted in current metabolomics studies. The increased frequency also results in the proton spectra being better resolved. NMR instruments with high magnetic fields are thus being increasingly used in the field of metabolomics.

Another aspect of importance when studying plant extracts is that metabolites can be present over a wide dynamic range and some major metabolome constituents maybe very dominant while minor ones can be difficult to detect or being masked by the major metabolites. In NMR, since the signals recorded are directly linked to the abundance of given compounds very abundant metabolites (e.g. sugars) may render the detection of minor constituents difficult. In MS, because of the specificity of the detection, such type of problem can be avoided but high amount of specific metabolites contaminate the ion source and biased the data. These issues can be partly avoided by adapted sample preparation of by using data mining methods (UV scaling) that give less weight to the intensities of the features in multivariate data analysis (see section 4.3).

3.1.2. 1D and 2D NMR Metabolomic Approaches

^{1}H NMR is routinely used in metabolomics because the NMR fingerprints provide a universal unbiased detection of all molecules that contain protons, and they can be generated with high throughput. A combination of chemical shift (the nature of the chemical environment in which a particular nucleus is located), spin–spin coupling (number and nature of nearby nuclei, and thus connectivity information) and peak intensity (concentration of protons), allow the rapid identification of any components of interest in the analysis. To obtain reproducible results on concentration, parameters such as relaxation delays, pulse widths, and acquisition times are important. The optimum range of the parameters for quantitative ^{1}H-NMR has been well documented [70].

Compared to the analysis of pure NPs, the fingerprinting of crude extract in plant metabolomics requires the optimisation of several NMR parameters. One issue in measuring ^{1}H-NMR spectra is the large signal of the residual non-deuterated water signal, which could overlap with the anomeric protons of sugars or glycosides (δ 4.8-5.2). To suppress this undesired water signal, several methods can be applied.

Popular water suppression techniques are weak radiofrequency irradiation (pre-saturation (pre-sat)), gradient-based methods such as WATERGATE (water suppression by gradient tailored excitation) and combinations of gradient and weak radiofrequency pulses such as WET (water suppression enhanced through T1 effects) [71]. Because of its simplicity and robustness, pre-saturation is the most widely used among these methods.

For accurate suppression, the temperature and pH of the samples should be carefully adjusted because the water signal is greatly affected by these factors. Depending on the molecular size of the metabolites, specific pulse sequences need to be applied. To filter out the signals of small molecules from those of large ones, spin diffusion differences can be utilised. This was shown in an NMR-based study on toxin-induced changes in lipoprotein profiles [72]. In the case of matrix-containing macromolecules (e.g., proteins or lipid vesicles), the application of a spin-echo sequence, such as the Meiboom-Gill modification of the Carr-Purcell (CPMG) pulse sequence, allows the attenuation of unwanted resonances from macromolecules [73].

While the ^{1}H nucleus is the most commonly detected atom in metabolomic studies, ^{13}C has also been employed for such analyses. In ^{13}C NMR, the chemical shift dispersion is twenty times greater, and spin–spin interactions are removed by decoupling. These properties can offer the potential to greatly simplify the acquired spectrum in terms of resonance overlap. In addition, for aqueous samples, there is no requirement for solvent suppression, which can result in the loss of some spectral information in ^{1}H NMR studies. Despite these advantages, the low sensitivity of ^{13}C NMR, due to its lower natural abundance and gyromagnetic ratio, prevents its routine use in metabolomics. One approach to avoid this limitation is to use ^{13}C-enriched samples and to follow metabolic flux. For example, ^{13}C can be transferred from ^{13}C labelled glucose to other endogenous metabolites [74].

The most useful 2D NMR experiment in metabolomics is most likely the ^{1}H-^{1}H *J*-resolved NMR. This experiment provides spectra that exhibit a coupling constant in a dimension that is orthogonal to the chemical shift axis. A sky-line projection of the obtained 2D plot onto the chemical-shift axis yields a fingerprint of peaks from the small molecules, with all spin-coupling peak multiplicities removed. (Fig. **4**) shows one of the examples of increased resolution of this spectrum compared to a standard ^{1}H NMR spectrum obtained on a grapevine leaf extract. The obtained high-resolution fingerprint was used to characterise both the resistant and the susceptible cultivars and their response against the pathogen. Relevant biomarkers were identified after multivariate data analysis [39].

3.2. MS-Based Metabolomics

MS is playing an increasingly important role in the progression of proteomics and metabolomics, and it enables the sensitive detection of most plant metabolites [23]. In simple terms, the MS platform measures of the mass-to-charge ratio (*m/z*) of elemental or molecular species via an experimental workflow that can be summarised as follows [22]: I) generation of charged species, which is necessary to enable ion manipulation in an ion source that is operated at atmospheric or vacuum pressure after sample introduction; II) separation, in space or time, of ions based on their *m/z* ratio in a

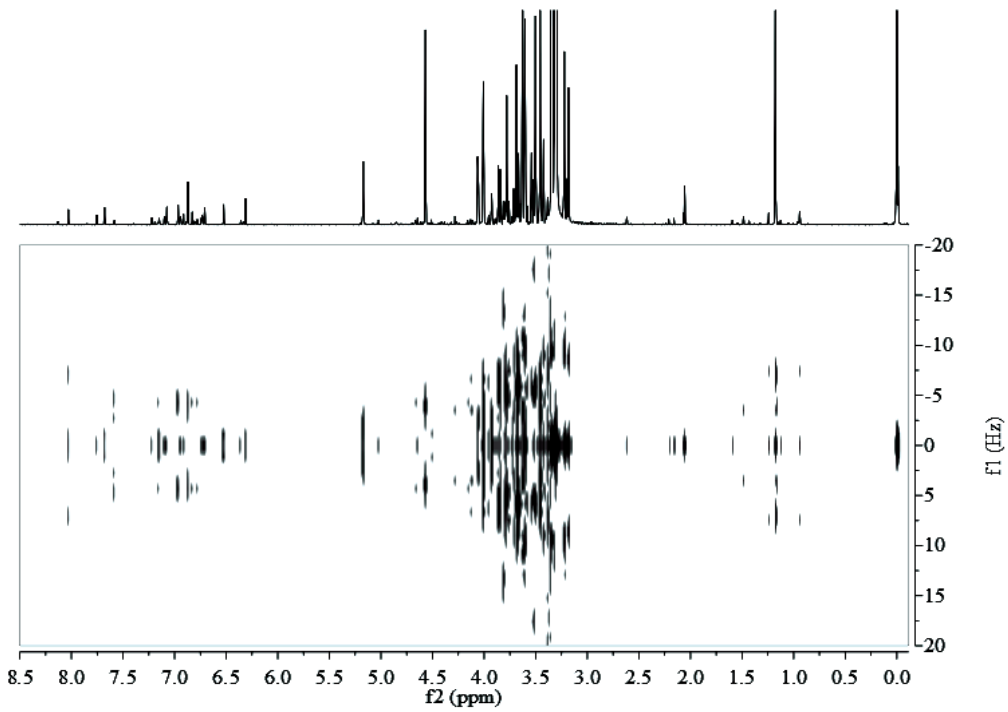

Fig. (4). *J*-resolved NMR spectrum and its skyline projection of grapevine leaf ('Regent' cultivar analysed after 48 hours of pathogen (*Plasmopara viticola)* inoculation). Modified from [39].

mass analyser; and III) detection of ions either physically at a detector as an ion current or by the detection of orbital frequencies as image currents. More details on the mode of operation of mass spectrometers can be found in dedicated reviews [7, 75, 76]. In metabolomic applications, two main steps are critical: metabolite ionisation (the correct ionisation method needs to be selected to observe the maximum of *m/z* features of a given sample) and ion separation by the mass analyser (high or low resolution MS instruments can be used and will strongly influence the number of features detected). Unlike NMR, MS detection is not universal, and MS ionisation is strongly compound-dependent; thus, intensities of the ions observed cannot be directly related to the concentration of the analytes.

MS can be used by the direct infusion of a crude extract solution in the mass spectrometer in an approach known as direct injection MS (DIMS). Alternatively, MS can be hyphenated to different chromatographic techniques that provide separation of the analytes prior to MS detection. For this, three main techniques have been applied: gas chromatography–MS (GC–MS), liquid chromatography–MS (LC-MS) and capillary electrophoresis-MS (CE-MS). The majority of plant metabolomics studies have been performed by GC-MS [54] or LC-MS [55] approaches. CE-MS represents an interesting alternative approach, particularly for the profiling of either polar or charged metabolites, offering high resolution of the analytes [77, 78]. However, it has been infrequently applied to plant tissues.

3.2.1. Ionisation

Ionisation is common to all MS approaches used in metabolomics. The majority of MS studies are based on electrospray ionisation (ESI). This ionisation method can be used with DIMS, LC-MS and CE-MS approaches. In ESI, a high voltage field (3–5 kV) is applied on a capillary through which the sample flow and produces the nebulisation of the infused solution or the liquid effluent, resulting in charged droplets that are directed toward the mass analyser. Various configurations and ion source geometries exist, but the ESI process can essentially be summarised as follows [75]: the electrostatic spraying of a sample solution initially generates an aerosol of charged droplets. Sometimes a concentric flow of gas, such as N_2, is used to facilitate this nebulisation process. The charged droplets consist of both solvent and analyte molecules with a net positive or negative charge, depending on the polarity of the applied voltage. Eventually, ions are released from their surrounding solvent, and these ions travel into the mass analyser of the spectrometer. With ESI, protonation/deprotonation is the main source of charging for biologically relevant ions. Different types of adducts with salts or solvent molecules may also be generated according to the nature of the analytes. It is important to note that analyte ionisation is largely compound-dependant and is mainly governed by either pKa (ESI) or gas-phase proton affinity (APCI). In a good approximation, acidic molecules (e.g., carboxylic acids or phenolics) will mainly produce $[M-H]^-$ in negative ionisation (NI), while bases (e.g., alkaloids, amines) will generate $[M+H]^+$ in positive ionisation (PI). Thus, the comprehensive coverage of most metabolites requires that PI and NI modes be used to survey of a maximal number of metabolites [55].

As an example, (Fig. **5**) displays a typical electrospray PI/NI profile obtained for the crude extract of maize leaves and illustrates the complementarity of both detection modes [79]. While most compounds were ionised in both the PI and NI modes, some were only detected with one polarity. The MS spectra of a specific compound show the diversity of molecular ion species that can be recorded. Thus, a combination of the qualitative information obtained in both models is very useful to determine the molecular mass (M) when an unknown analyte is detected. The molecular formula can be determined if high-resolution MS data are obtained. In this example, $C_{17}H_{22}NO_{13}$ was calculated from the ion at *m/z* 448.109 in NI and $C_{16}H_{21}NO_{11}Na$ was deduced for the *m/z* 426.114 ion in PI. The exact molecular formula that matches both ionisation modes was $C_{16}H_{21}NO_{11}$, which corresponded to DIM2BOA-Glc, a benzoxazinone derivative that is a well-known component of maize leaves [80].

ESI is widely used in proteomic and metabolomic studies. In proteomics, ESI mainly produces multiply charged ions of the

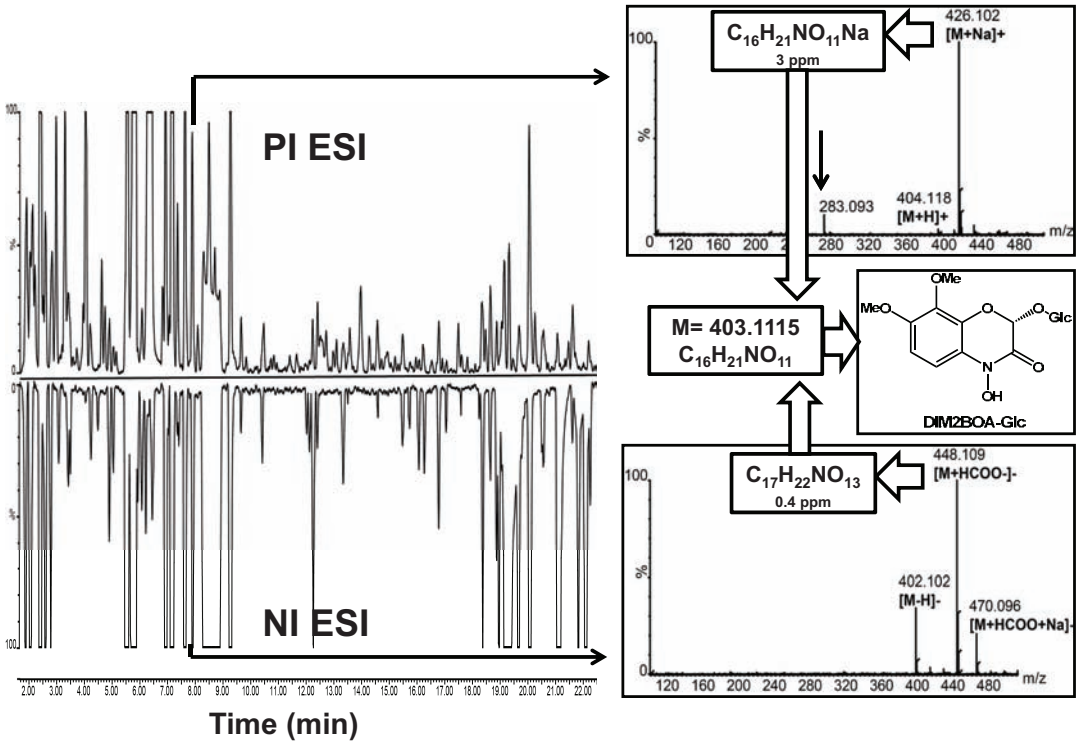

Fig. (5). PI ESI (upper side) and NI ESI (lower side) UHPLC-TOFMS profiling of a crude maize leaf extract. The comparison of both total ion chromatograms (TIC) shows the complementarity of detection of both ionisation modes. Arrows highlight one feature detected in both PI and NI modes. Right panels: PI and NI mass spectra extracted from the peak at RT= 8 minutes highlight several adducts encountered for DIM2BOA-Glc a known benzoxazinoid.

peptide and protein; however, in metabolomics, most of the analysed small molecules are singly charged. ESI provides the ionisation of a very broad array of analytes, which is advantageous for metabolomics studies, but this ionisation method is also prone to important ionisation suppression effects that can limit the reproducibility, accuracy, and sensitivity of relative quantification studies. An appropriate sample preparation method, the use of chromatographic methods with high resolution prior to detection and/or the selection of similar sample matrices decrease the impact of ion suppression on the acquired data [81].

In LC-MS based studies, other alternative ionisation methods can also be used. These methods include atmospheric pressure chemical ionisation (APCI) - ionisation of neutral molecules in the gas phase by chemical ionisation, and atmospheric pressure photoionisation (APPI) - UV light photons are used to ionise sample molecules [82]. These two techniques may be more efficient than ESI for the ionisation of relatively non-polar metabolites. As a rule of thumb, polar and high MW metabolites are more suitable for ESI, low MW compounds of medium to low polarity are better ionised in APCI, and APPI covers an intermediate region in terms of MW and polarity [6]. All these API methods produce a soft ionisation and induce little or no in-source fragmentation. Thus, the features detected in API-MS based metabolomics mainly correspond to singly charged molecular ion species of different forms: protonated or deprotonated molecule adducts, dimers or oligomers that are formed in-source. An example of these various PI and NI ion species is provided in (Fig. **5**).

Another ionisation method that is widely used but that is only commercially available with GC-MS based approaches is electron impact ionisation (EI). In EI, energetic electrons interact with analytes that are already present in the gas phase molecules to produce ions. These electrons are produced by heating a wire filament and are accelerated to 70 eV in the ion source. Close passage of highly energetic electrons induces ionisation by production of radical cations. Compared to the API method, EI is a hard ionisation method that instead of generating stable molecular ion species mainly yields fragments. One of the advantages of the technique is that it will yield reproducible fragmentation patterns in spectra that are directly searchable in GC-MS for biomarker identification. Furthermore, and contrary to LC-MS, ion suppression is very limited in GC-MS. A limitation of GC-MS is that it is only applicable to volatile compounds or compounds that need to be derivatised to be ionised.

Other ionisation methods that can be directly applied on solid matrices include desorption electrospray ionisation (DESI) and matrix-assisted laser desorption ionisation (MALDI) [75]. However, to our knowledge, desorption methods have been less frequently used in plant metabolomics, but they may be important for the spatial localisation of metabolites in imaging MS studies (see below).

3.2.2. Mass Spectrometry Platforms and Resolution

In each of the MS-based approaches described, many types of mass analysers can be used. A few of the key aspects that will affect the quality of the data obtained are the resolution, mass accuracy, spectral accuracy (isotopic pattern reproducibility) and sensitivity of the platform chosen. Accordingly, two main types of mass analysers can be considered: i) low-resolution (LR) MS, such as single quadrupole (Q) or ion trap (IT) mass spectrometers [75] and ii) high-resolution (HR) MS such as time-of-flight (TOF) instruments [75], Orbitraps [83], or FT-ICR-MS [84]. To detect the largest possible number of features and obtain a detailed picture of the plant metabolome, high sensitivity and high resolution are required. However, the cost associated with high sensitivity and resolution instruments should be considered to assess the true benefit in terms of the data generated in relation to a given biological issue. Furthermore, the compatibility of MS analysers, especially when used in combination with chromatography, should be taken into account.

LR-MS instruments such as quadrupoles and ion traps can be used in combination with chromatography approaches (GC-MS and

LC-MS). Such benchtop platforms are relatively cheap and the use of a separation technique before MS detection provides an orthogonal method for resolving the features. In such cases, even a low MS resolution might be sufficient for the acquisition of informative MS datasets. Accordingly, even with a simple quadrupole mass analyser, a GC-MS platform was successfully used to profile more than 300 metabolites in various *Arabidopsis* genotypes [9].

In the field of metabolomics, more expensive and efficient HR-MS instruments [85] should be used to increase the resolution of the analytes in the MS dimension. This is the case for time-of-flight (TOF) MS that provides resolution up to 20,000 and can be operated with a high acquisition frequency over a broad mass range that is compatible with rapid chromatographic methods without compromising sensitivity [86]. The possibility to acquire HR-MS data at high frequency is a key point for the recording of LC-MS metabolite fingerprints. Because a high throughput method is generally needed to handle the large number of samples and replicates that are analysed, fast chromatographic methods such as ultra-high pressure chromatography (UHPLC) are used prior to MS detection [87]. Such fast LC methods generate narrow LC-peaks (typically 3-4 sec width) and many MS scans across these peaks need to be recorded for satisfactory peak area integration in the peak picking process. Other HR-MS instruments, such as Orbitraps [83], also provide an attractive alternative for metabolomic studies. New generations of such instruments are capable of achieving a resolution of approximately 70,000 at acquisition speeds that are compatible with relatively fast chromatography. For the analysis of complex mixtures without separation, extreme MS resolution may be required. For this purpose, resolution that can exceed 1,000,000 can be obtained using Fourier Transform Ion Cyclotron Resonance (FTICR). Although these instruments are expensive, they are being used more frequently as reference instruments for numerous studies [84]. Another advantage of these different HR-MS instruments is that they provide high mass accuracies ranging from 5 to 1 ppm according to the type of instrument used. Such information is particularly useful for peak annotation of metabolite identification, as discussed later. Furthermore, the high mass accuracy ensures the high repeatability of the MS analysis and the precise alignment of the data for further comparison by data mining.

Many of these analysers are used either alone (ion trap type of MS: LIT, QIT, Orbitrap) or in combination (QqQ, Q-TOF) for performing tandem MS experiments (MS/MS or MS^n) that generate structural information by fragmentation of molecular ion species through collision induced dissociation (CID) [6] or that generate data in semi-targeted metabolomic approaches [88]. In this respect, triple-quadrupole (QqQ) MS–MS systems are the most commonly used. Ion-trap (IT) mass spectrometers have the unique capability of producing multiple-stage MS-MS (MS^n) data that may be essential for structural elucidation. For HR measurements in MS-MS, the Orbitrap Fourier transform (FT) MS instrument provides high quality spectra for metabolite identification [89]. In addition to these types of mass spectrometers, there are growing number of additional varieties, including hybrid systems that combine LR and HR analysers for specific applications [82]. Such information is also important for data analysis and will be discussed later.

The principle of operation of these different analysers can be found in several reviews [22, 75, 82, 90]. Some of the key features are summarised in (Table **1**).

3.2.3. Direct Injection MS (DIMS)

One of the easiest ways to rapidly generate a detailed metabolite fingerprint of crude extract is by direct injection or direct infusion MS methods. DIMS is a strategy to introduce a sample directly into an ESI-MS spectrometer for a short time period. In DIMS, the complex extracts are solubilised and injected through the ESI interface or in an automated manner to introduce the sample as a plug supplied by a LC system. Alternatively, samples can be introduced completely independently using chip-based infusion through nanoelectrospray ionisation with a robot that infuses the samples at a nL/min flow rate in a dedicated nano ESI spray nozzle that is used once for each sample [91]. The advantage of such an approach is that it minimises ion suppression and decreases source contamination [92]. The first application of this technique for fingerprinting crude extract was proposed in 1996 on a nominal mass resolution mass analyser for the analysis of fungal extracts [93]. However, as mentioned above, FT-ICR-MS [94] is one of the best detectors for this type of approach. Other less expensive benchtop HR-MS instruments such as TOF-MS [95] or Orbitrap [96] might also be considered. The recorded metabolite fingerprints consist of total mass spectra (TMS) that are summed over the different scans that are acquired during infusion. The features for data mining are referred to as 1D data and are represented by *m/z* ions vs. intensities. Such datasets are naturally well-aligned due to the high mass accuracy of HR-MS and can be easily compared with multivariate data analysis (MVDA) (see below).

This approach has the advantage of being rapid and not requiring chromatographic separation prior to MS detection. Thus, theoretically, all metabolites elute into the MS and none will be retained by the direct injection device. In this respect, it is comparable to NMR, which also generates1D data. DIMS is sensitive and enables the detection of a large number of features as shown in (Fig. **1**). The short analysis time increases the inter-sample reproducibility and improves the accuracy of subsequent cluster analysis [7]. One of the main disadvantages is that isomers, that are often present in plant extracts, will not be separated even if extreme HR-MS resolution is used because they share the same molecular formulae. For example, hexose sugars cannot be discriminated. Because all metabolites are injected simultaneously, significant ion suppression effects may occur, which increases the overall analytical variability and is detrimental for sample comparison. Additionally, contamination by non-volatile residues can cause a rapid deterioration of instrument performance. In comparison, in infusion experiments, great care needs to be taken to avoid sample carry-over. This problem can be overcome by using chip-based infusion technology where every sample is ionised in a separate nano ESI spray nozzle [91].

Because DIMS does not provide chromatographic separation, the ability to identify metabolites is limited [22]. However, the application of high mass accuracy (sup-ppm mass accuracies can be obtained [91]) and MS/MS experiments is useful in providing a higher degree of metabolite identification. The use of tandem mass spectrometry experiments in a more targeted detection manner and the application of multiple reaction monitoring (MRM) provide greater specificity and an improved S/N ratio for certain metabolites [22].

Total MS fingerprinting spectra can also be obtained with alternative ambient ionisation techniques methods such as desorption electrospray ionisation (DESI) [97], extractive electrospray ionisation (EESI) [98] or by direct analysis in real time (DART) [99, 100]. DESI and EESI employ charged liquid sprays from an electrospray source directly to sample biological systems without any or with minimal sample preparation and at ambient conditions [22]. DART, rely upon formation of a (distal) plasma discharge in a heated gas stream and thus is an APCI-related technique [101]. With all these approaches, no extraction is necessary and these methods can be directly applied on the biological material of interest [102] [100]. Similarly, matrix-assisted laser desorption ionisation (MALDI)[103] can also be applied. With MALDI, and unlike ESI where analyte ions are produced continuously, ions are produced by the pulsed-laser irradiation of a sample. The sample is co-crystallised with a solid matrix that can absorb the wavelength of light emitted by the laser [75]. The method is very sensitive and can directly extract MS information from discrete spots on an intact biological sample such as a plant leaf. The high spatial resolution of MALDI enables the possibility to localise metabolites on a biologi-

cal sample by rapid rastering across the sample to collect data from multiple areas in a relatively high-throughput manner. The collected MS information provides spatial mapping of all metabolites simultaneously in an approach defined as MS imaging (MSI) [104]. This approach has important potential in metabolomics. Although it has been used for the analysis of mammalian tissues [105], it has not been widely applied to plants [103]. For example, it was used for the differential mapping and localisation of glucosinolates in specific cells of reproductive organs of *Arabidopsis* [106]. This study demonstrated the potential of MSI for understanding the cell-specific compartmentalisation of plant metabolites and its regulation. Recently, such methods have been used to analyse the induction of metabolites in the confrontation zone of microorganisms in solid media culture [107].

On main disadvantage of MALDI, however, is the presence of the matrix, which causes a large degree of chemical noise to be observed at *m/z* ratios below 500 Da. As a result, samples with low molecular weights are usually difficult to analyse by MALDI [75]. Alternative methods such as matrix-free laser desorption ionisation applied (LDI) on silicon may circumvent such problems [108]. Additionally, very recently NanoDESI (MS), combined with the alignment of MS data and molecular networking, has enabled the direct monitoring of metabolite production of living microbial colonies grown on Petri dishes without the need for chemical tags, labels, or any sample preparation on the surface. This opened new possibilities to study metabolome modifications of microorganisms in a spatial or temporal manner [102].

3.2.4. Hyphenated Mass Spectrometry Methods

As previously indicated, direct MS methods such as DIMS enable the rapid and high-throughput screening of hundreds of samples, mainly for metabolite fingerprinting, but they have limited quantification and metabolite identification capabilities. The coupling of MS with separation techniques (GC, LC or CE; "hyphenated methods") is extremely powerful in terms of the detection, quantification and identification of a wide range of metabolites. Another advantage of the coupling of MS with separation methods concerns the decrease of matrix effects, which negatively affect the detection of low concentration analytes [6]. Hyphenation enables a bi-dimensional detection where each detected feature is resolved in both chromatographic (retention time, RT) and mass spectrometric (*m/z*) dimensions. The analytical signals obtained in this case have the following format: RT x *m/z* x peak aera. This generates intrinsic bi-dimensional (2D) data that render data mining more complex and that require alignment, especially for the chromatographic dimension (see data mining section). These approaches require rather longer analysis times because of the chromatographic separation (typically 10–60 min). The throughput can, however, be increased using faster chromatographic techniques such as Fast-GC [109] or UHPLC [50]. These methods can be used for both rapid metabolite fingerprinting and high-resolution profiling [50]. Another advantage is that once a biomarker is localised in a chromatogram, its physical isolation is possible after the optimisation of the chromatographic conditions. This is a key point for the *de novo* structural determination and/or bioactivity characterisation of a given biomarker [110].

3.2.4.1. Liquid Chromatography Mass Spectrometry (LC-MS)

LC-MS has been applied to various fields of plant science since its introduction as a practical analytical technique in the late 1980s [111]. Historically, this method has required more technological development than GC-MS because the main issue in LC-MS has been handling the relatively high liquid flow rates out of HPLC that are often not compatible with the high vacuum required for the MS detection. Since the early 1980s, many different interfaces have been developed to address this issue and overcome the inherent difficulties [112]. Currently, the overwhelming popularity of LC-MS is largely due to a series of interfaces operating at atmospheric pressure (API), including ESI and APCI and to a lesser extent APPI, as discussed above. LC-MS is currently the most widely used mass spectrometry technology, particularly in the life science and bioanalytical sectors [113].

Numerous types of mass spectrometers described above can be used for LC-MS applications. Low-resolution (LR) mass spectrometers, such as single quadrupoles (Q), can be used and are the least expensive. High-resolution (HR) mass spectrometers and those with exact-mass capabilities, such as the latest generation time-of-flight (TOF) instruments and Orbitraps, are becoming increasingly popular and are recognised for their excellent performance for such studies.

LC conditions

For metabolomics, LC-MS can be applied to an important range of metabolites presenting various physico-chemical properties. The sample preparation is minimal and is principally critical for ensuring good repeatability of the LC-MS analysis of a large number of samples to avoid column clogging or the irreversible adsorption of metabolites on the column stationary phase. Generally, no derivatisation is required. Compared to DIMS, however, the optimisation of chromatographic conditions prior to MS detection is necessary and the choice of the type of column eluent will influence the number of metabolites that are amenable to MS detection. To detect the maximum number of metabolites with various polarities, the LC-MS separation needs to be performed using broad organic-aqueous reversed phase (RP) gradients using either methanol-water or acetonitrile-water systems with mostly acidic modifiers such as formic acid to aid the ionisation and/or separation [114]. Other buffers might be considered according to the type of analysed compounds. The choice of column is strongly dependant on the polarity and the physicochemical properties of the metabolites that will be analysed. The majority of LC-MS plant metabolomics applications are performed using C18 or C8 RP columns, and several generic protocols have been developed for plant tissue analysis using C18 gradient separations [50, 55]. These types of separations are slightly biased toward semi-polar secondary metabolites. However, within the same extracts, many primary metabolites such as several organic acids, nucleotides, amino acids, sugars and their phosphorylated forms, can be detected, but they usually co-elute with other compounds in the injection peak [55].

To resolve very polar constituents, other types of stationary phase might be considered such as Hydrophilic Interaction Liquid Chromatography (HILIC) [115]. This type of partition chromatography occupies the opposite end of the partition spectrum from RP liquid chromatography. Contrary to separation on RP, HILIC separations are difficult to predict because the partition mechanism involves hydrophilic partition, electrostatic interactions and hydrogen bonding between a polar stationary phase and a highly organic mobile phase. In this case, the proportion of water is increased during the analytical run. HILIC might solve problems where the separation of very polar analytes is required as in the case of calystegines from *Schizanthus* species [116]. Another solution is to use ion-pair reagents for improving the retention in RP chromatography [117]; however, this might have a detrimental effect when MS detection is used. An evaluation of coupling RP, aqueous NP, and HILIC with MS for metabolomic studies of human urine has recently been conducted [118]. The study showed that an extended coverage of the metabolome could be obtained by combining these chromatographic methods. However, to our knowledge, such approaches have not yet been used in plant metabolomics. One practical way to rapidly obtain rather extensive metabolite coverage is to analyse the samples by both HILIC and RP chromatography. This option is straightforward and has been applied to the analysis of body fluids. The combined chromatographic approach generated two datasets that are unfortunately difficult to merge without the use of specific software programs [113].

For the separation of very apolar metabolites such as lipids, terpenes, carotenoids or pigments, normal-phase (NP) chromatography represents an attractive alternative; however, the solvents used for such separations are usually not compatible with MS. Some of these compounds (e.g., fatty acids or terpenes) are amenable to GC-MS and can be analysed by NP-GC-MS. Another interesting solution is the use of supercritical fluid chromatography (SFC), which is compatible with MS detection, and can provide a bridge between GC and LC applications. An additional analytical method termed convergence chromatography™ has recently been introduced by Waters company that is based on the use of sub-2 µm particles, as in UHPLC, and offers an alternative option for NP chromatography applications.

In metabolomic studies, access to fast chromatography methods that provide high separation, resolution and high repeatability is a key to generating high-quality datasets. In this respect, the introduction of ultra-high pressure liquid chromatography (UHPLC) systems operating at very high pressures and using sub-2 µm packing columns has allowed a remarkable decrease in analysis time and increases in peak capacity, sensitivity, and reproducibility compared to conventional HPLC. This technology has rapidly been widely accepted by the analytical community and is being gradually applied to various fields of plant analysis [87], especially metabolomics [119]. For this type of analysis, HPLC can be transferred to UHPLC conditions by applying gradient transfer methods while maintaining the same resolution [120]. By selecting the appropriate column length in UHPLC, it is theoretically possible to increase the throughput by a factor of 9 compared to conventional HPLC [121]. Such fast separations are critical for performing repeatable high throughput metabolite fingerprinting. Because UHPLC provides very narrow LC peaks, the use of MS detectors with very fast responses such as TOF-MS is generally required. Currently, UHPLC-TOF-MS is recognised as being very efficient for studies of plant and mammalian metabolomes [122-124].

UHPLC-TOF-MS can also provide very detailed high-resolution metabolite profiling. This type of analysis might be required for the separation of closely related isomers that need to be separated for specific studies. Gradient transfer can be applied from fast fingerprinting to high-resolution metabolite profiling on pooled samples to maintain the same separation selectivity. The HR separation allows a precise localisation of the biomarkers of interest in the context of their isolation or the recording of MS or MS/MS spectra of fully deconvoluted LC peaks [50]. Thus, by maintaining strictly identical HPLC and UHPLC column lengths, it is hypothetically possible to increase the theoretical plate number by a factor of 3 between columns packed with 5 and 1.7 µm particles and to reach up to 40,000 theoretical plates with a 150 mm, 1.7 µm packing [124]. By coupling several columns together and combining high temperature for a substantial reduction of the generated backpressure, very high peak capacities can be obtained for plant metabolites [124] and peptides [125].

In LC-MS, as is the case for DIMS, to maximise metabolome coverage, analyses in both PI and NI ion modes is recommended (see Fig. **5**). According to their duty cycle frequencies, some MS instruments can simultaneously perform this double detection in a single run. With high throughput fingerprinting, however, the recording of separated analyses is recommended to maintain optimal instrument performance. An example of metabolite fingerprinting and metabolite profiling of *A. thaliana* recorded by UHPLC-TOF-MS is shown in (Fig. **6**). Results from mass spectrometric analysis at low and high resolution for the fast fingerprinting and metabolite profiling of a given metabolite are illustrated.

3.2.4.2. Gas Chromatography Mass Spectrometry (GC-MS)

GC-MS is the oldest but most likely one of the most robust hyphenated MS methods. The first application of this technique was published in the early 1960s [126]. Because the analytes are already in the gas phase during the GC separation, the hyphenation with MS is straightforward and ionisation methods such as EI, as described above, can be directly applied. GC-MS has been one of the major analytical tools in the early development of metabolomics [127, 128] and currently remains a widely applied method [129]. In GC, the analytes are evaporated, separated and then carried through a column by an inert 'carrier' gas such as helium. The separation process of the analytes is mainly carried out between a liquid stationary phase and a gas mobile phase at an elevated temperature typically by applying a temperature gradient. All modern separations are carried out on capillary GC columns that provide high-resolution separation of complex mixtures. Most of the GC columns are silica capillaries 10–60 m in length with internal diameters ranging from 100 to 500 µm, externally coated with an imide layer to reduce column fragility and internally coated with a 10–50 µm thick liquid stationary phase. The stationary phase is generally obtained from siloxane or a compound with a similar chemical composition, with varying percentages of different chemical moieties. DB5 (95/5 mix of methyl and phenyl groups) or DB17 (50/50 mix) stationary phases are generally employed in metabolomics [22]. Separation of the constituents is optimised by an appropriate selectivity afforded by the selected column and can be modulated by gradient temperature [6].

When more physical resolution of the analytes is needed, comprehensive GC (GC x GC) can be employed. In GC x GC, two columns are connected sequentially. Typically, the first dimension is a conventional column and the second dimension is a fast GC column, with a modulator positioned between the two columns. This type of multi-dimensional chromatography is robust and has efficiently begun to be applied in metabolomics [129]. Initially, GC x GC-based metabolomics studies were hampered by the lack of appropriate data processing, alignment, and analysis tools. Today, numerous algorithms are available, but the entire procedure of obtaining biological knowledge from raw data still awaits complete automation [129]. Contrary to fast LC methods, fast GC methods are not yet used in metabolomics, but newer developments in high peak capacity separations in GC-TOF-MS might soon be applied to the field [109].

In GC-MS, volatile compounds are directly monitored, while other metabolites of molecular mass less than approximately 300-400 Da can be analysed after appropriate chemical derivatisation [22]. Various chemical derivatisation procedures are well established for the detection of non-volatile polar or non-polar metabolites [54]. For non-polar metabolites, the use of TMS enables the replacement of exchangeable hydrogens with trimethylsilyl groups, reduces intra- and intermolecular hydrogen bonding and increases volatility [51]. For polar analytes, a two-step process is necessary and before sylilation, oximation is performed to convert carbonyl moieties to an oxime as carbonyl groups react slowly with trimethylsilyl reagents. These powerful sample preparation steps allow the detection of various chemical classes including amino and organic acids, fatty acids and some lipids, sugars, sugar alcohols and phosphates, amines, amides and thiol-containing metabolites. However, derivatisation can introduce greater technical variability and complexity to the data, as a single metabolite can produce multiple derivatised metabolite peaks [22].

In GC-MS, the most widely used ionisation technique is EI, as described earlier. GC-MS provides high-resolution separation and has considerably less ion suppression effects than LC-MS. The main advantage of GC-EI-MS is that it provides reproducible fragmentation of the molecular ion at 70 eV, which is searchable in widely available libraries that greatly facilitate metabolite identification [130]. A disadvantage of EI is that many ions are related to a given metabolite, which complicates the mining of such data. This is in contrast to LC-MS where only very few features (molecular ion species) are linked to a given compound. These steps often require the deconvolution of co-eluting peaks [131] and peak

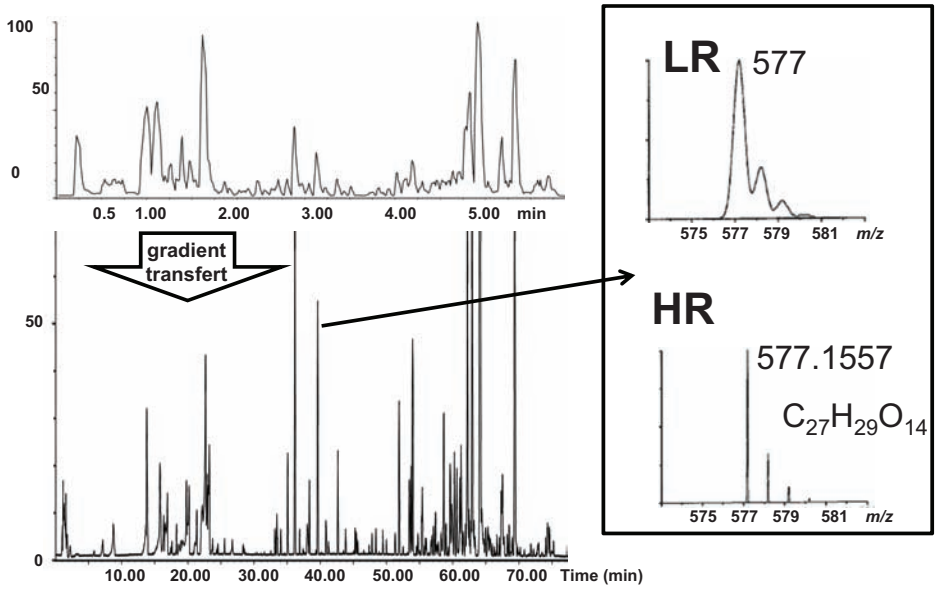

Fig. (6). Fast fingerprinting and high-resolution metabolite profiling of *Arabidopis thaliana* by UHPLC–TOFMS in NI ESI mode. This example demonstrate the difference in resolution that can be obtained by using a short gradient for fingerprinting on a short column (50 x 1.0 mm, 1.7 µm) or a long gradient on a long column (150 x 210 mm, 1.7 µm). The same selectivity can be maintain by applying a geometrical gradient transfer (adapted from [50]). The insets represent the increase in resolution that can be obtained on the isotopic pattern of a given metabolite of MW 578 (the $[M-H]^-$ is displayed) from low resolution MS data obtained on a quadrupole MS detector (R : 1'000) and by high resolution MS with a TOFMS detector (R :20'000). The high MS and spectral accuracy of the TOFMS also provide molecular formula information.

identification for further mining of the data [132]. GC-MS data can also be two-dimensional, as in the case of LC-MS.

GC-MS can be performed on a simple quadrupole instrument and is presently very efficient for providing relevant plant metabolomic data [9, 128]. However, most of the current work is performed on a GC-TOF-MS platform [131] because these platforms are very efficient for the profiling of primary metabolites after derivatisation [133].

3.2.5. Combined Analytical Approaches

Although each analytical platform presented can provide important information on metabolite composition, no individual platform is sufficient to grasp the complete chemical complexity of plant metabolomes. Therefore, the combination of information obtained from the same samples with multiple experimental platforms (e.g., NMR, DIMS, GC-, CE- or LC-MS) is expected to extend the coverage of the metabolites that characterise a biological system [134, 135]. In general the analyses are performed independently on the same samples with different methods. Sometimes in MS analysis, PI and NI detections can be alternated during the same analytical run using specific ion sources. ESI and APCI modes can also be switched in the same manner. The use of two MS detectors for the detection of features from a common UHPLC separation after equal splitting of the eluent has also been investigated [136]. Interestingly, it was shown that for the analysis of a urine sample, after multivariate statistical analysis, several ions were found to be unique to one data set or the other, a result that may have consequences for biomarker discovery and interlaboratory comparisons. The software package used for data analysis also had an effect on the outcome of the statistical analysis [136].

Similar types of data can be compared by data fusion. Otherwise, for analysis using orthogonal methods, e.g., NMR-MS, the intrinsic covariance between signal intensities of the same and related molecular fingerprints can be calculated. With such methods, features that are correlated in both NMR and MS datasets can be determined [137]. New chemometric methodologies (see below) are therefore needed to address the challenges that are associated with orthogonal or redundant data.

A series of combined approaches has recently been used for assessing key plant metabolome variations. For example, a multi-MS-based platform (GC-MS, LC-MS, and CE-MS) approach was used to assesses the substantial equivalence of tomatoes over-expressing the taste-modifying protein miraculin [138]. The approach was found to have an acceptable range of variation while simultaneously indicating a reproducible transformation-related metabolite signature. Similarly, combined NMR and LC-MS analyses revealed the metabolomic changes in *Salvia miltiorrhiza* induced by water depletion [139]. Metabolite fingerprints of ripe fruits from 50 different tomato cultivars were recorded by both ^{1}H-NMR and LC-QTOF-MS. NMR and LC-MS provided complementary datasets. Unsupervised multivariate analysis of the NMR and LC-MS datasets revealed a clear segregation between cultivars. Intra-method (NMR / NMR, LC-MS / LC-MS) and inter-method (NMR / LC-MS) correlation analyses were performed that enabled the annotation of metabolites from highly correlated metabolite signals. Inter-method correlation analysis produced informative and complementary information for the identification of metabolites, even in the case of low abundance NMR signals [140]. Combined NMR and MS approaches with covariance analysis were also recently applied to search for metabolites of resistance in various *Vitis* cultivars. Protocols that propose combined DIMS and NMR approaches are available for high-throughput plant metabolomic analysis [141].

4. DATA MINING AND MULTIVARIATE DATA ANALYSIS

Significant improvements are still required to address the analytical issues related to the global monitoring of the plant metabolome, and the automated processing of large metabolomics datasets remains an important challenge. Indeed, raw data from any of the 'omic' fields are an eventual source of information and in turn a source of *knowledge* [142]. However, to make the leap from one to the next requires considerable data processing, statistical analysis, and suitable data storage formats [143, 144]. Thus, the multivariate nature of data obtained from metabolomics experiments requires several steps to extract relevant information. The typical workflow includes the pre-processing and pre-treatment of the raw data, variable selection, modelling of the data and statistical validation.

The first important distinction that should be made concerns separative and non-separative analytical techniques. Spectral or non-separative methods, such as conventional NMR or DIMS, allow responses that correspond to a combination of multiple compound contributions to be obtained (e.g., multiple compounds may have a common NMR chemical shift, and isomers will have the same *m/z*). Spectral or pseudo-spectral information is obtained for each sample, and this structure is more trivial to manage because a one-dimensional vector of variables is obtained. Multiple individuals can be summarised in a conventional data table. Consequently, classical chemometric tools such as PCA and PLS-based algorithms can be applied. The remaining difficulty is to deconvolute the discriminant signal into pertinent information that can be exploited, e.g., unambiguous metabolite identification. Therefore, non-separative techniques are primarily used for fingerprinting, chemotaxonomy or simple classification studies, including developmental changes.

Although some important differences exist between the various hyphenated methods, such as LC-MS or GC-MS (i.e., ionisation mode, physicochemical properties of the detection compounds, etc.), the data structures could be considered similar because multi- or at least bi-dimensional structures are obtained. As an example, in LC-MS, the detected features correspond to analytes (or fragments) that present a certain *m/z* ratio at a certain retention time; a table is obtained for each sample with N variables (*m/z* value) X P retention time coordinates. In this case, when dealing with numerous samples, the complete data analysis generates a cube of data (N variables (*m/z* value) X P retention time coordinates X K individuals), which cannot be summarised in a conventional data table, but can be summarised in a (at least) three-dimensional data structure. Even today, some multi-way chemometric approaches, such as PARAFAC [145], Tucker [146] for unsupervised models and N-way projections to latent structures (N-PLS), are available to manage such high-dimensional data structures, the most common method consists of managing a two-dimensional matrix (i.e., a table) where the individual elements are identified in the first column and are defined by variables in the subsequent columns. For this type of structure, matricisation (also called unfolding or flattening) intends to reorganise the elements. Various matricisation alternatives are offered to reorder this tensor, such as frontal, horizontal or lateral slices from a three-way data cube that can be rearranged by a row-wise or column-wise concatenation to be analysed by conventional multivariate data analyses.

Finally, the combination of multiple data sources will undoubtedly provide a more comprehensive vision of metabolomics by extracting common traits from different datasets obtained from the same plant samples. Mining of the data should provide a powerful way to interrogate datasets from various viewpoints based on the specificity of the analytical platform used for the acquisition and for the biological question to solve. In this context, many efforts should integrate various dimensionality reductions strategies.

4.1. Non-Separative Or Spectral Methods (1D Structure)

When managing a 1D data structure (e.g., 1D NMR of DIMS), the proper correspondence of the variables across multiple samples constitutes a critical prerequisite for ensuring the validity of the data treatment. The alignment of detected features in different samples aims at removing the shifts between samples for a given signal. Numerous experimental factors, including the temperature, pH, sample carryover or degradation, may lead to differences that affect the overall signal (fingerprint). Although NMR appears quite robust in this context, shifts in the *m/z* dimension should be corrected with a permanent MS calibration. Several alignment techniques have been developed to minimise run-to-run shifts [22, 60-62]. As previously mentioned, NMR provides 1D data where the intensity of all features can be directly correlated to the abundance of the given metabolites. Shifts related to variations in the pH may occur, and the use of a buffer in the solvent avoids possible fluctuations of the signals in the NMR spectra. The effect of different alignment methods for NMR spectra on the classification results have been recently discussed [147].

The acquisition of NMR data is extremely useful when comparisons are performed based on the primary constituents of a plant extract or when samples must be compared during long-term studies. Indeed, NMR fingerprints are very reproducible, independently of when the analyses are performed, contrary to MS-based approaches where group clustering is easily biased by instrumental noise [40].

Contrary to NMR, MS can be used to detect minor compounds among complex plant extracts, but it is prone to more variation, such as ion suppression. Furthermore, detection might be strongly altered based on the level of contamination of a given instrument. In MS, the measured *m/z* will not shift based on the nature of the sample being analysed. In this case, features could be detected based on a peak picking strategy with tolerance limits provided by the analyst. The limit of tolerance will be sharper if HR-MS instruments are used rather than LR analysers.

For these 1D data structures, the binning of the data constitutes the simplest route to proceed. Misalignments are subsequently limited to the bin boundaries as the peaks can be alternatively allocated to neighbouring bins [148]. However, such a process induces a decrease of the initial resolution, which could be detrimental for the interpretation of the measured signal. Additionally, the variations could be non-linear [149] and more sophisticated methods may be required. Segments of adjustable width are defined and sequentially aligned [150]. Therefore, the size of the segments plays a central role for the adequacy of the process.

For NMR spectral data analysis, a digitalisation to numeric values for further statistical analysis has currently been achieved. In this procedure, the NMR spectrum is divided into a series of small bins (usually of 0.02 or 0.04 ppm). The sum of the signal intensities in each bin is then calculated. The disadvantage of equal size binning is related to peak splitting between two bins or peak collapsing within a bin. Therefore, the peak frequency may significantly influence the data analysis. Thanks to algorithms being able to account for peak positions and therefore focus on a better peak definition, it is possible to obtain binning where one bin covers only complete peaks [151].

Scaling to the sum of the total spectrum intensity helps to minimise the effect of variation between samples. In the case of DIMS, raw data are recorded as a single file that corresponds to a specific sample, and hyphenated data files are composed of a set of mass spectra that are recorded in sequence. Because a considerable number of variables are measured, the datasets not only become larger in size but are also more complex; therefore, pre-processing of the data is often required.

It has to be noted that structure elucidation is limited because there is no physical separation of the metabolites in spectral methods. However, with proper alignment of the features of the samples, the data treatment remains quite straightforward. The typical structure of the raw data obtained from a particular study with various samples converges to a simple data table where the values are the signal intensity at a given *m/z* or ppm. In this case, all of the conventional chemometric tools are directly available for data treatment (e.g., unsupervised: PCA, HCA, etc. or supervised: PLS, ANN, Decision trees, etc.). These tools are available in particular computing environments and commercial software dedicated for multivariate data analysis.

4.2. Hyphenation of MS to Separation Techniques (2D Structure)

As mentioned, the use of chromatographic techniques, such as LC or GC, before the MS analysis improves the resolution of fea-

tures by adding an additional dimension for the detection. With the use of such methods, isomers that could not be detected may be resolved by the chromatographic separation. However, the hyphenation generates multidimensional datasets that require specific algorithms to be deconvoluted.

All bi-dimensional datasets require specific data processing, and several steps are required to render the data meaningful and to obtain properly aligned features detected across multiple samples. Therefore, data processing is a crucial component of the metabolomics approach and is required to extract relevant signals from the raw data before data mining and interpretation [152]. Data processing includes data format conversion, noise filtering, normalisation, chromatographic alignment, peak detection, deconvolution and integration. A schematic that summarises how to convert and process 1D and 2D data is presented in (Fig. **7**).

The most common objective when managing an intrinsic 2D data structure is to first convert the initial data into a vector for each sample. This step (from 2D to 1D) allows an information-rich profile to be obtained that is suitable for pattern recognition and clustering. This step could be achieved by collapsing or signal detection (e.g., peak picking). For the former, two methods for dimensionality reduction are available, namely collapsing the signal axis or collapsing the time axis. As an example, raw LC-MS data could be collapsed into spectrometric profiles, (i.e., total mass spectra: TMS) where a fingerprint characterises each sample and only the MS information is retained. This approach differs from DIMS because sample infusion could lead to the ionisation suppression of several compounds when complex samples are directly introduced into the MS. When applying the time collapsing of initial 2D data, the chromatographic step performed during the data acquisition decreases the matrix effect by reducing the number of competing analytes that simultaneously enter the MS ion source, which leads to lower detection limits. Therefore, the obtained TMS is generally of higher quality for data treatment [153, 154].

The proper correspondence of the variables across multiple samples constitutes the same critical prerequisite to ensure the validity of further data analysis. The alignment of features detected in different samples aims at the removal of shifts between samples for a given compound and is similar as that achieved in the intrinsic 1D data structure. The primary difference remains that other alignment strategies should be employed on the raw data. As an example, several warping methods have been developed [155-157], and sophisticated algorithms have been proposed based on the experience of performing chromatography with conventional detection, such as correlation optimised warping (COW) [158] and dynamic time warping (DTW) [159]. Note that such iterative alignment procedures may be prohibitively time consuming when dealing with large datasets. Other alternatives exist, such as kernel density [148], component-resolving algorithms [160], progressive clustering [161] etc. After peak alignment, gap filling is usually applied to fill missing values when peaks could not be detected in all samples. This procedure avoids the inclusion of many zero values that would have detrimental effects on further data modelling.

The signal detection can be performed through the selection of signals that correspond to local maxima or fitting more elaborate mathematical models, e.g., Gaussian distributions. The apex and the inflection points are used for area integration [162]. Constraints on the peak shapes and criteria of minimal intensity, area or signal-to-noise are usually applied to distinguish true peaks from noise. Several parameters generally need to be adjusted to match the characteristics of the data (NMR or MS-based acquisition).

After this procedure, the information-rich profile is constituted for a sample by the detection of features that correspond to a signal intensity observed at two coordinates (e.g., *m/z* – time, ppm – time, etc.).

4.3. Data Pre-Processing

As shown in (Fig. **7**), the standard output of pre-processed raw data from a complete metabolomics study would appear as an aligned data table where each row corresponds to a particular sample that is described by a large series of columns, which correspond to the variables. The number and the nature of the variables represent a critical issue with respect to data modelling, specifically when dealing with multivariate data of high dimensionality where biologically relevant hypotheses are harder to find. Suitable variable selection and scaling methods are selected according to the characteristics of each analytical platform and/or the biological phenomena.

4.3.1. Variable Selection

Because the prediction accuracy is often decreased by highly correlated or irrelevant variables, supervised models can be improved by a prior variable selection [163]. Several other advantages could be obtained, including a decrease in computational time.

Variable selection is generally achieved using a two-step procedure, i.e., the generation of variable subsets and the estimation of their respective predictive or clustering ability. Individual variables or groups of potentially interacting features can be selected according to various criteria, such as redundancy removal or biological knowledge integration. Note that an individual evaluation is unable to account for putative interactions between variables; therefore, synergistic effects may be missed or lost. Data redundancy and correlations are undesirable and often detrimental to modelling. Computing correlations between variables constitutes the simplest method to evaluate redundancy.

The selection within a group of only one of the correlated features or the construction of new variables starting from the original correlated variables (variable construction/ transformation) could lead to a more compact and interpretable image of the study [164]. However, this selection should be handled with caution in metabolomics because redundant signals may also be informative to evaluate metabolic pathways or evidence networks of interrelated compounds [165].

4.3.2 Variable Normalisation

Normalisation strategies aim at reducing the effects of undesirable variability sources or systematic bias and to ensure the reliability of measurements and comparisons between samples over the entire dynamic range, from highly concentrated to low abundance metabolites.

Two main categories can be defined, namely methods dedicated to correct variations (i) between samples and (ii) between individual metabolites.

The former intends to reduce variations between samples, which can be due to analytical noise, experimental bias, biological variability or confounding factors (e.g., nutrition or medication). These procedures allow increasing differences between experimental groups (e.g., case vs. control) to render biological signals easily discernible. Such methods include the unit norm (scaling to the sum of the total spectrum), median intensities normalisation [166] or more sophisticated data transformation methods such as the cubic spline [167] and quantile normalisation [168].

Because many factors could increase data variability in biologic studies, including the intrinsic biological differences of the studied samples or the overall analytical variance (e.g., sample preparation, instrumentation, etc.), it was recently proposed that quality control samples (QC) should be included in each data acquisition set [169]. QCs are generally composed of a representative sample, which could be a pool of multiple individuals that present similar characteristics. While inter-individual differences are intrinsic to biological phenomena, undesirable variability and bias may also be added during the data processing. Therefore, numerous steps are crucial

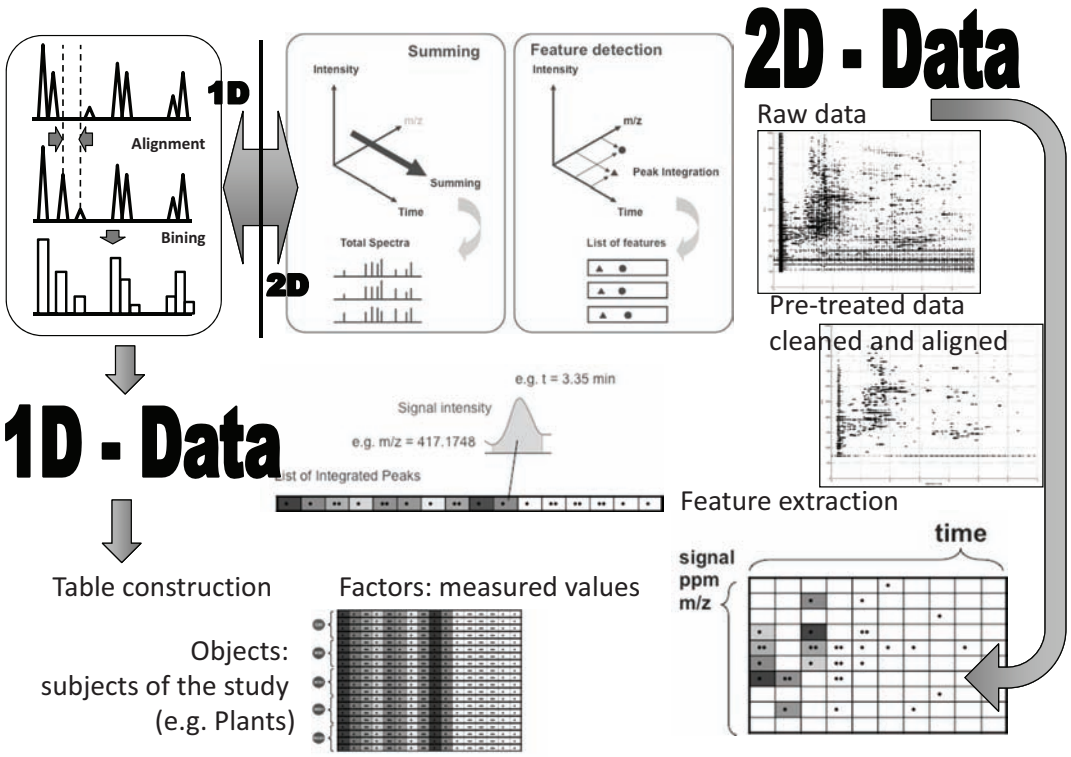

Fig. (7). Sample data table construction from 1D and 2D data. On left side, after relatively simple data treatment, e.g. alignment and binning, monodimensional data (1D-Data) will give a table of results that can be analysed by conventional multivariate data analyses. On the right side, specific pre-treatment for bi-dimensional data (2D-Data) are achieved. In this context, feature extraction is often mandatory (see text).

for the appropriate processing of these datasets to distinguish relevant biomarkers from the mass of recorded signals. QC samples are evaluated within the complete data set as an image of the global variability of the measurement system. Thanks to conventional statistics (relative standard deviation, student's t test, etc.) or multivariate data analyses such as PCA, the studied samples could be evaluated in terms of variability compared to QCs. The latter could present a relatively low RSD to observe a particular phenomenon in the data set. An initial overview of the quality of the run was obtained by the PCA of the complete data set, including all of the QC injections (i.e., including the initial conditioning injections). Based on the hypothesis described above, the closer the QC samples appear on the scores plot, the more reproducible the performance of the analytical system should be [170]. Because the biological effects related to concentration changes could vary greatly from one metabolite to another, the concentration variability of individual metabolites can fluctuate. This variation related to individual metabolites, namely heteroscedasticity, could be detrimental for observing a particular situation [171].

While a fine concentration tuning may be required for a given metabolite, drastic modifications can have very little phenotypic impact for others. Due to analytical reasons, low concentrations of minor metabolites are often more subject to measurement errors than high abundance molecules. It has often been demonstrated that compounds present at lower concentrations are more difficult to measure and consequently more altered by the analytical noise. Furthermore, highly abundant metabolites are not always the most related to the most biologically relevant phenomena under study. To normalise the variances of the different metabolites and make them comparable, scaling procedures are applied.

Different scaling methods, mainly, Unit Variance and Pareto could be performed. Unit Variance (UV) -scaling is the most standard method while Pareto-scaling is more often used with MS data, but also interesting in some case for NMR-based experiments. In UV scaling, each variable is weighted with the inverse of its standard deviation (1/SD). Each variable then has equal (unit) variance. Pareto is a softer scaling technique that increases the importance of low intensity ions without significant amplification of noise by weighting each variable with the inverse of the square root of its standard deviation (1/sqrt(SD)). Although UV scaling may emphasize analytical noise by giving importance to low intensity signals, some information may be lost due to Pareto scaling. Therefore, scaling is another important part of data preprocessing and should be carried out carefully.

Some reports suggest that such scaling approaches may deteriorate the signal-to-noise ratio leading to impaired data [172], and other strategies such as variance stabilisation normalisation have been proposed in the context of microarray data analysis as valuable alternatives [173]. Additionally, a mathematical transformation can be helpful to correct skewed data before modelling. The log function constitutes a well-known transformation that is applied to correct heteroscedasticity [174].

4.4. Data Mining

To extract the relevant information from the vast amount of data generated in metabolomics datasets, data mining in metabolomics constitute a great challenge from a statistical point of view, [152]. The choice of the data analysis strategy significantly depends on the questions that are asked. While a very few, highly reliable metabolites can be sufficient for diagnostic purposes (as in the case of disease evaluation), an extended set of compounds may be desired when a biochemical network is under examination. The choice of the most appropriate model for a given dataset therefore constitutes an important issue.

Univariate statistical tests, such as Student's t-test, initially used to identify the relevant variables remains of limited interest to provide trends and patterns among samples or relevant biomarkers within large metabolomics datasets. Therefore, in the last few years, considerable opportunities to increase the understanding of biologi-

cal phenomena through the use of multivariate data mining methods with respect to their strengths and applicability have been observed. Numerous algorithms, generic or dedicated to a particular data structure, have been applied in this context, and an overview of the current chemometric possibilities is out of the scope of the present review.

As shown in (Fig. **8**), the data mining workflow begins with an unsupervised analysis to provide the first contact with the dataset. This step is useful to assess the samples' distribution and detect potential outliers. Unsupervised statistical tools facilitate the first understanding of the relationship between the samples. The models can also provide information about the variables that are responsible for these relationships. Visualisation tools are mandatory to assess the interpretability and the usefulness of the model with respect to the data at hand. Some of the most common methods are Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA) and other agglomerative solutions, such as the K-means. In most of the cases, the next step consists of applying supervised methods to take advantage of prior information for the analysis of a set of observations. This information could be quantitative or qualitative in the context of classification. Several techniques have been developed for that purpose, originating from the statistical, chemometric or machine learning background. As previously mentioned, a discussion about the pros and cons of each supervised algorithms cannot be performed here, but note that over several years, partial least squares (PLS)-based methods have proven their usefulness in various applications. When a class attribute has to be predicted (e.g., known groups of observations), the PLS discriminant analysis (PLS-DA) has been demonstrated to be a potent tool for the classification of data from metabolomics experiments [175]. The orthogonal PLS algorithm (O-PLS) [176] and O2-PLS [177] have also recently been proposed to allow an easier interpretation of the models by separating the Y-predictive variability from the orthogonal one.

Other approaches, such as Decision Trees, Kernel Methods, Artificial Neural Network (ANN), and support vector machines (SVMs) could be applied as well as exploratory analysis, variable selection, and biomarkers discovery or classification issues in a supervised manner [178].

4.5. Software

Several commercially available or free software packages that have implemented specific parts or the complete metabolomics data processing procedure have been available for several years, and the relevant reviews are regularly updated. The use of software is more or less user-friendly, depending on the implementation of a graphical or a command line user interface. For MS based acquisition, free bioinformatics tools include MZmine [179], XCMS [148], MetAlign [180], MSFACTs [181], TagFinder [182], MET-IDEA [183], MathDAMP [184], msInspect [185], OpenMS [186] and MetaboliteDetector [187].

For NMR based approaches, a very recent review has compared several commercial or open source solutions for the development and analysis of the NMR metabolomics data set, including NMR spectra pre-processing and the primary chemometric methods used to analyse this type of data. Because not all of the processing or analysis steps are included in all of the packages, a useful comparison list has been reported [188].

4.6. Combined Approaches

Modern analytical platforms have proven their worth by providing insight into complex networks of metabolites. However, numerous authors have demonstrated that a single analytical method is not often sufficient to grasp the chemical complexity of plant metabolomes. Therefore, the combination of information obtained from the same samples with multiple experimental platforms (e.g., NMR, GC-, CE- or LC-MS) is expected to extend the coverage of the metabolites that characterise a biological system [189]. The original chemometric methodologies are mandatory for coping with the challenges associated with the fusion, the integration and the comparison of data generated using different, hopefully complementary, analytical techniques or data acquisition modes.

The first solution consists of processing and independently analysing the different datasets, manually comparing the results using high-level data integration and attempting to establish a relationship between the discriminant features [190]. However, such an approach is unable to uncover the possible interrelations between features or signals that are measured independently. Furthermore, a detrimental redundancy can occur as some classes of analytes may be well detected on the different platforms. Therefore, their related signals will have a greater impact on the subsequent models.

Computing correlation profiles between the two sets of variables constitutes another simple method for evaluating the relationship between multiple datasets. On the basis that the same samples are characterised by different data sources, the sample mode constitutes a common dimension across the tables, i.e., the matrices possess the same number of rows. This simple approach aims to correlate relevant metabolites highlighted in one of the data tables with related signals in the other matrices based on correlations, even if these signals would not have been detected when analysing the other tables separately. This approach can be particularly well suited to searching for biomarker candidates [134]. However, correlation does not necessarily imply causation. The statistical heterospectroscopy (SHY) method was proposed to combine NMR and LC-MS datasets based on the visual analysis of Pearson correlation profiles associated with highly statistically significant indices [191]. A similar approach was recently applied to associate data obtained with ^{1}H-NMR and GC-MS in the context of plant metabolomics [192].

Because an independent analysis of an individual dataset may be limited, other solutions are offered to integrate information from multiple sources, such as the horizontal concatenation of data matrices. Such data fusion is performed by considering variables of different origins to build a summary table such that multivariate classical methods, either unsupervised, e.g., PCA, or supervised, e.g., PLS-DA, can be applied to the merged data matrix [193]. Often, this approach aggravates the curse of dimensionality because it generates data structures with a prohibitive number of variables. In this context, applying variable selection or dimensionality reduction before concatenation can be useful to restrain the size of each data table [194].

Finally, more advanced chemometric methods, including multiblock and hierarchical component models, are attractive solutions for the simultaneous analysis of multiple data tables [195]. These data analyses tend to provide a simultaneous overall view of the major trends common to all tables (consensus status combining the information contained in all tables) but also highlight differences between tables and or individuals. These approaches were recently used to model multiple data sets in the context of microbial fermentation [195] and rat toxicity [196] but are still sparsely applied in the context of plant metabolomics [197].

5. BIOMARKER IDENTIFICATION

As discussed, with the current technology, metabolite fingerprints can be obtained in a very efficient and high throughput manner. In the entire metabolomics process, however, one of the main bottlenecks is the identification of biomarkers. This is even more challenging with plants because secondary metabolites are often species-specific and comprehensive freely accessible widespread databases do not exist. Biomarker identification plays a central role because the biomarker identity is mandatory to convert metabolomic results into biological knowledge.

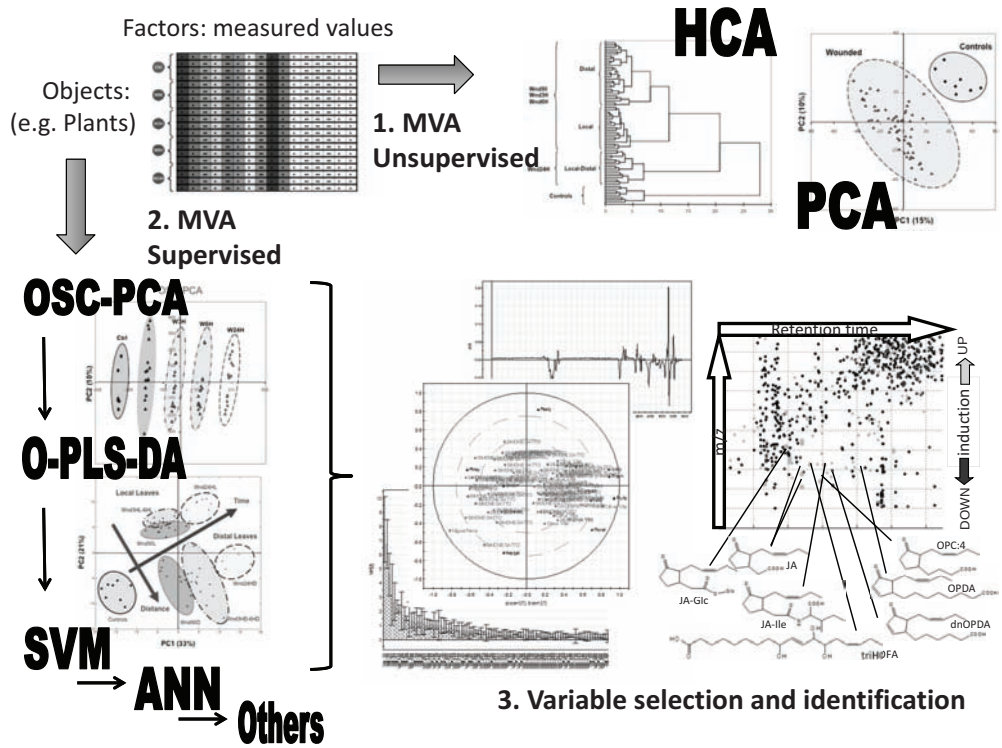

Fig. (8). From data table to information, examples of data mining outputs (OSC : Orthogonal Signal Correction, PCA : Principal Component Analysis, HCA : Hierarchical Cluster Analysis, O-PLS-DA : Orthogonal Partial Least Squares Discriminant Analysis, ANN artificial neural networks).

A number of guidelines are available regarding criteria to define the chemical identity of a metabolite [198, 199]. For high specificity, two orthogonal properties should be employed. In hyphenated MS methods for example, retention time depicts a physical property characteristic (volatility, hydrophobicity) and accurate mass or fragmentation mass spectra depict a characteristic of the structure of the metabolite [22].

5.1. Identification by NMR

In general, primary metabolites such as sugars, amino acids and organic acids are abundant and major metabolites that are found in plants. The signals of most of amino acids, except aromatic amino acids (phenylalanine, tyrosine and tryptophan), appear in the δ 2.0-1.0 region of ^{1}H NMR spectra. Organic acids are found in the δ 3.5-2.0 region, while the signals of sugars appear in the δ 5.5-3.0 region. The signals of phenolic compounds are shifted more downfield to δ 6.0 (δ 8.0-6.0). Signals can be identified by comparing them with reference data from the literature or in-house libraries of reference compounds. Most of the primary metabolites and simple secondary metabolites (e.g., phenolics) can be identified in this way. The identification of common metabolites that are found in plants has been well documented in several noteworthy reviews [40, 200, 201]. The different main regions that can be characterised directly from a ^{1}H-NMR spectrum are illustrated in (Fig. **9**) by the analysis of a direct methanol-d_4 /D_2O grapevine leaf extract [39]. NMR data of series of common phenolics (phenylpropanoids, flavonoids) were listed in (Tables **2-5**).

The more difficult and time-consuming part of the analysis is the identification of secondary metabolites that are present at very low concentrations or have complex structures. Diverse 2D NMR spectroscopy methods are necessary to identify and confirm their structures.

As discussed for the 2D-NMR experiments, 2D-^{1}H-^{1}H-*J*-resolved spectra are very useful tools to identify signals in the congested area of NMR fingerprints. They provide additional information including spin-spin coupling constants, together with chemical shifts. For example, when two signals are close to each other, it is unclear from the ^{1}H-NMR spectra whether they are doublets from one molecule or two singlets from the same or different molecules. The additional information of the spin-spin coupling in the *J*-resolved spectrum can solve this issue [69]. Using their typical coupling constants (*J* value) together with chemical shifts, some metabolites can be easily identified from complex ^{1}H-NMR spectra. For instance, H-8 of phenylpropanoids are well separated by characteristic coupling constants (15-16 Hz) in the 2D-*J*-resolved spectra from H-6 and H-8 of 5,7-dihydroxy flavonoids (2-3 Hz) in the δ 6.0 – δ 7.0 region [202].

Conversely, 2D NMR correlations between protons (^{1}H-^{1}H) or proton-carbons (^{1}H-^{13}C) can be useful to confirm whether the observed signals are derived from a single biomarker in an extract. A typical proton homonuclear experiment is correlation spectroscopy (COSY) that shows which signals in a ^{1}H-NMR have mutual spin-spin couplings. The resulting spectrum has the conventional 1D NMR spectrum along the diagonal cross-peaks at chemical shifts that correspond to pairs of coupled nuclei. In general, protons within three bonds are well-correlated in the COSY spectrum, but in sp_2 spin systems such as those found in olefinic and phenolic compounds, even the correlation beyond three bonds is readily detected. From the complex phenolic region (δ 6.0 – δ 8.5), flavonoids such as quercetin analogues are well identified in grape (*Vitis vinifera*) leaves using the proton-proton correlation pattern obtained from COSY spectra [203].

Another powerful 2D correlation technique is total correlation spectroscopy (TOCSY), which provides information such as the unbroken chains of coupled protons in the same molecule. For instance, the signals of sugars are mostly present in the crowded area (δ 3.0-4.0), except for the anomeric proton, which usually appears around δ 5.5-4.5. Once the anomeric proton of sugars is defined, the rest of the signals can be readily assigned using correlation in TOCSY spectrum. The correlations are detected largely based on the spin-lock duration and the adjustable parameters of the TOCSY

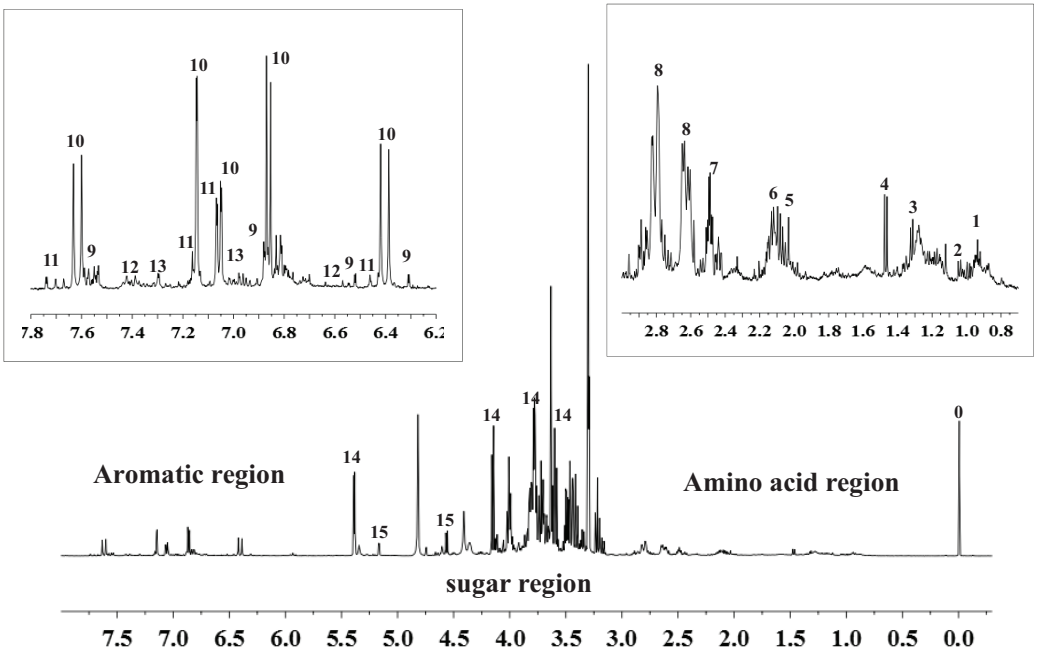

Fig. (9). ^{1}H-NMR spectrum of grapevine ('Regent' cultivar analyzed after 48 hours of pathogen inoculation). ^{1}H NMR can be divided into three distinct regions: amino acid (δ 2.5-0.0), sugar (δ 5.0-3.0) and aromatic (δ 8.5-5.5) region. Numbers indicate assignment of major signals to the metabolites. 0: TMSP (internal standard), 1: leucine, 2: valine, 3: threonine, 4: alanine, 5: glutamate, 6: proline and methionine, 7: glutamine, 8: malate, 9: quercetin glucoside, 10: caffeoyl moiety, 11: feruloyl moiety, 12: myricetin, 13: Tyrosine. 14: sucrose, 15: glucose (Modified from [39]).

experiment. Thus, the results are not dependent on the magnitude of the spin-spin couplings but rather on the duration of the spin-lock field [204, 205]. The longer the spin-lock period is applied, the further the magnetisation will be transferred down a chain of coupled nuclei. A short duration (less than 30 msec) gives a result that is similar to a COSY experiment, but longer times (typically in the range of 50 – 200 msec) result in the desired total correlation [204, 205]. For example, more correlations between signals in the sucrose, α-glucose, and β-glucose spectra are detected with longer duration.

An example of the ^{1}H-^{1}H *J*-resolved and the ^{1}H-^{1}H COSY spectra of grapevine leaf extract are illustrated in (Fig. **10**). As shown, the characteristic signals of various phenylpropanoids were assigned in the *J*-resolved experiment and their correlations are highlighted in the corresponding COSY 2D plot [39].

Pairs of 2D NMR spectra are often manually compared. However, a metabolomics approach involving the comparison of multiple 2DNMR statistical analysis methods for 2D NMR spectra has been recently developed. The hierarchical alignment of 2D spectra enabled pattern recognition of the exudates from two nematode species that were studied by 2D TOCSY NMR and chemical differences between the two species were efficiently assessed [206].

A direct correlation between carbons and protons can be detected by heteronuclear multiple quantum coherence (HMQC) or heteronuclear single quantum coherence (HSQC) [63]. HSQC gives better sensitivity than HMQC but requires phase correction. In HMQC and HSQC spectra any carbon that is directly attached to a proton is detected. Consequently, the chemical shift of a ^{13}C connected to a specific proton can be obtained. Recently, HSQC has been shown to be particularly useful in the metabolomics field. Indeed, the presence of specific metabolites within an extract was clearly distinguished by comparing two HSQC spectra – spectra of mixtures of known reference compounds and spectra of the extracts [207]. In addition, when comparing the HSQC spectra of interest, differences in the spectra of two similar susceptible/resistant plants species, *Ilex paraquariensis* and *Ilex dumosa* [208], were readily discernible. The cross-peaks corresponding to the main differences were linked to biomarkers that are characteristic of these specific plants.

Long-range correlations (two and three bonds) between carbons and protons can be observed by heteronuclear multiple bond correlation (HMBC) spectra [63]. This is an extremely powerful tool for structure elucidation as carbon-carbon correlations can be indirectly obtained with the aid of HMQC or HSQC. In addition, correlations between quaternary carbons such as carboxylic acid and nearby protons can be observed in the spectra. The long-range couplings in HMBC spectra can provide information on the connectivity between two moieties that are connected only by quaternary carbons or hetero atoms (e.g., *O*-glycosides). In metabolomic studies for example, the correlation between the anomeric proton of the glucose moiety and the carbon of C=N in glucosinolates is quite characteristic and were easily detected in the HMBC spectrum of *Brassica rapa* [209].

Using diverse combinations of NMR techniques, many secondary metabolites have been identified from crude extracts. Characteristics of ^{1}H NMR signals of several common secondary metabolites, particularly phenylpropanoids and flavonoids, are listed in (Tables **2-5**).

Tables **2-5**. Summary of chemical shifts used for NMR identification.

The use of 2D NMR provides an efficient means of enhancing the resolution of features in metabolomics and provides additional key information for structure determination. However, it should be taken into account that compared to ^{1}H NMR, the throughput is reduced because^{1}H-^{13}C correlation experiments need to be conducted for longer periods of time (several hours). The detection of ^{1}H-^{13}C is less sensitive than ^{1}H-NMR but more sensitive than direct ^{13}C-NMR because the 2D NMR experiment is conducted by indirect detection of the nucleus with the highest gamma value, which in this case is ^{1}H.

The role of 1D and 2D-NMR in comprehensive NMR-based metabolomic approaches for fingerprinting and metabolite identification are summarised schematically in (Fig. **11**).

Table 2. [1]H Chemical Shifts (ppm) and Coupling Constants (Hz) of Common Plant Phenylpropanoids (Phenolic Moiety).

Phenylpropanoids		H					solvent
		2	3	4	5	6	
cinnamic acid (**1**)		7.59 (m)	7.44 (m)	7.44 (m)	7.44 (m)	7.59 (m)	1
		7.57 (m)	7.39 (m)	7.39 (m)	7.39 (m)	7.57 (m)	2
caffeic acid (**2**)		7.12 (d, 2.0)	-	-	6.88 (d, 8.3)	7.03 (dd, 8.3, 2.0)	1
o-coumaric acid (**3**)		-	6.92 (dd, 8.0, 1.5)	7.28 (td, 8.0, 1.5)	6.93 (td, 8.0, 1.5)	7.55 (dd, 8.0, 1.5)	1
m-coumaric acid (**4**)		7.07 (brt, 2.0)	-	6.93 (brdd, 8, 1.6)	7.30 (t, 8.0)	7.13 (brd, 7.5)	1
p-coumaric acid (**5**)		7.50 (d, 8.6) [3]	6.89 (d, 8.6) [3]	-	6.89 (d, 8.6) [3]	7.50 (d, 8.6) [3]	1
ferulic acid (**6**)		7.19 (d, 1.9)	-	-	6.89 (d, 8.2)	7.10 (dd, 8.2, 1.9)	1
ferulic acid-4-*O*-glucoside (**7**)		7.28 (d, 1.5)	-	-	7.17 (d, 8.5)	7.20 (dd, 8.5, 1.5)	1
sinapic acid (**8**)		6.93 (s)	-	-	-	6.93 (s)	1
sinapic acid -4-*O*-glucoside (**9**)		6.98 (s)	-	-	-	6.98 (s)	1
coniferyl alcohol (**10**)		7.08 (brs)	-	-	6.83 (d, 8.0)	6.93 (dd, 8.5, 1.5)	1
sinapyl alcohol (**11**)		6.77 (s)	-	-	-	6.77 (s)	1
coniferin (**12**)		7.16 (d, 2.0)	-	-	7.13 (d, 8.0)	7.04 (dd, 8.0, 2.0)	1
syringin (**13**)		6.83 (s)	-	-	-	6.83 (s)	1
eugenol (**14**)		6.82 (d, 2.0)	-	-	6.80 (d, 8.0)	6.68 (dd, 8.0, 2.0)	1
		6.73 (d, 2.0)			6.69 (d, 8.0)	6.60 (dd, 8.0, 2.0)	2
isoeugenol (**15**)		6.99 (d, 1.5)	-	-	6.80 (d, 8.0)	6.84 (dd, 8.5, 2.0)	1
		6.91 (d, 2.0)	-	-	6.67 (d, 8.0)	6.74 (dd, 8.0, 2.0)	2
p-dihydrocoumaric acid (**16**)		7.09 (d, 8.6) [3]	6.78 (d, 8.6) [3]	-	6.78 (d, 8.6) [3]	7.09 (d, 8.6) [3]	1
chlorogenic acid (**17**)		7.12 (d, 2.1)	-	-	6.87 (d, 8.4)	7.01 (dd, 8.4, 2.0)	1
3-*O*-caffeoyl quinic acid (**18**)		7.15 (d, 2.0)	-	-	6.88(d, 8.4)	7.06 (dd, 8.4, 2.0)	1
4-*O*-caffeoyl quinic acid (**19**)		7.16 (d, 2.0)	-	-	6.89 (d, 8.4)	7.06 (dd, 8.4, 2.0)	1
cynarin (**20**)	1-*O*-caffeoyl	6.92 (d, 2.0)	-	-	6.67 (d, 8.0)	6.75 (dd, 8.0, 2.0)	1
	3-*O*-caffeoyl	6.82 (d, 2.0)	-	-	6.52 (d, 8.0)	6.61 (dd, 8.0, 2.0)	
3,5-dicaffeoyl quinic acid (**21**)	3-*O*-caffeoyl	7.19 (d, 2.0)	-	-	6.90 (d, 8.0)	7.10 (dd, 8.0, 2.0)	1
	5-*O*-caffeoyl	7.16 (d, 2.0)	-	-	6.89 (d, 8.0)	7.08 (dd, 8.0, 2.0)	
cichoric acid (**22**)[4]		7.08 (d, 1.5)	-	-	6.78 (d, 8.0)	6.98 (dd, 8.0, 1.5)	2
rosmarinic acid (**23**)	caffeoyl	7.11 (d, 2.0)	-	-	6.86 (d, 8.5)	7.02 (dd, 8.5, 2.0)	1
	8-hydroxy dihydrocaffeoyl	6.82 (d, 2.0)	-	-	6.77 (d, 8.0)	6.70 (dd, 8.0, 2.0)	

[1] CH_3OH-d_4-KH_2PO_4 buffer in D_2O (pH 6.0), calibrated to TSP at δ 0.00, [2] CH_3OH-d_4 calibrated to residual solvent at δ 3.30, [3] multiplet but seemingly doublet, [4] The same chemical shifts and coupling constants of two caffeoyl moieties.

5.2. MS-Based Identification

As discussed, one of the main pieces of information that can be obtained from MS is the molecular weight (MW) of the biomarkers. In approaches such as DIMS or LC-MS, soft ionisation methods are used that generate various molecular ion species. One of the first tasks in metabolite identification is to identify the MW of a given biomarker. This task is not straightforward for an unknown compound because different adducts, dimers or oligomers may be present in the MS spectra. An example of such spectra is illustrated in (Fig. **6**). Each ionisation mode (PI or NI) generates a defined type of adducts, and based on the relationships between molecular ion species and comparison of the PI and NI spectra obtained for a given peak, the MW can be unambiguously deduced [210]. With LR-MS approaches, the nominal mass that is deduced is a first element that can be used for database searches. However, this information is extremely limited and a large number of structures will match a given MW. When HR-MS instruments are used, the MW can be determined with high accuracy. Currently, most HR-MS instruments routinely provide mass accuracy < 5 ppm and this can reach 1 ppm or less on high-end FT-ICR-MS [85]. The accurate mass data together with high spectral accuracy (isotopic pattern determination) provide key information for molecular formula assignment. However, with < 5 ppm mass accuracy, the unambiguous determination of the molecular formula of a biomarker is difficult, especially for high MW (> 500 Da) compounds or for those that contain atoms other than CHO. To reduce the possibilities of molecular formula determination, different successive filters, known as heuristic filters [211], can be applied. This filtering process includes the following steps: (1) restrictions for the number of elements; (2) LEWIS and SENIOR chemical rules; (3) isotopic patterns; (4) hydrogen/carbon ratios; (5) elemental ratio of nitrogen, oxygen, phosphor, and sulphur versus carbon; and (6) element ratio probabilities. Molecular formulas identified in this way are usually determined

Table 3. ^{1}H Chemical Shifts (ppm) and Coupling Constants (Hz) of Common Plant Phenylpropanoids (Aliphatic Moiety).

Phenylpropanoids		H				solvent
		7	8	9	others	
cinnamic acid (**1**)		7.65 (d, 16.0)	6.50 (d, 16.0)	-	-	1
		7.66 (d, 16.0)	6.47 (d, 16.0)	-	-	2
caffeic acid (**2**)		7.52 (d, 15.9)	6.29 (d, 15.9)	-	-	1
o-coumaric acid (**3**)		7.88 (d, 16.0)	6.55 (d, 16.0)	-	-	1
m-coumaric acid (**4**)		7.59 (d, 16.0)	6.46 (d, 16.0)	-	-	1
p-coumaric acid (**5**)		7.59 (d, 15.9)	6.33 (d, 15.9)	-	-	1
ferulic acid (**6**)		7.56 (d, 15.9)	6.34 (d, 15.9)	-	3.89 (s, OCH_3)	1
ferulic acid-4-*O*-glucoside (**7**)		7.48 (d, 16.0)	6.42 (d, 15.5)	-	5.06 (d, 7.5)[3] 3.91 (s, OCH_3)	1
sinapic acid (**8**)		7.48 (d, 16.0)	6.37 (d, 16.0)	-	3.88 (s, OCH_3)	1
sinapic acid -4-*O*-glucoside (**9**)		7.32 (d, 16.0)	6.47 (d, 16.0)	-	4.99 (d, 7.5)[3] 3.90 (s, OCH_3)	1
coniferyl alcohol (**10**)		6.55 (brd, 16.0)	6.25 (dt, 15.5, 6.0)	4.23(dd, 6.0, 1.0)	3.88 (s, OCH_3)	1
sinapyl alcohol (**11**)		6.54 (brd, 16.0)	6.28 (dt, 16.0, 6.0)	4.24 (dd, 5.5, 1.0)	3.87 (s, OCH_3)	1
coniferin (**12**)		6.59 (brd, 16.0)	6.34 (dt, 16.0, 5.5)	4.25 (dd, 5.5, 1.0)	5.02 (d, 8.0)[3] 3.90 (s, OCH_3)	1
syringin (**13**)		6.58 (brd, 16.0)	6.38 (dt, 16.0, 5.5)	4.26 (dd, 6.0, 1.5)	4.94 (d, 7.5)[3] 3.88 (s, OCH_3)	1
eugenol (**14**)		3.30 (d, 7.0)	5.97 (ddt, 17.0, 10.0, 6.5)	5.06 (dq, 17.0, 2.0) 5.05 (dm, 10.0)	3.84 (s, OCH_3)	1
		3.27 (d, 6.5)	5.93 (ddt, 17.0, 10.0, 6.5)	5.03 (dq, 17.0, 2.0) 4.99 (dm, 10.0)	3.81 (s, OCH_3)	2
isoeugenol (**15**)		6.36 (dd, 15.5, 1.5)	6.16 (dq, 15.5, 6.5)	1.84 (dd, 6.5, 1.5)	3.86 (s, OCH_3)	1
		6.29 (dd, 15.5, 1.5)	6.06 (dq, 15.5, 6.5)	1.82 (dd, 6.5, 1.5)	3.84 (s, OCH_3)	2
p-dihydrocoumaric acid (**16**)		2.82 (t, 7.5)	2.59 (t, 7.5)	-	-	1
chlorogenic acid (**17**)		7.58 (d, 15.9)	6.33 (d, 15.9)	-	5.34 (ddd, 10.8, 9.8, 5.0)[4]	1
3-*O*-caffeoyl quinic acid (**18**)		7.61 (d, 15.9)	6.41 (d, 15.9)	-	5.37 (dt, 5.6, 3.1)[5]	1
4-*O*-caffeoyl quinic acid (**19**)		7.66 (d, 15.9)	6.44 (d, 15.9)	-	4.89 (dd, 8.3, 3.3)[6]	1
cynarin (**20**)	1-*O*-caffeoyl	7.39 (d, 16.0)	6.26 (d, 16.0)	-	-	1
	3-*O*-caffeoyl	7.43 (d, 16.0)	6.11 (d, 16.0)	-	5.32 (dt, 3.6, 3.2)	
3,5-dicaffeoyl quinic acid (**21**)	3-*O*-caffeoyl	7.65 (d, 16.0)	6.50 (d, 16.0)	-	5.43 (dt, 3.5, 3.0)[5]	1
	5-*O*-caffeoyl	7.64 (d, 16.0)	6.39 (d, 15.6)	-	5.49 (ddd, 11.6, 10.3, 5.0)[4]	
cichoric acid (**22**)[7]		7.65 (d, 16.0)	6.37 (d, 16.0)	-	5.79 (s)[8]	2
rosmarinic acid (**23**)	caffeoyl	7.52 (d, 16.0)	6.30 (d, 16.0)	-	-	1
	8-hydroxydihydrocaffeoyl	3.11 (dd, 14.0, 3.5) 2.95 (dd, 14.5, 10.0)	5.06 (dd, 9.5, 3.5)	-	-	

[1] CH_3OH-d_4-KH_2PO_4 buffer in D_2O (pH 6.0), calibrated to TSP at δ 0.00, [2] CH_3OH-d_4 calibrated to residual solvent at δ 3.30, [3] H-1 of β-glucose, [4] H-5 of quinic acid moiety, seemingly triplet-doublet in low resolution NMR, [5] H-3 of quinic acid moiety, [6] H-4 of quinic acid moiety, [7] The same chemical shifts and coupling constants of two caffeoyl moieties, [8] H-1 and H-2 of tartaric acid moiety.

unambiguously. Such rules are now implemented in freely available software for analysing mass spectrometry-based molecular profile data such as MZmine [212]. A search based on the molecular formula together with plant chemotaxonomic information in a natural product database, such as the dictionary of natural products [213] that contains more than 200,000 NPs, might already be very effective for performing efficient biomarker peak annotation [210]. This process is especially efficient when the biomarkers have already been reported in the same plant species or in species of the same genus.

To obtain additional structural information, molecular ion species generated by soft ionisation methods can be fragmented in tandem mass spectrometers using MS/MS or MS^n experiments. For this approach, triple-quadrupole (QqQ) MS/MS systems are the most commonly used mass spectrometers. Ion-trap (IT) mass spectrometers have the unique capability of producing multiple-stage MS/MS (MS^n) data that may be essential for structural elucidation. For high-resolution measurements in MS/MS, either a hybrid instrument such as a Q-TOF-MS [37] or an IT-MS such as an Orbitrap Fourier transform (FT) MS instrument provides high quality spectra for metabolite identification [89]. If in-house built MS/MS libraries exist, which is often the case in industry, efficient dereplication based on MS/MS spectra comparison can be performed. In the absence of standard or instrument-dedicated databases, which is often the case in non-industry research laboratories, the structural identification may rely on the interpretation of the fragmentation

Table 4. [1]H Chemical Shifts (ppm) and Coupling Constants (Hz) of Common Plant Flavonoids (A and C Rings).

Flavonoids	H						solvent
	2	3	4	5	6	8	
apigenin (**24**)	-	6.68 (s)	-	-	6.29 (d, 2.0)	6.56 (d, 2.0)	[1]
apigenin-7-*O*-glucoside (**25**)	-	6.74 (s)	-	-	6.57 (d, 2.0)	6.88 (d, 2.0)	[1]
luteolin (**26**)	-	6.65 (s)	-	-	6.29 (d, 2.0)	6.56 (d, 2.0)	[1]
luteolin-7-*O*-glucoside (**27**)	-	6.69 (s)	-	-	6.56 (d, 2.0)	6.86 (d, 2.0)	[1]
kaempferol (**28**)	-	-	-	-	6.28 (d, 2.0)	6.49 (d, 2.0)	[1]
rhamnetin (**29**)							[1]
quercetin (**30**)	-	-	-	-	6.27 (d, 1.5)	6.49 (d, 2.0)	[1]
quercitrin (**31**)	-	-	-	-	6.25 (d, 1.5)	6.41 (brs)	[1]
rutin (**32**)	-	-	-	-	6.28 (d, 2.0)	6.49 (d, 2.0)	[1]
myricetin (**33**)	-	-	-	-	6.26 (d, 2.0)	6.48 (d, 2.0)	[1]
vitexin (**34**)	-	6.69 (s)	-	-	6.36 (brs)[3]	-	[1]
isovitexin (**35**)	-	6.67 (s)	-	-	-	6.61 (s)	[1]
daidzein (**36**)	8.21 (s)	-	-	8.08 (d, 9.0)	7.04 (dd, 9.0, 2.0)	6.97 (d, 2.0)	[1]
daidzin (**37**)	8.28 (s)	-	-	8.17 (d, 9.0)	7.28 (dd, 9.0, 2.0)	7.32 (d, 2.0)	[1]
genistein (**38**)	8.12 (s)	-	-	-	6.31 (d, 2.0)	6.47 (d, 2.0)	[1]
genistin (**39**)	8.22 (s)	-	-	-	6.60 (d, 2.0)	6.82 (d, 2.0)	[1]
naringenin (**40**)	5.39 (dd, 12.5, 3.0)	3.16 (dd, 17.5, 15.0) 2.79 (dd, 17.5, 3.5)	-	-	5.94 (d, 2.5)[4]	5.94 (d, 2.5)[4]	[1]
naringin (**41**)	5.45 (dd, 10.0, 3.0)[3] 5.43 (dd, 10.0, 3.0)[3]	3.23 (dd, 17.5, 15.0)[3] 3.21 (dd, 17.5, 15.0)[3] 2.86 (dd, 17.5, 3.0)[3] 2.85 (dd, 17.5, 3.0)[3]	-	-	6.18 (d, 2.0)[4]	6.20 (d, 2.0)[4]	[1]
hesperetin (**42**)	5.40 (dd, 12.0, 3.5)	3.16 (dd, 17.5, 12.0) 2.83 (dd, 17.5, 3.5)	-	-	5.95 (d, 2.1)[4]	5.96 (d, 2.1)[4]	[2]
hesperidin (**43**)							[1]
neohesperidin (**44**)							[1]
(+)-catechin (**45**)	4.70 (d, 7.5)	4.13 (td, 7.5, 5.5)	2.84 (dd, 16.5, 5.5) 2.54 (dd, 16.5, 7.5)	-	5.96 (d, 2.0)	6.05 (d, 2.0)	[1]
gallocatechin (**46**)	4.65 (d, 7.0)	4.12 (td, 7.0, 5.5)	2.81 (dd, 16.0, 5.0) 2.53 (dd, 16.0, 7.5)	-	5.96 (d, 2.0)	6.04 (d, 2.0)	[1]
gallocatechin-3- gallate (**47**)	5.07 (d, 6.5)	5.38 (dd, 6.0, 5.5)	2.85 (dd, 16.5, 5.0) 2.74 (dd, 16.5, 6.0)	-	6.05 (d, 2.0)	6.07 (d, 2.0)	[1]
(-)-epicatechin (**48**)	[5]	4.26 (m)	2.90 (dd, 17.0, 4.5) 2.73 (dd, 17.0, 2.5)	-	6.03 (d, 2.0)	6.06 (d, 2.0)	[1]
epicatechin-3- gallate (**49**)	5.13(brs)	5.52 (brs)	3.05 (dd, 17.5, 4.5) 2.73 (brd, 17.5)	-	6.07 (d, 2.0)[3]	6.09 (d, 2.0)[3]	[1]
epigallocatechin (**50**)	[5]	4.26 (brs)	2.89 (dd, 17.0, 4.5) 2.73 (dd, 17.0, 2.5)	-	6.02 (d, 2.0)[3]	6.05 (d, 2.0)[3]	[1]
epigallocatechin-3- gallate (**51**)	5.06 (brs)	5.50 (brs)	3.04 (dd, 17.5, 4.5) 2.88 (brd, 17.5)	-	6.08 (d, 2.5)[3]	6.06 (d, 2.5)[3]	[1]

[1] CH_3OH-d_4-KH_2PO_4 buffer in D_2O (pH 6.0), calibrated to TSP at δ 0.00, [2] CH_3OH-d_4 calibrated to residual solvent at δ 3.30, [3] Signals from different rotamers, [4] Two doublets of H-6 and H-8 detected as AB system, [5] Hidden under HDO solvent.

Table 5. ^{1}H Chemical Shifts (ppm) and Coupling Constants (Hz) of Common Plant Flavonoids (B Ring).

flavonoids	2'	3'	5'	6'	others	solvent
apigenin (24)	7.93 (d, 9.0)[3]	7.02 (d, 9.0)[3]	7.02 (d, 9.0)[3]	7.93 (d, 9.0)[3]	-	[1]
apigenin-7-*O*-glucoside (25)	7.95 (d, 9.0)[3]	7.02 (d, 9.0)[3]	7.02 (d, 9.0)[3]	7.95 (d, 9.0)[3]	5.18 (d, 7.5)[4]	[1]
luteolin (26)	7.46 (d, 2.0)	-	7.01 (d, 8.5)	7.48 (dd, 8.0, 2.0)	-	[1]
luteolin-7-*O*-glucoside (27)	7.46 (d, 2.0)	-	6.99 (d, 8.0)	7.49 (dd, 8.0, 2.0)	5.18 (d, 7.5)[4]	[1]
kaempferol (28)	8.08 (d, 8.5)[3]	7.00 (d, 8.5)[3]	7.00 (d, 8.5)[3]	8.08 (d, 8.5)[3]	-	[1]
rhamnetin (29)						[1]
quercetin (30)	7.72 (d, 2.0)	-	6.99 (d, 8.5)	7.64 (dd, 8.5, 2.0)	-	[1]
quercitrin (31)	7.34 (d, 2.0)	-	7.00 (d, 8.5)	7.29 (dd, 8.5, 2.0)	5.27 (d, 1.0)[5] 0.91 (d, 6.0)[6]	[1]
rutin (32)	7.67 (d, 2.1)	-	6.98 (d, 8.5)	7.62 (dd, 8.5, 2.1)	5.01 (d, 7.7)[4] 4.54 (d, 1.3)[5]	[1]
myricetin (33)	7.34 (s)	-	-	7.34 (s)	-	[1]
vitexin (34)	8.00 (d, 8.0)[7] 7.92 (brs)[7]	7.03 (d, 8.5)	7.03 (d, 8.5)	8.00 (d, 8.0)[7] 7.92 (brs)[7]	5.02 (d, 10.0)[4]	[1]
isovitexin (35)	7.90 (d, 9.0)[3]	7.00 (d, 9.0)[3]	7.00 (d, 9.0)[3]	7.90 (d, 9.0)[3]	4.91 (d, 10.0)[4, 8]	[1]
daidzein (36)	7.39 (d, 9.0)[3]	6.95 (d, 9.0)[3]	6.95 (d, 9.0)[3]	7.39 (d, 9.0)[3]	-	[1]
daidzin (37)	7.41 (d, 8.5)[3]	6.96 (d, 8.5)[3]	6.96 (d, 8.5)[3]	7.41 (d, 8.5)[3]	5.23 (d, 7.5)[4]	[1]
genistein (38)	7.38 (d, 9.0)[3]	6.95 (d, 9.0)[3]	6.95 (d, 9.0)[3]	7.38 (d, 9.0)[3]	-	[1]
genistin (39)	7.40 (d, 8.5)[3]	6.96 (d, 8.5)[3]	6.96 (d, 8.5)[3]	7.40 (d, 8.5)[3]	5.17 (d, 7.0)[4]	[1]
naringenin (40)	7.36 (d, 8.5)[3]	6.91 (d, 8.5)[3]	6.91 (d, 8.5)[3]	7.36 (d, 8.5)[3]	-	[1]
naringin (41)	7.36 (d, 8.5)[3]	6.91 (d, 8.5)[3]	6.91 (d, 8.5)[3]	7.36 (d, 8.5)[3]	5.17 (d, 7.5)[4,7] 5.15 (d, 7.5)[4,7] 5.17 (d, 1.0)[5,7] 5.16 (d, 1.0)[5,7] 1.27 (d, 6.0)[6]	[1]
hesperetin (42)	7.00 (brs)	-	7.04 (d, 8.5)	6.99 (dd, 8.0, 1.5)	3.88 (s, OCH_3)	[2]
hesperidin (43)						[1]
neohesperidin (44)						[1]
(+)-catechin (45)	6.90 (d, 2.0)	-	6.88 (d, 8.5)	6.80 (dd, 8.5, 2.0)	-	[1]
gallocatechin (46)	6.48 (s)	-	-	6.48 (s)	-	[1]
gallocatechin-3- gallate (47)	6.38 (s)	-	-	6.38 (s)	6.96 (s)[9]	[2]
(-)-epicatechin (48)	7.02 (brs)	-	6.88 (brs)[10]	6.88 (brs)[10]	-	[1]
epicatechin-3- gallate (49)	7.01 (d, 2.0)	-	6.81 (d, 8.0)	6.90 (dd, 8.0, 2.0)	6.98 (s)[9]	[1]
epigallocatechin (50)	6.59 (s)	-	-	6.59 (s)	-	[1]
epigallocatechin-3- gallate (51)	6.60 (s)	-	-	6.60 (s)	6.98 (s)[9]	[1]

[1]CH_3OH-d_4-KH_2PO_4 buffer in D_2O (pH 6.0), calibrated to TSP at δ 0.00, [2]CH_3OH-d_4 calibrated to residual solvent at δ 3.30, [3]multiplet but seemingly doublet, [4]H-1 of β-glucose, [5]H-1 of α-rhamnose, [6]CH_3 of rhamnose, [7]Signals from different rotamers, [8]Hidden under HDO solvent, [9]H-2 and H-6 of gallic acid moiety, [10]Two doublets of H-5' and H-6' detected as AB system.

based on the existence of structure hypothesis or on the application of automated workflows [214]. MS peak annotation, however, sometimes remains putative. A scheme summarising the main steps that are used to annotate MS peaks is provided in (Fig. **12**).

Contrary to LC-MS, in GC-MS the electron ionisation MS spectra (EI-MS) that can be recorded are very reproducible between instruments (see above). This has the important advantage that large EI-MS databases can be directly searched for peak annotation. Despite the high resolution of GC separation, the co-elution of metabolites may occur and the deconvolution of these spectra is an important step. The latest advance in this area involves GC coupled to a TOF-MS instrument to take advantage of the high data acquisition rate and spectral continuity of TOF mass spectrometers to build libraries that take into account spectra/retention time-indexed data sets [130].

Despite these advances, very little progress has been reported for the identification of unknown peaks in GC-MS [215] because the method does not easily allow the isolation of the biomarker of interest for *de novo* characterisation.

5.3. *De Novo* Identification Based on Targeted Micro-Isolation

When unknown biomarkers need to be identified *de novo*, their complete characterisation requires NMR. In this respect, NMR signals associated with certain biomarkers might be highlighted in NMR-based metabolomic approaches using covariance methods such as Statistical Heterospectroscopy (SHY) if the same samples are analysed by NMR and MS [216]. SHY allows signals in one spectroscopic domain (NMR) to be dispersed in a second analytical spectroscopic domain (MS) to facilitate data co-analysis. Therefore, it is possible to generate meaningful NMR-to-*m*/*z* correlations to facilitate structure assignment. In addition, it could be helpful to derive connectivities between NMR signals and fragmentation patterns from the same molecules.

However, this approach is still rather complex and requires that the biomarker is present in a sufficient amount to be detected by

NMR in the crude mixture. Other alternatives include the synthesis of the putative compound and comparison with the peak detected in the sample or the targeted micro-isolation of the biomarkers of interest based on MS data. In the latter case, one advantage of LC-MS based metabolomics approaches is that because biomarkers of interest can be efficiently localised in the LC chromatogram, and provided that an adequate quantity of the sample is available, they can be isolated using a targeted procedure and subsequently identified by sensitive micro NMR methods [110].

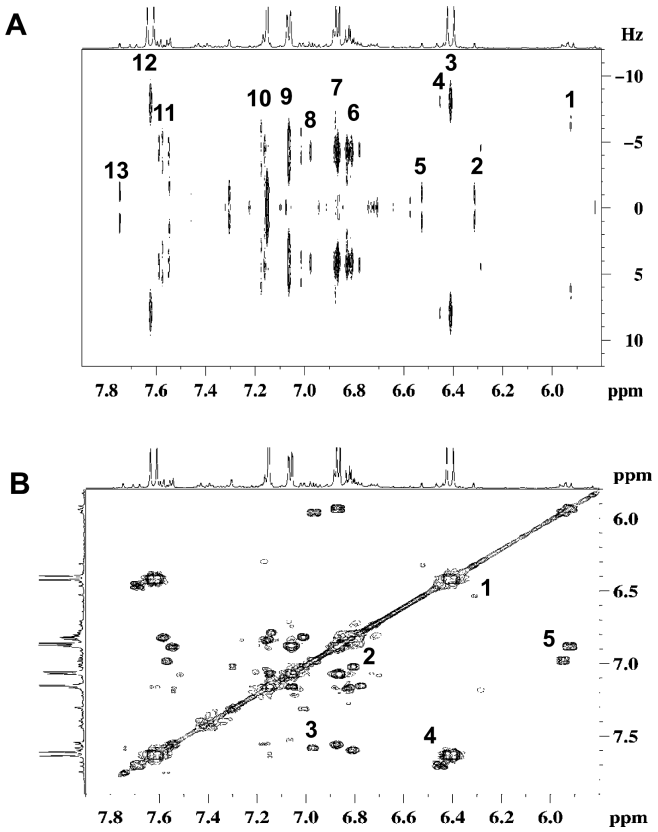

Fig. (10). ^{1}H-^{1}H *J*-resolved spectra (A) shows signals of phenylpropanoids and flavonoids. 1, H-8': *cis*-feruloyl moiety; 2, H-8: quercetin-3-*O*-glucoside; 3, H-8': ; 4, H-8': *trans*-feruloyl moiety; 5, H-6: quercetin-3-*O*-glucoside; 6, H-5': caffeoyl moiety; 7, H-3: *trans*-feruloyl moiety; 8, H-5': quercetin-3-*O*-glucoside; 9, H-6': caffeoyl moiety; 10, H-6': *cis*-feruloyl moiety; 11, H-6': quercetin-3-*O*-glucoside; 12, H-7: caffeoyl moiety; 13, H-2': quercetin-3-*O*-glucoside. ^{1}H-^{1}H COSY spectra (B) shows correlations among the signals of H-6 with H-8 (1) and H-5' with H-6' (3) of quercetin-3-*O*-glucoside; H-5 with H-6 (2) of *trans*- and *cis*-caffeoyl and feruloyl moiety; H-8' with H-7' (4, 5) of *trans*- and *cis*-caffeoyl and feruloyl moiety, respectively. (Adapted from Ali *et al*., 2012)

Due to the complexity of plant extracts, the purification of low concentration metabolites is critical. A rapid and rational strategy can be used that relies on optimisation of the chromatographic analysis using UHPLC-TOF-MS with modelling software. The optimised method can be transferred to semi-preparative LC conditions with MS detection. Complete characterisation of the isolated metabolites, often obtained in the low-microgram range from a few milligrams of plant extract, is then performed using micro-flow-NMR methods, such as CapNMR, that provide excellent sensitivity [217]. With such NMR probes, the biomarkers need only to be dissolved in a minimal amount of deuterated solvent (5 µL). With such a small volume, high concentrations can be obtained with an optimum filling factor, and high-quality 1D- and 2D-NMR spectra can be measured [218]. Such an approach has permitted the full characterisation of new phytohormones induced by wounding in the leaf of *Arabidopsis thaliana*. In this case, even minor key biomarkers were fully characterised by CapNMR after a two-step targeted LC-MS micro-isolation procedure [110]. This approach has also been used for the identification of various biomarkers that are induced upon herbivory attack in maize [80].

5.4. Databases for Biomarker Identification

The access to databases can considerably increase the speed of the biomarker identification process. Although great efforts have been made to create databases for researchers, these repositories remain incomplete in plant science, particularly in terms of secondary metabolites. Moreover, a large fraction of the metabolites included in complex plant extracts is still unknown. Two main categories of databases can be defined: spectroscopic databases dedicated to metabolites (MS-or NMR-based) and those that are pathway-oriented [37].

5.4.1. Spectroscopic Databases

Several types of NMR databases are available. The Madison Metabolomics Consortium Database (MMCD, http://mmcd.nmrfam.wisc.edu/) [219, 220] is a reference spectral database that contains chemical structures, NMR and LC-MS spectra of reference compounds, their physical properties and links to chemical and metabolomic databases. Databases for more specific metabolite profiles include the Biological Magnetic Resonance Data Bank (BMRD,www.bmrb.wisc.edu), the Human Metabolome Database (HMDB,www.hmdb.ca) [221], the NMR database of Linkoeping (MDL, http://www.liu.se/hu/mdl/main/), the Magnetic Resonance Metabolomics Database (http://www.metabolomics.bioc.cam.ac.uk/metabolomics), Prime [222] and the NMR Lab of Biomolecules (http://spinportal.magnet.fsu.edu). Other databases that might be useful for metabolite identification using NMR spectra include the NMRShiftDB (www.ebi.ac.uk/NMRshiftdb/), the Spectral Database for Organic Compounds (SDBS, www.riodb01.ibase.aist.go.jp) and the BioMagResBank (www.bmrb.wisc.edu). Recently, several studies dedicated to the establishment of databases have been published, including MeRy-B [223] and MetaboMiner [224, 225]. MetaboMiner automates the identification of metabolites from ^{1}H NMR data of biological origin.

Different types of MS databases exist. Those that are easy searchable are the GC-MS databases that are based on EI-MS spectra of standards. The database search is based on a computed score (similarity or distance function) that compares the queried spectrum and reference database entries to extracted putative hits. In this respect, large GC-MS databases exist for plant metabolomics such as the Golm Metabolomic Database (GMD) [226] and the FiehnLib [130]. Such databases contain primary metabolites below *ca* 500 Da and include lipids, amino acids, fatty acids, amines, alcohols, sugars, amino-sugars, sugar alcohols, sugar acids, organic phosphates, hydroxyl acids, aromatics, purines, and sterols as methoximated and trimethylsilylated mass spectra under EI ionisation. One advantage of GC-MS is that compound identification based on both mass spectral patterns can also integrate the retention index. Retention indices can be measured based on aliphatic carbon numbers [227] or alternatively, calculated from molecular properties [228]. Other generic databases exist such as NIST [229], METLIN [230] and MassBank [231]. They can be searched for EI-MS or CID MS/MS spectra. However, in this case, MS/MS spectra cannot be searched with the same level of confidence as EI-MS data. Significant variability exists between instruments with regard to ionisation and CID fragmentation due to the difficulty of applying standardised instrumental conditions. In this case, to ensure true positive matches, similar analytical platforms must be employed, and this constitutes a significant drawback of DIMS-based LC-MS approaches.

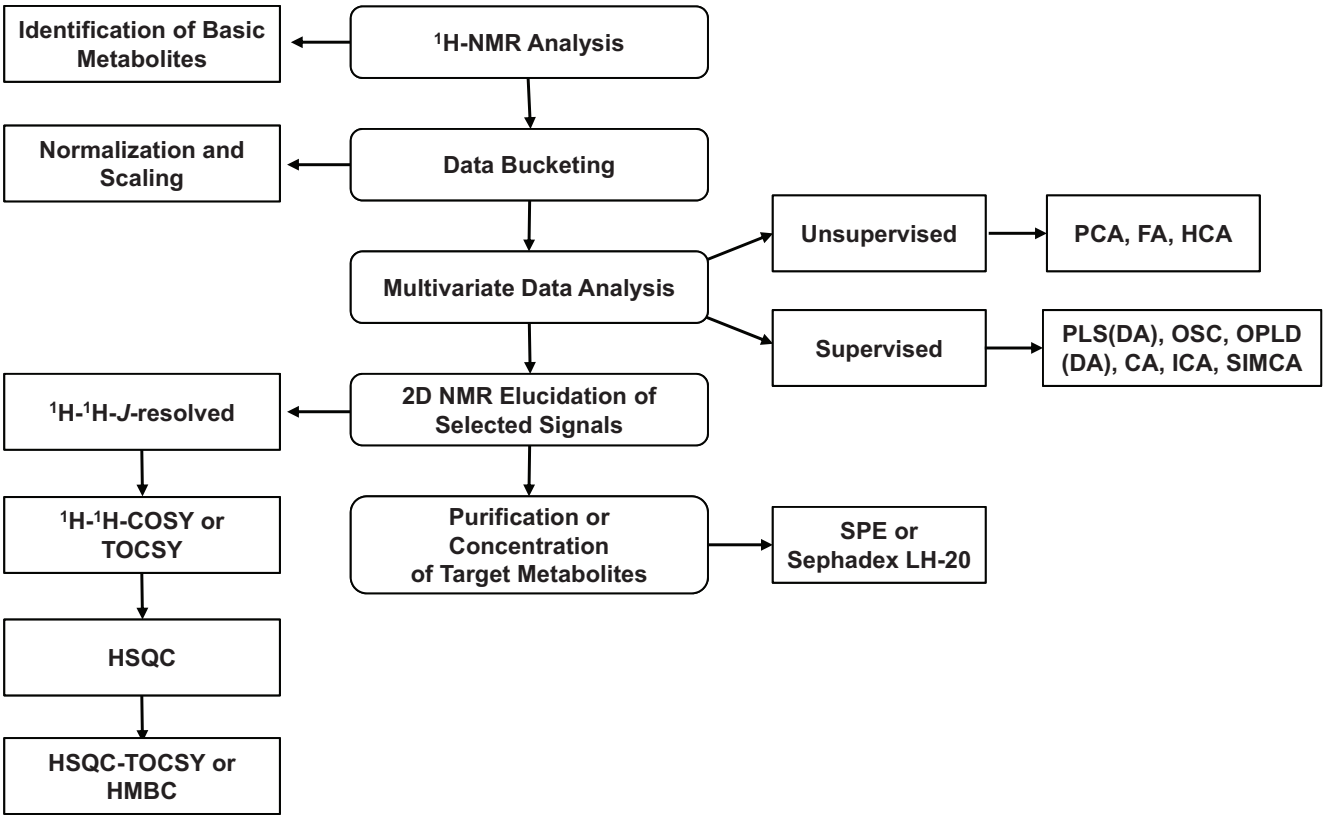

Fig. (11). Generic flow chart summarising the main steps of NMR identification.

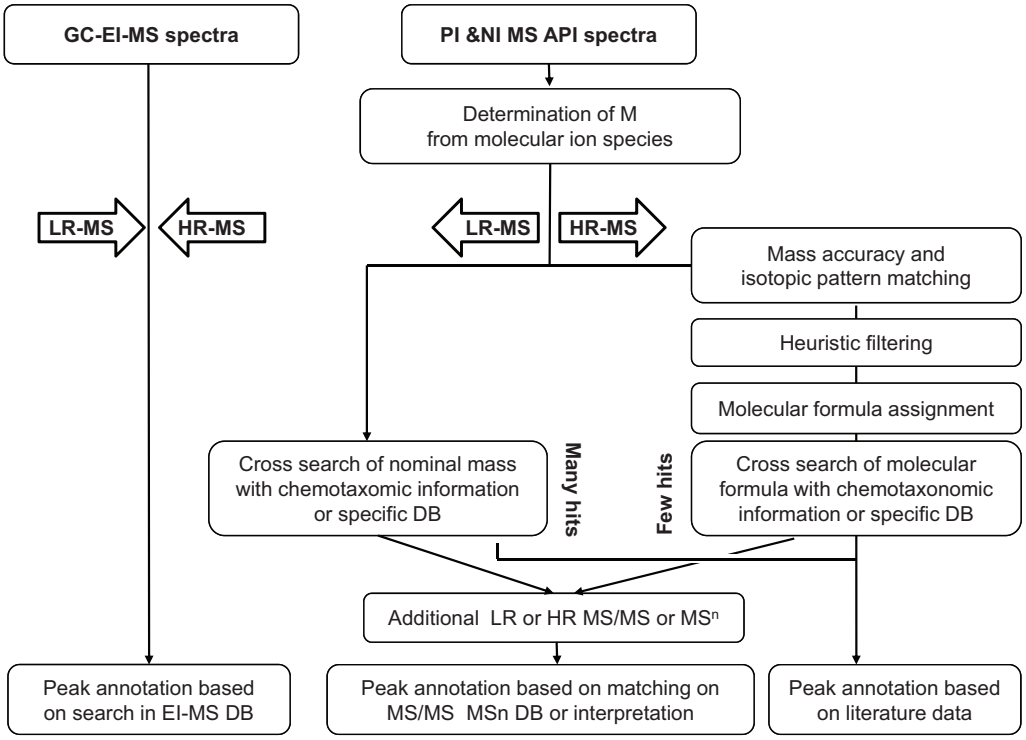

Fig. (12). Generic flow chart summarising the main steps of MS identification.

5.4.2. Pathway-Oriented Databases

Other databases provide information about annotated metabolic pathways (e.g., KEGG [232], MetaCyc [233] and Reactome [234]). They constitute key tools to extract biological information from metabolomic data by linking observed signals to existing knowledge [189]. They also provide information on enzymatic reactions with respect to the proteins or genes that are related to a given metabolic pathway. Among these, plant-specific databases including PlantCyc [235] and KNApSAcK [236] have been developed [37].

6. PLANT METABOLOMIC APPLICATIONS

During the last decade, metabolomics has been applied in several areas of research that are related to plant biology and chemistry. Its role differs according to the application, which ranges from the comprehensive and extensive profiling of a given plant metabolome [237] as an extension of what could be determined following a more traditional phytochemical investigation, to lead finding via the identification of biomarkers that can be statistically correlated to change in bioactivity and finally to the understanding of complex plant mechanisms as part of a systems biology approach. In particular, metabolomics was found to be helpful in solving issues related to (i) fingerprinting of species, genotypes or ecotypes for taxonomic, or biochemical (gene discovery) purposes [238-241]; (ii) comparing and contrasting the metabolite content of mutant or transgenic plants with that of their wild-type counterparts [23]; (iii) monitoring the behaviour of specific classes of metabolites in relation to applied exogenous chemical and/or physical stimuli [16, 242]; (iv) interaction of plants with the environment [243] or herbivores/pathogens [17, 80, 244]; (v) studying developmental processes, such as the establishment of symbiotic associations, fruit ripening, or germination [245]; (vi) quality control of medicinal herbs and phytopharmaceuticals [246, 247]; and (vii) and determining the activity of medicinal plants [248] and health-affecting compounds in food [79, 249, 250].

Many applications have been summarised in different useful field-specific reviews [6, 8, 17, 22, 23, 37, 40, 90, 143, 251, 252]. In this section, only a few selected applications in diverse fields have been summarised to provide a view of the various approaches that are used. These applications are listed in (Table **6**) and described in the following sections.

6.1. Chemotaxonomy and Classification of Plants

Several studies have demonstrated that metabolomics can be applied successfully to the classification of plants. For example, ^{1}H NMR metabolite fingerprinting in combination with PCA was applied to distinguish five different *Verbascum* species. Based on the metabolites, five species could be discriminated. Among these species, *V. xanthophoeniceum* and *V. nigrum* accumulated higher amounts of the pharmaceutically important harpagoside and verbascoside, forsythoside B and leucosceptoside B [253]. NMR-based metabolomics was applied to differentiate 11 *Ilex* species, of which only *I. paraguariensis* was used for the Yerba mate and the others were all adulterants. Metabolite analysis clearly showed that only *I. paraquariensis* contained xanthine alkaloids (caffeine, theobromine and theophylline) but it did not contain arbutin, which many of other species produced in large amounts. Hierarchical analysis of NMR data was well correlated to the phylogenetic data [208], which indicates that metabolome data can be efficiently used to establish a link between chemotaxonomy and phylogeny. However, great care should be taken in such analyses to avoid the influence of environments where plants have been grown. MS-based metabolomic studies were also found to be very useful for such applications. For example, an unbiased LC-MS-based metabolomic method utilising advanced data processing was applied to *Lonicera* species flower buds and provided efficient classification. Several biomarkers were identified with further MS-MS analysis. Iridoid glycosides in *L. japonica* were remarkably more abundant than those in other species, while saponins in *L. japonica* were limited, compared to the other 6 species [254].

6.2. Quality Control of Plant-Derived Medicines

The current quality control of medicinal plants or plant-derived medicines has some limitations because the analyses only target the main active ingredients. Metabolomics can provide a holistic view of metabolites in the extracts, which is necessary to ensure the pharmacological efficiency of phytopharmaceuticals and to reduce the toxicity caused by unknown ingredients. NMR-based metabolomics has been successfully applied for this purpose to Cannabis, Strychnos, Ephedra, Hypericum and Ginseng [274-278], and more recently to Scutellaria and Polygonum [259, 260]. Significantly, most of these studies have highlighted that not only the main active component was found but also other metabolites that were important to determine the quality of the medicinal plant or plant-derived medicines.

Recently UHPLC-qTOF-MS was found to be very efficient for the quality control of St. John's Wort (*Hypericum perforatum*) preparations [258]. A selection of batches from 9 commercially available *H. perforatum* products available on the German and Egyptian markets showed variable quality, particularly in hyperforin and fatty acid content. Multivariate data analysis showed discrimination between various preparations according to their global composition, including differentiation between various batches from the same supplier.

6.3. Lead Finding

A conventional method of finding leads of natural product origin is to integrate the bioactivity-guided fractionation of plant extracts until a pure active compound is obtained [279]. However, this approach is tedious and does not always lead to the identification of single bioactive compounds that may explain the clinical efficacy of a given botanical that is used as a drug. Furthermore, the various chromatography steps involved in bioactivity-guided fraction may degrade the bioactive compounds and bioactivity is sometimes not recovered after fractionation. These considerations have motivated the search for new alternative methods in lead finding [280]. Recently, metabolomics has been applied to lead finding, and it has been demonstrated to be a very promising tool. In particular, NMR-based approaches combined with chemometrics have provided valuable information about the active ingredients of plant extracts. Particularly, supervised multivariate data analysis methods, such as PLS or O2PLS, were found to be particularly useful for the identification of biomarkers in such studies. For example, 24 different extracts of four different accessions of St. John's Wort (*Hypericum perforatum*) extracted by six distinct solvents were evaluated [277]. Partial least squares analysis was used as a regression model and proved to be effective in identifying the resonances in the ^{1}H NMR spectrum that were correlated with the activity. The same NMR-based approach was used to predict the anti-plasmodial activity in different *Artemisia annua* extracts and allowed their classification based on this activity [263]. A similar approach was taken using the Mexican anxiolytic and sedative plant, *Galphimia glauca*. PLS-DA modelling discriminated active and non-active samples collected in different areas of Mexico and showed that the signals related to activity were associated with specific metabolites, the galphimines [264]. A comprehensive extraction of *Orthosiphon stamineus* combined NMR-based metabolomics with adenosine A1 receptor binding activity. Two flavonoids, among a large number of NPs, were clearly verified to be responsible for the activity without any further purification steps [265].

6.4. Plant physiology and Interaction with other Organisms

Plants respond upon stimuli and stress in various manners. A comprehensive analysis of metabolome modifications resulting from these effects has opened new avenues to study these complex plant chemical responses. Many very valuable studies of these aspects have been conducted, mainly using crop plants, and the results are summarised in different reviews [16, 17, 37]. In the field of abiotic stresses, hyphenated MS-based metabolomics has been applied to the study of the response of plants to dehydration [281], salt treatment [282], temperature changes [283] and light stress [284].

One recent example is the NMR metabolomics study of two cultivars of grapevines, 'Regent' (resistant) and 'Trincadeira'

Table 6. Selected Representative Plant Metabolomic Applications.

Method(s)	Instrumentation	Species/organ(s)	Factor(s)	Application	References
Chemotaxonomy /classification					
^{1}H-NMR	NMR-500 MHz	*Verbascum*	Five different species	Metabolite profiling	[253]
^{1}H-NMR	NMR-500 MHz	*Ilex*	Eleven different species/adulterant	Metabolite profiling	[208]
LC-MS	RRLC/qTOF-MS	*Lonicera*	Seven *Lonicera* species (flower buds)	Metabolite profiling	[254]
GC-MS, LC-MS, ^{1}H-NMR	LC–ESI-MS, EI-MS NMR-600 MHz	*Glycyrrhiza*	Classification based on genetic and or geographical origin	Metabolite profiling	[255]
GC-MS	EI-Q-MS	*Curcuma*	Three Curcuma species	Metabolite profiling	[256]
Quality control / Plant-derived medicine					
GC-MS	PY-GC-MS	*Angelica acutiloba*	cultivation area to quality evaluation	Metabolite fingerprinting	[257]
LC-MS	UPLC-qTOF-MS	*Hypericum perforatum*	Commercial preparations	Metabolite profiling	[258]
^{1}H-NMR	NMR-500 MHz	Polygonum	Commercial preparations	Metabolite fingerprinting	[259]
^{1}H-NMR	NMR-500 MHz	*Scutellaria baicalensis*	Plant from different origin	Metabolite profiling	[260]
HPLC-PDA-MS-SPE-NMR	NMR-600 MHz	*Ginkgo biloba*	Commercial preparations	Metabolite profiling	[261]
Lead finding					
^{1}H-NMR	NMR-500 MHz	*Echinacea*	Inhibitory Effects on Cytochrome P450 3A4	Metabolite profiling & PLS	[262]
^{1}H-NMR	NMR-500 MHz	*Artemisia*	Anti-plasmoidal activity	Metabolite profiling & PCA	[263]
^{1}H-NMR	NMR-500 MHz	*Galphimia*	Sedative activity	Metabolite profiling & PLS-DA	[264]
^{1}H-NMR	NMR-500 MHz	*Orthosiphon*	adenosine A1 receptor binding activity	Metabolite profiling & PLS	[265]
Plant physiology/Interaction with other organisms					
GC-MS; LC-MS	EI-TOF-MS; ESI-TOF-MS	Rice/leaves	Bacterial leaf blight disease	Metabolite profiling	[266]
^{1}H-NMR	NMR-600 MHz	Maize/roots and shoots	Salt stress	Metabolite fingerprinting	[267]
^{1}H-NMR and HR-MAS	NMR-500 and 600 MHz	Rice/seeds	Biotic and abiotic stress	Metabolite fingerprinting	[268]
^{1}H-NMR	NMR-400 MHz	Wheat/leaves and stem	Fusarium head blight disease	Metabolite fingerprinting	[269]
1D- and 2D-NMR	NMR-500 MHz	Pea/leaves	Drought stress	Metabolite fingerprinting	[270]
GC-MS	EI-Q-MS	Barley/roots and leaves	Salt stress	Metabolite profiling	[271]
ICP-AES; GC-MS	EI-TOF-MS	Lotus/shoots	Salt stress	Elements, transcript and metabolite profiling	[272]
GC-MS; 1D- and 2D-NMR	NMR-400 MHz; EI-Q-MS	Rice/roots	Chromium stress	Transcript and metabolite profiling	[273]

Abbreviations: EI, electronic ionization; ESI, electrospray ionization; GC, gas chromatography; LC, high performance liquid chromatography; MS, mass spectrometry; HR-MAS, high-resolution magic angle spinning NMR; ICP-AES, inductively coupled plasma-optical emission spectrometry; NMR, nuclear magnetic resonance; PDA, photodiode array; PLS, partial least squares regression analysis; Q, quadrupole; TOF, time of flight.

(susceptible), in relation to downy mildew pathogen (*Plasmopara viticola*) infection. Metabolites responsible for the discrimination were identified as a fertaric acid, caftaric acid, quercetin-3-O-glucoside, linolenic acid, and alanine in the resistant cultivar 'Regent', while the susceptible 'Trincadeira' showed higher levels of glutamate, succinate, ascorbate and glucose [39]. Another study related to the salt stress response of maize shoots and roots showed a clear separation of the growth and saline effects that was highlighted by supervised data mining methods of NMR data. Twenty-eight compounds were characterised based on proton chemical shifts. Among them, increased levels of alanine, glutamate, asparagine and glycine-betaine were detected upon salt stress [267].

Similar NMR approaches have been applied to Arabidopsis, Brassica, Senecio and Chrysanthemums to study the metabolite response of corresponding plants after infection or interaction with other organisms [36, 202, 209].

Very recently, an MS-based metabolomic study has revealed herbivore-induced metabolites of resistance and susceptibility in maize leaves and roots. Local and systemic herbivore-induced changes in maize leaves, sap, roots and root exudates were evaluated in *Spodoptera littoralis*-infested maize without any prior assumptions about their function. Thirty-two differentially regulated compounds from seedlings were highlighted by LC-MS and the structures of the biomarkers were further confirmed by microflow NMR analysis [80]. This last example revealed an interesting complementarity of MS and NMR in plant metabolomics studies.

CONCLUSIONS AND NEW TRENDS

Since the relatively recent first publications on plant metabolomics at the beginning of this millennium, this new field of science has progressed from an ambitious concept to a rapidly growing and valuable scientific strategy that provides a more global picture of the molecular organisation of various organisms [23].

Metabolomics has rapidly evolved during the last decade thanks to the impressive synergistic development of analytical technologies and powerful bioinformatics tools. In the case of MS, the widespread access to sensitive high-resolution bench-top instruments and the introduction of fast and efficient chromatographic methods represent important breakthroughs. Robust methods based on these technologies are now routinely used in numerous laboratories. In parallel, the development of ultrahigh-resolution instruments or imaging mass spectrometry methods has also opened new avenues. For NMR, impressive improvements in sensitivity with the introduction of micro coil and cryogenised probes have pushed the limits further. The advent of probes built with high temperature superconducting (HTS) material or the implementation of efficient polarisation transfer methods might also become very interesting means of attaining high sensitivity. The increase in NMR magnetic field strengths will certainly contribute to the constant improvement in spectral resolution.

Metabolomics can therefore be considered as a mature field that represents an important complement to the other "omics" technologies that are used in systems biology. As shown by many valuable applications, metabolomics has yielded significant new biological knowledge. This way of addressing new biological issues in many different fields of plant science by data-driven approaches has several advantages compared to the classical hypothesis-driven approach. Notwithstanding these impressive analytical and methodological developments, the size of a plant metabolome can still be only very roughly estimated and the comprehensive identification and quantitation of all metabolites in a given organism is far from being achieved. The current bottleneck of metabolomics continues to be the unambiguous identification of the numerous detected features that may contain key information about a biological system.

Clearly, more effort needs to be made in this area. To solve these problems, many attempts have been made, and many investigations are still underway. To raise plant metabolomics to the same level of identification as explored for the analysis of body fluids with the so-called "metabonomics" approach [285], it will be essential to create public databases and establish standardised protocols that normalise and facilitate the information exchange from different laboratories. While several rather generic protocols exist, unique methods and ways to store the data are still far from being established for the whole community. This is also likely related to the large diversity of natural products in the plant kingdom and the species-specific occurrence of secondary metabolites. Numerous research groups possessing similar metabolomic platforms have started to use standard protocols to allow the comparison and sharing of metabolomics information. Therefore, the development and coordination of databases will likely constitute a cornerstone to extracting original biological information from experimental data. Until now, a compromise between high quality chemotaxonomic information, MS/MS and NMR spectra and sophisticated algorithms (e.g., spectroscopic simulation MS^n spectral tree calculations, etc.) have helped natural product chemists to accelerate the pace at which biomarkers are identified.

The multivariate nature of data obtained from metabolomic experiments requires specific approaches to extract relevant information. As an increase in data accumulation is inevitable, automatically finding the valuable but hidden information in metabolomic data of high dimensionality is another crucial issue. Data mining, therefore, has a central role to play by interpreting the data and inferring the structures governing biological phenomena [152]. The combination of multiple data sources will undoubtedly provide a more comprehensive vision by extracting common traits from different datasets obtained from the same plant sample(s). However, data mining should not be considered a black box, and a thorough understanding of the raw data is required to prevent multidimensional data that have undergone data reduction from being inaccurately interpreted.

To study whole biological systems, knowledge generated by metabolomics has to be integrated with those from other 'omics' approaches such as genomics, transcriptomics and proteomics (inter-omic fusion). Even within metabolomics, the combination of data from complementary analytical platforms would be useful to study whole system (intra-omic fusion). This fusion of data from different omics level is challenging and requires the development of tools in order to display and interpret the vast amounts of data in an efficient manner and several attempts have been made in this direction (for a review see [286]). Despite some difficulties research effort in this direction are evolving very rapidly and will be key for further development of systems biology and functional genomics.

At present, most of the metabolomics studies that have been conducted in plant science have mainly focused on the characterisation of metabolome changes in entire plant organs that occur at a given time. This can provide interesting information regarding the metabolic status of an organism, but this only represents a static picture of the observed phenomena. A deeper understanding of subtle metabolome changes that may ultimately explain the functioning of the biological system would likely require a higher level of compartmentalisation (study of vacuoles, mitochondria, plastids and nuclei, or even single cells) that also takes metabolite fluxes into account. Therefore, spatiotemporal metabolomics studies performed on adequate timescales and, when possible, at the cellular level represent new challenges to be overcome in this constantly evolving and fascinating field of research.

CONFLICT OF INTEREST

The author(s) confirm that this article content has no conflicts of interest.

ACKNOWLEDGEMENTS

We thank Julien Boccard (Unige) Guillaume Marti (Unige) and Gaetan Glauser (Unine) for their fruitful discussion and advice in preparing the manuscript. JLW and SR are grateful to the Swiss National Science Foundation for the financial support of their metabolomics studies (Grant no. 205320-124667/1 and CRSII3_127187) and to the National Centre of Competence in Research (NCCR) Plant Survival, an additional research programme of the Swiss National Science Foundation. We are also grateful to the National Institute of Horticultural and Herbal Science, Rural Development Administration of Korea (Project No: PJ008700), for their financial support of YHC and HKK.

LIST OF ABBREVIATIONS

ANN	=	Artificial Neural Network
APCI	=	Atmospheric Pressure Chemical Ionisation
API	=	Atmospheric Pressure Ionisation
APPI	=	Atmospheric Pressure Photoionisation
CE-MS	=	Capillary Electrophoresis-MS
COSY	=	COrrelation SpectroscopY
COW	=	Correlation Optimised Warping
DART	=	Direct Analysis In Real Time
DESI	=	Desorption Electrospray Ionisation
DIMS	=	Direct Injection MS
DTW	=	Dynamic Time Warping
EESI	=	Extractive Electrospray Ionisation
EI	=	Electron Ionisation
ESI	=	Electrospray Ionisation

FT-ICR-MS	=	Fourier Transform - Ion Cylotron - Resonance Ms
GC	=	Gas Chromatography
GC-MS	=	Gas Chromatography - Ms
HCA	=	Hierarchical Cluster Analysis
HILIC	=	Hydrophilic Interaction Liquid Chromatography
HMBC	=	Heteronuclear Multiple Bond Correlation
HMQC	=	Heteronuclear Multiple Quantum Coherence
HPLC	=	High Performance Liquid Chromatography
HR	=	High-Resolution
HSQC	=	Heteronuclear Single Quantum Coherence Spectroscopy
HTS	=	High Temperature Superconducting
IT	=	Ion Trap
LC	=	Liquid Chromatography
LC-MS	=	Liquid Chromatography - MS
LC-NMR	=	Liquid Chromatography - NMR
LIT	=	Linear Ion Trap
LLE	=	Liquid-Liquid Extraction
LR	=	Low Resolution
MALDI	=	Matrix-Assisted Laser Desorption Ionisation
MS	=	Mass Spectrometry
MS/MS	=	Tandem Mass Spectrometry
MSI	=	MS Imaging
MSn	=	Multiple Stage Mass Spectrometry
MVDA	=	Multivariate Data Analysis
NI	=	Negative Ionisation
NMR	=	Nuclear Magnetic Resonance
NP	=	Normal Phase
NPs	=	Natural Products
O-PLS	=	Orthogonal PLS
PCA	=	Principal Component Analysis
PI	=	Positive Ionisation;
PLS	=	Partial Least Squares
PLS-DA	=	PLS Discriminant Analysis
Q	=	Quadrupole
QC	=	Quality Control
QIT	=	Quadrupole Ion Trap
QqQ	=	Triple Quadrupole
RP	=	Reverse Phase
SD	=	Standard Deviation
SHY	=	Statistical Heterospectroscopy
SPE	=	Solid-Phase Extraction
SVM	=	Support Vector Machine
TMS	=	Total Mass Spectra
TOCSY	=	TOtal Correlation SpectroscopY
TOF	=	Time-Of-Flight
UHPLC	=	Ultra-High Pressure Liquid Chromatography
UV-scaling	=	Unit Variance - Scaling

REFERENCES

[1] Fraenkel, G.S. The Raison d'Être of Secondary Plant Substances. *Science*, **1959**, *129*, 1466-1470.

[2] Bennett, R.N.; Wallsgrove, R.M. Secondary metabolites in plant defense-mechanisms. *New Phytol.*, **1994**, *127*, 617-633.

[3] Grotewold, E. Plant metabolic diversity: a regulatory perspective. *Trends Plant Sci.*, **2005**, *10*, 57-62.

[4] Wink, M. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry*, **2003**, *64*, 3-19.

[5] Kingston, D.G.I. Modern Natural Products Drug Discovery and Its Relevance to Biodiversity Conservation. *J. Nat. Prod.*, **2011**, *74*, 496-511.

[6] Wolfender, J.L.; Glauser, G.; Boccard, J.; Rudaz, S. MS-based Plant Metabolomic Approaches for Biomarker Discovery. *Nat. Prod. Comm.*, **2009**, *4*, 1417-1430.

[7] Dettmer, K.; Aronov, P.A.; Hammock, B.D. Mass spectrometry-based metabolomics. *Mass Spectrom. Rev.*, **2007**, *26*, 51-78.

[8] Sumner, L.W.; Mendes, P.; Dixon, R.A. Plant metabolomics:large-scale phytochemistry in the functional genomics era. *Phytochemistry*, **2003**, *62*, 817-836.

[9] Fiehn, O.; Kopka, J.; Dörmann, P.; Altmann, T.; Trethewey, R.N. Metabolite profiling for plant functional genomics. *Nature Biotechnol.*, **2000**, *18*, 1157-1161.

[10] Fukushima, A.; Kusano, M.; Redestig, H.; Arita, M.; Saito, K. Integrated omics approaches in plant systems biology. *Curr. Opin. Chem. Biol.*, **2009**, *13*, 532-538.

[11] Fukusaki, E.; Kobayashi, A. Plant meta-bolomics: Potential for practical operation. *J. Biosci. Bioeng.*, **2005**, *100*, 347-354.

[12] Hamilton, J.P.; Buell, C.R. Advances in plant genome sequencing. *Plant J*, **2012**, *70*, 177-190.

[13] Zhang, A.H.; Sun, H.; Wang, P.; Han, Y.; Wang, X.J. Recent and potential developments of biofluid analyses in metabolomics. *J. Proteomics*, **2012**, *75*, 1079-1088.

[14] Guy, C.; Kopka, J.; Moritz, T. Plant metabolomics coming of age. *Physiol. Plant.*, **2008**, *132*, 113-116.

[15] Shepherd, L.V.T.; Fraser, P.; Stewart, D. Metabolomics: a second-generation platform for crop and food analysis. *Bioanalysis*, **2011**, *3*, 1143-1159.

[16] Shulaev, V.; Cortes, D.; Miller, G.; Mittler, R. Metabolomics for plant stress response. *Physiol. Plant.*, **2008**, *132*, 199-208.

[17] Jansen, J.J.; Allwood, J.W.; Marsden-Edwards, E.; van der Putten, W.H.; Goodacre, R.; van Dam, N.M. Metabolomic analysis of the interaction between plants and herbivores. *Metabolomics*, **2009**, *5*, 150-161.

[18] Verpoorte, R.; Choi, Y.H.; Mustafa, N.R.; Kim, H.K. Metabolomics: back to basics. *Phytochem. Rev.*, **2008**, *7*, 525-537.

[19] D'Auria, J.C.; Gershenzon, J. The secondary metabolism of *Arabidopsis thaliana*: growing like a weed. *Curr. Opin. Plant Biol.*, **2005**, *8*, 308-316.

[20] Dixon, R.A. Natural products and plant disease resistance. *Nature*, **2001**, *411*, 843-847.

[21] Dixon, R.A.; Strack, D. Phytochemistry meets genome analysis, and beyond. *Phytochemistry*, **2003**, *62*, 815-816.

[22] Dunn, W.B. Current trends and future requirements for the mass spectrometric investigation of microbial, mammalian and plant metabolomes. *Phys. Biol.*, **2008**, *5*.

[23] Hall, R.D. Plant metabolomics: from holistic hope, to hype, to hot topic. *New Phytol.*, **2006**, *169*, 453-468.

[24] Kim, H.K.; Verpoorte, R. Sample preparation for plant metabolomics. *Phytochem. Anal.*, **2010**, *21*, 4-13.

[25] Abdel-Farid, B.B.; Kim, H.K.; Choi, Y.H.; Verpoorte, R. Metabolic characterization of Brassica rapa leaves by NMR spectroscopy. *J. Agric. Food Chem.*, **2007**, *55*, 7936-7943.

[26] Urbanczyk-Wochniak, E.; Baxter, C.; Kolbe, A.; Kopka, J.; Sweetlove, L.J.; Fernie, A.R. Profiling of diurnal patterns of metabolite and transcript abundance in potato (*Solanum tuberosum*) leaves. *Planta*, **2005**, *221*, 891-903.

[27] Glauser, G.; Grata, E.; Dubugnon, L.; Rudaz, S.; Farmer, E.; Wolfender, J.L. Spatial and temporal dynamics of Jasmonate synthesis and accumulation in Arabidopsis in response to wounding. *J. Biol. Chem.*, **2008**, *283*, 16400-16407.

[28] Harbourne, N.; Marete, E.; Jacquier, J.C.; O'Riordan, D. Effect of drying methods on the phenolic constituents of meadowsweet (*Filipendula ulmaria*) and willow (*Salix alba*). *Lwt-Food Sci. Technol.*, **2009**, *42*, 1468-1473.

[29] Schmitke, J.L.; Wescott, C.R.; Klibanov, A.M. The mechanistic dissection of the plunge in enzymatic activity upon transition from water to anhydrous solvents. *J. Am. Chem. Soc.*, **1996**, *118*, 3360-3365.

[30] Fiehn, O. Metabolomics--the link between genotypes and phenotypes. *Plant Mol. Biol.*, **2002**, *48*, 155-171.

[31] Gibon, Y.; Usadel, B.; Blaesing, O.E.; Kamlage, B.; Hoehne, M.; Trethewey, R.; Stitt, M. Integration of metabolite with transcript and enzyme activity profiling during diurnal cycles in Arabidopsis rosettes. *Gen. Biol.*, **2006**, *7*.

[32] Liang, Y.S.; Choi, Y.H.; Kim, H.K.; Linthorst, H.J.M.; Verpoorte, R. Metabolomic analysis of methyl jasmonate treated *Brassica rapa* leaves by 2-

dimensional NMR spectroscopy. *Phytochemistry*, **2006**, *67*, 2503-2511.

[33] Schneider, B.; Holscher, D. Laser microdissection and cryogenic nuclear magnetic resonance spectroscopy: an alliance for cell type-specific metabolite profiling. *Planta*, **2007**, *225*, 763-770.

[34] Shroff, R.; Vergara, F.; Muck, A.; Svatos, A.; Gershenzon, J. Nonuniform distribution of glucosinolates in *Arabidopsis thaliana* leaves has important consequences for plant defense. *Proc. Natl. Acad. Sci. USA*, **2008**, *105*, 6196-6201.

[35] Robert, C.A.M.; Veyrat, N.; Glauser, G.; Marti, G.; Doyen, G.R.; Villard, N.; Gaillard, M.D.P.; Köllner, T.G.; Giron, D.; Body, M.; Babst, B.A.; Ferrieri, R.A.; Turlings, T.C.J.; Erb, M. A specialist root herbivore exploits defensive metabolites to locate nutritious tissues. *Ecol. Lett.*, **2012**, *15*, 55-64.

[36] Leiss, K.A.; Choi, Y.H.; Abdel-Farid, I.B.; Verpoorte, R.; Klinkhamer, P.G.L. NMR Metabolomics of Thrips (*Frankliniella occidentalis*) Resistance in Senecio Hybrids. *J. Chem. Ecol.*, **2009**, *35*, 219-229.

[37] Glauser, G.; Boccard, J.; Wolfender, J.L.; Rudaz, S. Metabolomics : Application in plant sciences. In *Metabolomics in Practice*; Laemmerhofer, M., Weckwerth, W., Eds.; Wiley-VCH **2012**; Vol. In press.

[38] Le Gall, G.; Colquhoun, I.J.; Defernez, M. Metabolite profiling using H-1 NMR spectroscopy for quality assessment of green tea, *Camellia sinensis* (L.). *J. Agric. Food Chem.*, **2004**, *52*, 692-700.

[39] Ali, K.; Maltese, F.; Figueiredo, A.; Rex, M.; Fortes, A.M.; Zyprian, E.; Pais, M.S.; Verpoorte, R.; Choi, Y.H. Alterations in grapevine leaf metabolism upon inoculation with *Plasmopara viticola* in different time-points. *Plant Sci.*, **2012**, *191*, 100-107.

[40] Kim, H.K.; Choi, Y.H.; Verpoorte, R. NMR-based metabolomic analysis of plants. *Nat. Protocols*, **2010**, *5*, 536-549.

[41] Namiesnik, J.; Gorecki, T. Sample preparation for chromatographic analysis of plant material. *Jpc-J. Planar Chromat.*, **2000**, *13*, 404-413.

[42] Sargenti, S.R.; Vichnewski, W. Sonication and liquid chromatography as a rapid technique for extraction and fractionation of plant material. *Phytochem. Anal.*, **2000**, *11*, 69-73.

[43] Jonsson, P.; Gullberg, J.; Nordstrom, A.; Kusano, M.; Kowalczyk, M.; Sjostrom, M.; Moritz, T. A strategy for identifying differences in large series of metabolomic samples analyzed by GC/MS. *Anal. Chem.*, **2004**, *76*, 1738-1745.

[44] Benthin, B.; Danz, H.; Hamburger, M. Pressurized liquid extraction of medicinal plants. *J. Chromatogr. A*, **1999**, *837*, 211-219.

[45] Castioni, P.; Christen, P.; Veuthey, J.L. Supercritical-fluid extraction of compounds from plant-origin. *Analusis*, **1995**, *23*, 95-106.

[46] Glauser, G.; Grata, E.; Rudaz, S.; Wolfender, J.-L. High-resolution profiling of oxylipin-containing galactolipids in Arabidopsis extracts by ultra-performance liquid chromatography/time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.*, **2008**, *22*, 3154-3160.

[47] Kim, H.; Choi, Y. Metabolomic analysis of *Catharanthus roseus* using NMR and principal component analysis. In *Plant Metabolomics: Biotechnology in Agricultural and Forestry* Saito, K., Dixon, R.A., Willmitzer, L., Eds.; Springer Berlin Heidelberg, **2006**, pp 261-276.

[48] Ali, K.; Maltese, F.; Fortes, A.M.; Pais, M.S.; Verpoorte, R.; Choi, Y.H. Pre-analytical method for NMR-based grape metabolic fingerprinting and chemometrics. *Anal. Chim. Acta*, **2011**, *703*, 179-186.

[49] Khatib, A.; Wilson, E.G.; Kim, H.K.; Lefeber, A.W.M.; Erkelens, C.; Choi, Y.H.; Verpoorte, R. Application of two-dimensional J-resolved nuclear magnetic resonance spectroscopy to differentiation of beer. *Anal. Chim. Acta*, **2006**, *559*, 264-270.

[50] Grata, E.; Boccard, J.; Guillarme, D.; Glauser, G.; Carrupt, P.-A.; Farmer, E.E.; Wolfender, J.-L.; Rudaz, S. UPLC-TOF-MS for plant metabolomics: A sequential approach for wound marker analysis in *Arabidopsis thaliana*. *J. Chromatogr. B*, **2008**, *871*, 261-270.

[51] Halket, J.M.; Zaikin, V.G. Derivatization in mass spectrometry - 1. Silylation. *Eur. J. Mass Spectrom.*, **2003**, *9*, 1-21.

[52] Glassop, D.; Roessner, U.; Bacic, A.; Bonnett, G.D. Changes in the sugarcane metabolome with stem development. Are they related to sucrose accumulation? *Plant Cell Physiol.*, **2007**, *48*, 573-584.

[53] Fiehn, O.; Wohlgemuth, G.; Scholz, M.; Kind, T.; Lee, D. Y.; Lu, Y.; Moon, S.; Nikolau, B. Quality control for plant metabolomics: reporting MSI-compliant studies. *Plant J.*, **2008**, *53*, 691-704.

[54] Lisec, J.; Schauer, N.; Kopka, J.; Willmitzer, L.; Fernie, A.R. Gas chromatography mass spectrometry-based metabolite profiling in plants. *Nat. Protoc.*, **2006**, *1*, 387-396.

[55] De Vos, R.C.H.; Moco, S.; Lommen, A.; Keurentjes, J.J.B.; Bino, R.J.; Hall, R.D. Untargeted large-scale plant metabolomics using liquid chromatography coupled to mass spectrometry. *Nat. Protoc.*, **2007**, *2*, 778-791.

[56] Want, E.J.; Wilson, I.D.; Gika, H.; Theodoridis, G.; Plumb, R.S.; Shockcor, J.; Holmes, E.; Nicholson, J. K. Global metabolic profiling procedures for urine using UPLC-MS. *Nat. Protoc.*, **2010**, *5*, 1005-1018.

[57] Kruger, N.J.; Troncoso-Ponce, M.A.; Ratcliffe, R.G. 1H NMR metabolite fingerprinting and metabolomic analysis of perchloric acid extracts from plant tissues. *Nat. Protoc.*, **2008**, *3*, 1001-1012.

[58] Leiss, K.A.; Choi, Y.H.; Verpoorte, R.; Klinkhamer, P.G.L. An overview of NMR-based metabolomics to identify secondary plant compounds involved in host plant resistance. *Phytochem. Rev.*, **2011**, *10*, 205-216.

[59] Wolfender, J.L. LC-NMR and Related Techniques for the Rapid Identification of Plant Metabolites. In *High Performance Liquid Chromatography in Phytochemical Analysis* Waksmundzka-Hajnos, M.J.S. Eds.; **2010**; Vol. 102, pp 287-330.

[60] Choi, Y.H.; Kim, H.K.; Linthorst, H.J.M.; Hollander, J.G.; Lefeber, A.W.M.; Erkelens, C.; Nuzillard, J.M.; Verpoorte, R. NMR metabolomics to revisit the tobacco mosaic virus infection in Nicotiana tabacum leaves. *J. Nat. Prod.*, **2006**, *69*, 742-748.

[61] Colquhoun, I.J. Use of NMR for metabolic profiling in plant systems. *J. Pest. Sci.*, **2007**, *32*, 200-212.

[62] Wishart, D.S. Quantitative metabolomics using NMR. *TrAC*, **2008**, *27*, 228-237.

[63] Reynolds, W.F.; Enriquez, R.G. Choosing the best pulse sequences, acquisition parameters, postacquisition processing strategies, and probes for natural product structure elucidation by NMR spectroscopy. *J. Nat. Prod.*, **2002**, *65*, 221-244.

[64] Wolfender, J.-L.; Bohni, N.; Ndjoko-Ioset, K.; Edison, A.S. Advanced Spectroscopic Detectors for Identification and Quantification: Nuclear Magnetic Resonance. In *Handbooks in Separation Science: Liquid Chromatography*; Fanali, S.P.H. Lloyd, D.; Poole, C.F.; Schoenmakers, P.J. Eds., **2012**, pp (in press).

[65] Hoult, D.I.; Richards, R.E. Signal-to-Noise Ratio of Nuclear Magnetic-Resonance Experiment. *J. Magn. Reson.*, **1976**, *24*, 71-85.

[66] Aramini, J.M.; Rossi, P.; Anklin, C.; Xiao, R.; Montelione, G.T. Microgram-scale protein structure determination by NMR. *Nat. Meth.*, **2007**, *4*, 491-493.

[67] Schroeder, F.C.; Gronquist, M. Extending the scope of NMR spectroscopy with microcoil probes. *Angew. Chem. Int. Ed.*, **2006**, *45*, 7122-7131.

[68] Kovacs, H.; Moskau, D.; Spraul, M. Cryogenically cooled probes - a leap in NMR technology. *Prog. Nucl. Magn. Reson. Spectrosc.*, **2005**, *46*, 131-155.

[69] Viant, M.R. Improved methods for the acquisition and interpretation of NMR metabolomic data. *Biochem. Biophys. Res. Commun.*, **2003**, *310*, 943-948.

[70] Pauli, G.F.; Jaki, B.U.; Lankin, D.C. Quantitative H-1 NMR: Development and potential of a method for natural products analysis. *J. Nat. Prod.*, **2005**, *68*, 133-149.

[71] McKay, R.T. Recent advances in solvent suppression for solution NMR: a practical reference. *Annu. Rep. NMR Spectr.*, **2009**, *66*, 33-76.

[72] Beckwith-Hall, B.M.; Thompson, N.A.; Nicholson, J.K.; Lindon, J.C.; Holmes, E.A metabonomic investigation of hepatotoxicity using diffusion-edited H-1 NMR spectroscopy of blood serum. *Analyst*, **2003**, *128*, 814-818.

[73] Phalaraksh, C.; Lenz, E.M.; Lindon, J.C.; Nicholson, J.K.; Farrant, R.D.; Reynolds, S.E.; Wilson, I.D.; Osborn, D.; Weeks, J.M. NMR spectroscopic studies on the haemolymph of the tobacco hornworm, Manduca sexta: assignment of H-1 and C-13 NMR spectra. *Insect Biochem. Mol. Biol.*, **1999**, *29*, 795-805.

[74] Kikuchi, J.; Shinozaki, K.; Hirayama, T. Stable isotope labeling of *Arabidopsis thaliana* for an NMR-based metabolomics approach. *Plant Cell Physiol.*, **2004**, *45*, 1099-1104.

[75] Glish, G.L.; Vachet, R.W. The basics of mass spectrometry in the twenty-first century. *Nat. Rev. Drug Discov.*, **2003**, *2*, 140-150.

[76] Villas-Boas, S.G.; Mas, S.; Aakesson, M.; Smedsgaard, J.; Nielsen, J. Mass spectrometry in metabolome analysis. *Mass Spectrom. Rev.*, **2005**, *24*, 613-646.

[77] Soga, T.; Ohashi, Y.; Ueno, Y.; Naraoka, H.; Tomita, M.; Nishioka, T. Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J. Prot. Res.*, **2003**, *2*, 488-494.

[78] Lino, K.; Sugimoto, M.; Soga, T.; Tomita, M. Profiling of the charged metabolites of traditional herbal medicines using capillary electrophoresis time-of-flight mass spectrometry. *Metabolomics*, **2012**, *8*, 99-108.

[79] Marti, G.; Erb, M.; Rudaz, S.; Turlings, T.C.J.; Wolfender, J.L. Search for Low-Molecular-Weight Biomarkers in Plant Tissues and Seeds using Metabolomics: Tools, Strategies, and Applications. In *SEED DEVELOPMENT: OMICS Technologies toward Improvement of Seed Quality and Crop Yield*; Agrawal, G.K.; Rakwal, R. Eds.; Springer, **2012**; Vol. In press.

[80] Marti, G.; Erb, M.; Boccard, J.; Glauser, G.; Doyen, G.R.; Villard, N.; Turlings, T.C.J.; Rudaz, S.; Wolfender, J.L. Metabolomics reveals herbivore-induced metabolites of resistance and susceptibility in maize. *Plant Cell Environ.*, **2012**, *(in press) doi: 10.1111/pce.12002*.

[81] Bottcher, C.; von Roepenack-Lahaye, E.; Willscher, E.; Scheel, D.; Clemens, S. Evaluation of matrix effects in metabolite profiling based on capillary liquid chromatography electrospray ionization quadrupole time-of-flight mass spectrometry. *Anal. Chem.*, **2007**, *79*, 1507-1513.

[82] Korfmacher, W.A. Principles and applications of LC-MS in new drug discovery. *Drug Discov. Today*, **2005**, *10*, 1357-1367.

[83] Hu, Q.Z.; Noll, R.J.; Li, H.Y.; Makarov, A.; Hardman, M.; Cooks, R.G. The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.*, **2005**, *40*, 430-443.

[84] Feng, X.; Siegel, M.M. FTICR-MS applications for the structure determination of natural products. *Anal. Bioanal. Chem.*, **2007**, *389*, 1341-1363.

[85] Tsybin, Y.O.; Fornelli, L.; Kozhinov, A.N.; Vorobyev, A.; Miladinovic, S.M.

High-Resolution and Tandem Mass Spectrometry - the Indispensable Tools of the XXI Century. *Chimia*, **2011**, *65*, 641-645.

[86] Guillarme, D.; Schappler, J.; Rudaz, S.; Veuthey, J.-L. Coupling ultra-high-pressure liquid chromatography with mass spectrometry. *TrAC*, **2010**, *29*, 15-27.

[87] Eugster, P.; Guillarme, D.; Rudaz, S.; Veuthey, J.L.; Carrupt, P.A.; Wolfender, J.L. Ultra high pressure liquid chromatography for crude plant extracts profiling. *J. AOAC Int.*, **2011**, *94*, 51-70.

[88] D'Alessandro, A.; Gevi, F.; Zolla, L. Targeted mass spectrometry-based metabolomic profiling through multiple reaction monitoring of liver and other biological matrices. *Meth. Mol. Biol.*, **2012**, *909*, 279-294.

[89] van der Hooft, J.J.J.; Vervoort, J.; Bino, R.J.; Beekwilder, J.; de Vos, R.C.H. Polyphenol Identification Based on Systematic and Robust High-Resolution Accurate Mass Spectrometry Fragmentation. *Anal. Chem.*, **2011**, *83*, 409-416.

[90] Zhou, B.; Xiao, J.F.; Tuli, L.; Ressom, H.W. LC-MS-based metabolomics. *Mol. Biosystems*, **2012**, *8*, 470-481.

[91] Southam, A.D.; Payne, T.G.; Cooper, H.J.; Arvanitis, T.N.; Viant, M.R. Dynamic range and mass accuracy of wide-scan direct infusion nanoelectrospray Fourier transform ion cyclotron resonance mass spectrometry-based metabolomics increased by the spectral stitching method. *Anal. Chem.*, **2007**, *79*, 4595-4602.

[92] Scholz, M.; Gatzek, S.; Sterling, A.; Fiehn, O.; Selbig, J. Metabolite fingerprinting: detecting biological features by independent component analysis. *Bioinformatics*, **2004**, *20*, 2447-2454.

[93] Smedsgaard, J.; Frisvad, J.C. Using direct electrospray mass spectrometry in taxonomy and secondary metabolite profiling of crude fungal extracts. *J. Microbiol. Met.*, **1996**, *25*, 5-17.

[94] Oikawa, A.; Nakamura, Y.; Ogura, T.; Kimura, A.; Suzuki, H.; Sakurai, N.; Shinbo, Y.; Shibata, D.; Kanaya, S.; Ohta, D. Clarification of pathway-specific inhibition by Fourier transform ion cyclotron resonance/mass spectrometry-based metabolic phenotyping studies. *Plant Physiol.*, **2006**, *142*, 398-413.

[95] Boernsen, K.O.; Gatzek, S.; Imbert, G. Controlled protein precipitation in combination with chip-based nanospray infusion mass spectrometry. An approach for metabolomics profiling of plasma. *Anal. Chem.*, **2005**, *77*, 7255-7264.

[96] Breitling, R.; Pitt, A.R.; Barrett, M.P. Precision mapping of the metabolome. *Trends Biotechnol.*, **2006**, *24*, 543-548.

[97] Venter, A.; Nefliu, M.; Cooks, R.G. Ambient desorption ionization mass spectrometry. *TrAC*, **2008**, *27*, 284-290.

[98] Chen, H.; Wortmann, A.; Zenobi, R. Neutral desorption sampling coupled to extractive electrospray ionization mass spectrometry for rapid differentiation of bilosamples by metabolomic fingerprinting. *J. Mass Spectrom.*, **2007**, *42*, 1123-1135.

[99] Cody, R.B.; Laramee, J.A.; Durst, H.D. Versatile new ion source for the analysis of materials in open air under ambient conditions. *Anal. Chem.*, **2005**, *77*, 2297-2302.

[100] Lee, S.M.; Kim, H.J.; Jang, Y.P. Chemometric Classification of Morphologically Similar Umbelliferae Medicinal Herbs by DART-TOF-MS Fingerprint. *Phytochem. Anal.*, **2012**, *23*, 508-512.

[101] Hajslova, J.; Cajka, T.; Vaclavik, L. Challenging applications offered by direct analysis in real time (DART) in food-quality and safety analysis. *TrAC*, **2011**, *30*, 204-218.

[102] Watrous, J.; Roach, P.; Alexandrov, T.; Heath, B.S.; Yang, J.Y.; Kersten, R.D.; van der Voort, M.; Pogliano, K.; Gross, H.; Raaijmakers, J.M.; Moore, B.S.; Laskin, J.; Bandeira, N.; Dorrestein, P.C. Mass spectral molecular networking of living microbial colonies. *Proc. Nat. Acad. Sci. U.S.A.*, **2012**, *109*, E1743-E1752.

[103] Lee, Y.J.; Perdian, D.C.; Song, Z.; Yeung, E.S.; Nikolau, B.J. Use of mass spectrometry for imaging metabolites in plants. *Plant J.*, **2012**, *70*, 81-95.

[104] Smeal, J.; Wilkins, C.L. Imaging Mass Spectrometry. *Appl. Spectr. Rev.*, **2011**, *46*, 425-439.

[105] Balluff, B.; Schoene, C.; Hoefler, H.; Walch, A. MALDI imaging mass spectrometry for direct tissue analysis: technological advancements and recent applications. *Histochem. Cell Biol.*, **2011**, *136*, 227-244.

[106] Sarsby, J.; Towers, M.W.; Stain, C.; Cramer, R.; Koroleva, O. A. Mass spectrometry imaging of glucosinolates in Arabidopsis flowers and siliques. *Phytochemistry*, **2012**, *77*, 110-118.

[107] Mohimani, H.; Liu, W.-T.; Mylne, J.S.; Poth, A.G.; Colgrave, M.L.; Tran, D.; Selsted, M.E.; Dorrestein, P.C.; Pevzner, P.A. Cycloquest: Identification of Cyclopeptides via Database Search of Their Mass Spectra against Genome Databases (vol 10, pg 4505, 2010). *J. Proteome Res.*, **2011**, *10*, 5575-5575.

[108] Peterson, D.S. Matrix-free methods for laser desorption/ionization mass spectrometry. *Mass Spectrom. Rev.*, **2007**, *26*, 19-34.

[109] Wilson, R.B.; Hoggard, J.C.; Synovec, R.E. Fast, High Peak Capacity Separations in Gas Chromatography-Time-of-Flight Mass Spectrometry. *Anal. Chem.*, **2012**, *84*, 4167-4173.

[110] Glauser, G.; Guillarme, D.; Grata, E.; Boccard, J.; Thiocone, A.; Carrupt, P.-A.; Veuthey, J.-L.; Rudaz, S.; Wolfender, J.-L. Optimized liquid chromatography-mass spectrometry approach for the isolation of minor stress biomarkers in plant extracts and their identification by capillary nuclear magnetic resonance. *J. Chromatogr. A*, **2008**, *1180*, 90-98.

[111] Wolfender, J.L.; Maillard, M.; Marston, A.; Hostettmann, K. Mass-Spectrometry of Underivatized Naturally-Occurring Glycosides. *Phytochem. Anal.*, **1992**, *3*, 193-214.

[112] Niessen, W.M.A. *Liquid chromatography - mass spectrometry*, 3rd ed.; Taylor & Francis: Boca Raton, **2006**.

[113] Theodoridis, G.A.; Gika, H.G.; Want, E.J.; Wilson, I.D. Liquid chromatography-mass spectrometry based global metabolite profiling: A review. *Anal. Chim. Acta*, **2012**, *711*, 7-16.

[114] Wolfender, J.-L. HPLC in Natural Product Analysis: The Detection Issue. *Planta Med.*, **2009**, *75*, 719-734.

[115] Antonio, C.; Larson, T.; Gilday, A.; Graham, I.; Bergstrom, E.; Thomas-Oates, J. Hydrophilic interaction chromatography/electrospray mass spectrometry analysis of carbohydrate-related metabolites from *Arabidopsis thaliana* leaf tissue. *Rapid Commun. Mass Spectrom.*, **2008**, *22*, 1399-1407.

[116] Russo, R.; Guillarme, D.; Rudaz, S.; Bicchi, C.; Veuthey, J.-L. Evaluation of the coupling between ultra performance liquid chromatography and evaporative light scattering detector for selected phytochemical applications. *J. Sep. Sci.*, **2008**, *31*, 2377-2387.

[117] Coulier, L.; Bas, R.; Jespersen, S.; Verheij, E.; van der Werf, M.J.; Hankemeier, T. Simultaneous quantitative analysis of metabolites using ion-pair liquid chromatography - Electrospray ionization mass spectrometry. *Anal. Chem.*, **2006**, *78*, 6573-6582.

[118] Zhang, T.; Creek, D.J.; Barrett, M.P.; Blackburn, G.; Watson, D.G. Evaluation of Coupling Reversed Phase, Aqueous Normal Phase, and Hydrophilic Interaction Liquid Chromatography with Orbitrap Mass Spectrometry for Metabolomic Studies of Human Urine. *Anal. Chem.*, **2012**, *84*, 1994-2001.

[119] Wang, X.J.; Sun, H.; Zhang, A.H.; Wang, P.; Han, Y. Ultra-performance liquid chromatography coupled to mass spectrometry as a sensitive and powerful technology for metabolomic studies. *J. Sep. Sci.*, **2011**, *34*, 3451-3459.

[120] Guillarme, D.; Nguyen, D.T.T.; Rudaz, S.; Veuthey, J.-L. Method transfer for fast liquid chromatography in pharmaceutical analysis: Application to short columns packed with small particle. Part II: Gradient experiments. *Eur. J. Pharm. Biopharm.*, **2008**, *68*, 430-440.

[121] De Villiers, A.; Lestremau, F.; Szucs, R.; Gelebart, S.; David, F.; Sandra, P. Evaluation of ultra performance liquid chromatography. Par I. Possibilities and limitations. *J. Chromatogr. A*, **2006**, *1127*, 60-69.

[122] Wilson, I.D.; Nicholson, J.K.; Castro-Perez, J.; Granger, J.H.; Johnson, K.A.; Smith, B.W.; Plumb, R.S. High resolution "Ultra performance" liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies. *J. Proteome Res.*, **2005**, *4*, 591-598.

[123] Plumb, R.S.; Rainville, P.D.; Potts, W.B., III; Johnson, K.A.; Gika, E.; Wilson, I.D. Application of Ultra Performance Liquid Chromatography-Mass Spectrometry to Profiling Rat and Dog Bile. *J. Proteome Res.*, **2009**, *8*, 2495-2500.

[124] Grata, E.; Guillarme, D.; Glauser, G.; Boccard, J.; Carrupt, P.-A.; Veuthey, J.-L.; Rudaz, S.; Wolfender, J.-L. Metabolite profiling of plant extracts by ultra-high-pressure liquid chromatography at elevated temperature coupled to time-of-flight mass spectrometry. *J. Chromatogr. A*, **2009**, *1216*, 5660-5668.

[125] Eugster, P.J.; Biass, D.; Guillarme, D.; Favreau, P.; Stöcklin, R.; Wolfender, J.-L. Peak capacity optimisation for high resolution peptide profiling in complex mixtures by liquid chromatography coupled to time-of-flight mass spectrometry: Application to the Conus consors cone snail venom. *J. Chromatogr. A*, **2012**, *1259*, 187-199.

[126] Gohlke, R.S. Time-of-Flight mass spectrometry and gas-liquid partition chromatography. *Anal. Chem.*, **1959**, *31*, 535-541.

[127] Fiehn, O.; Kopka, J.; Trethewey, R.N.; Willmitzer, L. Identification of uncommon plant metabolites based on calculation of elemental compositions using gas chromatography and quadrupole mass spectrometry. *Anal. Chem.*, **2000**, *72*, 3573-3580.

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

[129] Almstetter, M.F.; Oefner, P.J.; Dettmer, K. Comprehensive two-dimensional gas chromatography in metabolomics. *Anal. Bioanal. Chem.*, **2012**, *402*, 1993-2013.

[130] Kind, T.; Wohlgemuth, G.; Lee, D.Y.; Lu, Y.; Palazoglu, M.; Shahbaz, S.; Fiehn, O. FiehnLib: Mass Spectral and Retention Index Libraries for Metabolomics Based on Quadrupole and Time-of-Flight Gas Chromatography/Mass Spectrometry. *Anal. Chem.*, **2009**, *81*, 10038-10048.

[131] Lu, H.; Dunn, W.B.; Shen, H.; Kell, D.B.; Liang, Y. Comparative evaluation of software for deconvolution of metabolomics data based on GC-TOF-MS. *TRAC*, **2008**, *27*, 215-227.

[132] Behrends, V.; Tredwell, G.D.; Bundy, J.G. A software complement to AMDIS for processing GC-MS metabolomic data. *Anal. Biochem.*, **2011**, *415*, 206-208.

[133] Steinbrenner, A.D.; Gomez, S.; Osorio, S.; Fernie, A.R.; Orians, C.M. Herbivore-Induced Changes in Tomato (*Solanum lycopersicum*) Primary Metabolism: A Whole Plant Perspective. *J. Chem. Ecol.*, **2011**, *37*, 1294-1303.

[134] Liland, K.H. Multivariate methods in metabolomics - from pre-processing to dimension reduction and statistical analysis. *TRAC*, **2011**, *30*, 827-841.

[135] Ni, Y.; Su, M.M.; Qiu, Y.P.; Chen, M.J.; Liu, Y.M.; Zhao, A.H.; Jia, W. Metabolic profiling using combined GC-MS and LC-MS provides a systems understanding of aristolochic acid-induced nephrotoxicity in rat. *FEBS Lett.*, **2007**, *581*, 707-711.

[136] Gika, H.G.; Theodoridis, G.A.; Earll, M.; Snyder, R.W.; Sumner, S.J.; Wilson, I.D. Does the Mass Spectrometer Define the Marker? A Comparison of Global Metabolite Profiling Data Generated Simultaneously via UPLC-MS on Two Different Mass Spectrometers. *Anal. Chem.*, **2010**, *82*, 8226-8234.

[137] Crockford, D.J.; Maher, A.D.; Ahmadi, K.R.; Barrett, A.; Plumb, R.S.; Wilson, I.D.; Nicholson, J.K. 1H NMR and UPLC-MSE Statistical Heterospectroscopy: Characterization of Drug Metabolites (Xenometabolome) in Epidemiological Studies. *Anal. Chem.*, **2008**, *80*, 6835-6844.

[138] Kusano, M.; Redestig, H.; Hirai, T.; Oikawa, A.; Matsuda, F.; Fukushima, A.; Arita, M.; Watanabe, S.; Yano, M.; Hiwasa-Tanase, K.; Ezura, H.; Saito, K. Covering Chemical Diversity of Genetically-Modified Tomatoes Using Metabolomics for Objective Substantial Equivalence Assessment. *PLoS ONE*, **2011**, *6*, e16989.

[139] Dai, H.; Xiao, C.; Liu, H.; Tang, H. Combined NMR and LC-MS Analysis Reveals the Metabonomic Changes in *Salvia miltiorrhiza* Bunge Induced by Water Depletion. *J. Proteome Res.*, **2010**, *9*, 1460-1475.

[140] Moco, S.; Forshed, J.; De Vos, R.; Bino, R.; Vervoort, J. Intra- and inter-metabolite correlation spectroscopy of tomato metabolomics data obtained by liquid chromatography-mass spectrometry and nuclear magnetic resonance. *Metabolomics*, **2008**, *4*, 202-215.

[141] Baker, J.M.; Ward, J.L.; Beale, M.H. Combined NMR and flow injection ESI-MS for Brassicaceae metabolomics. *Meth. Mol. Biol.*, **2012**, *860*, 177-191.

[142] Goodacre, R. Making sense of the metabolome using evolutionary computation: seeing the wood with the trees. *J. Exp. Bot.*, **2005**, *56*, 245-254.

[143] Allwood, J.W.; Ellis, D.I.; Goodacre, R. Metabolomic technologies and their application to the study of plants and plant-host interactions. *Physiol. Plant.*, **2008**, *132*, 117-135.

[144] Issaq, H.J.; Van, Q.N.; Waybright, T.J.; Muschik, G.M.; Veenstra, T.D. Analytical and statistical approaches to metabolomics research. *J. Sep. Sci.*, **2009**, *32*, 2183-2199.

[145] Harshman, R.A.; Lundy, M.E. PARAFAC: Parallel factor analysis. *Comput. Stat. Data Anal.*, **1994**, *18*, 39-72.

[146] Geladi, P. Analysis of multi-way (multi-mode) data. *Chemom. Intell. Lab. Syst.*, **1989**, *7*, 11-30.

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

[148] Smith, C.A.; Want, E.J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. XCMS: Processing mass spectrometry data for metabolite profiling using Nonlinear peak alignment, matching, and identification. *Anal. Chem.*, **2006**, *78*, 779-787.

[149] Nordstrom, A.; O'Maille, G.; Qin, C.; Siuzdak, G. Nonlinear data alignment for UPLC-MS and HPLC-MS based metabolomics: Quantitative analysis of endogenous and exogenous metabolites in human serum. *Anal. Chem.*, **2006**, *78*, 3289-3295.

[150] Wang, C.P.; Isenhour, T.L. Time-warping algorithm applied to chromatographic peak matching gas-chromatography fourier-transform infrared mass-spectrometry. *Anal. Chem.*, **1987**, *59*, 649-654.

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

[152] Boccard, J.; Veuthey, J.L.; Rudaz, S. Knowledge discovery in metabolomics: An overview of MS data handling. *J. Sep. Sci.*, **2010**, *33*, 290-304.

[153] Boccard, J.; Grata, E.; Thiocone, A.; Gauvrit, J.Y.; Lanteri, P.; Carrupt, P.A.; Wolfender, J.L.; Rudaz, S. Multivariate data analysis of rapid LC-TOF/MS experiments from *Arabidopsis thaliana* stressed by wounding. *Chemom. Intell. Lab. Syst.*, **2007**, *86*, 189-197.

[154] Grata, E.; Boccard, J.; Glauser, G.; Carrupt, P.-A.; Farmer, E.E.; Wolfender, J.-L.; Rudaz, S. Development of a two-step screening ESI-TOF-MS method for rapid determination of significant stress-induced metabolome modifications in plant leaf extracts: The wound response in *Arabidopsis thaliana* as a case study. *J. Sep. Sci.*, **2007**, *30*, 2268-2278.

[155] Bylund, D.; Danielsson, R.; Malmquist, G.; Markides, K.E. Chromatographic alignment by warping and dynamic programming as a pre-processing tool for PARAFAC modelling of liquid chromatography-mass spectrometry data. *J. Chromatogr. A*, **2002**, *961*, 237-244.

[156] Prince, J.T.; Marcotte, E.M. Chromatographic alignment of ESI-LC-MS proteomics data sets by ordered bijective interpolated warping. *Anal. Chem.*, **2006**, *78*, 6140-6152.

[157] Tomasi, G.; van den Berg, F.; Andersson, C. Correlation optimized warping and dynamic time warping as preprocessing methods for chromatographic data. *J. Chemom.*, **2004**, *18*, 231-241.

[158] Nielsen, N.P.V.; Carstensen, J.M.; Smedsgaard, J. Aligning of single and multiple wavelength chromatographic profiles for chemometric data analysis using correlation optimised warping. *J. Chromatogr. A*, **1998**, *805*, 17-35.

[159] Pravdova, V.; Walczak, B.; Massart, D.L. A comparison of two algorithms for warping of analytical signals. *Anal. Chim. Acta*, **2002**, *456*, 77-92.

[160] Andreev, V.P.; Rejtar, T.; Chen, H.S.; Moskovets, E.V.; Ivanov, A.R.; Karger, B.L.A universal denoising and peak picking algorithm for LC-MS based on matched filtration in the chromatographic time domain. *Anal. Chem.*, **2003**, *75*, 6314-6326.

[161] De Souza, D.P.; Saunders, E.C.; McConville, M.J.; Likic, V.A. Progressive peak clustering in GC-MS metabolomic experiments applied to Leishmania parasites. *Bioinformatics*, **2006**, *22*, 1391-1396.

[162] Castillo, S.; Gopalacharyulu, P.; Yetukuri, L.; Oresic, M. Algorithms and tools for the preprocessing of LC-MS metabolomics data. *Chemom. Intell. Lab. Syst.*, **2011**, *108*, 23-32.

[163] Yu, L.; Liu, H. Efficient feature selection via analysis of relevance and redundancy. *J. Mach. Learn. Res.*, **2004**, *5*, 1205-1224.

[164] Glauser, G.; Boccard, J.; Rudaz, S.; Wolfender, J.-L. Mass Spectrometry-based Metabolomics Oriented by Correlation Analysis for Wound-induced Molecule Discovery: Identification of a Novel Jasmonate Glucoside. *Phytochem. Anal.*, **2010**, *21*, 95-101.

[165] Steuer, R. On the analysis and interpretation of correlations in metabolomic data. *Brief. Bioinf.*, **2006**, *7*, 151-158.

[166] Wang, W.X.; Zhou, H.H.; Lin, H.; Roy, S.; Shaler, T.A.; Hill, L.R.; Norton, S.; Kumar, P.; Anderle, M.; Becker, C.H. Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. *Anal. Chem.*, **2003**, *75*, 4818-4826.

[167] Workman, C.; Jensen, L.; Jarmer, H.; Berka, R.; Gautier, L.; Nielser, H.; Saxild, H.-H.; Nielsen, C.; Brunak, S.; Knudsen, S. A new non-linear normalization method for reducing variability in DNA microarray experiments. *Gen. Biol.*, **2002**, *3*, research0048.0041 - research0048.0016.

[168] Bolstad, B.M.; Irizarry, R.A.; Astrand, M.; Speed, T.P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **2003**, *19*, 185-193.

[169] Gika, H.G.; Theodoridis, G.; Extance, J.; Edge, A.M.; Wilson, I.D. High temperature-ultra performance liquid chromatography-mass spectrometry for the metabonomic analysis of Zucker rat urine. *J. Chromatogr. B*, **2008**, *871*, 279-287.

[170] Gika, H.G.; Macpherson, E.; Theodoridis, G.A.; Wilson, I.D. Evaluation of the repeatability of ultra-performance liquid chromatography-TOF-MS for global metabolic profiling of human urine samples. *J. Chromatogr. B*, **2008**, *871*, 299-305.

[171] van den Berg, R.A.; Hoefsloot, H.C.J.; Westerhuis, J.A.; Smilde, A.K.; van der Werf, M.J. Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics*, **2006**, *7*.

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

[173] Huber, W.; von Heydebreck, A.; Sultmann, H.; Poustka, A.; Vingron, M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, **2002**, *18 Suppl 1*, S96-104.

[174] Kvalheim, O.M.; Brakstad, F.; Liang, Y.Z. Preprocessing of analytical profiles in the presence of homoscedastic or heteroscedastic noise. *Anal. Chem.*, **1994**, *66*, 43-51.

[175] Jonsson, P.; Bruce, S.J.; Moritz, T.; Trygg, J.; Sjostrom, M.; Plumb, R.; Granger, J.; Maibaum, E.; Nicholson, J.K.; Holmes, E.; Antti, H. Extraction, interpretation and validation of information for comparing samples in metabolic LC/MS data sets. *Analyst*, **2005**, *130*, 701-707.

[176] Trygg, J.; Wold, S. Orthogonal projections to latent structures (O-PLS). *J. Chemom.*, **2002**, *16*, 119-128.

[177] Trygg, J. O2-PLS for qualitative and quantitative analysis in multivariate calibration. *J. Chemom.*, **2002**, *16*, 283-293.

[178] Boccard, J.; Kalousis, A.; Hilario, M.; Lanteri, P.; Hanafi, M.; Mazerolles, G.; Wolfender, J.-L.; Carrupt, P.-A.; Rudaz, S. Standard machine learning algorithms applied to UPLC-TOF/MS metabolic fingerprinting for the discovery of wound biomarkers in *Arabidopsis thaliana*. *Chemom. Intell. Lab. Syst.*, **2010**, *104*, 20-27.

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

[180] Lommen, A. MetAlign: Interface-Driven, Versatile Metabolomics Tool for Hyphenated Full-Scan Mass Spectrometry Data Preprocessing. *Anal. Chem.*, **2009**, *81*, 3079-3086.

[181] Duran, A.L.; Yang, J.; Wang, L.J.; Sumner, L.W. Metabolomics spectral formatting, alignment and conversion tools (MSFACTs). *Bioinformatics*, **2003**, *19*, 2283-2293.

[182] Luedemann, A.; Strassburg, K.; Erban, A.; Kopka, J. TagFinder for the quantitative analysis of gas chromatography - mass spectrometry (GC-MS)-based metabolite profiling experiments. *Bioinformatics*, **2008**, *24*, 732-737.

[183] Lei, Z.; Li, H.; Chang, J.; Zhao, P.; Sumner, L. MET-IDEA version 2.06; improved efficiency and additional functions for mass spectrometry-based metabolomics data processing. *Metabolomics*, **2012**, *8*, 105-110.

[184] Baran, R.; Kochi, H.; Saito, N.; Suematsu, M.; Soga, T.; Nishioka, T.; Robert, M.; Tomita, M. MathDAMP: a package for differential analysis of metabolite profiles. *BMC Bioinformatics*, **2006**, *7*.

[185] Bellew, M.; Coram, M.; Fitzgibbon, M.; Igra, M.; Randolph, T.; Wang, P.; May, D.; Eng, J.; Fang, R.; Lin, C.; Chen, J.; Goodlett, D.; Whiteaker, J.; Paulovich, A.; McIntosh, M. A suite of algorithms for the comprehensive analysis of complex protein mixtures using high-resolution LC-MS. *Bioinformatics*, **2006**, *22*, 1902-1909.

[186] Kohlbacher, O.; Reinert, K.; Groepl, C.; Lange, E.; Pfeifer, N.; Schulz-Trieglaff, O.; Sturm, M. TOPP - the OpenMS proteomics pipeline. *Bioinformatics*, **2007**, *23*, E191-E197.

[187] Hiller, K.; Hangebrauk, J.; Jager, C.; Spura, J.; Schreiber, K.; Schomburg, D. MetaboliteDetector: Comprehensive Analysis Tool for Targeted and Nontargeted GC/MS Based Metabolome Analysis. *Anal. Chem.*, **2009**, *81*, 3429-3439.

[188] Izquierdo-Garcia, J.L.; Villa, P.; Kyriazis, A.; del Puerto-Nevado, L.; Perez-Rial, S.; Rodriguez, I.; Hernandez, N.; Ruiz-Cabello, J. Descriptive review of current NMR-based metabolomic data analysis packages. *Prog. Nucl. Magn. Reson. Spectrosc.*, **2011**, *59*, 263-270.

[189] Karp, P.D.; Caspi, R. A survey of metabolic databases emphasizing the MetaCyc family. *Arch. Toxicol.*, **2011**, *85*, 1015-1033.

[190] Roussel, S.; Bellon-Maurel, W.; Roger, J.M.; Grenier, P. Authenticating white grape must variety with classification models based on aroma sensors, FT-IR and UV spectrometry. *J. Food Eng.*, **2003**, *60*, 407-419.

[191] Crockford, D.J.; Holmes, E.; Lindon, J.C.; Plumb, R.S.; Zirah, S.; Bruce, S.J.; Rainville, P.; Stumpf, C.L.; Nicholson, J.K. Statistical heterospectroscopy, an approach to the integrated analysis of NMR and UPLC-MS data sets: Application in metabonomic toxicology studies. *Anal. Chem.*, **2006**, *78*, 363-371.

[192] Jones, O.A.H.; Maguire, M.L.; Griffin, J.L.; Jung, Y.-H.; Shibato, J.; Rakwal, R.; Agrawal, G.K.; Jwa, N.-S. Using metabolic profiling to assess plant-pathogen interactions: an example using rice (*Oryza sativa*) and the blast pathogen *Magnaporthe grisea*. *Eur. J. Plant Pathol.*, **2011**, *129*, 539-554.

[193] Tikunov, Y.M.; de Vos, R.C.H.; Gonzalez Paramas, A.M.; Hall, R.D.; Bovy, A.G. A Role for Differential Glycoconjugation in the Emission of Phenylpropanoid Volatiles from Tomato Fruit Discovered Using a Metabolic Data Fusion Approach. *Plant Physiol.*, **2010**, *152*, 55-70.

[194] Brereton, R.G. Consequences of sample size, variable selection, and model validation and optimisation, for predicting classification ability from analytical data. *TRAC*, **2006**, *25*, 1103-1111.

[195] Van Mechelen, I.; Smilde, A.K. A generic linked-mode decomposition model for data fusion. *Chemom. Intell. Lab. Syst.*, **2010**, *104*, 83-94.

[196] Forshed, J.; Idborg, H.; Jacobsson, S.P. Evaluation of different techniques for data fusion of LC/MS and H-1-NMR. *Chemom. Intell. Lab. Syst.*, **2007**, *85*, 102-109.

[197] Biais, B.; Allwood, J.W.; Deborde, C.; Xu, Y.; Maucourt, M.; Beauvoit, B.; Dunn, W. B.; Jacob, D.; Goodacre, R.; Rolin, D.; Moing, A. H-1 NMR, GC-EI-TOFMS, and Data Set Correlation for Fruit Metabolomics: Application to Spatial Metabolite Analysis in Melon. *Anal. Chem.*, **2009**, *81*, 2884-2894.

[198] Sumner, L.W.; Amberg, A.; Barrett, D.; Beale, M.H.; Beger, R.; Daykin, C.A.; Fan, T.W.M.; Fiehn, O.; Goodacre, R.; Griffin, J.L.; Hankemeier, T.; Hardy, N.; Harnly, J.; Higashi, R.; Kopka, J.; Lane, A.N.; Lindon, J.C.; Marriott, P.; Nicholls, A.W.; Reily, M. D.; Thaden, J.J.; Viant, M.R. Proposed minimum reporting standards for chemical analysis. *Metabolomics*, **2007**, *3*, 211-221.

[199] Fernie, A.R.; Aharoni, A.; Willmitzer, L.; Stitt, M.; Tohge, T.; Kopka, J.; Carroll, A.J.; Saito, K.; Fraser, P.D.; DeLuca, V. Recommendations for Reporting Metabolite Data. *Plant Cell*, **2011**, *23*, 2477-2482.

[200] Verpoorte, R.; Choi, Y.; Kim, H. NMR-based metabolomics at work in phytochemistry. *Phytochem. Rev.*, **2007**, *6*, 3-14.

[201] Fan, T.W.M. Metabolite profiling by one- and two-dimensional NMR analysis of complex mixtures. *Prog. Nucl. Magn. Reson. Spectrosc.*, **1996**, *28*, 161-219.

[202] Hendrawati, O.; Yao, Q.Q.; Kim, H.K.; Linthorst, H.J.M.; Erkelens, C.; Lefeber, A.W. M.; Choi, Y.H.; Verpoorte, R. Metabolic differentiation of Arabidopsis treated with methyl jasmonate using nuclear magnetic resonance spectroscopy. *Plant Sci.*, **2006**, *170*, 1118-1124.

[203] Ali, K.; Maltese, F.; Zyprian, E.; Rex, M.; Choi, Y.H.; Verpoorte, R. NMR Metabolic Fingerprinting Based Identification of Grapevine Metabolites Associated with Downy Mildew Resistance. *J. Agric. Food Chem.*, **2009**, *57*, 9599-9606.

[204] Reily, M.D.; Lindon, J.C. NMR spectroscopy: principles and instrumentation. In *Metabonomics in toxicity assessment*; Robertson, D.G.; Lindon, J.; Nicholson, J.K.; Holmes, E. Eds.; CRC Press: Boca Raton, **2005**, pp 75-104.

[205] Berger, S.; Braun, S. *200 and more NMR experiments*; Wiley-VCH: Weinheim, **2004**.

[206] Robinette, S.L.; Ajredini, R.; Rasheed, H.; Zeinomar, A.; Schroeder, F.C.; Dossey, A.T.; Edison, A.S. Hierarchical Alignment and Full Resolution Pattern Recognition of 2D NMR Spectra: Application to Nematode Chemical Ecology. *Anal. Chem.*, **2011**, *83*, 1649-1657.

[207] Xi, Y.; de Ropp, J.S.; Viant, M.R.; Woodruff, D.L.; Yu, P. Improved identification of metabolites in complex mixtures using HSQC NMR spectroscopy. *Anal. Chim. Acta*, **2008**, *614*, 127-133.

[208] Kim, H.K.; Saifullah; Khan, S.; Wilson, E.G.; Kricun, S.D.P.; Meissner, A.; Goraler, S.; Deelder, A.M.; Choi, Y.H.; Verpoorte, R. Metabolic classification of South American Ilex species by NMR-based metabolomics. *Phytochemistry*, **2010**, *71*, 773-784.

[209] Widarto, H.T.; Van der Meijden, E.; Lefeber, A.W.M.; Erkelens, C.; Kim, H.K.; Choi, Y.H.; Verpoorte, R. Metabolomic differentiation of *Brassica rapa* following herbivory by different insect instars using two-dimensional nuclear magnetic resonance spectroscopy. *J. Chem. Ecol.*, **2006**, *32*, 2417-2428.

[210] Funari, C.S.; Eugster, P.J.; Martel, S.; Carrupt, P.-A.; Wolfender, J.-L.; Silva, D.H.S. High resolution ultra high pressure liquid chromatography–time-of-flight mass spectrometry dereplication strategy for the metabolite profiling of Brazilian Lippia species. *J. Chromatogr. A*, **2012**, *1259*, 167-178.

[211] Kind, T.; Fiehn, O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. *BMC Bioinformatics*, **2007**, *8*.

[212] Pluskal, T.; Castillo, S.; Villar-Briones, A.; Oresic, M. MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *Bmc Bioinformatics*, **2010**, *11*.

[213] Hall, C. Dictionary of Natural Products on CD-ROM; Chapman & Hall, **2012**.

[214] Brown, M.; Wedge, D.C.; Goodacre, R.; Kell, D.B.; Baker, P.N.; Kenny, L.C.; Mamas, M.A.; Neyses, L.; Dunn, W.B. Automated Workflows for Accurate Mass-based Putative Metabolite Identification in LC/MS-derived Metabolomic Datasets. *Bioinformatics*, **2011**, *27*(8), 1108-12

[215] Boroczky, K.; Laatsch, H.; Wagner-Dobler, I.; Stritzke, K.; Schulz, S. Cluster analysis as selection and dereplication tool for the identification of new natural compounds from large sample sets. *Chem. Biol.*, **2006**, *3*, 622-634.

[216] Crockford, D.J.; Maher, A.D.; Ahmadi, K.R.; Barrett, A.; Plumb, R.S.; Wilson, I.D.; Nicholson, J.K. H-1 NMR and UPLC-MSE Statistical Heterospectroscopy: Characterization of drug metabolites (xenometabolome) in epidemiological studies. *Anal. Chem.*, **2008**, *80*, 6835-6844.

[217] Wolfender, J.L.; Queiroz, E.F.; Hostettmann, K. Phytochemistry in the microgram domain - a LC-NMR perspective. *Magn. Reson. Chem.*, **2005**, *43*, 697-709.

[218] Wolfender, J.L.; Marti, G.; Queiroz, E.F. Advances in techniques for profiling crude extracts and for the rapid identification of natural products : dereplication, quality control and metabolomics. *Curr. Org. Chem.*, **2010**, *14*, 1808-1832.

[219] Cui, Q.; Lewis, I.A.; Hegeman, A.D.; Anderson, M.E.; Li, J.; Schulte, C.F.; Westler, W.M.; Eghbalnia, H.R.; Sussman, M.R.; Markley, J.L. Metabolite identification via the Madison Metabolomics Consortium Database. *Nat. Biotechnol.*, **2008**, *26*, 162-164.

[220] Markley, J.L.; Anderson, M.E.; Cui, Q.; Eghbalnia, H.R.; Lewis, I.A.; Hegeman, A.D.; Li, J.; Schulte, C.F.; Sussman, M.R.; Westler, W.M.; Ulrich, E.L.; Zolnai, Z. New bioinformatics resources for metabolomics. *Pacific Symp. Biocomput.*, **2007**, 157-168.

[221] Wishart, D.S.; Tzur, D.; Knox, C.; Eisner, R.; Guo, A.C.; Young, N.; Cheng, D.; Jewell, K.; Arndt, D.; Sawhney, S.; Fung, C.; Nikolai, L.; Lewis, M.; Coutouly, M.-A.; Forsythe, I.; Tang, P.; Shrivastava, S.; Jeroncic, K.; Stothard, P.; Amegbey, G.; Block, D.; Hau, D.D.; Wagner, J.; Miniaci, J.; Clements, M.; Gebremedhin, M.; Guo, N.; Zhang, Y.; Duggan, G.E.; MacInnis, G.D.; Weljie, A.M.; Dowlatabadi, R.; Bamforth, F.; Clive, D.; Greiner, R.; Li, L.; Marrie, T.; Sykes, B.D.; Vogel, H.J.; Querengesser, L. HMDB: the human metabolome database. *Nucleic Acids Res.*, **2007**, *35*, D521-D526.

[222] Akiyama, K.; Chikayama, E.; Yuasa, H.; Shimada, Y.; Tohge, T.; Shinozaki, K.; Hirai, M.Y.; Sakurai, T.; Kikuchi, J.; Saito, K. PRIMe: a Web site that assembles tools for metabolomics and transcriptomics. *In Silico Biol.*, **2008**, *8*, 339-345.

[223] Ferry-Dumazet, H.; Gil, L.; Deborde, C.; Moing, A.; Bernillon, S.; Rolin, D.; Nikolski, M.; de Daruvar, A.; Jacob, D. MeRy-B: a web knowledgebase for the storage, visualization, analysis and annotation of plant NMR metabolomic profiles. *BMC Plant Biology*, **2011**, *11*.

[224] Xia, J.; Bjorndahl, T.C.; Tang, P.; Wishart, D.S. MetaboMiner - semi-automated identification of metabolites from 2D NMR spectra of complex biofluids. *BMC Bioinformatics*, **2008**, *9*.

[225] Alm, E.; Slagbrand, T.; Aberg, K.M.; Wahlstrom, E.; Gustafsson, I.; Lindberg, J. Automated annotation and quantification of metabolites in H-1 NMR data of biological origin. *Anal. Bioanal. Chem.*, **2012**, *403*, 443-455.

[226] Kopka, J.; Schauer, N.; Krueger, S.; Birkemeyer, C.; Usadel, B.; Bergmuller, E.; Dormann, P.; Weckwerth, W.; Gibon, Y.; Stitt, M.; Willmitzer, L.;

Fernie, A.R.; Steinhauser, D. GMD@CSB.DB: the Golm Metabolome Database. *Bioinformatics*, **2005**, *21*, 1635-1638.

[227] Kováts, E. Gas-chromatographische Charakterisierung organischer Verbindungen. Teil 1: Retentionsindices aliphatischer Halogenide, Alkohole, Aldehyde und Ketone. *Helv. Chim. Acta*, **1958**, *41*, 1915-1932.

[228] Stein, S.E.; Babushok, V.I.; Brown, R.L.; Linstrom, P.J. Estimation of Kovats retention indices using group contributions. *J. Chem. Inf. Mod.*, **2007**, *47*, 975-980.

[229] NIST/EPA/NIH Mass Spectral Database. **2011**.

[230] Smith, C.A.; O'Maille, G.; Want, E.J.; Qin, C.; Trauger, S.A.; Brandon, T.R.; Custodio, D.E.; Abagyan, R.; Siuzdak, G. METLIN - A metabolite mass spectral database. *Ther. Drug Monit.*, **2005**, *27*, 747-751.

[231] Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; Oda, Y.; Kakazu, Y.; Kusano, M.; Tohge, T.; Matsuda, F.; Sawada, Y.; Hirai, M. Y.; Nakanishi, H.; Ikeda, K.; Akimoto, N.; Maoka, T.; Takahashi, H.; Ara, T.; Sakurai, N.; Suzuki, H.; Shibata, D.; Neumann, S.; Iida, T.; Tanaka, K.; Funatsu, K.; Matsuura, F.; Soga, T.; Taguchi, R.; Saito, K.; Nishioka, T. MassBank: a public repository for sharing mass spectral data for life sciences. *J. Mass Spectrom.*, **2010**, *45*, 703-714.

[232] Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.*, **2000**, *28*, 27-30.

[233] Caspi, R.; Altman, T.; Dreher, K.; Fulcher, C.A.; Subhraveti, P.; Keseler, I. M.; Kothari, A.; Krummenacker, M.; Latendresse, M.; Mueller, L.A.; Ong, Q.; Paley, S.; Pujar, A.; Shearer, A.G.; Travers, M.; Weerasinghe, D.; Zhang, P.; Karp, P.D. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.*, **2012**, *40*, D742-D753.

[234] Croft, D.; O'Kelly, G.; Wu, G.; Haw, R.; Gillespie, M.; Matthews, L.; Caudy, M.; Garapati, P.; Gopinath, G.; Jassal, B.; Jupe, S.; Kalatskaya, I.; Mahajan, S.; May, B.; Ndegwa, N.; Schmidt, E.; Shamovsky, V.; Yung, C.; Birney, E.; Hermjakob, H.; D'Eustachio, P.; Stein, L. Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res.*, **2011**, *39*, D691-D697.

[235] Zhang, P.; Dreher, K.; Karthikeyan, A.; Chi, A.; Pujar, A.; Caspi, R.; Karp, P.; Kirkup, V.; Latendresse, M.; Lee, C.; Mueller, L.A.; Muller, R.; Rhee, S.Y. Creation of a Genome-Wide Metabolic Pathway Database for Populus trichocarpa Using a New Approach for Reconstruction and Curation of Metabolic Pathways for Plants. *Plant Physiol.*, **2010**, *153*, 1479-1491.

[236] Shinbo, Y.; Nakamura, Y.; taf-Ul-Amin, M.; Asahi, H.; Kurokawa, K.; Arita, M.; Saito, K.; Ohta, D.; Shibata, D.; Kanaya, S. Plant Metabolomics: Biotechnology in Agricultural and Forestry In *Plant Metabolomics*; Saito, K.; Dixon, R.A.; Willmitzer, L. Eds.; Springer Berlin Heidelberg, **2006**, pp 165-181.

[237] Mohn, T.; Plitzko, I.; Hamburger, M. A comprehensive metabolite profiling of *Isatis tinctoria* leaf extracts. *Phytochemistry*, **2009**, *70*, 924-934.

[238] Bottcher, C.; von Roepenack-Lahaye, E.; Schmidt, J.; Schmotz, C.; Neumann, S.; Scheel, D.; Clemens, S. Metabolome analysis of Biosynthetic mutants reveals a diversity of metabolic changes and allows identification of a large number of new compounds in Arabidopsis. *Plant Physiol.*, **2008**, *147*, 2107-2120.

[239] Macel, M.; van Dam, N.M.; Keurentjes, J.J.B. Metabolomics: the chemistry between ecology and genetics. *Mol. Ecol. Res.*, **2010**, *10*, 583-593.

[240] Reynolds, T. The evolution of chemosystematics. *Phytochemistry*, **2007**, *68*, 2887-2895.

[241] Duan, L.-X.; Chen, T.-L.; Li, M.; Chen, M.; Zhou, Y.-Q.; Cui, G.-H.; Zhao, A.-H.; Jia, W.; Huang, L.-Q.; Qi, X. Use of the Metabolomics Approach to Characterize Chinese Medicinal Material Huangqi. *Mol. Plant*, **2012**, *5*, 376-386.

[242] Ruan, C.-J.; da Silva, J.A.T. Metabolomics: Creating new potentials for unraveling the mechanisms in response to salt and drought stress and for the biotechnological improvement of xero-halophytes. *Crit. Rev. Biotechnol.*, **2011**, *31*, 153-169.

[243] Ahuja, I.; de Vos, R.C.H.; Bones, A.M.; Hall, R.D. Plant molecular stress responses face climate change. *Trends Plant Sci.*, **2010**, *15*, 664-674.

[244] Aliferis, K.A.; Jabaji, S. Deciphering plant-pathogen interactions applying metabolomics: principles and applications. *Can. J. Plant Pathol.*, **2012**, *34*, 29-33.

[245] Toffali, K.; Zamboni, A.; Anesi, A.; Stocchero, M.; Pezzotti, M.; Levi, M.; Guzzo, F. Novel aspects of grape berry ripening and post-harvest withering revealed by untargeted LC-ESI-MS metabolomics analysis. *Metabolomics*, **2011**, *7*, 424-436.

[246] Satheeshkumar, N.; Nisha, N.; Sonali, N.; Nirmal, J.; Jain, G.K.; Spandana, V. Analytical Profiling of Bioactive Constituents from Herbal Products, using Metabolomics - A Review. *Nat. Prod. Comm.*, **2012**, *7*, 1111-1115.

[247] Zhang, H.-M.; Li, S.-L.; Zhang, H.; Wang, Y.; Zhao, Z.-L.; Chen, S.-L.; Xu, H.-X. Holistic quality evaluation of commercial white and red ginseng using a UPLC-QTOF-MS/MS-based metabolomics approach. *J. Pharm. Biomed. Anal.*, **2012**, *62*, 258-273.

[248] Urich-Merzenich, G.; Zeitler, H.; Jobst, D.; Panek, D.; Vetter, H.; Wagner, H. Application of the "-omic-" technologies in phytomedicine. *Phytomedicine*, **2007**, *14*, 70-82.

[249] Hall, R.D.; Brouwer, I.D.; Fitzgerald, M.A. Plant metabolomics and its potential application for human nutrition. *Physiol. Plant.*, **2008**, *132*, 162-175.

[250] Cevallos-Cevallos, J.M.; Reyes-De-Corcuera, J.I.; Etxeberria, E.; Danyluk, M.D.; Rodrick, G. E. Metabolomic analysis in food science: a review. *Trends Food Sci. Technol.*, **2009**, *20*, 557-566.

[251] Holmes, E.; Tang, H.R.; Wang, Y.L.; Seger, C. The assessment of plant metabolite profiles by NMR-based methodologies. *Planta Med.*, **2006**, *72*, 771-785.

[252] van der Kooy, F.; Maltese, F.; Hae Choi, Y.; Kyong Kim, H.; Verpoorte, R. Quality Control of Herbal Material and Phytopharmaceuticals with MS and NMR Based Metabolic Fingerprinting. *Planta Med.*, **2009**, *75*, 763-775.

[253] Georgiev, M.I.; Ali, K.; Alipieva, K.; Verpoorte, R.; Choi, Y.H. Metabolic differentiations and classification of *Verbascum* species by NMR-based metabolomics. *Phytochemistry*, **2011**, *72*, 2045-2051.

[254] Gao, W.; Yang, H.; Qi, L.-W.; Liu, E.H.; Ren, M.-T.; Yan, Y.-T.; Chen, J.; Li, P. Unbiased metabolite profiling by liquid chromatography-quadrupole time-of-flight mass spectrometry and multivariate data analysis for herbal authentication: Classification of seven Lonicera species flower buds. *J. Chromatogr. A*, **2012**, *1245*, 109-116.

[255] Farag, M.A.; Porzel, A.; Wessjohann, L.A. Comparative metabolite profiling and fingerprinting of medicinal licorice roots using a multiplex approach of GC-MS, LC-MS and 1D NMR techniques. *Phytochemistry*, **2012**, *76*, 60-72.

[256] Xiang, Z.; Wang, X.Q.; Cai, X.J.; Zeng, S. Metabolomics study on quality control and discrimination of three Curcuma species based on gas chromatograph–mass spectrometry. *Phytochem. Anal.*, **2011**, *22*, 411-418.

[257] Tianniam, S.; Bamba, T.; Fukusaki, E. Pyrolysis GC-MS-based metabolite fingerprinting for quality evaluation of commercial *Angelica acutiloba* roots. *J. Biosci. Bioengineer.*, **2010**, *109*, 89-93.

[258] Farag, M.A.; Wessjohann, L.A. Metabolome Classification of Commercial *Hypericum perforatum* (St. John's Wort) Preparations via UPLC-qTOF-MS and Chemometrics. *Planta Med.*, **2012**, *78*, 488-496.

[259] Frederich, M.; Wauters, J.-N.; Tits, M.; Jason, C.; de Tullio, P.; Van der Heyden, Y.; Fan, G.; Angenot, L. Quality Assessment of *Polygonum cuspidatum* and *Polygonum multiflorum* by (1)H NMR Metabolite Fingerprinting and Profiling Analysis. *Planta Med.*, **2011**, *77*, 81-86.

[260] Kang, J.; Choi, M.-Y.; Kang, S.; Kwon, H.N.; Wen, H.; Lee, C.H.; Park, M.; Wiklund, S.; Kim, H.J.; Kwon, S.W.; Park, S. Application of a H-1 Nuclear Magnetic Resonance (NMR) Metabolomics Approach Combined with Orthogonal Projections to Latent Structure-Discriminant Analysis as an Efficient Tool for Discriminating between Korean and Chinese Herbal Medicines. *J. Agric. Food Chem.*, **2008**, *56*, 11589-11595.

[261] Agnolet, S.; Jaroszewski, J.W.; Verpoorte, R.; Staerk, D. (1)H NMR-based metabolomics combined with HPLC-PDA-MS-SPE-NMR for investigation of standardized Ginkgo biloba preparations. *Metabolomics*, **2010**, *6*, 292-302.

[262] Modarai, M.; Yang, M.; Suter, A.; Kortenkamp, A.; Heinrich, M. Metabolomic profiling of liquid *Echinacea* medicinal products with in vitro inhibitory effects on cytochrome P450 3A4 (CYP3A4). *Planta Med.*, **2010**, 378-385.

[263] Liu, N.Q.; Cao, M.; Frederich, M.; Choi, Y.H.; Verpoorte, R.; van der Kooy, F. Metabolomic investigation of the ethnopharmacological use of Artemisia afra with NMR spectroscopy and multivariate data analysis. *J. Ethnopharmacol.*, **2010**, *128*, 230-235.

[264] Cardoso-Taketa, A.T.; Pereda-Miranda, R.; Choi, Y.H.; Verpoorte, R.; Villarreal, M.L. Metabolic profiling of the Mexican anxiolytic and sedative plant *Golphimia glauca* using nuclear magnetic resonance spectroscopy and multivariate data analysis. *Planta Med.*, **2008**, *74*, 1295-1301.

[265] Yuliana, N.D.; Khatib, A.; Verpoorte, R.; Choi, Y.H. Comprehensive Extraction Method Integrated with NMR Metabolomics: A New Bioactivity Screening Method for Plants, Adenosine A1 Receptor Binding Compounds in *Orthosiphon stamineus* Benth. *Anal. Chem.*, **2011**, *83*, 6902-6906.

[266] Sana, T.R.; Fischer, S.; Wohlgemuth, G.; Katrekar, A.; Jung, K.H.; Ronald, P.C.; Fiehn, O. Metabolomic and transcriptomic analysis of the rice response to the bacterial blight pathogen Xanthomonas oryzae pv. oryzae. *Metabolomics*, **2010**, *6*, 451-465.

[267] Gavaghan, C.L.; Li, J.V.; Hadfield, S.T.; Hole, S.; Nicholson, J.K.; Wilson, I.D.; Howe, P.W.A.; Stanley, P.D.; Holmes, E. Application of NMR-based metabolomics to the investigation of salt stress in maize (Zea mays). *Phytochem. Anal.*, **2011**, *22*, 214-224.

[268] Fumagalli, E.; Baldoni, E.; Abbruscato, P.; Piffanelli, P.; Genga, A.; Lamanna, R.; Consonni, R. NMR Techniques Coupled with Multivariate Statistical Analysis: Tools to Analyse Oryza sativa Metabolic Content under Stress Conditions. *J. Agric. Crop Sci.*, **2009**, *195*, 77-88.

[269] Browne, R.A.; Brindle, K.M. 1H NMR-based metabolite profiling as a potential selection tool for breeding passive resistance against Fusarium head blight (FHB) in wheat. *Mol. Plant Pathol.*, **2007**, *8*, 401-410.

[270] Charlton, A.; Donarski, J.; Harrison, M.; Jones, S.; Godward, J.; Oehlsch-

lager, S.; Arques, J.; Ambrose, M.; Chinoy, C.; Mullineaux, P.; Domoney, C. Responses of the pea (*Pisum sativum* L.) leaf metabolome to drought stress assessed by nuclear magnetic resonance spectroscopy. *Metabolomics*, **2008**, *4*, 312-327.

[271] Widodo; Patterson, J.H.; Newbigin, E.; Tester, M.; Bacic, A.; Roessner, U. Metabolic responses to salt stress of barley (*Hordeum vulgare* L.) cultivars, Sahara and Clipper, which differ in salinity tolerance. *J. Exp. Bot.*, **2009**, *60*, 4089-4103.

[272] Sanchez, D.H.; Pieckenstain, F.L.; Szymanski, J.; Erban, A.; Bromke, M.; Hannah, M.A.; Kraemer, U.; Kopka, J.; Udvardi, M.K. Comparative Functional Genomics of Salt Stress in Related Model and Cultivated Plants Identifies and Overcomes Limitations to Translational Genomics. *PLoS ONE*, **2011**, *6*, e17094.

[273] Dubey, S.; Misra, P.; Dwivedi, S.; Chatterjee, S.; Bag, S.; Mantri, S.; Asif, M.; Rai, A.; Kumar, S.; Shri, M.; Tripathi, P.; Tripathi, R.; Trivedi, P.; Chakrabarty, D.; Tuli, R. Transcriptomic and metabolomic shifts in rice roots in response to Cr (VI) stress. *BMC Genomics*, **2010**, *11*, 648.

[274] Choi, Y.H.; Kim, H.K.; Hazekamp, A.; Erkelens, C.; Lefeber, A.W.M.; Verpoorte, R. Metabolomic differentiation of Cannabis sativa cultivars using H-1 NMR spectroscopy and principal component analysis. *J. Nat. Prod.*, **2004**, *67*, 953-957.

[275] Frederich, M.; Choi, Y.H.; Angenot, L.; Harnischfeger, G.; Lefeber, A.W.M.; Verpoorte, R. Metabolomic analysis of Strychnos nux-vomica, Strychnos icaja and Strychnos ignatii extracts by H-1 nuclear magnetic resonance spectrometry and multivariate analysis tectiniques. *Phytochemistry*, **2004**, *65*, 1993-2001.

[276] Kim, H.K.; Choi, Y.H.; Erkelens, C.; Lefeber, A.W.M.; Verpoorte, R. Metabolic fingerprinting of Ephedra species using H-1-NMR spectroscopy and principal component analysis. *Chem. Pharm. Bull.*, **2005**, *53*, 105-109.

[277] Roos, G.; Roseler, C.; Buter, K.B.; Simmen, U. Classification and correlation of St. John's Wort extracts by nuclear magnetic resonance spectroscopy, multivariate data analysis and pharmacological activity. *Planta Med.*, **2004**, *70*, 771-777.

[278] Yang, S.Y.; Kim, H.K.; Lefeber, A.W.M.; Erkelens, C.; Angelova, N.; Choi, Y.H.; Verpoorte, R. Application of two-dimensional nuclear magnetic resonance spectroscopy to quality control of ginseng commercial products. *Planta Med.*, **2006**, *72*, 364-369.

[279] Queiroz, E.F.; Wolfender, J.L.; Hostettmann, K. Modern Approaches in the Search for New Lead Antiparasitic Compounds from Higher Plants. *Curr. Drug Targets*, **2009**, *10*, 202-211.

[280] Kim, H.K.; Wilson, E.G.; Choi, Y.H.; Verpoorte, R. Metabolomics: A Tool for Anticancer Lead-Finding from Natural Products. *Planta Med.*, **2010**, *76*, 1094-1102.

[281] Urano, K.; Maruyama, K.; Ogata, Y.; Morishita, Y.; Takeda, M.; Sakurai, N.; Suzuki, H.; Saito, K.; Shibata, D.; Kobayashi, M.; Yamaguchi-Shinozaki, K.; Shinozaki, K. Characterization of the ABA-regulated global responses to dehydration in Arabidopsis by metabolomics. *Plant J.*, **2009**, *57*, 1065-1078.

[282] Kim, J.K.; Bamba, T.; Harada, K.; Fukusaki, E.; Kobayashi, A. Time-course metabolic profiling in *Arabidopsis thaliana* cell cultures after salt stress treatment. *J. Exp. Bot.*, **2007**, *58*, 415-424.

[283] Guy, C.; Kaplan, F.; Kopka, J.; Selbig, J.; Hincha, D.K. Metabolomics of temperature stress. *Physiol. Plant.*, **2008**, *132*, 220-235.

[284] Bino, R.J.; de Vos, C.H.R.; Lieberman, M.; Hall, R.D.; Bovy, A.; Jonker, H.H.; Tikunov, Y.; Lommen, A.; Moco, S.; Levin, I. The light-hyperresponsive high pigment-2(dg) mutation of tomato: alterations in the fruit metabolome. *New Phytol.*, **2005**, *166*, 427-438.

[285] Nicholson, J.K.; Lindon, J.C. Systems biology - Metabonomics. *Nature*, **2008**, *455*, 1054-1056.

[286] Richards, S.E.; Dumas, M.-E.; Fonville, J.M.; Ebbels, T.M.D.; Holmes, E.; Nicholson, J.K. Intra- and inter-omic fusion of metabolic profiling data in a systems biology framework. *Chemom. Intell. Lab. Syst.*, **2010**, *104*, 121-131.

Received: October 16, 2012 Revised: November 15, 2012 Accepted: November 15, 2012