Analytical Aspects of Plant Metabolite Profiling Platforms: Current Standings and Future Aims

ARTICLE *in* JOURNAL OF PROTEOME RESEARCH · MARCH 2007

Impact Factor: 4.25 · DOI: 10.1021/pr0604716 · Source: PubMed

2 AUTHORS:

Christoph Seger
University of Innsbruck
101 PUBLICATIONS **1,574** CITATIONS
SEE PROFILE

Sonja Sturm
University of Innsbruck
66 PUBLICATIONS **958** CITATIONS
SEE PROFILE

reviews Journal of proteome research

Analytical Aspects of Plant Metabolite Profiling Platforms: Current Standings and Future Aims

Christoph Seger* and Sonja Sturm

Institute of Pharmacy/Pharmacognosy, Center of Molecular Biosciences, University of Innsbruck, Innrain 52, A-6020 Innsbruck, Austria

Received September 11, 2006

Over the past years, metabolic profiling has been established as a comprehensive systems biology tool. Mass spectrometry or NMR spectroscopy-based technology platforms combined with unsupervised or supervised multivariate statistical methodologies allow a deep insight into the complex metabolite patterns of plant-derived samples. Within this review, we provide a thorough introduction to the analytical hard- and software requirements of metabolic profiling platforms. Methodological limitations are addressed, and the metabolic profiling workflow is exemplified by summarizing recent applications ranging from model systems to more applied topics.

Keywords: metabolic profiling • metabolomics • mass spectrometry • NMR • hyphenated techniques • multivariate statistical methods • plant metabolism • primary metabolites • secondary metabolites • phytopharmaceuticals • genetically modified organisms

1. The Position of Metabolic Profiling within the Systems Biology Framework

Systems biology (Figure 1) can be considered as one of the most important novel scientific concepts in natural sciences.[1−4] On the basis of the tremendous achievements in genomics, RNAomics, and proteomics, for example, several successful genomewide sequencing efforts, access to comprehensive expression level profiles at the transcript levels (e.g., RNA microarray techniques), and introduction of a vast array of 2D gel and chromatography-based methodologies to decipher (i.e., identify and quantify) expressed proteins, a novel understanding of the complexity of interactions within a living organism emerged. Confronted with the vast variety of often contradicting and hardly interconnected bits and pieces of information emerging on different information levels, we are certainly encountering the dawn of a new scientific era.[5,6] It will take major efforts in all branches of natural sciences, including rapidly expanding areas like bioinformatics and biostatistics, to handle these vast amounts of data and to decipher useable regularities and patterns within it.[5−10]

If we consider the fact that most proteins do catalyze chemical reactions, the well-understood information flow, from DNA sequence over messenger RNA to proteins, analyzed by the above-mentioned technologies, usually does not stop at this level but yields small organic molecules (metabolites) of astonishing structural variety. Recently, analytical technologies and technology combinations (platforms) allowing the systematic qualitative and quantitative analysis of these molecules have been increasingly associated with buzzwords like metabonomics, metabolomics, or metabolic profiling. Metabonomics is—as defined by Nicholson et al. in 1999—a whole systems approach reaching for the "quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification".[13] In contrast, metabolomics[14] , a term originally tightly linked to the definition of the "metabolome" (the sum of metabolites produced by an investigated biological system)[15−17], is (although lacking a clear definition in some of the first publications[18]) nowadays best described as an approach which "seeks to identify and quantify the complete set of metabolites in a cell or tissue type and to do so as quickly as possible and without bias".[19] Metabolic profiling is defined as a biased approach, which addresses only a limited number of often structurally

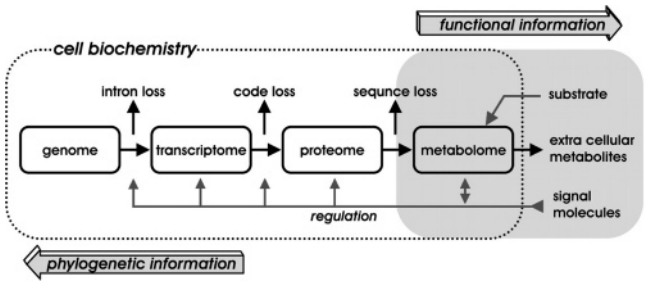

Figure 1. The information flow from an inherited genotype to the observed phenotype is accompanied by an increase in functional information and a decrease in coded (mostly phylogenetic) information, expressing the organism's interaction with its environment. The metabolome is in a dynamic equilibrium with the environment, both through the availability of substrates and the exchange of chemical information molecules interweaving metabolomes of different organisms.[11] Reprinted with permission from ref 12. Copyright 2005, Society for Experimental Botany.

* To whom correspondence should be addressed. Phone: +43 512 504 81155. Fax: +43 512 507 2939. E-mail: christoph.seger@uibk.ac.at.

10.1021/pr0604716 CCC: $37.00
Published on Web 12/02/2006

closely related analytes due to methodological limitations of the analytical platform used.[20,21] Metabolic fingerprinting, the term metabolomics originally derived from,[14,18] is presently understood as "high-throughput, rapid, global analysis of samples to provide sample classification".[21] Analyte identification and quantification is said not to be used in this context.

In an alternative approach to classify the multitude of approaches encountered in the field of comprehensive small molecule analysis, Nielsen and co-workers completely avoided the -omics definitions given above and divided the field of metabolite analysis into target analysis ("identification and quantitative analysis of predefined metabolites") and metabolic profiling ("scanning of all metabolites detectable by a selected analytical technique") in a recent overview.[22] The latter term was subdivided into fingerprinting, dealing with "intercellular metabolites", and footprinting, dealing with "extracellular metabolites". Unfortunately, these definitions may be too closely related to microbiological laboratory practice and not very useful for other, more bioanalytically orientated science branches. Hence, although also highly artificial and not completely satisfying, the differentiation of the current analytical efforts into metabolic profiling dedicated to a deeper understanding of the molecular biochemistry of the investigated systems and metabolic fingerprinting of organisms characterizing differences or similarities of investigated entities (without *a priori* knowledge of the molecular cause) seems to be more appropriate.

Currently, none of the analytical platforms used, neither NMR nor any of the mass spectrometry-based approaches, have matured enough to meet the prestigious aim of comprehensive metabolite analysis as requested for an "-omics" technique. In addition to limitations introduced by analyte preselection due to the chromatographic assay chosen (e.g., the need of analyte volatility in GC), the most severe bottleneck is still the challenging process to structurally characterize analytes found to discriminate samples, the putative biomarkers. In this context, Glinski and Weckwerth stated that "a major drawback of metabolomic technologies yet to be overcome is the vast number of unknown compound structures. Here, HPLC−MS techniques using MS^n, high-accuracy mass spectrometers like FTICR−MS, offline NMR, and coupling of HPLC/NMR are required for structure elucidation and identification of complex metabolite structures".[23] Nielsen and Oliver did point out that combination of metabolic profiling data with other systems biology data sets as transcriptomics profiles is hardy possible, if analytes cannot be annotated. They state that this situation leaves only two strategies to follow, either to restrain the quantitative data exploitation to metabolites known by external referencing with the possibility to miss the discriminating analytes or to identify analytes having a key role in sample discrimination first hand (e.g., by a biomarker discovery approach) and focusing on their unequivocally characterization.[24] In a critical overview of the current challenges in the field of GC−MS-based metabolite profiling, Joachim Kopka did stress that "increasing the number of metabolite identifications within existing profiling platforms is prerequisite for a substantial improved scope of profiling studies. Clear, reproducible strategies for metabolite identification and exchange of identification between laboratories will facilitate further developments".[25] It has to be strongly emphasized that mass spectrometry alone, even if the named high-resolution technologies are employed, can hardly provide more than the molecular formula at its best.[26,27] Electron ionization and secondary fragmentation mass spectra may allow a deeper insight into the molecular structure of the analyte under investigation, but clearly fail to determine its molecular scaffold unequivocally (not to speak of its 3D structure).[28] *De novo* structure elucidation of secondary natural products in solution can currently only be achieved by the extensive use of 1D and 2D homo- and heteronuclear NMR measurements[29] assisted by spectroscopical methods (e.g., IR, UV) and high-resolution mass spectrometry. In addition, when starting from complex mixtures as encountered within the discussed applications, analyte separation prior to NMR experiments is strongly desired to simplify the process of analyte characterization. Thus, combining HPLC-based separations with NMR spectroscopy in an on-line or at-line mode, as realized in HPLC−NMR,[30,31] HPLC−SPE−NMR,[32,33] and capNMR,[34] is a most promising analytical approach to aid metabolic profiling.

Proton NMR spectroscopy-based profiling approaches are, compared to mass spectrometry-based assays, limited by low analytical resolution and relative insensitivity.[19,35−37] Whereas resolution problems can be overcome by the application of 2D NMR methods, especially by the use of the *J*−RES experiment,[38] the lack of analytical sensitivity makes NMR spectroscopy and mass spectrometry platforms complementary on the analyte concentration scale. Mass spectrometry is considered to be a trace analysis methodology, capable to detect and quantify low molecular weight metabolites at concentrations well below the nanogram per milliliter (ng/mL) range. However, this only holds true for thoroughly optimized MS conditions which can be hardly realized whenever structurally diverse analytes are encountered, as is the case in the holistic approaches discussed in this review. Hence, the sensitivity of mass spectrometry platforms is shifted at least 2 to 3 orders of magnitudes to higher analyte concentrations. Ion suppression effects are another well-known problem in mass spectrometry of biological materials.[39−41] This complex phenomenon leading to a decreased ion yield strongly depends on the sample workup strategy used and is not necessarily constant over a chromatographic run. If comparing samples of variable origin, that is, different plant species, organs, or development stages, care must be taken that analyte responses are not unpredictably influenced by this effect. A potential approach to circumvent this analytical pitfall is to perform spiking experiments to evaluate the recovery rate for each analyte (or at least analyte class).

2. Plants: Demanding but Rewarding Research Objects

Plants and products derived thereof are the source of a multitude of foods including modern day's functional foods and nutriceuticals. Furthermore, they also serve mankind as medicines, fragrances, or poisons since primeval times.[42,43] Intended cultivation of wild plants by domestication and selection of wanted traits is as old as human civilization. Plant breeding, either performed by conventional interbreeding or by modern techniques as chemical mutagenesis, wide cross techniques, or modern day generation of genetically modified organisms (GMOs), led to countless novel cultivars with highly improved traits. Breeders successfully increased crop yield and qualities, selected more vigorous strains sustaining environmental stress as droughts or soil salinity, or introduced resistance against different plant pests including viruses, bacteria, fungi, or insects. However, generating plant hybrids bears the risk of introducing unintended alterations in gene expression profiles with the likely consequence of qualitative

and quantitative changes in metabolite patterns. This may lead to the loss of desired traits, for example, a decrease of the nutritional value of a crop or an altered composition of volatile constituents to be utilized as fragrance. Additionally, the reintroduction of unwanted or even toxicological relevant metabolite groups (i.e., *Solanum* alkaloids in potatoes or cucurbitacin derivatives in melons and pumpkins) cannot be ruled out. Hence, monitoring levels of metabolites in the course of plant breeding experiments is essential and has been well-performed during the last decades. As it will be shown in detail, metabolic profiling efforts can take the aim of controlling intended and unintended changes in the breeding progress by addressing hundreds of analytes in parallel and studying their shifting patterns in detail.

Plant breeding is only one aspect of the current still successful bioprospecting efforts.[44] Exploiting natural resources usually starts with a discovering or rediscovering process of a desired trait, followed by an optimization phase (here plant domestication and plant breeding takes place) leading to an optimized cultivar, whose properties have to be documented and maintained. In addition, whenever selecting a natural resource (i.e., a plant, a fungus, a marine organism, etc,) as source for a desired product, (e.g., a pharmacological potent extract or purified substance), the quality of the starting material has to be monitored. When authentication and quality control of biological materials used as foods, herbal medicinal products, or nutraceuticals are performed, observed changes of metabolite profiles are often found to be associated with deliberate sample adulteration or microbial spoilage.[45] In this context, the reliable discrimination between properly treated materials and contaminated samples is mandatory. Classical analytical techniques as HPTLC, HPLC, CE, and GC coupled with different detector systems such as FID, DAD, or MS have been used for decades to cover these purposes.[46,47] However, these methods often show limited separation and/or detection capabilities owing to the physicochemical properties of the analytes under investigation (e.g., poor volatility, poor ionizability, lack of chromophores, etc.). Hence, the derived metabolic profiles are of limited extend, and analyte signatures do strongly depend on the analytical equipment used.[22] Additionally, for most analytical platforms, confirmative qualitative measurements are needed for analyte identification due to the low information content of detection methods used in screening assays. A notable exception (also in terms of separation capabilities) is GC−MS, with retention time index formation (e.g., the Kovats index[48]) and database comparison supporting analyte identification. Combining this analytical platform with statistical pattern classification procedures has been used for a broad variety of applications. The use of screening techniques unbiased in both analytical and statistical terms will take these analytical efforts to a higher level of quality, since unintended changes in the multivariate metabolite space are more likely to be identified. Environmental adaptation of organisms, one of the major optimization goals in breeding efforts, is an important aspect in ecological research. This includes studies on the influence of abiotic and biotic environmental stressors on the vitality of the organism under investigation. In this regard, metabolomics or metabolic profiling experiments as unbiased holistic approaches will unravel the metabolic effects of complex interactions scenarios as presented in "real world" (i.e., ecological) situations. Research topics to be envisioned include reinvestigations of ecological "hot" topics as the interaction of herbivores, parasites, and pathogens with plants[49,50] as well as novel emerging worldwide ecological problems as the aggressiveness of invasive plant species and their obvious successful adaptation to new ecological niches.[51−54]

Analytes encountered in plant metabolic profiling and fingerprinting can be subdivided into primary and secondary metabolites based on their biogenetical source and their physiological role. Primary metabolites arise from highly conserved universal metabolic pathways which can be found throughout nature, for example, glycolysis, the Krebs cycle, or the pentose phosphate pathway. They show rather limited structural variety and belong to a limited set of substance classes, for example, carbohydrates, amino acids, organic acids, and fatty acid derivatives. The spanned metabolic network is indispensable for sustaining vital functions as growth and development. Studying the generation and interconversion of these metabolites is covered by flux analysis. This has been a major topic in plant biochemistry over the last decades and has become an integrative part of systems biology. Currently, the use of stable isotope-enriched analytes, for example, partially ^{13}C labeled glucose, is the state of the art approach in this context. Mass spectrometry and NMR spectroscopy including their hyphenations to different chromatographic devices are utilized as major analytical techniques in this field.[19,55−58] In contrast, secondary metabolites are stemming from biogenetical pathways not involved in primary life supporting functions.[42,49,59] Major substance classes are terpenoids, alkaloids, polyphenols, and polyketides.[60] An astonishing structural diversity is encountered within these very general molecular scaffold types, and even structural hybrids combining unrelated biogenetical pathways are frequently encountered. It can be estimated that hundreds of thousands of secondary metabolites exist within the Plant Kingdom, and many of them have been assigned to certain bioactivities.[44,61] The found metabolite groups can be either confined to certain taxa,[59] supporting chemotaxonomical research efforts, or are widespread over unrelated species.[62] Within a single species, accumulations of several dozen of chemically highly diverse secondary metabolites are commonly encountered, especially if essential oils and volatiles are involved.[63]

3. Analytical Platforms Utilized for Metabolic Profiling

3.1. Sample Preparation. Samples encountered in plant biology can stem from very different sources, presenting a high diversity of matrices ranging from tender leaves and sturdy lignified roots to fat-rich fruits and carbohydrate-based tubers. Hence, sampling strategies and work-up procedures have to be adapted on a case-by-case basis. Preanalytical sample treatment in metabolic profiling does not differ from strategies already evolved and optimized over the last decades in analytical plant chemistry (Figure 2).

Generally, it starts with the correct specimen collection and ends with the transfer of the analyte cocktail in a preparation form suitable for introduction into the analytical platform chosen (e.g., a deuterated solvent for NMR spectroscopy or a volatile solvent for GC−MS). It has to be carried out with utmost care; artifact introduction must not occur. Unwanted matrix components have to be removed, and sample adulteration (e.g., glycoside hydrolysis promoted by pH shifts) has to be excluded or at least minimized and monitored. Usually, sample workup includes one or several extraction steps using a broad variability of solvent systems or solid phases. In mass spectrometry-based approaches, different derivatization techniques are employed to increase analyte volatility (in GC−MS)

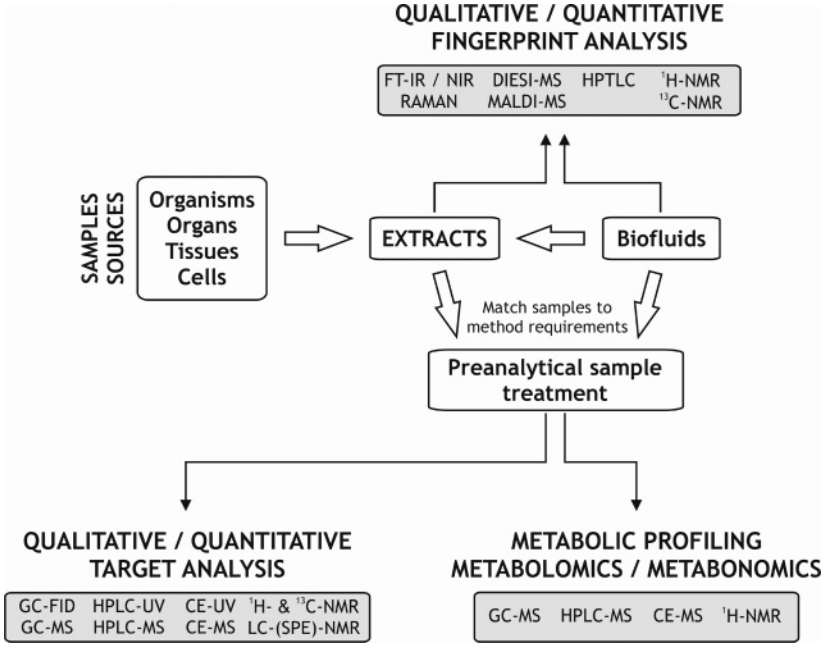

Figure 2. Flowchart summarizing preanalytical sample workup and analytical technologies utilized in metabolome analysis.

or ionization efficiency (in HPLC−MS). Care must be taken to keep the physicochemical characteristics of the final samples (e.g., pH, ion strength, solvent composition) constant over the whole sample set. Current efforts to unify sampling strategies will have to be limited to specific plant organs/plant parts or to limited sample sources, for example, cultivars/interbreeds of a species or genus. For further details of currently applied strategies, the readership is referred to recently published reviews.[22,64]

3.2. Mass Spectrometry. Whereas in NMR spectroscopy the predominance of one-dimensional (1D) liquid state ^{1}H NMR as detection tool for metabolite profiles is evident (see below), a different situation is presented in mass spectrometry. Within the past decades, different methodologies for ion formation and ion selection of analytes have been developed. As a consequence, a variety of strategies to combine analyte separation techniques with mass spectrometry have evolved. Next, we will give a short overview on the analytical platforms which are currently mostly used in the field of metabolite profiling, where a connection to chromatographic devices, fast scanning rates, and high mass resolution is desired, if not essential. If further information is needed, we refer to several comprehensive reviews allowing a deeper insight into mass spectrometry applications in bioanalytics.[23,65−71]

3.2.1. Analyte Ionization. As ions are actually analyzed in the vacuum of the mass spectrometer, the most important reaction in the MS might be the one that converts analytes of interest into gas−phase ions. The oldest, most used and probably best understood ionization technique is electron ionization (EI), formerly better known as electron impact ionization, in which the analyte vapor is subjected to a bombardment by energetic electrons (typically 70 eV). EI is performed in a high-vacuum source avoiding intermolecular collisions. Hence, spectra are highly reproducible, and as a consequence, several comprehensive databases are available. Since in EI the analytes have to be presented as vapors, it is the ionization of choice for GC−MS analysis. As an alternative approach in GC−MS, a reagent gas can be applied leading to chemical ionization (CI), which results in spectra showing less fragmentation but a clear molecule ion, which is not always the case in EI. Both techniques can be combined with any mass analyzer. Nowadays, quadrupole (Q) or quadrupole-time-of-flight (Q-TOF) mass spectrometers are mostly used in combination with gas chromatographs.[65,69]

Over the past decades, a number of ionization techniques have been developed for analyses of nonvolatile and thermally labile compounds with electrospray ionization (ESI) and matrix-assisted laser desorption ionization (MALDI) emerging as prime methods of choice. In ESI, ions are generated from a liquid at atmospheric pressure. The electrospray setup is technically simple, a hollow needle through which the eluent flows (usually 1−1000 μL/min) is charged with a potential. The high field at the tip of the needle produces a cone-shaped liquid meniscus from which a spray of highly charged droplets emerges. Subsequent evaporation of the droplets results in ion formation. ESI has some very impressive attributes allowing its application to a wide variety of biological problems. Almost all ions in solution are amenable to analyses with ESI. The production of multiply charged species extends the dynamic range of the mass spectrometer to the range of 10^{6} u. Furthermore, it can be considered as a very "soft" ionization technique which does not destroy noncovalently bound biomolecular complexes. ESI is probably the most utilized ionization source combined with any sort of analyzer for HPLC−MS and CE−MS analysis.[65,69] A considerable shortcoming of ESI, particularly in direct infusion experiments (DIESI), is its susceptibility to matrix and ion suppression effects, for example, by high salt concentrations in biological samples or by high concentrated analytes which might hinder the ionization of depleted analytes. A common alternative to ESI, often combined within one source housing, is atmospheric pressure chemical ionization (APCI). The ionization is initiated by a corona discharge from a needle, producing a fine plasma of reagent ions stemming from volatiles in the source (auxiliary gases and vaporized solvents). A complex reaction cascade leads to a most efficient ionization of analytes. APCI is known for its sensitivity, robust-

ness, ruggedness and reliability, and much less susceptibility to chemical interferences than ESI.[65] A further emerging alternative to ESI is atmospheric pressure photo ionization (APPI), based on absorption of a high-energy photon, which is produced in a discharge lamp (usually Krypton lamps). Compounds with an ionization potential below the energy of the lamp, can be directly ionized. For all others, an easily ionisable dopant reagent (e.g., toluene) is added, which facilitates the charge transfer to the analyte. Frequently used carrier solvents are water, methanol, and acetonitrile with ionization potentials <10 eV, thus, remaining unaffected and causing no background signals.[70,72] Although all atmospheric pressure ionization techniques, that is, ESI, APCI, and APPI have impressive attributes, one has always to bear in mind that all three of them show compound class specific sensitivities; very low ion yields cannot be excluded in certain cases.

A distinct different concept of introducing analytes into the mass analyzer forms the basis of MALDI. The analyte is cocrystallized with a matrix, and ions are formed by laser irradiations. An outstanding feature of MALDI is the pulsed nature of the laser producing ions packages in discrete events of high sensitivity. Because of the short duration of these pulses, the number of utilizable mass analyzers is limited, as high scanning rates are required. MALDI is the method of choice for high-throughout analysis, since target plates loaded with hundreds of samples can be used. Its relatively high tolerance to salts and buffers makes it the method of choice for analyzing complex biogenic matrices. The major drawback of MALDI lies in the nature of the utilized matrices, especially if low-molecular weight compounds have to be analyzed. In the case of one of the currently most popular applications of MALDI, the analysis of tryptic digests of proteins in proteomics, a high degree of chemical noise at low masses has often been observed. Various attempts have been undertaken to overcome this problem, for example, by utilizing matrices consisting of high-molecular weight compounds or by the derivatization of analytes. Furthermore, much progress has been made in the development of novel sample supports (SELDI)[73] and desorption/ionization techniques from silica supports (DIOS).[74] However, the complete and reproducible MALDI of a metabolic profiling sample still remains a challenge.[75,76]

3.2.2. Mass Analyzers: Low Resolution Instruments. Transmission quadrupoles are used as mass filters by applying precisely controlled voltages to opposing sets of poles. By ramping the voltages on each side of poles, a complete range of masses can be passed to the detector in a time-dependent manner. Quadrupoles were the most common analyzers in the past, particularly in combination with an ion source which produces structurally significant fragments as EI in GC−MS platforms. Although quadrupoles can operate up to masses >4000 u, they are best suited for applications <1000 u. One notable disadvantage of a quadrupole mass analyzer is the relatively slow scanning rate (<4000 u/s). Thus, in full-scan mode, only a small fraction of available ions is used, a fact which makes them not suitable for modern high-speed separations with appreciably smaller peaks widths. They also cannot be used as final mass selector in combination with MALDI. However, the situation appears totally different, when three quadrupoles are arranged such that a central nonresolving quadrupole is flanked by two transmission quadrupoles forming a tandem mass spectrometer, also known as triple quadrupole (QqQ) mass spectrometer. This setup allows controlled ion fragmentation in the second quadrupole which is used as collision cell. The obtained secondary ion distribution serves as a set of structural parameters in analyte identification. Furthermore, the QqQ setup allows combining ion selection, ion reaction, and ion scan processes to novel experiments, for example, the highly specific product ion, precursor ion, or neutral loss scan modes, which are already indispensable tools in proteomics, glycomics, and lipidomics and will certainly find their way to routine applications in metabolic profiling (e.g., in the target analysis of amino acids).[77] The multiple reaction monitoring (MRM), a targeted analysis experiment of highest sensitivity and accuracy is the methodological backbone of the application of mass spectrometry in clinical chemistry[68] and forensics, as well as in the pharmaceutical industry.[65,69]

Paul ion traps are variations of the general quadrupole mass filter with performance parameters (resolving power, mass accuracy, sensitivity, and mass range) comparable to transmission quadrupoles. Ions are confined ("trapped") in these devices by electrodynamic focusing. In contrast to a quadrupole, where ions have a stable trajectory on their way to the detector, trapped ions are forced to leave the trap by putting them in unstable orbits by increasing the rf-voltage. The capability of ion traps to perform MS^n (theoretically up to $n = 10$), their ease in use, and their relatively low cost has made them into rugged workhorses in many laboratories.[65,69] Notably, increased quantitative performance and dynamic range have been gained by the development of linear ion traps, two-dimensional Paul ion traps confining ions in the axial dimension by means of an electric field at the ends of the trap. Here, benefits are mainly associated with the optimized volume of the mass analyzer.[67,71]

3.2. Mass Analyzers: High-Resolution Instruments. TOF mass analyzers, in principle one of the simplest devices in mass spectrometry, separate ions based on their velocity, which is mass-dependent. All ions are formed at once in the ion source and subsequently accelerated through a fixed potential. Hence, small mass ions arrive to the detector earlier relative to large mass ions. A common variation of TOF incorporates an electrostatic mirror, a reflectron, into the field-free region to compensate for effects of kinetic energy distribution of the initial ions. The TOF analyzer gained great popularity in the early days of MS, but, as it needs a pulse of ions, it was not compatible with the available ionization methods. The development of new pulsed-ionization techniques, such as MALDI, led to a TOF detector renaissance in the 1990s. Currently, it is mainly used as a "readout" mass analyzer, combined with either another TOF, an ion trap, or, most frequently, with one or two quadrupoles (Q-TOF or Qq-TOF) preceding it. These hybrid mass spectrometers stand out through an unlimited mass range, exceptional high scan rates up to 10^6 u/s, and high resolving power resulting in a mass accuracy in the parts per million (ppm) range. Therefore, they are the technique of choice for applications demanding high-throughput at utmost efficiency as it can be expected to be the case for large-scale routine metabolic profiling.[65,69,71,78]

A different type of mass analyzer which provides unequalled resolving power (up to 10^6 fwhm) is the Fourier transform ion cyclotron mass spectrometer (FT−ICR MS). Like the sector field analyzer, the FT−ICR MS uses a magnetic field to determine the m/z of an ion. The instrument consists of a cell in a vacuum chamber ($<10^{-6}$ Torr) centered in a high magnetic field. Ions are trapped in the cell by a combination of the magnetic field and electric potentials. In the presence of the magnetic field, the ions will take circular trajectories with a frequency of

rotation inversely proportional to mass. Rf pulse triggered ion passage over detector plates results in a time domain signal, which is converted to frequency domain by Fourier transformation. This signal is subsequently transformed to a mass spectrum on the basis of the inverse relationship of frequency and *m/z*. Combination of FT−ICR MS with an ionization device like ESI/APCI requires ion sources outside of the magnetic field. Although the capabilities of FT−ICR MS are undeniable, one has to be aware of size, cost, and efforts to upkeep an instrument based on a high-powered multi Tesla magnet.[65,69,71] In 2005, an interesting alternative to the FT−ICR was introduced to the market, the orbitrap. This new mass analyzer technology in the size of a walnut works on the principle of injecting ions for storage into an electrostatic field. The orbitrap is the last mass analyzer in a complex hybrid instrument: a linear ion trap is used to selectively fill an intermediate ion storage device named C-trap with ions of interest which are subsequently injected into the orbitrap. This instrument uses image current detection and Fourier transform similar to the FT−ICR MS but does not require a superconducting magnet which significantly reduces initial costs and maintenance requirements. It shows a reported nominal mass resolving power of at least 30 000 fwhm in the mass range of typical metabolomics applications and a 5 ppm mass accuracy in a dynamic range of >5000.[79−81]

3.2.4. Combination with Chromatographic Separation Systems. Choosing an appropriate MS platform remains—despite the variety of possible combinations of techniques and new promising developments—challenging. Neither the on-line combination of separation techniques with MS nor direct-infusion MS can cover the analysis of the complete metabolome of an organism. In theory, the optimal way to obtain a mass spectrum of all metabolites in a sample in an unbiased way would be the direct application into the MS instrument without any purification or separation, usually via ESI (DIESI) or by MALDI and related techniques as SELDI and DIOS. When high-performance mass analyzers such as FT−ICR, orbitarp, or TOF are used, a screening of metabolites can be performed within a few seconds, providing accurate masses and some structural information of analytes as well as metabolite quantities, if response factors are known by appropriate method validation. However, if working with complex biogenic matrices, these approaches are associated not only with demanding instrument maintenance tasks (e.g., routine cleaning of ion source and transfer optics), but also with severe drawbacks due to unspecific matrix effects and ion suppression by competitive ionization. Time-consuming liquid/liquid extraction procedures or alternative sample preparation strategies have to be applied to reduce these effects. A second limitation of direct infusion MS approaches is the incapability to distinguish between chromatographically separable isobaric structural isomers, making unambiguous discrimination of many metabolites impossible.[22]

GC−EI/CI−MS can be considered as the most mature platform as underlined by its extensive use during the past decades.[25,82] GC, in combination with fast high resolution analyzers as TOF, stands out for its extreme chromatographic resolution combined with highly reproducible spectra. The desired identification of analytes is eased by numerous comprehensive databases made available either as joint efforts of the scientific community[83,84] (e.g., the GMD@CSB.DB database at Golm[85] and Oliver Fiehn's database[86]), or by some commercial sources. GC−MS is the method of choice for low-molecular weight and volatile compounds. Nonvolatile components have to be derivatized, which makes extensive time-consuming sample preparation steps necessary. Furthermore, the structural identification of unknown derivatized anaytes is complicated, and thermolabile metabolites cannot be analyzed at all. An unquestionable benefit of GC analysis is the possibility to correct retention time shifts easily by calculating retention time indices (Kovats index) or even by reanalyzing samples on a second stationary phase resulting in a second confirmative index.

The combination of HPLC and API−MS allows the analysis of a wide variety of compounds including the above-mentioned thermolabile and high-molecular weight analytes which cannot be addressed by GC−MS.[22,23] A broad range of stationary phases, both in terms of column dimensions and surface modifications, are commercially available, allowing the optimization of the chromatographic separation to the matrix under investigation. The latest developments in this field include the recent introduction of "ultraperformance" hardware, which leads to a significant reduction in analysis time and the use of hydrophilic interaction chromatography (HILIC) columns, addressing the analytical challenges of highly polar metabolites, often encountered in metabolic profiling.[87,88] Coupling liquid chromatography with mass spectrometry limits HPLC solvents to volatile buffers and organic additives to avoid salt precipitation in the mass spectrometer. Hence, sodium/potassium phosphate buffers, frequently used as eluents in HPLC−UV, had to be replaced by ammonium formiate or acetate systems. As ESI, the most prominent API source, is a "soft" ionization technique producing mainly spectra with hardly any fragmentation, multistage analyzers as QqQ, Qq-TOF, or ion-traps have to be used to obtain fragment ions aiding analyte identification by comparison with reference material. Complications in the analysis of complex "real world" samples might arise from chromatographic interferences, matrix effects, adduct and cluster formation, and ion suppression by competing metabolites. Hence, thorough sample preparation is most crucial for gathering sensitive and reliable results. Even then, the comparability of ESI−MS spectra recorded with different mass spectrometers (not necessarily from different vendors) is limited. Therefore, the generation of ESI−MS-based metabolite databases was not pursued until recently, when Moco et al. did present the MoToDB dedicated to the metabolome of *Solanum lycopersicum*.[89]

A complementary promising combination of a highly efficient separation technique and MS as detector is CE−MS.[90−93] Capillary electrophoresis (CE) offers a number of methodological benefits such as high-resolution capability due to the flat flow profile, low sample and solvent consumption, and inexpensive replacement of capillaries. A wide range of analytes can be analyzed by CE, from polar low-molecular weight compounds up to proteins or even bacteria and viruses. The implementation of CE−MS was not straight forward, since the mandatory closed electrical circuit required proper grounding of the hardware. Currently, mainly two types of ESI CE−MS interfaces are employed, the more robust coaxial sheath-flow interface and the liquid-junction interface. Both require an additional sheath flow, although sheathless interfaces have been already realized in nanospray devices. The adaptation of APCI and APPI interfaces for CE coupling are currently in the state of commercial development. Utilized mass analyzers for CE−MS cover the full range of possible instrumentations, from single quadrupoles to high-performing FT−ICR instruments.

Considering the complexity of samples in metabolic studies, the choice of fast analyzers seems to be advantageous. Despite the advantages of CE, particularly for the analysis of charged molecules as organic acids and organic bases, development of a robust CE−MS method is still challenging, since CE principally differs in theory and in practice from HPLC and GC. Thus, it is absolutely necessary to "think different, to think CE" as Schmitt-Kopplin and Frommberger put it in their excellent and comprehensive review on CE−MS.[91]

Within the past decades, almost all available mass spectrometry platforms have been applied to target analysis and, more recently, metabolic profiling of plant-derived sample materials. Mass spectrometry-based metabolic profiling investigations have been mostly restricted to model plant systems such as *Arabidopsis thaliana*, *Solanum tuberosum*, and *S. lycopersicum*, which is a noticeable contrast to NMR-based approaches. Selected examples will be discussed further below.

3.3. Nuclear Magnetic Resonance Spectroscopy. Since NMR-based metabolic profiling has been recently reviewed in a comprehensive manner (coauthored by one of us),[64] it is only summarized briefly in this contribution. NMR spectroscopy has a long tradition in natural product chemistry, as it is the major analytical technique used to structurally characterize small organic compounds. The development of modern day NMR hard- and software has been tightly connected to the growing needs of analytical chemists unraveling the molecular signatures of life, including 3D structures of macromolecules as proteins, DNA, and RNA. In addition to the use of 1D and 2D NMR methods in the field of natural product analysis, NMR proved to be useful in the direct quantitative assessment of crude plant extracts (qNMR)[36] if well-separated analyte signals are present.[94−97] This is a methodological advantage over mass spectrometry, where varying ionization efficiency and unpredictable ion suppression effects do complicate analyte quantification. SNIF NMR, analyzing the naturally occurring distribution of NMR detectable stable isotopes in ethanol (^{2}H, ^{13}C), had been successfully introduced nearly two decades ago to trace the geographical origin and/or adulteration of wines.[98] The deliberate use of a stable isotope enriched compound is widely applied in plant biochemistry and has replaced the use of radioactive isotopes (i.e., ^{3}H, ^{14}C). Applications of this technique include primary metabolite flux analysis investigations[58] as well as successful attempts to unravel the biogenesis of prominent secondary natural products (e.g., hyperforin[99]), which have culminated in the discovery of novel biogenesis pathways for terpenoid-derived secondary metabolites.[100] Since the introduction of biofluids' metabolite profiling little more than two decades ago,[101−103] which later evolved further to the concept of metabonomics,[13] high-field ^{1}H NMR spectroscopy has played a central role in assessing the complexity of biogenic matrices. To obtain the spectral resolution and sensitivity needed for these kinds of applications, the NMR instrumentation used should operate at a proton field strength equaling or exceeding 400 MHz. The use of an autosampler (if working with tubes) or a sample robot (in the flow setup) in conjunction with automated shimming and tuning routines is strongly recommended to minimize artifact introduction by manual interactions. In NMR-based metabolic profiling, sample pH has to be monitored closely, as well as temperature equilibration, prior to measurement. Within natural product analysis, the first applications of combining NMR spectroscopy and multivariate data analysis were published less than a decade ago.[104] Currently, several dozen contributions dealing with NMR-based metabolic profiling of plant-derived samples have been published; outstanding examples will be given below.

3.4. Post-Analysis Data Treatment and Evaluation. One of the most challenging problems in analyzing large sample sets of high complexities in a high-throughput mode is the reproducibility of the measurements. Although qualitative and quantitative aspects of the analytical results of single analytes can be easily addressed by appropriate validation procedures, that is, the determination of sample stabilities and the use of internal standards as well as the assessment of intra- and interbatch errors, the stable performance of the assay in an automated multicomponent analysis used is harder to address. Here, the most critical factor for any chromatographical setup is the stability of the analyte retention time, a prerequisite for unequivocal automated analyte identification. It is influenced not only by changes in the sample matrix (e.g., highly enriched coeluting metabolite groups like fatty acids or alkanes as well as high salt loads) disturbing the phase exchange process of the analyte, but also by hardly controllable nonsystematic system parameter variations as increasing column back pressures, fluctuations of the column compartment temperature, air bubbles trapped within HPLC systems, drifts in the gradient-forming process, and so forth. The same holds true for NMR spectroscopy measurements where ^{1}H signal positions (chemical shifts) may strongly depend on solvent conditions (i.e., pH and salt concentration of the extract or residual amounts of sample water and extraction solvent present) and measurement temperature. However, since, for any of the used analytical platforms, residual fluctuations in retention times or chemical shifts have been observed to occur even under tight control of the experimental parameters, several algorithms (including commercial products supplied by instrument vendors) have been developed to allow corrective measures.[20,105]

3.4.1. Mass Spectrometry. The most recent achievements in the field of chromatography-based assays used for peak detection and mass spectra deconvolution are MET−IDEA,[106] XCMS,[107] and an approach combining hierarchical multivariate curve resolution (MCR) for data deconvolution with either partial least-square discriminant analysis (PLS−DA)[108] or orthogonal partial least-squares (O−PLS)[109] for data filtering. MET−IDEA, a freeware to academic users upon request, covers the whole data processing starting from raw data IRt (ion/retention time) lists obtained by the application of ADMIS[105] or manual data entry. Hence, it is compatible with several chromatographic platforms (most efficiently for GC−MS) and can be used via the Microsoft.NET framework. MET−IDEA allows correcting both deviations in retention times and mass values. Peak areas are calculated as sum of selected ion intensity values falling within a distinct peak range provided by the deconvolution process. In the given application example, a GC−MS analysis of a plant cell suspension culture treated with elicitors of secondary metabolism, more than 200 chromatographic peaks were successfully extracted from about 180 analytical runs.[106]

XCMS, an open-source freeware software, has been developed by the team of Gary Siuzdak to aid HPLC−MS platforms. It starts from raw data files; thus, it includes the peak identification process provided by ADMIS in the MET−IDEA case. Peaks detected in single-ion chromatograms (recorded at a resolution of 0.1 m/z) are combined, filtered if applicable, integrated, and selected using a signal-to-noise cutoff factor of 10. The peak matching algorithm across samples is assuming high stability of the mass spectrometer, thus, using fixed

interval bins in the mass spectrometry domain with a 0.25 m/z bandwidth. Retention time alignment performed thereafter is based on peak groups with high probability of correct match ("well-behaved peak groups") which are used to determine nonlinear profiles of retention time drift throughout the sample set. In a first quantitative application using spiked serum samples, XCMS did prove to be applicable to deconvolute UPLC−MS and HPLC−MS runs of complex matrices.[110] Hence, the use of XCMS will be easily extendable to plant-derived sample sets.

A Swedish team including Henrik Antti and Thomas Moritz did recently publish a software package (its script is available to academic and noncommercial users) combining MCR used to deconvolute GC−MS data with multivariate statistical methods (e.g., PLS−DA) to facilitate the identification of sample discriminators.[111] Within the first data processing step, smoothing, background reduction, time domain slicing, and MCR are performed. Thereafter, integrated peak areas of the chromatographic profiles from all samples in all time windows are combined with a data matrix used as input for further multivariate statistical analysis as PCA, PLS−DA, or O−PLS. Visual or statistical investigation of the variable weights allows a straightforward identification of variable groups (an IRt list) explaining differences within the investigated samples. Since mass spectra and retention times of each compound are listed in a second data matrix, export of the mass spectra of discriminating analytes to external databases for analyte identification is easily executable. Predictive metabolic profiling based on GC−MS data can be performed with this approach. Once model systems are set up, independent samples can be added at a later time point using the processing parameters established in the training data set. The feasibility studies included rat urine samples, aspen leave extracts, and human plasma samples.[109]

Once the overall metabolic profile of a mass spectrometry-derived sample set has been recorded, transformed, and exploited in terms of analyte identification, it has to be further evaluated in terms of biomarkers discovered. Often, simple comparative tools as difference profiles (e.g., treated vs untreated) and analysis of variance (ANOVA) of single analytes are used.[112−117] However, these comparisons have to be assessed carefully, since interdependencies of significantly changed analytes are usually not assessed. Hollywood et al. did correctly state in the context of assessing difference profiles: "In the unlikely event that these have worked, it probably means that the biomarker is obvious; however, single biomarkers are unlikely".[118] Therefore, the use of multivariate statistical methodologies, either unsupervised multivariate profiling techniques or more sophisticated supervised learning algorithms ("the chemometric zoo"[119]) can be used.[118,120] Unsupervised techniques as principal component analysis (PCA) or hierarchical cluster analysis (HCA) help to reduce data complexity. PCA is usually used as a first explorative step in the analysis of NMR data; cluster analysis approaches are more widespread in mass spectrometry-based contributions.[18,83,121,122] Supervised algorithms can be classified in many ways. A recent overview[118] distinguished discriminate analysis (DA) algorithms, multivariate regression models such as partial least-square-based approaches (e.g., PLS1, PLS−DA), artificial neural networks (ANNs), rule induced (RI) algorithms (e.g., CART, FuRES), and evolutionary computation (EC) techniques such as genetic programming and genetic algorithms.[123,124] Much progress has been made in the application of PLS−DA or O−PLS type algorithms as biomarker discovery tools in the NMR field (see below). With the currently available novel software tools of Antti and Moritz,[108,109,111] they will certainly start to contribute to mass spectrometry-based metabolic profiling.

A different approach to metabolic data set mining is devoted to unravel metabolic network topologies to obtain a deeper understanding of the overall regulation of metabolic processes as the ultimate goal.[118,125] Here, multivariate correlation analyses are the statistical tools of choice. This includes correlations with other -omics data sets, allowing the "visualization of biological and molecular coherence".[125] Most often, the distance information (e.g., Pearsons correlation coefficient) is displayed by the application of the Fruchterman−Reingold algorithm available through the Pajek software package.[126,127] However, utmost care must be taken in interpreting such data presentations. Highly correlated metabolite pairs, which have been often observed, are not necessarily neighbors in underlying metabolic networks.[128,129] Hence, metabolic correlations do not necessarily depict reaction cascades, but flux networks often extending over several well-balanced enzymatic analyte interconversion steps affected by chemical equilibrium, mass conservation, asymmetric control distribution, and high variances in gene expression.[128]

3.4.2. Nuclear Magnetic Resonance Spectroscopy. Although NMR can be considered more robust than chromatography-based assays and has proven to give low measurement uncertainty in large-scale investigations such as the INTERMAP study[130] or under inter-laboratory study conditions as realized in the COMET project,[131,132] imperfect sample preparation techniques might introduce inhomogeneities to NMR spectra. Defernez and Colquhoun have summarized some pitfalls to be encountered in plant-related projects and discussed possible solutions, including signal alignment methods, that can be applied post-acquisition to suboptimal data sets.[133] Applications of their efforts have been demonstrated in several case studies to be described further below. Recently, the problems of using data reduction protocols ("binning data"), as well as the effects of different data preprocessing procedures such as normalization and scaling operations, have been highlighted for NMR[134] and GC−MS data.[135] It has been strongly emphasized in both contributions that, as Craig et al. express it, "it is important to be aware of the effects of preprocessing on statistical outcome and to be mindful of the consequences of a chosen method of preprocessing and the limitations that this will place on the interpretation of any chemometric model".[134] More recent approaches do include curve fitting algorithms allowing to avoid usually employed data reduction procedures.[132] Among other statistical benchmark tests, it was shown that higher data consistency was obtained if more than one NMR signal of an analyte was used for quantification purposes.

Until recently, NMR-based approaches relied mainly on pattern recognition techniques such as unsupervised PCA and supervised PLS−DA for multivariate data exploration. These methods, which have been comprehensively documented in the literature,[136−139] will be discussed only briefly here. PCA is a linear transformation of initial variables (NMR descriptors) associated with the samples into a smaller set of variables termed principal components (PCs). These are orthogonal to each other and explain progressively less variance in the data set. Hence, variance in the data set can be displayed without including irrelevant spectral and metabolic background information, often summarized as "noise". Usually PCs are displayed in two- or three-dimensional "scores plots", allowing the

analysis of the distribution and grouping of the samples in the new variable space. PCs are linear combinations of original input variables, the NMR signals or signal regions. Each of these variables is given a weight (loading) which indicates the influence strength of that variable on the overall profile of a set of samples. These weights can also be visually inspected in one- or two-dimensional "loading plots" allowing the identification of strongly influencing variables.

Supervised methods such as the PLS−DA make use of known sample classes (e.g., treated vs untreated plants) to maximize the difference between these classes.[137] This provides a mean to identify discriminative candidate biomarkers. However, if enough components are used, it will always be possible to separate two classes perfectly.[138] Therefore, in supervised analysis, data have to be divided into training and validation subsets. A model is generated from the samples in the training set and is subsequently used to predict the class of independent samples in the validation set. Within the past years, more advanced methods based on PLS have been developed to accommodate the large amount of extraneous variation unrelated to sample classes which is often an inherent part of biological data. An orthogonal filter incorporated into the PLS algorithm (O−PLS) allows to calculate, examine, and remove (if warranted!) spectral components orthogonal to the sample classification, which may arise from metabolic fluctuations unrelated to the sample classes chosen.[140,141] It was demonstrated that, if O−PLS algorithms are utilized to analyze NMR data at full spectral resolution, the influence of slightly shifting resonances can be neglected.[142] Furthermore, the interpretation of the loading data, hampered by the high number of variables at full resolution can be simplified by O−PLS. Their number can be reduced, and a spectrum-like representation can be achieved. The most recent breakthrough is based on the calculation of correlation matrices from metabonomics data sets.[143] Statistical total correlation spectroscopy (STOCSY) analysis takes advantage of the multicolinearity of intensities in a set of spectra. Pseudo 2D NMR spectra displaying the correlation of peaks intensities across the whole sample set are generated. These display not only intramolecular connectivities as in conventional TOCSY type 2D NMR spectra (chemical covariance), but also intermolecular correlations based on coregulation of analytes (biological covariance). It has been exemplified that, if STOCSY is combined with a supervised pattern recognition method performed on full resolution NMR data, analyte identification from the loading plot is simplified, since intramolecular correlations are highlighted (Figure 3).[143]

In another hallmark paper, this approach has been extended to the combined analysis of NMR and UPLC−MS spectra. It was successfully demonstrated in a toxicological study that direct cross-correlation of chemical shifts from NMR and m/z data from mass spectrometry was achieved, thus, improving the odds in structurally characterizing putative biomarkers.[144] This approach can be extended to plant-based metabolic profiling, as we have shown recently for a *Hypericum perforatum* sample set (Figure 4).

3.5. Method and Assay Validation. Evaluation and validation of bioanalytical methods has to be a well-described and tightly controlled process, especially if an industrial or clinical environment is involved. A broad range of international standards and guidelines have been published in this field,[145−147] including key documents provided by the FDA,[148] the ICH,[149] and the EURACHEM.[150] It cannot be easily comprehended, why these standards should not be applicable to metabolic profiling,

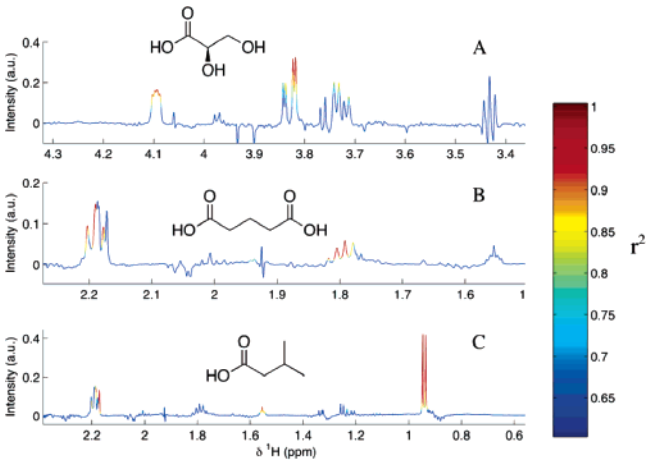

Figure 3. Combination of O-PLS model coefficients plotted as function of their corresponding chemical shifts (allowing their interpretation in the same way as a conventional NMR spectrum) with the STOCSY correlations corresponding to three significant resonances (A, δ_H = 3.82 ppm, glycerate; B, δ_H = 1.79 ppm, isovalerate; C, δ_H = 0.947 ppm, glutarate) clearly revealing the spin systems of these metabolites. Peak intensities mirror the contribution of the spectral regions to data discrimination, whereas the superimposed color coding allows identifying intramolecular spin systems of influential peak groups.[143]

especially if analyte response pattern are, as it is the current practice, mainly reduced to metabolites readily identified by the use of available external reference materials. One could even argue that metabolic profiling is still methodologically closer related to targeted analysis than to metabolomics. This should not be a concern; on the contrary, it might be even an advantage that analyses are currently mostly reduced to "the known", since "known ones" are often available at high purities (reference standards), an inevitable requisite for analytical method validation. Recent application examples have shown impressively that, with state of the art GC−MS, HPLC−MS, or CE−MS instrumentation, quantitative analysis of vast arrays of primary[151−155] and secondary[156] analytes as well as of xenobiotics[157] is accomplishable, even at data densities experienced in metabolic profiles. However, since thoroughly validated targeted analysis approaches covering comprehensively several dozen or even hundreds of commonly encountered primary and secondary metabolites are still missing or in a premature state, there is an urgent need of further development to reach the state of broad applicability and marketability.

4. Mass Spectrometry-Based Applications

The first publications in the field of mass spectrometry-based metabolic profiling were dedicated to prove the concept of this novel methodology combination. In their fist two contributions, Oliver Fiehn et al. structurally characterized about 170 metabolites from *A. thaliana* by the application of GC−MS.[18,158] In the more analytical contribution,[158] the combination of molecular masses (a mass accuracy of 0.005 u was achieved) and isotope ratios together with database searches and comparison with reference substances was used to confirm 94 of 102 analytes in a polar extract. On the basis of this knowledge, an extended investigation including the lipophilic extract phase was performed on leaves of two homozygous *A. thaliana* ecotypes and mutants thereof.[18] Semiquantitative analysis of more than 300 analytes was performed; about half of them were

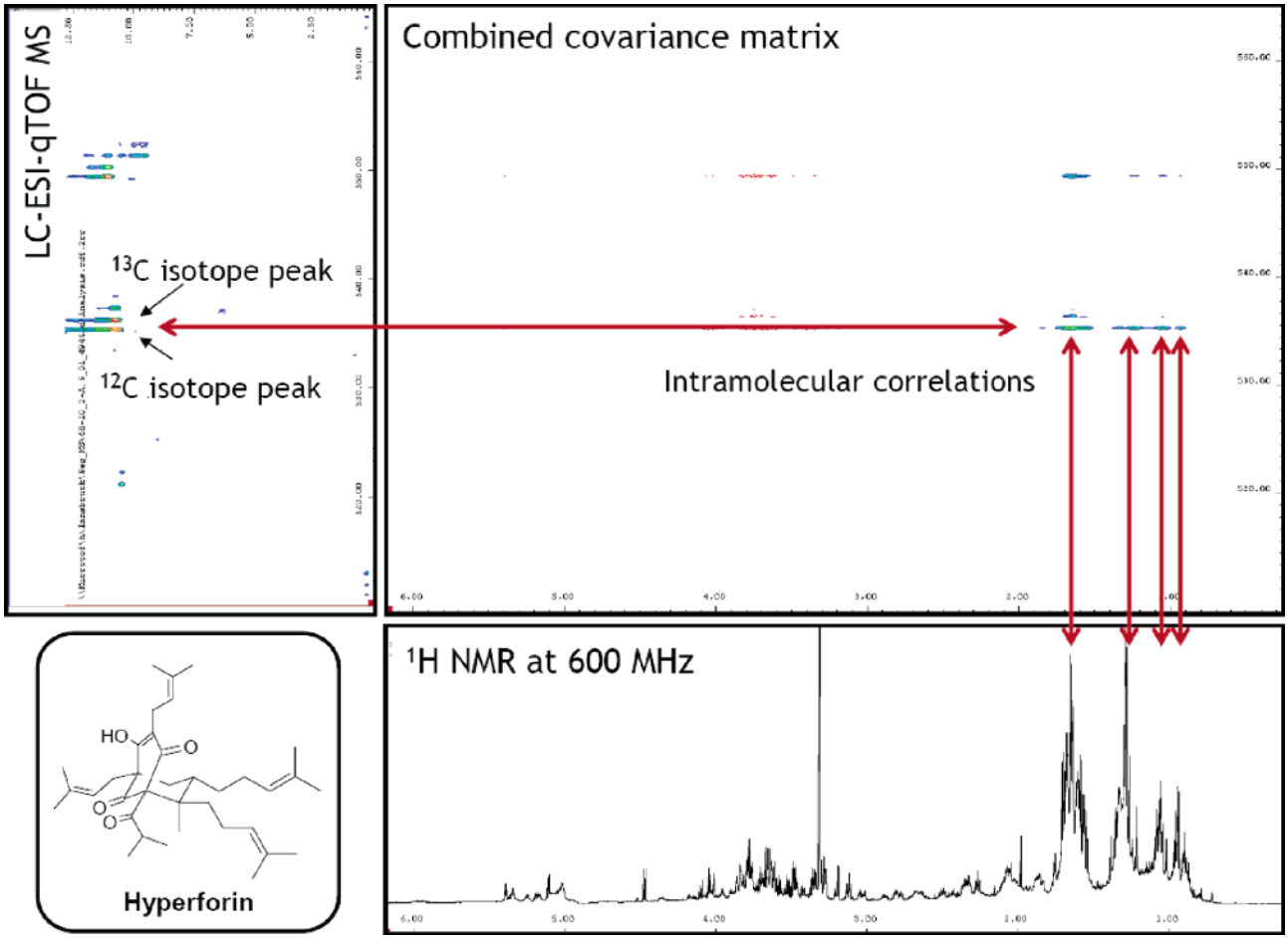

Figure 4. Detail of the combined covariance matrix generated from the metabolic profiling data of 14 *H. perforatum* lots. Intramolecular correlation signals of hyperforin, a phloroglucin derivative showing high discriminative power in the investigated samples, are highlighted. By this approach, ^{1}H NMR and MS data of analytes can be easily combined, even if the assays are not directly correlated. Hence, NMR data can be linked with mass spectrometry-derived information and chromatographic parameters as retention times and UV spectra, easing the access to reference data. NMR and mass spectra were obtained on a Metabolic Profiler setup comprised of an ESI-TOF mass spectrometer (Bruker Daltonics, Bremen) coupled with an Agilent HPLC system and a 600 MHz NMR spectrometer equipped with a 5 mm TCI cryoprobe (Bruker Biospin, Rheinstetten). Data handling, multivariate data analysis, and data visualization were carried out with the AMIX software package by Bruker Biospin.

identified as stated above. Relative quantification was supported by checking the linearity of detector response using dilution experiments. Absolute quantification was facilitated by the use of stable isotope labeled reference materials available for 11 analytes. Ion suppression effects up to 50% were observed, which showed no difference between wild-type and mutant strains. Analytical reproducibilities were addressed for 149 polar metabolites and found to be 8 ± 6%. A PCA performed on log-transformed data of 151 samples resulted in four clusters, identical to the investigated genotypes.

Similar contributions published in parallel or shortly thereafter did prove the applicability of GC−MS to analyze the metabolome of *S. tuberosum* (potato) and *S. lycopersicum* (tomato). It was concluded that, at least in the case of the *S. tuberosum* tubers, the biological variation exceeded the analytical error, and that analyte identification can only be facilitated by the use of several hundred reference compounds or reference libraries (NIST, WILEY).[112] The subsequent application of the established platform to the analysis of different genotypes was aided by PCA and HCA.[113] In the first paper on *S. lycopersicum*, the above-mentioned technology platform was successfully transferable to the matrix of leaves and fruits.[115]

In a continuative study, the GC−MS survey was extended to several wild species known to be used in crossbreeding. However, beyond listing several hundreds of metabolites and their relative distribution in the investigated species, no attempt was undertaken in this contribution to use multivariate statistical tools for deepened data analyses.[116]

The applicability of FT−ICR MS for nontargeted metabolome analysis was investigated by flow injection of strawberry (fruit ripening) and tobacco (flower color mutant) extracts into the ESI or APCI source of a 7.0 T machine.[159] This platform was performing at a mass resolution of 0.1−0.2 ppm, allowing the complete resolution of the metabolome.[27] Several hundred masses were observed with any of the ionization techniques, and putative analyte identifications based solely on the molecular formulas were carried out for dozens of metabolites. A couple of masses were found to significantly change between the sample sets. The authors concluded, that ion suppression and adduct formation do hamper the use of FT−ICR MS, which however is a high-throughput instrument providing high information content data.

In one of the most comprehensive metabolome analysis efforts ever undertaken, Bino et al. did present data on the fruit

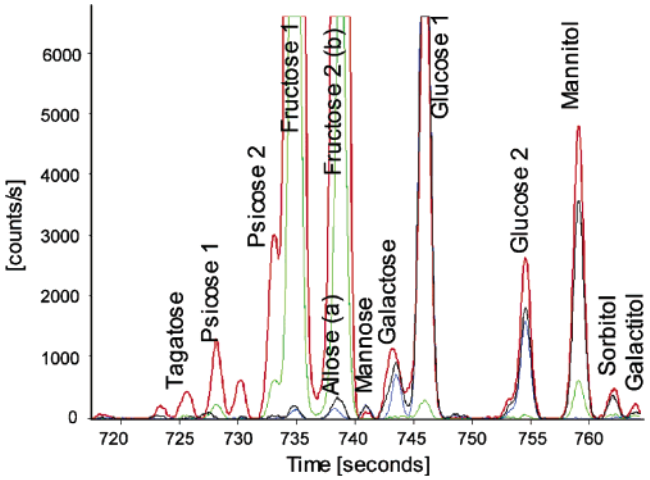

Figure 5. A 40 s time-slice of a GC-TOF−MS spectrum of a polar *S. tuberosum* (potato) leaf extract featuring common and rare monosaccharides and sugar alcohols. Ion traces such as *m/z* 103 (red), 307 (green), 160 (blue), and 319 (black) are used for relative quantification. Reprinted with permission from ref 160. Copyright 2004, The National Academy of Sciences of the USA.

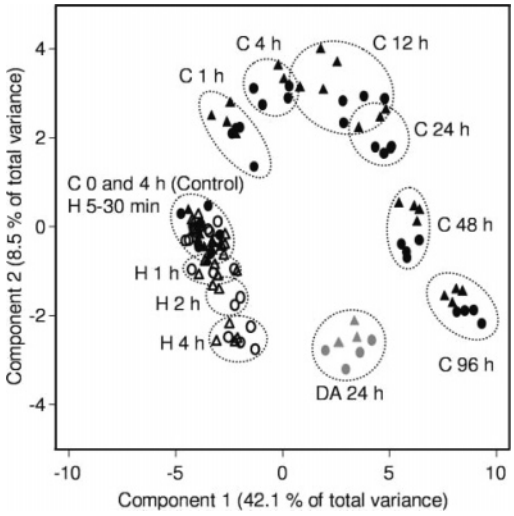

Figure 6. Principal component analysis of two temperature stress experiments performed with *A. thaliana* plants. In the first PC (component 1), samples differentiate according to the performed experiment (CS cold shock, HS heat shock), whereas in the second PC (component 2), the time series within the experiments are differentiating. Upon deacclimatization (DA), a new metabolic state is reached. Reprinted with permission from ref 61. Copyright 2002, The American Society of Plant Biologists.

metabolome of a genetically modified (GM) tomato strain.[5] The team from Wageningen used HPLC−DAD (carotinoids, phenolics, and ascorbic acid), HPLC−ESI−QTOF−MS (untargeted, nonvolatiles), and GC−MS (volatiles) to comprehensively address the metabolome of a high-pigment mutant. More than 11 000 features in the positive and negative ion HPLC−MS chromatograms (total run time about 50 min) were identified by MetAlign as mass signals, with only less of 5% being at least 2-fold different between the strains. From these, only eight have been presented by the authors as putatively assigned (about 2% of the discriminative masses). This clearly shows the urgent need for (i) more efficient methods to reduce the list of putative biomarker analytes, for example, by using more stringent statistical approaches (e.g., PCA and related methods), (ii) improved methods of analyte identification (e.g., by extract breakdown combined with NMR platforms), and (iii) comprehensive databases in the field of HPLC-based metabolic profiling. First steps in this direction have been recently presented.[89]

The potential of GC−TOF−MS-based metabolic profiling to detect silent phenotypes has been investigated on a *S. tuberosum* system.[160] By the use of the TOF mass spectrometer, the number of detectable analytes was increased dramatically compared to previous reports (Figure 5).[112,113] Analytes were identified by comparison with reference chromatograms and matched against a reference spectrum database. As expected, match quality did decrease dramatically for low-abundance peaks (e.g., 66% match at signal/noise <25). In the leaf metabolome, only about 20% of the metabolites did show significant differences between the silent transgenic and the parental genotype. None of the identified analytes changed more than 2-fold. Fourteen out of 15 analytes showing the strongest discrimination power had to be classified as "primary candidates for *de novo* structural elucidation". PCA did not discriminate the sample sets, whereas discriminant function analysis did reproduce the results of the univariate ANOVA. Topology network representation of the pairwise correlation data were investigated also. It was concluded that, at least within this case study, mass spectrometry-based metabolic profiling is not suited to detect metabolic differences in the silent phenotype. The authors summarized that "a mechanistic interpretation of such observations is still hampered by two important features: first, a high number of metabolic peaks in MS are as yet unknown, disabling a link to known biochemical networks. Second, regulations that can cause metabolic correlations can be based on all system levels, i.e., transcription, translation, and ultimately enzyme activities including protein interaction clusters".[160]

A PCA approach was used to analyze a flow injection mass spectrometry data set obtained from potato tuber extracts of wild-type and GM plants.[87] The loading plot allowed assigning a small number of masses as discriminators. They were identified as fructan congeners with different chain lengths by hydrophilic interaction chromatography (HILIC) coupled with mass spectrometry. Di- and trisaccharides were finally confirmed by GC−MS analysis. This contribution is one of the first applications of PCA in mass spectrometry-based metabolomics. It represents a successful combination of unsupervised biomarker discovery and targeted analysis for unequivocal biomarker identification. Furthermore, the level of potentially toxic glycoalkaloid derivatives in transgenic potato plants was found to be normal, a finding which was confirmed by an extended investigation.[88]

Heat and freezing conditions as environmental stressor were assessed in the model plant *A. thaliana*.[161] Samples were drawn at several time points, and GC−MS-based metabolite profiles were recoded based on previously published methods.[112,113] Electrolyte leakage was used as control assay to asses the kinetics of thermal stress. About 500 mass spectral tags (MSTs) with 20% identified as metabolites were obtained from analyzing the GC−MS data. A PCA based on relative response data was used to test metabolic differences between heat shock (HS) and cold shock (CS) treatments (Figure 6). The first PC explaining 42% of the total variance in the data set differentiated HS from CS treatment, whereas the second PC (9% variance) did span the time series in both cases. The starting points of both CS and HS series grouped tightly, proving reproducibility of the setup. Upon deacclimatization of a

sample set (DA), a new metabolic state was established which was readily separable in the third PC. In all cases, time points did group tightly, proving the robustness of the experimental setup. The analysis of metabolic changes causing the sample spreading in the PCA was aided by nonparametric ANOVA on all MSTs and metabolites. Most of these (~90%) were found to show significant level changes in at least one time point, with >50% showing sustained influence. About 38% of all MSTs and metabolites reacted specifically to the CS, whereas only 4% were assignable to the HS conditions. Nearly 20% of metabolites were found under both conditions. Comparison of the ANOVA data with the PCA loadings did unveil that the 19 MSTs and metabolites responsible for the differentiation of the first component were identified by ANOVA as increased in response to CS. Component 2 was differentiated by HS and CS correlated metabolites.

One of the first parallel analyses combining metabolomics with gene expression data has been performed by Fernie and co-workers on the potato tuber system established in the MPI at Golm.[162] In addition to a gene expression analysis based on more than 1000 genes represented by expressed sequence tags (ESTs), metabolite profiles were recorded based on previously developed methodologies.[113] Independent PCA of both platforms did show that transgenic lines differed from wild-type samples only by their metabolite pattern but not by the transcription levels. Pairwise correlations analysis did result in a small number of strong relationships between enzymes and substrates (e.g., sucrose transporter and sucrose). However, a deeper insight into the metabolic network was not achieved by this approach, since "unexpected results only reveal correlated parameters and do not allow the identification of causality between transcript and metabolite".

A. thaliana served as model system for integrated investigations covering the whole range of systems biology, genomics (realized as transcriptomics), proteomics, and metabolomics (realized as metabolic profiling). In 2004, Saito and co-workers did report the first combination of transcriptomics and metabolic profiling investigating plant response to sulfur and nitrogen depletion.[163] Samples derived from roots and leaves of plants exposed to seven different cultivation conditions (a total of 14 samples) were examined. The *Arabidopsis* genome was covered by a cDNA array containing more than 13 000 ESTs. Analytical platforms used included direct infusion FT−ICR MS[159] (no details on the putative metabolite identification by exact masses as single "structural" parameters are disclosed) and targeted analysis by HPLC and CE. PCA and batch-learning self-organizing maps (BL−SOM) analysis were used for separate data analysis of both the metabolite profiles and ESTs. Superimposing primary and secondary metabolic maps with significant ESTs (about 10% of total ESTs) allowed a deeper understanding of metabolic pathways affected by nutrient depletion. Targeted analysis of amino acids, sugars, organic acids, and thiols aided this analytical approach. However, any of the conclusions and findings are hampered by the fact that, at least according to the publication, only single samples stemming from different cultivation conditions were investigated. Biological and analytical variability as measure for experimental repeatability has not been addressed.

Weckwerth et al. presented a novel sample preparation protocol for *A. thaliana* allowing sequential extraction of metabolites, proteins, and RNA from the same sample (less than 100 mg of plant material required), thus, providing a convenient protocol for generating replicates, a prerequisite to address assay stability as well as inherent biological variability in a sample set.[121] The analytical setup included a GC−TOF−MS platform for metabolite analysis (>650 metabolites identified and quantified with a mean coefficient of variation (CV) of 10%) and a two-dimensional HPLC coupled with an ion trap mass spectrometer used for protein analysis (shotgun proteomics) based on a standard trypsin digest (about 600 peptides allowed to identify approximately 300 proteins at a precision of 25% CV). RNA purity checks were performed by gel electrophoresis. Northern blot hybridization of the yielded RNA to *Arabidopsis* isopropyl−malate synthase probes were used as further proof of concept. The new platform was applied to differentiate two genotypes; the PCA and HCA classifications were performed on combined protein/metabolite data sets allowing the investigation of individual metabolite−protein coregulation pairs. For both protein and metabolites, the biological variability clearly exceeded the measurement uncertainty, although developmental stage and growth conditions of the used plants were strictly monitored.

Within a subsequent contribution, the time-related regulation of the *Arabidopsis* protein−metabolite network was investigated.[164] On the basis of the above-mentioned protocols, the day/night cycle of wild-type (wt) and starchless mutant plants were investigated. The analytical robustness of the GC−TOF−MS platform was assured by several means. A retention time index similar to the Kovats index was introduced to compensate for fluctuations in column temperature and flow rates. Using stable isotope labeled internal standards (sorbitol, leucine, and aspartic acid) ensured a decent data quality in the quantification of sugars, sugar acids, amino acids, and TCA intermediates (CVs below 24%). Furthermore, the repeatability was assessed by replicate measurements on pooled samples, showing no influence of the harvesting time point. The biological variation in the data set exceeded the analytical error severalfold. On the basis of these data, a variance threshold of 30% was defined as prerequisite for metabolite correlation analysis. Pattern recognition by independent component analysis (ICA) using the principal components of PCAs did show that combining metabolites and proteins in one data set resulted in an improved discrimination of wt and mutant samples compared to separated analyses. Within the first independent component, wt/mutant discrimination was observed; diurnal variations were detectable in the second component. Visual inspection of pairwise correlation data by metabolite correlation networks and metabolite correlation matrices did unveil the coregulation of an amino acid cluster connected to carbohydrates indicating a shift in the C/N partitioning. This finding was corroborated by alterations in the urea cycle, suggesting enhanced protein degradation.

The genetic basis of naturally occurring metabolite fluctuations in 14 *Arabidopsis* accessions was addressed by combining HPLC−TOF−MS-based metabolic profiling with quantitative trait locus (QTL) analysis.[165] By the use of a novel alignment software, the combination of about 1600 high-resolution mass signals with a genetic map derived from available genotype data on the investigated populations was facilitated. Most of the masses were assignable to QTLs, and visual inspection allowed to detect cold and hot spots for the regulation of metabolite compositions. Since aliphatic glucosinolate derivatives are easily identifiable by high-resolution mass spectrometry, this congener group was used for proof of concept. Indeed, analytes did map to only two trait loci, MAM and AOP. In the second case study presented, a series of masses, putatively identified

as quercetin, kaempferol, and isorhamnetin−glycosides (based on exact masses, MS/MS fragmentation pattern, and DAD-derived UV spectra) were found to map onto a single locus. Closer investigations including the analysis of QTL likelihood profiles and relative level of involved metabolites suggested that the discovered QTL has to carry a glycosyl transferase, which has not been identified before.

A similar approach was applied to identify the flux control points of flavonoid biosynthesis in the model tree *Populus* (poplar).[166] The strategy of combining metabolite profiling with QTL analysis was christened "genetical metabolomics" by the authors. An AFLP linkage map was used as genomical data source and, because of the rather small and well-defined set of secondary metabolites under investigation, a targeted analysis based on HPLC−UV chromatograms was chosen to generate the metabolite profile. Data exploration by metabolite level correlation and trait locus search allowed mapping of four trait loci involved in the flux control of only one or two flavonoid derivatives, each. Tentative functions were assigned to the trait loci based on structure comparison of the involved metabolites, and corresponding candidate genes were mapped.

A more comprehensive analysis was pursued to establish the "gene-to-metabolite" network of indole alkaloid biosynthesis in periwinkle (*Catharanthus roseus*) cell cultures.[167] Periwinke is a well-established medicinal plant, since its secondary metabolites vinblastin and vincristin, both bisindole alkaloids, are of utmost clinical importance due to their well-established antineoplastic activities. Within this study, cell cultures were dosed with the auxin NAA and additionally exposed to methyl jasmonate, a plant hormone known to induce the accumulation of indole alkaloids. A genomewide transcription profile was established by using quantitative cDNA-AFLP technology. It resulted in about 470 sequenced fragments which were subjected to BLAST analysis using different databases. It was demonstrated that a vast majority of tags were new to *C. roseus*. Furthermore, about 40% of all sequences did not even display sequence similarity to any known plant gene. Nontargeted metabolic profiles obtained from HPLC−ESI−MS using a triple quadrupole mass spectrometer operating under full scan conditions resulted in about 4000 mass features, which were reduced to a list of about 180 peaks by normalization and peak filtering. Only nine of these analytes (5%) were tentatively identified with the aid of reference masses and retention times. Furthermore, six selected alkaloids were addressed quantitatively in a separate experimental setup by HPLC−ESI−MS using a triple quadrupole mass spectrometer recording single ion traces. Application of HCA of the semiquantitative metabolite profiles revealed several subclusters. More than half of the peak list entries did group within two clusters, potentially displaying auxin-mediated analyte accumulation or auxin-mediated analyte depletion (Figure 7). Evaluation of the combined gene and metabolite data sets by PCA revealed a clear separation of auxin-treated cells from untreated ones with the incubation time contributing to the second principal component. In addition, correlation network analysis based on Pearson's correlation coefficients was used to establish gene-to-gene and gene-to-metabolite coregulation patterns.

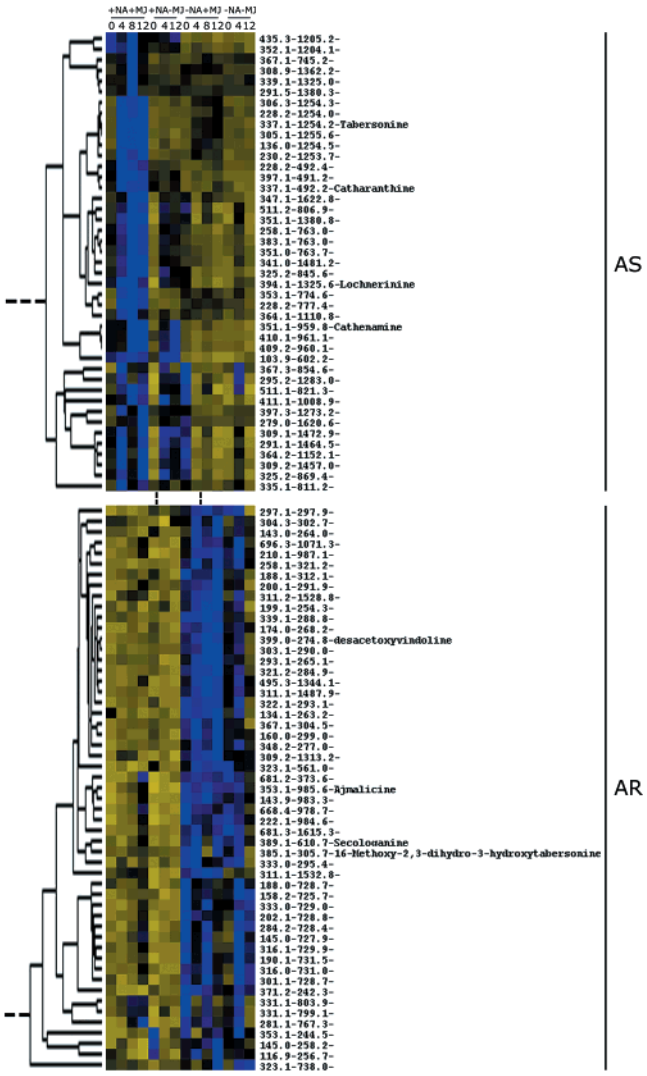

Figure 7. Metabolome analysis of elicited *C. roseus* cells. Nontargeted metabolite profiling represented by average linkage hierarchical clustering of metabolite accumulation profiles. Two subclusters grouping metabolites affected by auxin treatments (AR, auxin repressed; AS, auxin stimulated), are shown. The treatments, the presence and absence of auxin (NA) and MeJA (MJ), and the time points (in hours) are indicated at the top, and the retention times and the *m/z* values are on the right. Blue and yellow boxes correspond to stimulated or repressed accumulation of metabolites, relative to the average accumulation level of all samples, respectively. Reprinted with permission from ref 167. Copyright 2006, The National Academy of Sciences of the USA.

5. Nuclear Magnetic Resonance-Based Applications

Whereas currently mass spectrometry-based metabolic profiling contributions are mostly limited to model organisms or crop plant systems, NMR-based profiling publications do cover a broader range of plant systems. In addition to a focal point on food chemistry applications and a number of crop plant related contributions, taxa with phytomedicinal relevance were investigated. Because of the character of secondary metabolites (e.g., their structural diversity and accumulation in certain species) with relevant bioactivities, NMR-based work often includes the identification of these analytes. Within the following paragraphs, a selection of recent key publications in this field will be given. For a more comprehensive overview of this field, the already mentioned recent review of Holmes and coauthors is recommended.[64]

The consumption of tea, especially unfermented green tea, has been associated with different health benefits, including antioxidative, anticarcinogenic, and antitumorigenic properties,

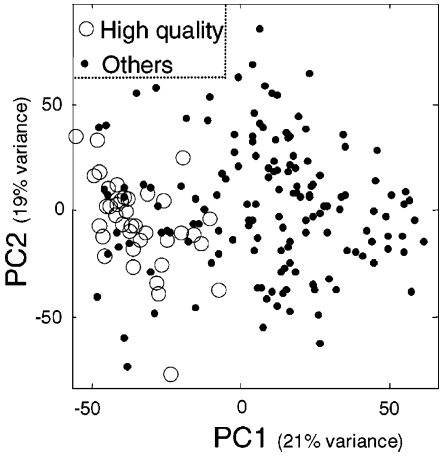

Figure 8. Principal component analysis on the ^{1}H NMR spectra of 184 teas showing a separation of the high-quality batches in the first principal component.[172]

as well as protection against cardiovascular diseases. At least some of these bioactivities can be associated with certain groups of polyphenols like theaflavin and epigallocatechin congeners.[168−171] Comparing more than 190 commercially available tea samples using ^{1}H NMR spectroscopy, LeGall et al. were able to distinguish "high quality" Chinese teas from other samples stemming from all over South East Asia by the application of a PCA (Figure 8).[172] Eighteen putative biomarker molecules were identified in the loading plot, including several structurally assignable tea polyphenols, some unidentified sugar derivatives, and caffeine. Their discriminative power was validated by the application of ANOVA (analysis of variance) to NMR peak intensities. Significant differences were found for 17 of these metabolites (Figure 9). Chinese teas showed higher levels of theanine, gallic acid, caffeine, epigallocatechin gallate, and epicatechin gallate; levels of epigallocatechin, sucrose, and fatty acids were lowered compared to the other samples. In addition, a peak alignment algorithm, reducing the impact of sample inhomogeneities producing artifacts in the multivariate statistical data processing,[173] was successfully applied to this data set.

A series of publications has been devoted to the ^{1}H NMR based metabolic profiling of olive oils, which have been already the topic of several previous ^{13}C NMR-based studies.[174] More than 200 "*extra vergine*" olive oil samples collected in different Italian regions have been investigated. Unsupervised tree clustering resulted in sample groupings according to the location of their production. Furthermore, different production years from identical locations were distinguishable.[175] The broad applicability of this methodology for olive oil classification was proven by several subsequent studies of similar design.[176−180] In addition, adulteration of olive oil with hazelnut oil[181] or olive oil of low quality batches[182] has been investigated, too. The adulteration detection limit was found to be better than 8%.

Tracking bioactive principles in a complex plant matrix can be assisted by correlating bioactivities of extracts of different polarities with the metabolite spectrum using a partial least-square (PLS) analysis. With this approach, *Artemisia annua* lots from different sources were classified accordingly to their antiplasmodial activity and their toxicity to cell cultures. The validity of the derived model was checked by correctly predicting the activity of independent samples not used in the model

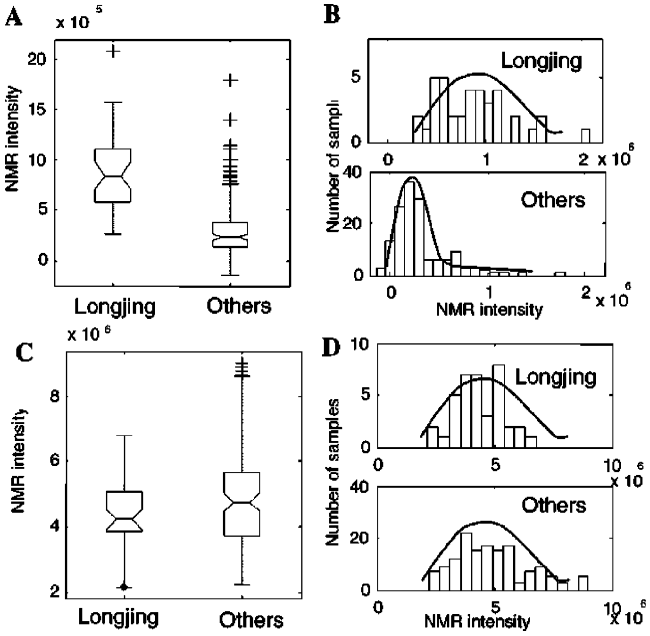

Figure 9. NMR signal intensities of the discriminating spectral regions, which can be regarded to be at least semiquantitative, were tested on their discriminative power by ANOVA. Whereas the top row (A and B) shows significant differences for an unidentified sugar derivative, quinic acid (C and D) did not discriminate between low-quality and high-quality tea accessions.[172]

generation.[183] In a similar study, ^{1}H NMR spectra of various extracts of St. John's Wort (*H. perforatum*) samples obtained by using six different extraction solvent mixtures were calibrated against nonselective binding to opioid receptors using the partial least-square (PLS) algorithm. Subsequently, multivariate data analysis was used to predict the pharmacological efficacy of St. John's Wort extracts on the basis of their proton NMR spectra.[184]

Since phytopharmaceuticals and especially traditional Chinese medicines (TCMs) are complex mixtures of several matrices holding often dozens of bioactive constituents, special care has to be taken for the quality control of both raw and final formulated products. Currently used technologies such as TLC, HPLC−UV, HPLC−MS, and GC−MS provide only a limited fingerprint type insight into the complexity of the samples. They are tailored to fulfill the control needs of both authorities and manufacturers to track known adulterations. However, species mixups or other adulterations of unknown origin, as well as microbial lot spoilage, can be easily overlooked. Here, metabolite profiling with its capability to assess similarities and differences in a sample set in an unbiased way can be successfully employed. Several case studies have been performed which address either samples stemming from different species of a certain genus or lots obtained from within a plant species (geotypes, ecotypes, and cultivars).

Mate tea is a coffein- and theobromin-containing herbal drink produced from *Ilex paraguariensis* and used throughout South America. Five hundred *Ilex* species are known, some of which are used as substitutes or adulterants of *I. paraguariensis*. Furthermore, the metabolite profile of this product is known to vary depending on preparation techniques, age of plant materials at the point of harvest, and the manufacturer. Usually employed HPLC−DAD assays allow the quantification of

polyphenols but cannot address the presence of multiple *Ilex* species in crude drug mixtures. Verpoorte and co-workers were able to characterize 11 different *Ilex* species from both aqueous and organic extracts using an NMR-based method. Of these, only *I. paraguariensis* was found to contain caffeine and theobromine, whereas other species contained high levels of arbutin, absent from *I. paraguariensis*.[185] In the case of chamomile (*Matricaria recutita*), different accessions of the flower heads stemming from Easter Europe and Northern Africa were distinguishable. In addition, the percentage of stalk material in the samples was assessable also.[186] Both chemotaxonomical and quality control aspects have been covered by a study dealing with the genus *Ephedra*. Three *Ephedra* species and commercial *Ephedra*-based products were investigated. In addition to the successful identification of a broad range of metabolites from crude *Ephedra* extracts achieved without any chromatographic separation, this method revealed that the differentiation of *Ephedra sinica*, *Ephedra intermedia*, and *Ephedra distachya* was due to some undisclosed benzoic acid analogues in the aqueous fraction and ephedrine-type alkaloids in the organic fraction. On the basis of these findings, one of nine evaluated commercial *Ephedra* materials was shown to be a mixture of *Ephedra* species.[187] Extracts of ginseng (*Panax ginseng*) products (roots and capsules) were recently assessed using skyline projections of two-dimensional *J*-resolved NMR spectra. The efficacy of using this fast 2D NMR method, which allows to improve both the resolution and the sensitivity of ^{1}H NMR spectra by reducing complex multiplet signal patterns to single lines, has been proven by successfully differentiating the samples under investigation in a PCA. The most discriminating analytes have been found to be primary as well as secondary metabolites.[188] Commercially available preparations of *H. perforatum* (St. John's Wort) have been very recently distinguished by a PCA applied to ^{1}H NMR spectra obtained from extracting tablets and capsule and redissolving the extracts in a NMR solvent. In addition to different formulation excipients showing up as sample discriminators (>95% of total data set variance), different flavonoid congeners were responsible for sample discrimination. The use of full resolution (30 000 variables) instead of integrated (200 variables) NMR data as proposed previously by Nicholson's group for metabonomics[142] allowed a direct assignment of discriminating variables from the loading plots.[189] The significance of the detected sample discrimination was not addressed, neither changes in the signal intensities have been evaluated (as in LeGall's study on tea lots), nor a comparison to a reference method has been presented. We have been recently able to show in a preliminary case study performed on extracts of seven *H. perforatum* crude drug lots provided by commercial suppliers that HPLC−DAD and ^{1}H NMR fingerprinting had similar discriminating power in the PCA. Grouping within and separation between the lots were comparable; the most influential discriminators were found to be the phloroglucinol-type secondary metabolites such as hyperforin, adhyperforin, and flavonoids.[64,190] Hence, NMR-based metabolic profiling and fingerprinting can be predicted to show a discriminatory power similar to routinely used HPLC−DAD assays. Similar results obtained for chamomile[183] and tomato[191,192] do support our findings. In the latter studies performed by LeGall et al., the secondary metabolite profile of GM tomatoes was assessed. By the use of NMR analysis of the extracted whole fruits in combination with uni- and multivariate data analyses, the metabolite composition was compared with that of controls. Differences were observed in several primary and secondary metabolite groups. Flavonoids (e.g., kaempferol and naringenin glycosides) were upregulated more than 10-fold in transgenic fruits. Other smaller (greater than 2- to 3-fold) still statistically significant differences between wild-type and modified fruits included trigonelline, glutamine, and asparagine, which were present in relatively higher concentration in the transgenic tomatoes, and citrate, phenylalanine, sucrose, malate, branched chain amino acids, and cinnamic acids, which showed higher concentrations in the wild-type fruits. These metabolic differences were present at all levels of maturity of the fruit; differences increased with ripeness. Levels of glutamic acid, fructose, and some nucleosides and nucleotides gradually increased from the immature to the ripe stage, whereas the levels of valine and γ-aminobutyric acid decreased during the same process. A quantitative HPLC−DAD assay performed in parallel and targeting flavonoid congeners did allow independent proof of the capabilities of ^{1}H NMR-based metabolite profiling. The authors stated that "NMR analysis confirmed the main changes in metabolite levels already identified by HPLC".[192] This approach of applying two independent analytical methodologies was extended to several other applications. Investigations on GM potato plants resulted in the detection of significant differences between two of four plant lines with modified polyamine metabolism and the controls. Proline, trigonelline, and numerous phenolics were identified as discriminators.[193] Moing and co-workers used ^{1}H NMR analysis of metabolite composition in a large feasibility study including the assessment of strawberry fruit quality, tomato strains overexpressing hexokinase, and *A. thaliana* transformants with decreased phosphoenolpyruvate carboxylase activity.[194] In the case of the strawberry fruits, electronic referencing of the NMR spectra by the ERETIC method[195] calibrated against the anomeric proton of glucose was used for analyte quantification. NMR-based metabolite concentrations were compared to enzymatically or HPLC-derived analyte levels. Correlation coefficients of most linear regressions exceeded 0.7 in the case of the enzyme assays and 0.9 in the case of HPLC analyses. Regression slopes ranged between 0.81 and 1.15; repeated analyses (8 samples, 14 analytes) showed a mean coefficient of variation of 3.2%. Qualitatively, the results unveiled that biochemical phenotyping holds the potential to detect quantitative trait loci. The comparison of metabolic profiles of the roots of tomato plants revealed that environmental factors, that is, culture conditions, can significantly modify the metabolic status of plants and, thus, hide or emphasize the expression of a given genetic background. An *A. thaliana* strain hosting an antisense chalcone synthase gene was investigated by Ian Colquhoun and co-workers.[196] Both HPLC−DAD and ^{1}H NMR were utilized as analytical platforms. 2D NMR experiments were performed in combination with HPLC−MS measurements to identify most of the secondary metabolites (flavonoids and glucosinolates) used in the HPLC−DAD-based quantitative analysis. A PCA of the ^{1}H NMR spectra resulted in a list of putative discriminators spanning both primary and secondary metabolites, which were confirmed in the latter case by the quantitative HPLC analyses.

6. Conclusions and Outlook

The full potential of metabolic profiling and metabolomics in plant sciences has not been realized yet. Current applications are merely first feasibility studies to evaluate different hardware setups and to develop software tools allowing handling the enormous data amounts produced by high-resolution instru-

mentation. Despite current methodological limitations, the unequalled potential of this technology for phenotyping arrays of organisms, for example, libraries of genetically modified model plant lines, has been impressively proven. Recent efforts combining metabolomics data sets with RNA microarray or shotgun proteomics data offer new opportunities to strive toward a comprehensive understanding of biological processes. However, currently, several methodological limitations hamper swift advancement toward this prestigious goal and have to be addressed in the near future. To strengthen the confidence in any metabolic profiling techniques, standard validation protocols will have to be applied in the future. Analytical reliability will have to be proven; especially intra- and inter-laboratory repeatabilities have to be demonstrated if the broad applicability of the assays is desired. The current low rate of analyte identification, because of the lack of structural information obtainable by any kind of mass spectrometer, presents the major hurdle in MS-based metabolic profiling. In addition to using arrays of reference compound to generate comprehensive spectral databases for currently used assays, strategies have to be developed to allow the rapid characterization of "general unknowns". Since NMR spectroscopy has the highest potential in identifying small organic molecules, miniaturized or hyphenated NMR setups as capNMR or HPLC−(SPE)−NMR will soon have to be added to the arsenal of metabolic profiling technologies. In contrast, NMR-based metabolic profiling will have to expand toward the chromatographic dimension, both to overcome spectral overlap and to obtain additional structure information such as UV spectra and mass spectrometry-derived molecular formulas. Furthermore, if an interesting analyte (e.g., a metabolite discriminating sample subset) can be separated from the matrix by the application of chromatographic techniques, subsequent HPLC−(SPE)−NMR experiments will strongly aid its structural characterization. Hence, it is evident that only combining NMR and MS-based technologies coupled with state of the art chromatographic assays presents a chance to fulfill the strived high-flying aim of developing metabolic profiling toward a sturdy, robust analytical platform, indispensable in the comprehensive analysis of complex biological systems.

Acknowledgment. The authors express their gratitude toward Markus Godejohann, Eberhard Humpfer, Hartmut Schäfer, and Manfred Spraul of Bruker Biospin/Rheinstetten for their experimental and intellectual efforts in supporting our progressing research in NMR-based secondary plant metabolite analysis. Furthermore we acknowledge Franz Hadacek for stimulating discussions regarding the complexity of secondary metabolites actions in the "real world" of plant−environment interactions.

References

(1) Kitano, H. *Science* **2002**, *295*, 1662−1664.
(2) Ideker, T.; Galitski, T.; Hood, L. *Annu. Rev. Genomics Hum. Genet.* **2001**, *2*, 343−372.
(3) Oltvai, Z. N.; Barabasi, A. L. *Science* **2002**, *295*, 1662−1664.
(4) Weckwerth, W. *Ann. Rev. Plant Biol.* **2003**, *54*, 669−689.
(5) Bino, R. J.; Hall, R. D.; Fiehn, O.; Kopka, J.; Saito, K.; Draper, J.; Nikolau, B. J.; Mendes, P.; Roessner-Tunali, U.; Beale, M. H.; Trethewey, R. N.; Lange, B. M.; Wurtele, E. S.; Sumner, L. W. *Trends Plant Sci.* **2004**, *9*, 418−425.
(6) Barabasi, A. L.; Oltvai, Z. N. *Nat. Rev. Genet.* **2004**, *5*, 101−113.
(7) Jenkins, H.; Hardy, N.; Beckmann, M.; Draper, J.; Smith, A. R.; Taylor, J.; Fiehn, O.; Goodacre, R.; Bino, R. J.; Hall, R.; Kopka, J.; Lane, G. A.; Lange, B. M.; Liu, J. R.; Mendes, P.; Nikolau, B. J.; Oliver, S. G.; Paton, N. W.; Rhee, S.; Roessner-Tunali, U.; Saito, K.; Smedsgaard, J.; Sumner, L. W.; Wang, T.; Walsh, S.; Wurtele, E. S.; Kell, D. B. *Nat. Biotechnol.* **2004**, *22*, 1601−1606.
(8) Lindon, J. C.; Nicholson, J. K.; Holmes, E.; Keun, H. C.; Craig, A.; Pearce, J. T.; Bruce, S. J.; Hardy, N.; Sansone, S. A.; Antti, H.; Jonsson, P.; Daykin, C.; Navarange, M.; Beger, R. D.; Verheij, E. R.; Amberg, A.; Baunsgaard, D.; Cantor, G. H.; Lehman−McKeeman, L.; Earll, M.; Wold, S.; Johansson, E.; Haselden, J. N.; Kramer, K.; Thomas, C.; Lindberg, J.; Schuppe−Koistinen, I.; Wilson, I. D.; Reily, M. D.; Robertson, D. G.; Senn, H.; Krotzky, A.; Kochhar, S.; Powell, J.; Van der Ouderaa, F.; Plumb, R.; Schaefer, H.; Spraul, M. *Nat. Biotechnol.* **2005**, *23*, 833−838.
(9) Fernie, A. R.; Trethewey, R. N.; Krotzky, A. J.; Willmitzer, L. *Nat. Rev. Mol. Cell. Biol.* **2004**, *5*, 1−7.
(10) Fiehn, O.; Kristal, B.; Van Ommen, B.; Sumner, L. W.; Sansone, S. A.; Taylor, C.; Hardy, N.; Kaddurah-Daouk, R. *OMICS* **2006**, *10*, 158−163.
(11) Nicholson, J. K.; Holmes, E.; Lindon, J. C.; Wilson, I. D. *Nat. Biotechnol.* **2004**, *10*, 1268−1274.
(12) Smedsgaard, J.; Nielsen, J. *J. Exp. Bot.* **2005**, *56*, 273−286.
(13) Nicholson, J. K.; Lindon, J. C.; Holmes, E. *Xenobiotica* **1999**, *29*, 1181−1189.
(14) Fiehn, O. *Plant Mol. Biol.* **2002**, *48*, 155−171.
(15) Tweedale, H.; Notley-McRobb, L.; Ferenci, T. *J. Bacteriol.* **1998**, *180*, 5109−5116.
(16) Oliver, S. G.; Winson, M. K.; Kell, D. B.; Baganz, F. *Trends Biotechnol.* **1998**, *16*, 373−378.
(17) Raamsdonk, L. M.; Teusink, B.; Broadhurst, D.; Zhang, N.; Hayes, A.; Walsh, M. C.; Berden, J. A.; Brindle, K. M.; Kell, D. B.; Rowland, J. J.; Westerhoff, H. V.; Van Dam, K.; Oliver, S. G. *Nat. Biotechnol.* **2001**, *19*, 45−50.
(18) Fiehn, O.; Kopka, J.; Dormann, P.; Altmann, T.; Trethewey, R. N.; Willmitzer, L. *Nat. Biotechnol.* **2000**, *18*, 1157−1161.
(19) Krishnan, P.; Kruger, N. J.; Ratcliffe, R. G. *J. Exp. Bot.* **2005**, *56*, 255−265.
(20) Hall, R. D. *New Phytol.* **2006**, *169*, 453−468.
(21) Dunn, W. B.; Bailey, N. J. C.; Johnson, H. E. *Analyst* **2005**, *130*, 606−625.
(22) Villas-Boas, S. G.; Mas, S.; Akesson, M.; Smedsgaard, J.; Nielsen, J. *Mass Spectrom. Rev.* **2005**, *24*, 613−646.
(23) Glinski, M.; Weckwerth, W. *Mass Spectrom. Rev.* **2006**, *25*, 173−214.
(24) Nielsen, J.; Oliver, S. *Trends Biotechnol.* **2005**, *23*, 544−546.
(25) Kopka, J. *J. Biotechnol.* **2006**, *124*, 312−322.
(26) Kind, T.; Fiehn, O. *BMC Bioinf.* **2006**, *7*, 234.
(27) Sunghwan, K.; Rodgers, R. P.; Marshall, A. G. *Int. J. Mass Spectrom.* **2006**, *251*, 260−265.
(28) Tolstikov, V.; Costisella, B.; Weckwerth, W.; Zhang, B.; Fiehn, O. *Proceedings of the 50st ASMS Conference on Mass Spectrometry and Allied Topics*, Orlando, FL, June, 2002.
(29) Reynolds, W. F.; Enriquez, R. G.; *J. Nat. Prod.* **2002**, *65*, 221−244.
(30) Wolfender, J. L.; Queiroz, E. F.; Hostettmann, K. *Expert Opin. Drug Discovery* **2006**, *1*, 237−260.
(31) Jaroszewski, J. W. *Planta Med.* **2005**, *71*, 691−700.
(32) Jaroszewski, J. W. *Planta Med.* **2005**, *71*, 795−802.
(33) Seger, C.; Godejohann, M.; Spraul, M.; Girtler, A.; Sturm, S.; Stuppner, H. *Anal. Chem.* **2005**, *77*, 878−885.
(34) Gronquist, M.; Meinwald, J.; Eisner, T.; Schroeder, F. C. *J. Am. Chem. Soc.* **2005**, *127*, 10810−10811.
(35) Choi, Y. H.; Choi, H. K.; Hazekamp, A.; Bermejo, P.; Schilder, Y.; Erkelens, C.; Verpoorte, R. *Chem. Pharm. Bull.* **2003**, *51*, 158−161.
(36) Pauli, G. F.; Jaki, B. U.; Lankin, D. C. *J. Nat. Prod.* **2005**, *68*, 133−149.
(37) Noteborn, H. P.; Lommen, A.; Van der Jagt, R. C.; Weseman, J. M. *J. Biotechnol.* **2000**, *77*, 103−114.
(38) Viant, M. R. *Biochem. Biophys. Res. Commun.* **2003**, *310*, 943−948.
(39) Matuszewski, B. K.; Constanzer, M. L.; Chavez-Eng, C. M. *Anal. Chem.* **2003**, *75*, 3019−3030.
(40) Annesley, T. M. *Clin. Chem.* **2003**, *49*, 1041−1044.
(41) Taylor, P. J. *Clin. Biochem.* **2005**, *38*, 328−334.
(42) Nakanishi, K. An historical perspective of natural products chemistry. In *Comprehensive Natural Products Chemistry*; Barton, D., Nakanishi, K., Eds.; Elsevier: Amsterdam, 1999; p XXIII−XL.
(43) Cragg, G. M.; Newman, D. J. *Pharm. Biol.* **2001**, *39* (Suppl.), 8−17.
(44) Verpoorte, R. *Drug Discovery Today* **1998**, *3*, 232−238.
(45) Filtenborg, O.; Frisvad, J. C.; Thrane, U. *Int. J. Food Microbiol.* **1996**, *33*, 85−102.
(46) Hostettmann, K.; Marston, A.; Hostettmann, M. *Preparative Chromatography Techniques. Applications in Natural Product Isolation*, 2nd ed.; Springer: Berlin, 1997.
(47) Hostettmann, K.; Wolfender, J. L.; Terreaux, C. *Pharm. Biol.* **2001**, *39* (Suppl.), 18−32.

(48) Kovats, E. *Helv. Chim. Acta* **1958**, *41*, 1915−1932.
(49) Hadacek, F. *Crit. Rev. Plant Sci.* **2002**, *21*, 273−322.
(50) Harborne, J. B. *Introduction to Ecological Biochemistry*, 4th ed.; Academic Press: London, New York, 1994.
(51) Stinson, K. A.; Campbell, S. A.; Powell, J. R.; Wolfe, B. E.; Callaway, R. M.; Thelen, G. C.; Hallett, S. G.; Prati, D.; Klironomos, J. N. *PLoS Biol.* **2006**, *4*, 727−731.
(52) Hierro, J. L.; Callaway, R. M. *Plant Soil* **2003**, *256*, 29−39.
(53) Callaway, R. M.; Aschehoug, E. T. *Science* **2000**, *290*, 521−523.
(54) Reigosa, M. J.; Sanchez-Moreiras, A.; Gonzalez, L. *Crit. Rev. Plant Sci.* **1999**, *18*, 577−608.
(55) Bligny, R.; Douce, R. *Curr. Opin. Plant Biol.* **2001**, *4*, 191−196.
(56) Kruger, N. J.; Ratcliffe, R. G.; Roscher, A. *Phytochem. Rev.* **2003**, *2*, 17−30.
(57) Ratcliffe, R. G.; Shachar-Hill, Y. *Biol. Rev.* **2005**, *80*, 27−43.
(58) Ratcliffe, R. G.; Shachar-Hill, Y. *Plant J.* **2006**, *45*, 490−511.
(59) Wink, M. *Phytochemistry* **2003**, *64*, 2−19.
(60) Dewick, P. M. *Medicinal Natural Products: A Biosynthetic Approach*, 2nd ed.; John Wiley & Sons: New York, 2002.
(61) Tulp, M.; Bohlin, L. *Trends Pharmacol. Sci.* **2002**, *23*, 225−231.
(62) Tulp, M.; Bohlin, L. *Bioorg. Med. Chem.* **2005**, *13*, 5274−5282.
(63) Pichersky, E.; Noel, J. P.; Dudareva, N. *Science* **2006**, *311*, 808−811.
(64) Holmes, E.; Tang, H.; Wang, Y.; Seger, C. *Planta Med.* **2006**, *72*, 771−785.
(65) Niessen, W. M. A. *Liquid Chromatography−Mass Spectrometry*; Chromatographic Science Series; Marcel Dekker: New York, 1999; Vol. 79, pp 31−70.
(66) Niessen, W. M. A. *J. Chromatogr., A* **2003**, *1000*, 413−436.
(67) Hopfgartner, G.; Varesio, E.; Tschappat, V.; Grivet, C.; Bourgogne, E.; Leuthold, L. A. *J. Mass Spectrom.* **2004**, *39*, 845−855.
(68) Chace, D. H. *Chem. Rev.* **2001**, *101*, 445−478.
(69) Glish, G. L.; Vachet, R. W. *Nat. Rev. Drug Discovery* **2003**, *2*, 140−150.
(70) Raffaelli, A.; Saba, A. *Mass Spectrom. Rev.* **2003**, *22*, 318−331.
(71) Sparkman, O. D. *Spectrosc. Eur.* **2006**, *18*, 3−4.
(72) Hanold, K. A.; Fischer, S. M.; Cormia, P. H.; Miller, C. E.; Syage, J. A. *Anal. Chem.* **2004**, *76*, 2842−2851.
(73) Merchant, M.; Weinberger, S. R. *Electrophoresis* **2000**, *2*, 1164−1177.
(74) Wie, J.; Buriak, J.; Siuzdak, G. *Nature* **1999**, *399*, 243−246.
(75) Vaidyanathan, S.; Jones, D.; Broadhurst, D. I.; Ellis, J.; Jenkins, T.; Dunn, W. B.; Hayes, A.; Burton, N.; Oliver, S. G.; Kell, D. B; Goodacre, R. *Metabolomics* **2005**, *1*, 243−250.
(76) Vaidyanathan, S.; Gaskell, S.; Goodacre, R. *Rapid Commun. Mass Spectrom.* **2006**, *20*, 1192−1198.
(77) Domon, B.; Aebershold, R. *Science* **2006**, *312*, 212−217.
(78) Balogh, M. P. *LC−GC Eur.* **2004**, *17*, 152−159.
(79) Makarov, A.; Denisov, E.; Lange, O.; Horning, S. *J. Am. Soc. Mass Spectrom.* **2006**, *17*, 977−982.
(80) Makarov, A.; Denisov, E.; Kholomeev, A.; Balschun, W.; Lange, O.; Strupat, K.; Horning, S. *Anal. Chem.* **2006**, *78*, 2113−2120.
(81) Olsen, J. V.; DeGodoy, L. M.; Li, G.; Macek, B.; Mortensen, P.; Pesch, R.; Makarov, A.; Lange, O.; Horning, S.; Mann, M. *Mol. Cell. Proteomics* **2005**, *4*, 2010−2021.
(82) Lehotay, S. J.; Hajslova, J. *Trends Anal. Chem.* **2002**, *21*, 686−697.
(83) Wagner, C.; Sefkow, M.; Kopka, J. *Phytochemistry* **2003**, *62*, 887−900.
(84) Schauer, N.; Steinhauser, D.; Strelkov, S.; Schomburg, D.; Allison, G.; Moritz, T.; Lundgren, K.; Roessner−Tunali, U.; Forbes, M. G.; Willmitzer, L.; Fernie, A. R.; Kopka, J. *FEBS Lett.* **2005**, *579*, 1332−1337.
(85) http://csbdb.mpimp-golm.mpg.de/ as of 2006-09-10.
(86) http://fiehnlab.ucdavis.edu/ as of 2006-09-10.
(87) Catchpole, G. S.; Beckmann, M.; Enot, D. P.; Mondhe, M.; Zywicki, B.; Taylor, J.; Hardy, N.; Smith, A.; King, R. D.; Kell, D. B.; Fiehn, O.; Draper, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 14458−14462.
(88) Zywicki, B.; Catchpole. G.; Draper, J.; Fiehn, O. *Anal. Biochem.* **2005**, *336*, 178−186.
(89) Moco, S.; Bino, R. J.; Vorst, O.; Verhoeven, H. A.; De Groot, J.; Van Beek, T. A.; Vervoort, J.; De Vos, C. H. R. *Plant Physiol.* **2006**, *141*, 1205−1218.
(90) Servais, A. C.; Crommen, J.; Fillet, M. *Electrophoresis* **2006**, *27*, 2616−2629.
(91) Schmitt-Kopplin, P.; Frommberger, M. *Electrophoresis* **2003**, *24*, 3837−2867.
(92) Von Brocke, A.; Nicholson, G.; Bayer, E. *Electrophoresis* **2001**, *22*, 1251−1266.
(93) Klampfl, C. W. *Electrophoresis* **2006**, *27*, 3−34.
(94) Hazekamp, A.; Choi, Y. H.; Verpoorte, R. *Chem. Pharm. Bull.* **2004**, *52*, 718−721.
(95) Kim, H. K.; Choi, Y., H.; Chang, W. T.; Verpoorte, R. *Chem. Pharm. Bull.* **2003**, *51*, 1382−1385.
(96) Choi, Y. H.; Choi, H. K.; Hazekamp, A.; Bermejo, P.; Schilder, Y.; Erkelens, C.; Verpoorte, R. *Chem. Pharm. Bull.* **2003**, *51*, 158−161.
(97) Bilia, A. R.; Bergonzi, M. C.; Lazari, D.; Vincieri, F. F. *J. Agric. Food Chem.* **2002**, *50*, 5016−5025.
(98) Martin, G. J.; Guillou, C.; Martin, M. L.; Cabanis, M. T.; Tep, Y.; Aerny, J. *J. Agric. Food Chem.* **1988**, *36*, 316−322.
(99) Adam, P.; Arigoni, D.; Bacher, A.; Eisenreich, W. *J. Med. Chem.* **2002**, *45*, 4786−4793.
(100) Eisenreich, W.; Bacher, A.; Arigoni, D.; Rohdich, F. *Cell. Mol. Life Sci.* **2004**, *61*, 1401−1426.
(101) Bock, J. L. *Clin. Chem.* **1982**, *28*, 1873−1877.
(102) Bales, J. R.; Higham, D. P.; Howe, I.; Nicholson, J. K.; Sadler, P. J. *Clin. Chem.* **1984**, *30*, 426−432.
(103) Bales, J. R.; Nicholson, J. K.; Sadler, P. J. *Clin. Chem.* **1985**, *31*, 757−762.
(104) Belton, P. S.; Colquhoun, I. J.; Kemsley, E. K.; Delgadillo, I.; Roma, P.; Dennis, M. J.; Sharman, M.; Holmes, E.; Nicholson, J. K.; Spraul, M. *Food Chem.* **1998**, *61*, 207−213.
(105) Halket, J. M.; Waterman, D.; Przyborowska, A. M.; Patel, R. K. P.; Fraser, P. D.; Bramley, P. M. *J. Exp. Bot.* **2005**, *56*, 219−243.
(106) Broeckling, C. D.; Reddy, I. R.; Duran, A. L.; Zhao, X.; Sumner, L. W. *Anal. Chem.* **2006**, *78*, 4334−4341.
(107) Smith, C. A.; Want, E. J.; O'Maille, G.; Abagyan, R.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 779−787.
(108) Jonsson, P.; Gullberg, J.; Nordström, A.; Kusano, M.; Kowalczyk, M.; Sjöström, M.; Moritz, T. *Anal. Chem.* **2004**, *76*, 1738−1745.
(109) Jonsson, P.; Johansson, E. S.; Wuolikainen, A.; Lindberg, J.; Schuppe-Koistinen, I.; Kusano, M.; Sjöström, M.; Tyrgg, J.; Moritz, T.; Antti, H. *J. Proteome Res.* **2006**, *5*, 1407−1414.
(110) Nordström, A.; O'Maille, G.; Qin, C.; Siuzdak, G. *Anal. Chem.* **2006**, *78*, 3289−3295.
(111) Jonsson, P.; Johansson, A. I.; Gullberg, J.; Tyrgg, J. A. J.; Grung, B.; Marklund, S.; Sjöström, M.; Antti, H.; Moritz, T. *Anal. Chem.* **2005**, *77*, 5635−5642.
(112) Roessner, U.; Wanger, C.; Kopka, J.; Trethewey, R. N.; Willmitzer, L. *Plant J.* **2000**, *23*, 131−142.
(113) Roessner, U.; Luedemann, A.; Brust, D.; Fiehn, O.; Linke, T.; Willmitzer, L.; Fernie, A. R. *Plant Cell* **2001**, *13*, 11−29.
(114) Bino, R. J.; De Vos, C. H. R.; Liebermann, M.; Hall, R. D.; Bovy, A. G.; Jonker, H. H.; Tikunov, Y.; Lommen, A.; Moco, S.; Levin, I. *New Phytol.* **2005**, *166*, 427−438.
(115) Roessner-Tunali, U.; Hegemann, B.; Lytovchenko, A.; Carrari, F.; Bruedigam, C.; Granot, D.; Fernie, A. R. *Plant Physiol.* **2003**, *133*, 84−99.
(116) Schauer, N.; Zamir, D.; Fernie, A. R. *J. Exp. Bot.* **2005**, *56*, 297−307.
(117) Fiehn, O. *Phytochemistry* **2003**, *62*, 875−886.
(118) Hollywood, K.; Brison, D. R.; Goodacre, R. *Proteomics* **2006**, *6*, 4716−4723.
(119) Goodacre, R.; Vaidyanathan, S.; Dunn, W. B.; Harrigan, G. G.; Kell, D. B. *Trends Biotechnol.* **2004**, *22*, 245−252.
(120) Kell, D. B. *FEBS J.* **2006**, *273*, 873−894.
(121) Weckwerth, W.; Wenzel, K.; Fiehn, O. *Proteomics* **2004**, *4*, 78−83.
(122) Tikunov, Y.; Lommen, A.; De Vos, C. H. R.; Verhoeven, H. A.; Bino, R. J.; Hall, R. D.; Bovy, A. G. *Plant Physiol.* **2005**, *139*, 1125−1137.
(123) Goodacre, R. *J. Exp. Bot.* **2005**, *56*, 245−254.
(124) Fukusaki, E.; Kobayashi, A. *J. Biosci. Bioeng.* **2005**, *100*, 347−354.
(125) Weckwerth, W.; Morgenthal, K. *Drug Discovery Today* **2005**, *10*, 1551−1558.
(126) Batagelj, V.; Mrvar, A. *Connections* **1998**, *21*, 47−57.
(127) http://vlado.fmf.uni-lj.si/pub/networks/pajek/ as of 2006-09-10.
(128) Camacho, D.; De la Fuente, A.; Mendes, P. *Metabolomics* **2005**, *1*, 53−63.
(129) Steuer, R.; Kurths, J.; Fiehn, O.; Weckwerth, W. *Bioinformatics* **2003**, *19*, 1019−1026.
(130) Dumas, M. E.; Maibaum, E. C.; Teague, C.; Ueshima, H.; Zhou, B.; Lindon, J. C.; Nicholson, J. K.; Stamler, J.; Elliott, P.; Chan, Q.; Holmes, E. *Anal. Chem.* **2006**, *78*, 2199−2208.
(131) Lindon, J. C.; Nicholson, J. K.; Holmes, E.; Antti, H.; Bollard, M. E.; Keun, H.; Beckonert, O.; Ebbels, T. M.; Reily, M. D.; Robertson, D.; Stevens, G. J.; Luke, P.; Breau, A. P.; Cantor, G. H.; Bible, R. H.; Niederhauser, U.; Senn, H.; Schlotterbeck, G.; Sidelmann, U. G.; Laursen, S. M.; Tymiak, A.; Car, B. D.; Lehman-McKeeman, L.; Colet, J. M.; Loukaci, A.; Thomas, C. *Toxicol. Appl. Pharmacol.* **2003**, *187*, 137−146.

(132) Crockford, D. J.; Keun, H. C.; Smith, L. M.; Holmes, E.; Nicholson, J. K. *Anal. Chem.* **2005**, *77*, 4556−4562.
(133) Defernez, M.; Colquhoun, I. J. *Phytochemistry* **2003**, *62*, 1009−1117.
(134) Craig, A.; Cloarec, O.; Holmes, E.; Nicholson, J. K.; Lindon, J. C. *Anal. Chem.* **2006**, *78*, 2262−2267.
(135) Van den Berg, R. A.; Hoefsloot, H. C. J.; Westerhuis, J. A.; Smilde, A. K.; Van der Werf, M. J. *BMC Genomics* **2006**, *7*, 142.
(136) Holmes, E.; Antti, H. *Analyst* **2002**, *127*, 1549−1557.
(137) Wold, S.; Sjostrom, M.; Eriksson, L. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109−130.
(138) Defernez, M.; Kemsley, E. K. *Trends Anal. Chem.* **1997**, *16*, 216−221.
(139) Eriksson, L.; Antti, H.; Gottfries, J.; Holmes, E.; Johansson, E.; Lindgren, F.; Long, I.; Lundstedt, T.; Trygg, J.; Wold, S. *Anal. Bioanal. Chem.* **2004**, *380*, 419−429.
(140) Trygg, J. *J. Chemom.* **2002**, *16*, 283−293.
(141) Beckwith-Hall, B. M.; Brindle, J. T.; Barton, R. H.; Coen, M.; Holmes, E.; Nicholson, J. K.; Antti, H. *Analyst* **2002**, *127*, 1283−1288.
(142) Cloarec, O.; Dumas, M. E.; Trygg, J.; Craig, A.; Barton, R. H.; Lindon, J. C.; Nicholson, J. K.; Holmes, E. *Anal. Chem.* **2005**, *77*, 517−526.
(143) Cloarec, O.; Dumas, M. E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J. K. *Anal. Chem.* **2005**, *77*, 1282−1289.
(144) Crockford, D. J.; Holmes, E.; Lindon, J. C.; Plumb, R. S.; Zirah, S.; Bruce, S. J.; Rainville, P.; Stumpf, C. L.; Nicholson, J. K. *Anal. Chem.* **2006**, *78*, 363−371.
(145) Spence, D. M. *Analyst* **2004**, *129*, 102−104.
(146) Taverniers, I.; De Loose, M.; Van Bockstaele, E. *Trends Anal. Chem.* **2004**, *23*, 480−490.
(147) Taverniers, I.; De Loose, M.; Van Bockstaele, E. *Trends Anal. Chem.* **2004**, *23*, 535−552.
(148) Guidance for Industry; Bioanalytical Method Validation U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER) and Center for Veterinary Medicine (CVM), Rockville, MD, 2001 (http://www.fda.gov/cder/guidance/index.htm).
(149) Guideline for Industry: Text on Validation of Analytical Procedures, ICH−Q2A, 1995 (http://www.fda.gov/cder/guidance/index.htm).
(150) Eurachem Guide: The fitness for purpose of analytical methods. A laboratory guide to method validation and related topics, LGC, Teddington, Middlesex, U.K., 1998 (http://www.Eurachem.bam.de).
(151) Sato, S.; Soga, T.; Nishioka, T.; Tomita, M. *Plant J.* **2004**, *40*, 151−163.
(152) Villas-Boas, S. G.; Delicado, D. G.; Akesson, M.; Nielsen, J. *Anal. Biochem.* **2003**, *322*, 134−138.
(153) Soga, T.; Kakazu, Y.; Robert, M.; Tomita, M.; Nishioka, T. *Electrophoresis* **2004**, *25*, 1964−1972.
(154) Shurubor, Y. I.; Paolucci, U.; Krasinov, B. F.; Matson, W. R.; Kristal, B. S. *Metabolomics* **2005**, *1*, 75−85.
(155) Soga, T.; Ohashi, Y.; Ueno, Y.; Naraoka, H.; Tomita, M.; Nishioka, T. *J. Proteome Res.* **2003**, *2*, 488−494.
(156) Sakakibara, H.; Honda, Y.; Nakagawa, S.; Ashida, H.; Kanazawa, K. *J. Agric. Food Chem.* **2003**, *51*, 571−581.
(157) Müller, C. A.; Weinmann, W.; Dresen, S.; Schreiber, A.; Gergov, M. *Rapid Commun. Mass Spectrom.* **2005**, *19*, 1332−1338.
(158) Fiehn, O.; Kopka, J.; Trethewey, R. N.; Willmitzer, L. *Anal. Chem.* **2000**, *72*, 3573−3580.
(159) Aharoni, A.; De Vos, C. H. R.; Verhoeven, H. A.; Maliepaad, C. A.; Kruppa, G.; Bino, R.; Goodenowe, D. B. *OMICS* **2002**, *6*, 217−233.
(160) Weckwerth, W.; Loureiro, M. E.; Wenzel, K.; Fiehn, O. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 7809−7814.
(161) Kaplan, F.; Kopka, J.; Haskell, D. W.; Zhao, W.; Schiller, K. C.; Gatzke, N.; Sung, D. Y.; Guy, C. L. *Plant Physiol.* **2004**, *136*, 4159−4168.
(162) Urbanczyk-Wochinak, E.; Luedermann, A.; Kopka, J.; Selbig, J.; Roessner−Tunail, U.; WIllmitzer, L.; Fernie, A. R. *EMBO Rep.* **2003**, *4*, 1−5.
(163) Hirai, M. Y.; Yano, M.; Goodenowe, D. B.; Kanaya, S.; Kimura, T.; Awazuhara, M.; Arita, M.; Fujiwara, T.; Saito, K. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10205−10210.
(164) Morgenthal, K.; Wienkoop, S.; Scholz, M.; Selbig, J.; Weckwerth, W. *Metabolomics* **2005**, *1*, 109−121.
(165) Keurentjes, J. J. B.; Fu, J.; De Vos, C. H. R.; Lommen, A.; Hall, R. D.; Bino, R. J.; Van den Plas, L. H. W.; Jansen, R. C.; Vreugdenhil, D.; Koornneef, M. *Nat. Genet.* **2006**, *38*, 842−849.
(166) Morreel, K.; Goeminne, G.; Storme, V.; Sterck, L.; Ralph, J.; Coppieters, W.; Breyne, P.; Steenackers, M.; Georges, M.; Messens, E.; Boerjan, W. *Plant J.* **2006**, *47*, 224−237.
(167) Rischer, H.; Oresic, M.; Seppänen-Laasko, T.; Katajamaa, M.; Lammertyn, F.; Ardiles-Diaz, W.; Van Montagu, M. C. E.; Inze, D.; Oksman-Caldentery, K. M.; Goossens, A. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 5614−5619.
(168) Pan, M. H.; Liang, Y. C.; Lin-Shiau, S. Y.; Zhu, N. Q.; Ho, C. T.; Lin, J. K. *J. Agric. Food Chem.* **2000**, *48*, 6337−6346.
(169) Kazi, A.; Smith, D. M.; Zhong, Q.; Dou, Q. P. *Mol. Pharmacol.* **2002**, *62*, 765−771.
(170) Vergote, D.; Cren-Olive, C.; Chopin, V.; Toillon, R. A.; Rolando, C.; Hondermarck, H.; Le Bourhis, X. *Breast Cancer Res. Treat.* **2002**, *76*, 195−201.
(171) Leone, M.; Zhai, D.; Sareth, S.; Kitada, S.; Reed, J. C.; Pellecchia, M. *Cancer Res.* **2003**, *63*, 8118−8121.
(172) Le Gall, G.; Colquhoun, I. J.; Defernez, M. *J. Agric. Food Chem.* **2004**, *52*, 692−700.
(173) Defernez, M.; Colquhoun, I. J. *Phytochemistry* **2003**, *62*, 1009−1017.
(174) Vlahov, G. *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *35*, 341−357.
(175) Mannina, L.; Patumi, M.; Proietti, N.; Bassi, D.; Segre, A. L. *J. Agric. Food Chem.* **2001**, *49*, 2687−2696.
(176) Mannina, L.; Patumi, M.; Proietti, N.; Segre, A. L. *Ital. J. Food Sci.* **2001**, *13*, 53−63.
(177) Brescia, M. A.; Alviti, G.; Liuzzi, V.; Sacco, A. *J. Am. Oil Chem. Soc.* **2003**, *80*, 945−950.
(178) Vlahov, G.; Del Re, P.; Simone, N. *J. Agric. Food Chem.* **2003**, *51*, 5612−5615.
(179) Mannina, L.; Dugo, G.; Salvo, F.; Cicero, L.; Ansanelli, G.; Calcagni, C.; Segre, A. L. *J. Agric. Food Chem.* **2003**, *51*, 120−127.
(180) Rezzi, S.; Axelson, D. E.; Heberger, K.; Remiero, F.; Mariani, C.; Guillou, C. *Anal. Chim. Acta* **2005**, *552*, 13−24.
(181) Garcia-Gonzalez, D. L.; Mannina, L.; D'Imperio, M.; Segre, A. L.; Aparicio, R. *Eur. Food Res. Technol.* **2004**, *219*, 545−548.
(182) Fragaki, G.; Spyros, A.; Siragakis, G.; Salivaras, E.; Dais, P. *J. Agric. Food Chem.* **2005**, *53*, 2810−2816.
(183) Bailey, N. J. C.; Wang, Y. L.; Sampson, J.; Davis, W.; Whitcombe, I.; Hylands, P. J.; Croft, S. L.; Holmes, E. *J. Pharm. Biomed. Anal.* **2004**, *35*, 117−126.
(184) Roos, G.; Röseler, C.; Berger−Büter, K.; Simmen, U. *Planta Med.* **2004**, *70*, 771−777.
(185) Choi, Y. H.; Sertic, S.; Kim, H. K.; Wilson, E. G.; Michopoulos, F.; Lefeber, A. W. M.; Erkelens, C.; Kricun, S. D. P.; Verpoorte, R. *J. Agric. Food Chem.* **2005**, *53*, 1237−1245.
(186) Wang, Y.; Tang, H.; Nicholson, J. K.; Hylands, P. J.; Sampson, J.; Whitcombe, I.; Stewart, C. G.; Caiger, S.; Oru, I.; Holmes, E. *Planta Med.* **2004**, *70*, 250−255.
(187) Kim, H. K.; Choi, Y. H.; Erkelens, C.; Lefebre, A. W. M.; Verpoorte, R. *Chem. Pharm. Bull.* **2005**, *53*, 105−109.
(188) Yang, S. Y.; Kim, H. K.; Lefeber, A. W. M.; Erkelens, C.; Angelova, N.; Choi, Y. H.; Verpoorte, R. *Planta Med.* **2006**, *72*, 364−369.
(189) Rasmussen, B.; Cloarec, C.; Tang, H.; Staerk, D.; Jaroszewski, J. W. *Planta Med.* **2006**, *72*, 556−563.
(190) Seger, C.; Sturm, S.; Humpfer, E.; Schäfer, H.; Spraul, M.; Stuppner, H. *Differentiation of Hypericum perforatum (St. John's wort)* lots by NMR based metabonomics. Presented as poster at the 53rd Congress of the Society for Medicinal Plant Research, Florence, Italy, 2005; P372, book of abstracts, p 298). The presentation is available at http://homepage.uibk.ac.at/~c74058/Florenz_GA2005_Hypericum.pdf (accessed November 5, 2006).
(191) Le Gall, G.; Colquhoun, I. J.; Davis, A. L.; Collins, G. J.; Verhoeyen, M. E. *J. Agric. Food Chem.* **2003**, *51*, 2447−2456.
(192) Le Gall, G.; DuPont, M. S.; Mellon, F. A.; Davis, A. L.; Collins, G. J.; Verhoeyen, M. E.; Colquhoun, I. J. *J. Agric. Food Chem.* **2003**, *51*, 2438−2446.
(193) Defernez, M.; Gunning, Y. M.; Parr, A. J.; Shepherd, L. V. T.; Davies, H. V.; Colquhoun, I. J. *J. Agric. Food Chem.* **2004**, *52*, 6075−6085.
(194) Moing, A.; Maucourt, M.; Renaud, C.; Gaudillere, M.; Brouquisse, R.; Lebouteiller, B.; Gousset-Dupont, A.; Vidal, J.; Granot, D.; Denoyes-Rothan, B.; Lerceteau-Köhler, E.; Rolin, D. *Funct. Plant Biol.* **2004**, *31*, 889−902.
(195) Akoka, S.; Baratin, L.; Tierweiler, M. *Anal. Chem.* **1999**, *71*, 2554−2557.
(196) Le Gall, G.; Metzdorff, S. B.; Pedersen, J.; Benett, R. N.; Colquhoun, I. J. *Metabolomics* **2005**, *1*, 181−198.

PR0604716