TUTORIAL

Basics of mass spectrometry based metabolomics

Frédérique Courant[1,2], Jean-Philippe Antignac[2,3], Gaud Dervilly-Pinel[2] and Bruno Le Bizec[2]

[1] Department of Environmental Sciences and Public Health, University of Montpellier 1, UMR 5569 Hydrosciences, Montpellier, France
[2] Laboratoire d'Etude des Résidus et Contaminants dans les Aliments (LABERCA), LUNAM Université Oniris, USC INRA 1329, Nantes, France
[3] INRA, Nantes, France

The emerging field of metabolomics, aiming to characterize small molecule metabolites present in biological systems, promises immense potential for different areas such as medicine, environmental sciences, agronomy, etc. The purpose of this article is to guide the reader through the history of the field, then through the main steps of the metabolomics workflow, from study design to structure elucidation, and help the reader to understand the key phases of a metabolomics investigation and the rationale underlying the protocols and techniques used. This article is not intended to give standard operating procedures as several papers related to this topic were already provided, but is designed as a tutorial aiming to help beginners understand the concept and challenges of MS-based metabolomics. A real case example is taken from the literature to illustrate the application of the metabolomics approach in the field of doping analysis. Challenges and limitations of the approach are then discussed along with future directions in research to cope with these limitations. This tutorial is part of the International Proteomics Tutorial Programme (IPTP18).

Received: June 4, 2014
Revised: July 18, 2014
Accepted: August 26, 2014

Keywords:
Challenges / Future directions / Mass spectrometry / Metabolomics / Systems biology / Workflow

Additional supporting information may be found in the online version of this article at the publisher's web-site

1 Historical background

The high complexity of living organisms and biological systems imposes new more integrated and global (namely untargeted) characterization approaches that would allow addressing, in a comprehensive manner, complex situations that are currently dealt in a piecemeal manner. Indeed, the efficiency of conventional targeted approaches is well established even if some limitations are acknowledged. Targeted methods are then sensitive and specific but they do focus only on particular compounds or activities. With such methods, we search for the known and we only find what we have been searching. Unidentified compounds and potentially novel biomarkers remain undetected, unless unbiased (nontargeted) approaches are employed.

The concept of biological phenotyping has then emerged and comprehensive "omics" approaches have become for the last years a new way for addressing life complexity. There are three main characterization levels of biological systems, namely genomics–transcriptomics that have emerged in the 80s [1], proteomics in the 90s [2], and the most recent one, metabolomics, introduced about 15 years ago and in constant evolution since [3] (Fig. 1). Transcriptomics refers to the global study of gene expression at a given time point and the transcriptome therefore corresponds to the complete set of RNA transcripts present in particular cells. Proteomics relates to the study of protein expression in a given biological system. Metabolomics corresponds to the study of small molecules as ultimate cellular signaling events resulting from transcriptional and translational changes [4]. While

Correspondence: Dr. Frédérique Courant, Département Sciences de l'Environnement et Santé Publique, Faculté de Pharmacie, UMR Hydrosciences, Université Montpellier 1, 15, Av Charles Flahault, 34093 Montpellier Cedex 05, France
E-mail: frederique.courant@univ-montp1.fr

Abbreviations: EI, electron ionization; **IMS**, ion-mobility spectrometry; **SFC**, supercritical fluid chromatography

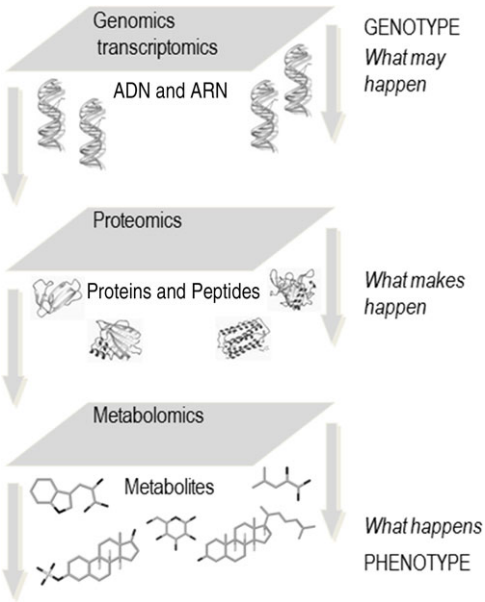

Figure 1. The three different biological levels of omics approaches.

the genome, transcriptome, and proteome can be seen as mediums in the flow of gene expression, the metabolome represents the most downstream level that reflects changes in phenotype and function [5].

The terms metabonomics [6] and metabolomics [7] appeared at the end of the 1990s and early 2000s, respectively. Although sometimes similarly used in practice, the exact definitions of both terms may be differentiated. "Metabonomics is defined as the quantitative measurement of the dynamic multi-parametric metabolic response of living systems to pathophysiological stimuli or genetic modification" [4], while "Metabolomics corresponds to the study of the complete set of metabolites/low-molecular-weight intermediates, which are context dependent, varying according to the physiology, developmental or pathological state of the cell, tissue, organ or organism" [3]. In most cases, the approach consists in a differential study of metabolomes generated from "control" and "test" subgroups of observations to find differences in their profiles in response to external stimuli (pathologies, effect of biochemical or environmental stresses, food processing, etc.) (Fig. 2).

Even if comprehensive quantitative analysis of metabolites had already been achieved in the 1970s through GC-MS [8, 9], at the end of the 1990s, the combined improvement in analytical instrumentation and informatics have enabled the acquisition of comprehensive metabolic profiles and their mining using appropriate multivariate statistical tools. These tools were subsequently used for metabolomics proof-of-principle studies in a variety of species and applications. Historically, the most widely used technique for metabolomics purposes is NMR. NMR spectroscopy provides a rapid, non-destructive, high-throughput method that requires minimal sample preparation. Moreover, robustness of the NMR equipments is a key parameter that has not been equaled to date by other types of instruments. However, MS-based methods have proved now for years to be valuable for such studies, especially thanks to recent technological advances. Indeed, MS offers higher performance in terms of sensitivity than NMR, which is extremely useful for measuring species with low abundance put potentially valuable information. Moreover,

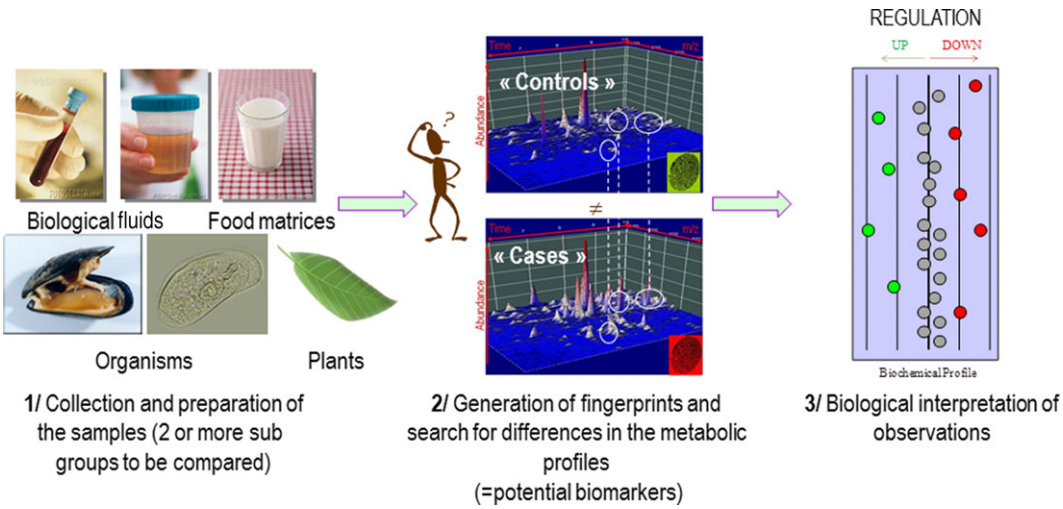

Figure 2. Global principle of a typical metabolomics experiment. Samples from two or more subgroups are collected, prepared, and characterized through a comprehensive analytical technique. The raw data are then processed and differences that correspond to potential biomarkers are revealed. The structural elucidation of the differential signals (biomarker identification) is expected to improve the knowledge regarding the metabolic differences between the two subgroups from a descriptive, explicative/mechanistic, and/or predictive/diagnostic point of view.

the specificity of MS (through high-resolution and/or multidimensional MSn techniques) can help and even facilitate elucidation of the chemical structures of potential metabolites of interest. One drawback of MS may be its need for higher sample preparation than NMR and its possible hyphenation to a separation technique (either LC or GC), to separate metabolic components from the samples, which may extend the time of analysis.

The purpose of this article is to guide the reader through the main steps of the MS-based metabolomics workflow, from study design to structure elucidation, and help the reader to understand the key phases of a metabolomics investigation and the rationale underlying the protocols and techniques used. This article is not intended to give standard operating procedures as several papers related to this topic were already provided, but is designed as a tutorial aiming to help beginners understanding the concept and challenges of MS-based metabolomics.

2 Basic concepts

2.1 What is the metabolome?

The metabolome is defined as the complete set of metabolites present in an organism. It is admitted that the metabolome basically refers to small molecular species (from 50 to 1500 Da), which are endogenous organic substances (intermediates and products of metabolism, which is the set of chemical reactions that occur in living organisms), although inorganic and elemental species can also be studied [10], and the metabolome includes as well xenobiotics. Endogenous metabolites can be classified as primary and secondary metabolites. Primary metabolites are directly involved in processes essential to life such as normal growth, development, and reproduction (amino acids, organic acids, etc.). Secondary metabolites are not directly involved in those processes and are not essential to sustain cells life. They are synthesized for particular biological functions and are present in a taxonomically restricted set of organisms or cells (e.g., alkaloids in plants or antibiotics in fungi). Xenobiotics are foreign chemical substances found within an organism that is not naturally produced by or expected to be present within that organism. They may come from external sources such as diet, medications, etc. The "xenometabolome" refers to metabolites generated by the biotransformation of xenobiotics, mainly in liver for mammals [11]. It consists in phase I reactions mostly corresponding to oxidation–reduction reactions acting to detoxify the compounds and phase II reactions corresponding to conjugation to polar moieties for elimination purposes.

The metabolome represents a vast number of components that belong to a wide variety of compound classes. This provides wide variations in chemical (molecular weight, polarity, solubility) and physical (volatility) properties ranging from ionic inorganic species to hydrophilic carbohydrates and secondary natural products to hydrophobic lipids. Moreover,

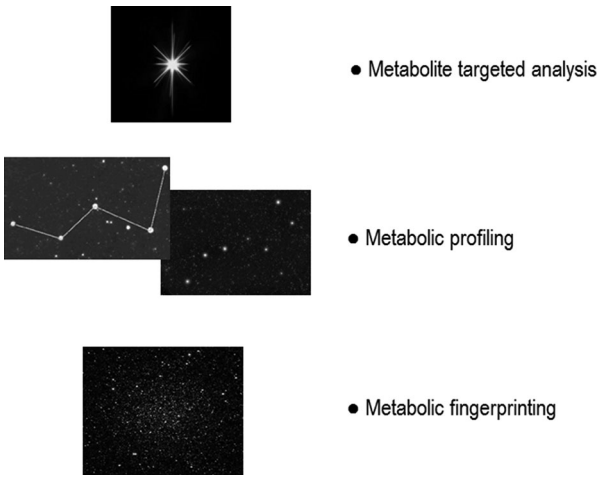

Figure 3. The three different approaches that are encountered in MS-based metabolomics. The metabolite targeted analysis corresponds to the quantification of one or a restricted number of compounds (e.g., one star in the sky). The metabolic profiling corresponds to the monitoring and (semi)quantification of a metabolic pathway or a particular class of compounds (e.g., a constellation in the sky, e.g., Cassiopeia constellation or Ursa major constellation). Metabolic fingerprinting corresponds to the (semi)quantification of all the metabolites accessible to the analysis (e.g., the whole sky).

metabolites occur in a wide concentration range, over an estimated 7–9 magnitude of concentrations (pmol–mmol) [12]. The size of the metabolome varies greatly, depending on the organism studied; *Saccharomyces cerevisiae* contains approximately 600 metabolites [13] and the plant kingdom has an estimated 200 000 primary and secondary metabolites [7]. The human metabolome may be narrower than the plant one. However, both the hormone complexity and the input of food (and medicine) metabolites may lead to a much wider metabolome in human [14]. At present, more than 4200 compounds have been annotated in human metabolite databases [15]. Consequently, comprehensive metabolomics investigations are primarily a challenge for analytical chemistry and specifically MS has vast potential as a tool for this type of investigation [16].

2.2 MS-based metabolomics: Different approaches

Current MS-based metabolomics investigations can be categorized as three approaches roughly classified according to the data quality and the number of metabolites that can be detected [12] (Fig. 3). First is the metabolite targeted analysis that refers to the detection and precise quantification of a single or small set of target compounds. Second is metabolic profiling, which focuses on the analysis of a group of metabolites either related to a specific metabolic pathway or a class of compounds. In most cases, metabolic profiling is a hypothesis-driven approach: metabolites identities (or chemical family)

are usually known a priori. For these two approaches, particular metabolites are selected for analysis depending on the investigated study, and specific analytical methods are developed for their determination. As an example in metabolic profiling, endogenous glucuronides (corresponding to phase II metabolites, see Section 2.1) excreted in human urine can be selectively monitored for classification and prediction of gender [17]. The advantage of both approaches is that the structure of the monitored metabolites is partly (or fully) known that facilitates their structure elucidation and associated biological interpretation. Numerous quantitative metabolic profiling methods analyzing different metabolite subsets have already been developed and are routinely used. If these methods measuring key metabolites from different biochemical pathways are assembled as building blocks to study the metabolome, a powerful metabolomics approach can evolve [16]. The third approach toward metabolomics is metabolic fingerprinting used for an extended metabolome comparison. The objective is to compare patterns of all the metabolites accessible to the analysis that change in response to the studied factor. This approach is not driven by a priori hypothesis, therefore it is open to new findings. Its disadvantage may be that the identity of the metabolites of interest is established a posteriori and this remains a challenge (see Section 2.3.6). Metabolic fingerprinting can also be used strictly as an investigation tool when the intention is not to structurally elucidate each observed signals, but to use the combination of differential signals for diagnosis. For example, in horse racing doping analysis, a metabolic fingerprinting approach was developed as a new effect-based screening tool to tackle the illegal use of recombinant equine growth hormones [18].

2.3 How do we perform metabolomics analyses?

Beyond the concept of biological signature, or more particularly those of metabolic phenotypes, metabolomics requires a practical implementation of a global methodological framework [19]. This methodological framework is composed of several steps that will all affect the quantity and quality of the data obtained (Fig. 4). Briefly, a metabolomic study implies samples collection, preparation, and generation of metabolic profiles through a given analytical tool. The acquired data are then processed using dedicated deconvolution softwares and analyzed using appropriate statistical techniques to isolate a smaller number of potentially relevant metabolites that are believed to differentiate the subgroups of observations to be compared (differential metabolites and/or potential biomarkers). Follow-up experiments then focus on identification of these differential metabolites and further biological explanation of the observed changes.

Metabolomics represents then a multi- and interdisciplinary approach. It relies on three main disciplines: analytical chemistry, chemometrics, and biology. As mentioned already, the metabolome is characterized by a high diversity in terms of physicochemical properties and concentrations

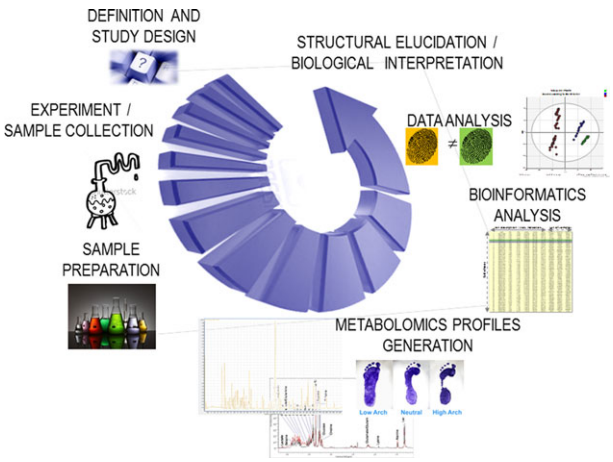

Figure 4. Typical analytical workflow constituting metabolomics studies. Each step, relying either on biology (study design, sample collection, biological interpretation), chemistry (sample preparation, metabolomics profiles generation, structure elucidation), or bioinformatics (bioinformatics analysis, structure elucidation) must be designed carefully to maximize the quantity and the quality of the information obtained from an experiment.

expected. Therefore, the sample preparation, metabolic profiles generation, and structure elucidation steps will imply several challenges that will need analytical chemistry to be solved. Considering the large number of signals detected in metabolomics approaches and especially in metabolic fingerprinting approach, chemometrics then appears necessary for extracting the most relevant information from the generated data. Finally, in the objective to contribute to biochemical knowledge and understanding of underlying mechanisms of action, biology is essential to make sense to the observations as well as understand the metabolic pathways of interest.

Within the scope of this article, we will describe the metabolomics workflow in a stepwise fashion with the objective to help the reader to understand the key phases of a metabolomics analysis, the challenges associated to each step, and the rationale underlying the different methodologies existing. This article is dedicated to MS-based metabolomics approach, therefore MS-related protocols will be particularly emphasized.

2.3.1 Study experimental design

If the different steps constituting a metabolomics study should all be carefully considered, the study design is particularly crucial since all subsequent results from the study will rely on the successful achievement of this step [20]. As it is the case with the quality of analysis of the experimental data, the study design directly influences the amount and quality of information obtained from a given experiment. A good experimental design must enable to distinguish specific from unspecific biological variations, in order to access more clearly to

the meaning of observed metabolic changes. A successful design of experiment must be designed or at least checked by the different disciplines involved in a metabolomic investigation. Biologists may clearly state the research question and may inform about biological factors and possible confounding ones. Indeed, since genetics [21], gender [22], environment [23], diet [24] (etc.) will dramatically influence the metabolic profiles, these factors have to be considered in study design and attention should be paid to homogenize controls and tests groups and to minimize possible confounders. Statisticians may validate the statistical power of the study to ensure that a sufficient number of samples are acquired and to reduce the influence of biological variability and obtain statistically validated data. The sample size required to get validated data is dependent on the experiment conducted, that is, the expected effect size, SD of the effect of interest, measurement variability, biological variance, etc. Therefore, there is no gold number regarding the number of replicates that should be included in a metabolomics experiment. For instance, five individuals may be sufficient in a study evaluating the effects of high-dose exposures to phthalates in rats [25], whereas tens of samples are required to highlight urinary biomarkers of citrus consumption in human [26]. Finally, analytical chemists may compile the recommendations of biologists and statisticians and ensure that (i) the technical variance is under control, and (ii) the number of samples to analyze is in agreement with the capabilities of the analytical laboratory.

When considering the design of a metabolomics study, an issue to consider is the time of the sampling. Indeed, intraindividual variations may obscure the metabolic profiles. For example, diurnal variations in metabolite content were observed in two mouse strains from which urine samples were collected in the morning and afternoon. Interpretation of the metabolic profiles established that the metabolic differences between the two mouse strains were confounded by the influence of diurnal variation, illustrating how normal physiological variation factors may hinder metabonomics analyses [27]. The metabolome may as well vary rapidly after intake of food. Analysis of the influence of a single intake of almond skin polyphenols on the 24-h kinetic trajectory of the human urinary metabolic profiles was assessed using LC MS. Modifications in the urinary metabolome were observable as soon as 6 h after the intake [28].

Two other questions must be answered during the study design, that is, (i) what type of sample will provide the most information for the particular investigation? and (ii) what types of samples are most feasible and easy to collect? [29] The biological samples most relevant in the study of mammalian biochemistry are biofluids, cells, or tissues. Each biofluid is unique in the type of information it has to offer: urine contains the highest number of water-soluble metabolites and provides a rich source of metabolic information. The composition of blood is well maintained through homeostatic control, which makes it less variable than urine [30]. However, there are significant chemical differences between serum and plasma due to their collection and therefore the choice between both biofluids may not be arbitrary [31]. Moreover, for plasma samples, EDTA is recommended as anticoagulant in sample collection, because peaks derived from heparin might overlap with endogenous metabolites (especially for GC-MS analysis), which may induce intersample variation [32]. Depending on the application, solid samples (e.g., tissues, food products, or plant material) may be collected and a rational choice has to be made. For example, concerning plant metabolomics, there is a large variety of available samples (e.g., leaves, roots, sap, fruits, flowers, etc.), which do not provide the same information in terms of accessible metabolism.

Samples collection must follow a single and reproducible procedure that has been defined during the study design. Indeed, these sampling conditions appear crucial in terms of representativeness and homogeneity of the samples [30, 32]. In particular, samples from cases and controls need to be treated identically as certain metabolites may be very susceptible to slight deviations from standard protocol. For instance, the devices into which the samples are collected should be chosen so as to minimize unwanted variation (use the same tubes for all the collected samples) and to avoid introducing external contaminants such as surfactants and plasticizers that can cause serious interference during the analysis. More globally each applied sample pretreatment can have direct consequences on the metabolic fingerprints generated (content and variability) [33]. As an example, for blood sample collection, details such as clotting time, clotting temperature, and treatment conditions prior to centrifugation need to be considered and standardized. It was demonstrated that variation due to clotting time caused changes in energy metabolites, which were delayed by clotting on ice [34]. Finally, following sample collection, special care must be taken to stop the formation or degradation of metabolites by quenching step in order to stabilize the sample by stopping metabolic reactions. Sample storage at −80°C or at least −20°C is recommended, and urine samples were shown to be stable for up to 6 months when stored in these conditions [35], while storage at 4°C without addition of preservative for prolonged periods induced a change in metabolic content due to microbial contamination [36].

2.3.2 Sample preparation

The extremely wide diversity of potential metabolites present in a biological sample in terms of physicochemical properties (hydrophilicity/hydrophobicity, volatility, chemical reactivity/stability) and concentrations makes the goal of measuring them all in metabolomics unrealistic. No single method of sample collection (sampling, quenching, storage) and preparation (extraction, dilution, clean-up) is applicable to all metabolites because conditions that stabilize one type of compound may destroy other types or interfere with their analysis. Methods are developed to be appropriate for the majority of metabolites while removing matrix components that will interfere with the analysis even if many crucial

metabolites, particularly minor or unstable ones, will be missed in metabolomics analyses. Indeed, any kind of sample preparation will cause analyte losses either total or partial. In the case of a partial loss of some analytes, the issue of sample preparation is not in the recovery but in the reproducibility of the protocol. Metabolomic studies typically comparing the relative levels of metabolites in a large number of samples, good reproducibility for a wide range of metabolites is absolutely necessary to ensure that small changes in these metabolite levels can be accurately determined. This may be simple for biofluids when little or no sample preparation is needed but may be much more complex for other matrices. Moreover, metabolism is in constant flux, the concentration of metabolites can change rapidly if appropriate measures are not taken during sampling and extraction, these measures must not denature or modify the sample. Considering those challenges, an ideal sample preparation method for global metabolomics should (i) incorporate a metabolism quenching step to represent true metabolome composition at the time of sampling; (ii) be as nonselective as possible to ensure adequate metabolite coverage; (iii) be simple and fast to prevent metabolite loss and/or degradation during the preparation procedure; and (iv) be reproducible [37].

Special care must be taken to stop the formation or degradation of metabolites by adequate sample preparation and storage conditions. A preliminary quenching step is therefore important in order to stabilize the sample by stopping metabolic reactions [38]. This can be done by use of low temperatures (cold solvent addition, freezing in liquid nitrogen), addition of acid, or fast heating [39]. Sample storage at −80°C and sample aliquoting at collection time in order to avoid multiple freeze-thawing cycles of the samples between storage and measurement may be recommended [40]. The need for an unselective sample-preparation procedure is dictated by the need to analyze quickly as wide a range of metabolites as possible. Regarding liquid samples, simple unselective methods such as dilution and solvent precipitation predominate in the metabolomics of biological fluids because they enable high metabolite coverage. Those different approaches are well reviewed in the literature [37–39, 41]. Typical dilution factors for urine between 1:1 and 1:10 with purified water are encountered [42–44]. The advantage of such method is its fastness, whereas major inconvenience may be ionization suppression occurring during the analysis because of the competitive nature of the electrospray process if LC-related methods are used (for more information regarding ion suppression phenomenon, look at Antignac et al. [45]). A freeze-drying step followed by a reconstitution at a defined dry matter concentration is reported for urine, as a way to avoid the dilution factor issue typically encountered with this matrix [46]. Normalization to specific gravity providing a fair estimation of urine osmolality can also be used [47]. Regarding serum or plasma, the high protein content of blood necessitates protein removal. Precipitation with ACN or acetone is reported as efficient [48], and precipitation with methanol, ethanol, or a mixture of methanol and ethanol results in a good metabolite coverage [49]. One drawback of such methods is that none of these enables complete protein removal, and the amounts of proteins remaining in the final extract are estimated to be 2–10%, depending on the solvent and precipitant ratios selected [50]. Ultrafiltration may also be employed for liquid matrices with the simple use of a filter that allows passage of molecules of specific molecular weight (common molecular weight cutoffs of 3000, 10 000, and 30 000 Da) [51]. One drawback may be the possible loss of hydrophobic species.

Besides, the level of detail to which the metabolome should be covered depends on the scope of the study and on prior knowledge. Some applications can use relatively selective sample preparation protocols if some preliminary data or knowledge suggests that the most useful information to look for is present in one or more particular fractions of the sample to be targeted, as is the case with metabolic targeted analysis or metabolic profiling approaches for which preselected metabolic pathways or groups of metabolites with similar chemical properties may be monitored. SPE is one example among the most important sample preparation techniques used for this purpose. SPE can be very selective by selectively washing the matrix interferences and eluting the analytes of interest [52]. The advantage of such techniques lies in the good repeatability of the protocol and effectiveness of removal of interferences, but its main disadvantage is that the sorbent material increases the selectivity of the preparation procedure that can lead to reduced metabolite coverage and impair the scope of metabolic fingerprinting.

Regarding solid matrices, an extraction step is required for transferring the metabolome compounds into a liquid phase. Samples can be freeze-dried prior to extraction to allow for better homogeneity, repeatability, and extraction capabilities. Different solvent systems with various polarities can be used (e.g., methanol, ACN, ether, acetone, hexane, cyclohexane), which lead not surprisingly to the production of different and complementary fractions of the studied metabolome (Fig. 5). Methods that can simultaneously extract both hydrophilic and lipophilic species, for example, chloroform–methanol or chloroform–methanol–water extraction are also popular [53, 54]. Subsequent purification steps may then be employed (e.g., liquid–liquid partitioning, SPE) as described for liquid matrices depending on degree of selectivity wished to be reached for the particular application.

Finally, if a gas chromatographic separation method is used for analysis, the addition of a derivatization reaction step prior to injection may be considered for low volatiles analytes in order to reduce polarity and increase volatility. The most common derivatization procedures are alkylation, acylation, or silylation, the active hydrogen in functional groups (–COOH, –OH, –NH, and –SH) are replaced by alkyl-, acyl-, or silyl-groups to form esters or ethers [55, 56]. Carbonyl groups may as well be derivatized in their corresponding oximes, which stabilize a-ketoacids and locks sugars in opening conformation [57]. Chloroformates have been proven to be strong and rapid derivatizing reagents, and in contrast to trimethylsilylation derivatization, alkyl chloroformate

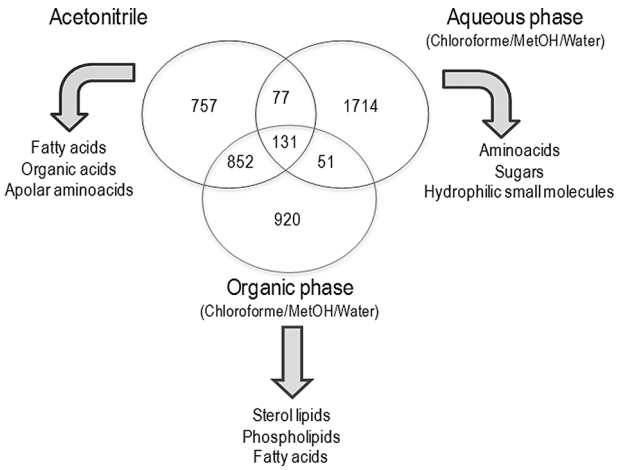

Figure 5. Venn diagram representing the results obtained from the comparison of the metabolic information (common and unique ions detected in MS) generated from the extraction of bovine tissue with ACN or a biphasique chroloroforme/methanol/water mixture (Bligh and Dyer protocol). The use of different procedures leads to specific information being generated: unsurprisingly, the aqueous and organic phases generated from the Bligh and Dyer extraction have only a few ions in common; the ACN extract and the organic phase of the Bligh and Dyer extraction have around 50% information in common, corresponding mostly to lipids, while the ACN and the aqueous phase generated from the Bligh and Dyer protocol have less than 10% ions in common.

derivatization reactions occurred directly in aqueous media without the requirement of heating, thereby simplifying the sample pretreatment and derivatization procedure [58].

2.3.3 Metabolomics profiles generation

Again, considering the large number of possible metabolites, their chemical diversity and the large range of possible concentrations, there is no single analytical technique that can achieve full coverage of all metabolites simultaneously. GC-MS may be suitable for small and volatile compounds while analysis of polar or ionic metabolites may be achieved with LC-MS. The size of the metabolome is unknown but supposed to be very large with tens to hundreds metabolites having the same molecular weight. This is problematic for MS-based approaches and the ability of the mass spectrometer to separate compounds having close m/z is particularly challenged. Ideally, the techniques used for metabolic profiling should then be capable of providing correct mass resolving power. A good mass accuracy is as well appreciated for structure elucidation purposes (see Section 2.3.6) with analyzers with sufficient dynamic range to be able to cope with the wide variations in metabolite concentration likely to be encountered. As mentioned in the sample preparation section, metabolomic analysis being semiquantitative, such techniques should also provide reliable and reproducible performance to enable the comparison and analysis of large number of samples.

As a starting point of this section, it appears worth describing the different signal-acquisition modes that may be encountered for metabolome analyses with MS. The most common acquisition mode is certainly the full-scan mode, particularly for metabolic fingerprinting [59]. In this mode, a target range of ions is monitored by the MS analyzer, for example, from m/z 50–1000 (Fig. 6A). Instruments with high mass resolving power ($R > 10\ 000$) are the most popular instruments when working in full-scan mode (e.g., instruments such as FT-ICR or Orbitrap ™ and to a lesser extent TOF instruments). The mass resolving power is the capacity of a mass spectrometer to separate ions of close m/z ratios, that is, isobars, chemicals that do not have the same raw formula but identical nominal masses (when integer masses are used for the atoms). It is defined as the ratio of the measured mass "m" to "Δm", the full width of the peak at half its maximum height (i.e., $m/\Delta m$, FWHM). Such analyzers allow the separation of compounds having close m/z and allow therefore the generation of highly informative data and are of immense help for structure elucidation purposes. An excellent review dealing with the use of high-resolution instruments for metabolome analysis has been recently released [60]. For metabolic profiling or targeted analysis, more selective signal-acquisition modes may be preferred. MS/MS uses two analyzers (or more) in a single instrument and allow precursor ion and neutral loss scanning (e.g., triple quadrupole instruments). Precursor ion scanning corresponds to the analysis of all precursors of a single charged product. The first mass analyzer is scanning a range of ions, while the second mass analyzer is static at the m/z of a product ion known to be common to the analytes to be monitored (Fig. 6B). Neutral loss scanning corresponds to the analysis of all precursors of a single uncharged product. The two mass filters are scanning synchronously at a user-defined offset. The uncharged product is known to be common to the metabolites (Fig. 6C). Both signal acquisition modes may be used when the aim is to monitor a selected number of predefined metabolites belonging to a same family and presenting common structural particularities (metabolic profiling) [61]. As an example, phase II metabolites show characteristic fragmentation patterns in MS: diagnostic product ions m/z 97 corresponds to the sulfate moiety of sulfoconjugated metabolites. Using the precursor ion scan mode, it is possible to scan for the precursors of ion m/z 97 thus getting access to the family of compounds that are sulfoconjugated. In the scope of revealing potential biomarkers signing a fraudulent administration of 4-androstenedione, this strategy was successfully conducted with UPLC-MS/MS and the diagnostic ion m/z 97 was used to fish for the precursor ions of any potential conjugated substance leading to a sulfate moiety after fragmentation, thus revealing potential marker metabolites of 4-androstenedione administration [52]. Finally, the MRM analyses specific precursors producing specific products. Both the first and second mass analyzers are held static at the m/z of the precursor

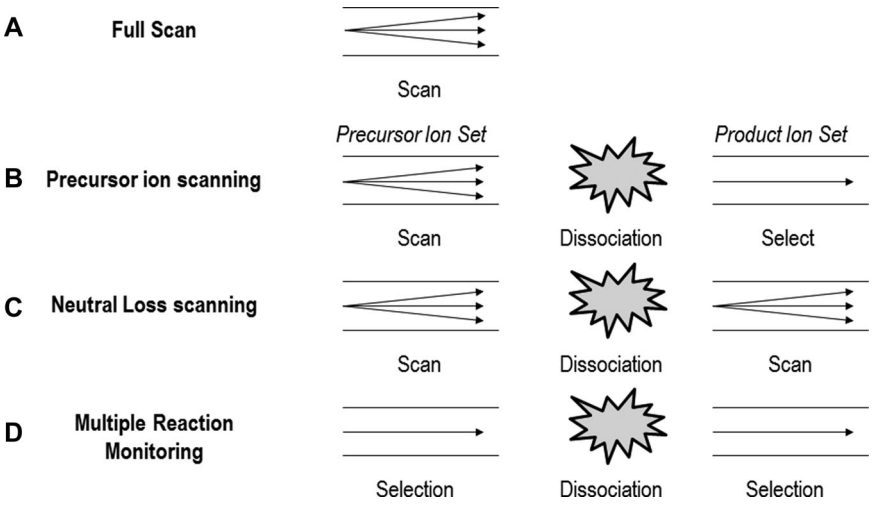

Figure 6. Description of the different signal-acquisition modes that may be encountered for metabolome analyses.

ion and the product ion, respectively (Fig. 6D). This signal acquisition is particularly used when the objective is to monitor several known metabolites (targeted approach). As an example, a MRM method was developed for measurement of 141 water-soluble cellular metabolites including components of central carbon, amino acid, and nucleotide metabolism among which 69 could be successfully quantified [62].

Regarding the sample introduction into the MS systems, direct introduction of the samples (infusion or flow-injection analysis) is commonly used with atmospheric pressure ionization techniques, particularly electrospray. Considering the above, such introduction technique is usually coupled with high mass resolving power MS-analyzers operating in full-scan mode. The application of this approach has been proven to be powerful for low complex matrices [63] or after efficient purification of the samples [64]. Its advantage lies in the high-throughput sample processing offered, but it can be dramatically impaired by matrix effects and in particular ion suppression phenomenon [65].

To circumvent this, coupling the MS analyzers (simple MS or MS/MS) with a separation technique is relevant. Indeed, optimized LC or GC separation then becomes essential in order to reduce matrix effects and to separate isobaric compounds, compounds having exactly the same m/z but different physicochemical properties inducing different retention times. The use of GC with electron ionization (EI) is well suited for volatiles or nonpolar compounds [56]. Compounds screened by GC-MS cover large parts of primary metabolism [55]. Due to the highly energetic process of ionization in EI, GC-MS provides informative MS spectra of the metabolites detected and as a result, the capability to identify unknowns. One drawback may be the high redundancy of the data (several m/z corresponding to one metabolite) that makes the data processing a tricky issue. Then, chemical ionization, known to be less energetic and offering the possibility to observe the intact molecule (molecular ion) can be preferred even if this ionization process is not often encountered in metabolomics. Nevertheless, GC analysis may be limited to thermally stabile compounds with a sufficient vapor pressure for volatilization during the injection. The introduction of ionic or nonvolatiles species may then be performed with LC [66]. Hydrophobic components will then be well separated with the used of RP chromatography, while hydrophilic and neutral compounds are best suited for hydrophilic separation (hydrophilic interaction chromatography) [67] ESI (soft ionization) remains the most commonly employed techniques in LC, while atmospheric pressure chemical ionization and photoionization may lead to complementary information being generated [68], but these alternative techniques have not yet extensively been used in the field. From this, we understand that there is not one single technique suited for metabolomics since GC and LC have their own advantages/disadvantages (Table 1), but the combination of different techniques for profiles generation represents one means to achieve the widest coverage of the metabolome [69].

2.3.4 Data processing and bioinformatics

Typically, an MS-metabolomic experiment produces a huge amount of raw data. The handling of such complex datasets manually is practically impossible. Hence, specific software tools and algorithms are needed. The objective is to convert these so-called raw, instrumental data into extracted data (e.g., peak tables) that can easily be processed by statistical tools. In MS-based metabolomics, metabolite fingerprints are described by m/z values and corresponding intensities of detected ions. For direct injection MS, spectral bins may be the most straightforward method to deconvolute the spectrum. It corresponds to the transformation of the mass spectrum into vectors of uniform length (e.g., in nominal mass resolution spectra, m/z are converted to integer values and the bin sizes are usually 1 amu) and the report of the intensity of the MS signals detected. If high resolution is utilized for acquisition, the issue is trickier since small drifts in mass measurement can occur and lead to a metabolite being assigned to different

Table 1. Advantages and disadvantages of GC and LC for metabolomics application

	LC	GC
Advantages	Suited for polar to apolar compounds Minimal sample preparation Access to the molecular ion (intact molecule)	Suited for apolar and volatiles compounds Structure information obtained through in-source fragmentation Existence of universal databank (facilitates structure elucidation)
Disadvantages	Small and very polar molecules needs specific chromatographic column Reduced fragmentation (need MSn for structure identification) Subjected to ion suppression phenomenon	Requires higher sample preparation Polar compounds need derivatization Extensive fragmentation (leads to redundancy in the information generated and molecular ions hardly identifiable)

bins from one sample to another. This problem can be minimized by the use of internal calibration (a reference standard introduced continuously into the MS) that allows subsequent alignment of mass spectra.

If the MS instrument is hyphenated to a separation technique, retention times are also used to index metabolites (Fig. 7A). Thus, the challenge lies in the correct and complete restitution of all the information contained in the raw data, that is, m/z values, retention times, and intensities while handling issues such as baseline drift, retention time shifts, noise, and artifacts generated by the instrument. Comprehensive overviews of many of the existing tools for MS data processing in metabolomics have been recently reviewed [70–72]. Basically, the first step is the filtration of the analytical background, which has no interest for the study in itself and may confuse the data. The second step corresponds to the peak picking step, that is, the report of the signal abundance observed for each ion $[m/z;\mathrm{rt}]_i$ in each of the analyzed samples. The same issue as the one encountered for spectral binning may occur here for high-resolution data and internal calibration is as well a mean of minimizing it. Depending on the software used, peak picking may proceed differently (Fig. 7B). Indeed, some softwares extract successively each specific ion chromatogram corresponding to each m/z value and integrates (either height or area) the corresponding chromatographic peaks. Some other softwares extract successively the mass spectra at each retention time unit and report the signal intensities of each ion observed on these mass spectra. The advantage of the first option is certainly to minimize the proportion of unwanted noise considering that the presence of a chemical entity behind the detected signal is preliminary checked when the existence of a chromatographic peak is demonstrated. The third step of such data processing corresponds to a peak alignment process for correcting retention time shifts that may occur from one sample to another. As depicted in Fig. 7C, an ion at a given m_i/z_i (named Mx) and a given retention time rt_i (named Ty) may be subjected to small drifts in retention time (epsilon). These drifts have to be corrected to ensure that each ion $[m_i/z_i]$ of the MS fingerprint appears at identical retention time (rt_i) across all the analyzed samples and is reported as being the same ion for all the samples. Finally, missing values may be treated during a last step "fill the gap." Missing values may arise because the concentration of a particular metabolite is below the LOD of the instrument used. In that case, and considering that in metabolomics we do not have the knowledge of the LOD for each existing metabolites, it may be reasonable not to assign a value of zero but rather half (or one-third) the lower value detected for respective ion. Some softwares integrate the proximal background noise when no chromatographic peak is present. The last step of the data processing is the generation of the final report. Integrated peaks (MxTy) are sorted out to generate a 2D data table, in which rows represent the different ions and columns report some characteristics of the detected ions (i.e., m/z, retention time, intensities in the samples, etc.). This table can then be easily processed by statistical tools in order to extract the relevant information from this huge dataset.

Secondary signals such as isotopes, adducts, dimers, and fragments may further lead to data redundancy because they account for the same metabolite. This may disrupt the statistical analysis and slow down the identification procedure by increasing the number of variables that have to be investigated and also by leading to unsuccessful database queries. Therefore, these signals can be automatically annotated by correlation analysis on both signal shape and intensity patterns using software tools such as CAMERA [73], PUTMEDID-LCMS [74], and mzMatch [75]. Such peaks are not discarded, but only flagged, so that their assigned annotations can be taken into account in the metabolite identification step.

2.3.5 Data analysis

Data generated through metabolomics approaches are characterized by high dimensionality, where the number of variables measured per subject vastly exceeds the number of subjects in the study, especially for metabolic fingerprinting approaches [76]. The previous data-processing stage usually

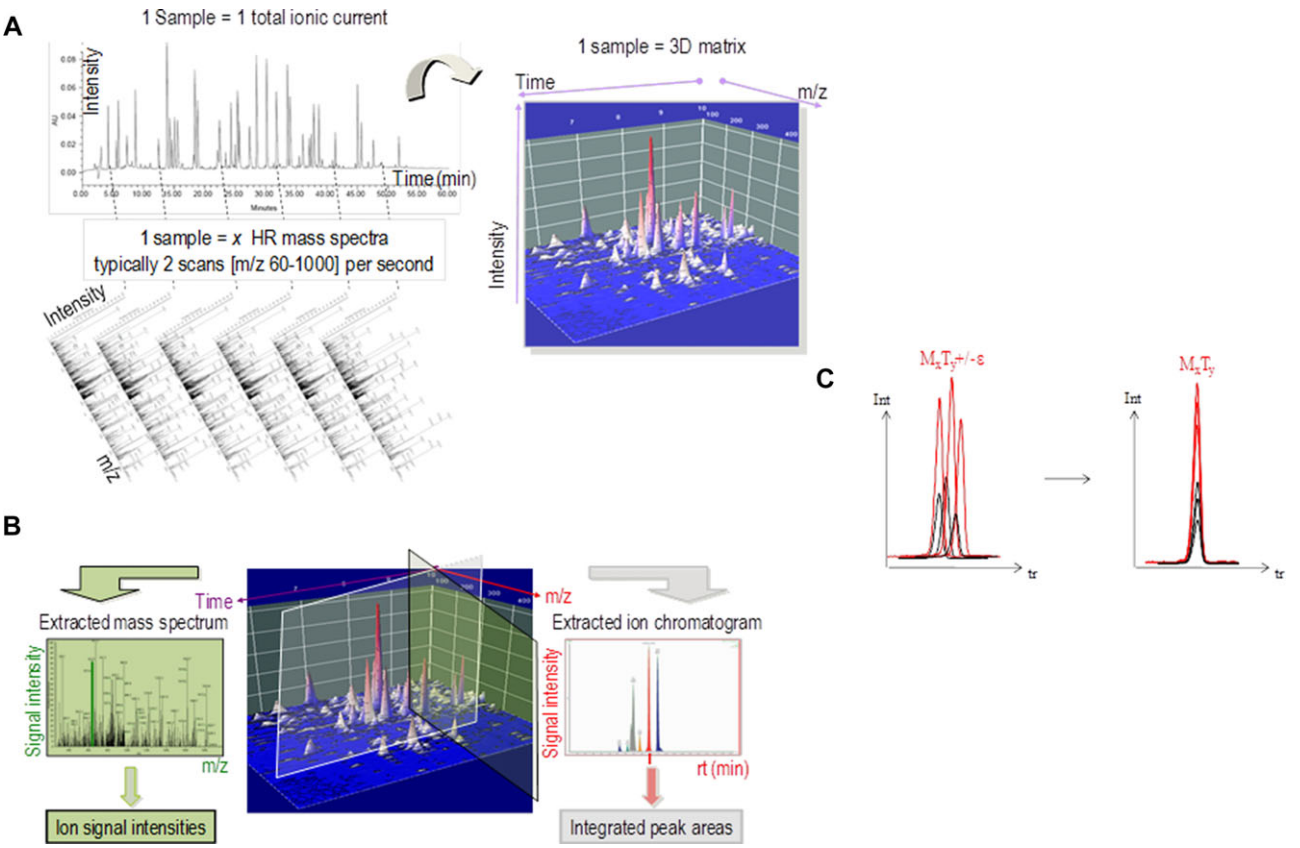

Figure 7. (A) Complexity of a metabolic profile. One sample corresponds to one total ionic current. At each retention time unit is associated a full-scan mass spectrum. The raw data can then be visualized as a 3D map. (B) Peak picking step: it may proceed by the extraction of mass spectra corresponding to each unit of retention time or by the extraction of specific ion chromatograms corresponding to *m/z* values scanned during the acquisition. (C) Analytical shifts that may occur in retention time from one sample to another.

generates data tables that are difficult to analyze comprehensively without appropriate and hyphenated statistical tools. Before looking into the more complex multivariate methods, it may be desirable to look at the statistical properties (mean and variance) of individual metabolites and the relations between them (and each other) and the other measured properties of interest. Univariate methods as, for example, the classical Student's *t*-test can be used as a first means of revealing potential candidate compounds presenting significant differences in terms of abundance between two subgroups of samples. However, a limitation of this approach is the absence of consideration of correlations between variables that could be at least or even more informative than the variables considered separately. Moreover, because the biological differences between samples sometimes arises from comparatively small differences in many metabolite concentrations, recognizing the pattern and interpreting is not straightforward. Such correlation patterns can be studied using multivariate techniques that aim to reduce the complexity of the datasets and to highlight the analytical information of biological relevance [77, 78]. PCA is the most commonly used method to explore relationships between samples in metabolic fingerprinting studies. It aims to extract a small number of latent components that summarize the measured data with minimal information loss by taking advantage of the correlation structure of peak intensities. The new latent components are derived from the original data in such a way that the greatest variation in the data is captured in the first group (first component, or PC1), the second greatest variation in the second (PC2), and so on. This method is particularly useful as a first step in data analysis to visualize trends and outliers. PCA is expected to reveal the most important factors of variability characterizing the considered data set. However, the main source of variability characterizing the dispersion of the original data is rarely the source of variability associated to the studied factor. In this case, supervised methods such as partial least squares-discriminant analysis or orthogonal partial least squares-discriminant analysis may be used in order to extract from the whole dataset the useful information, for explaining and predicting the membership of the analyzed samples to different subgroups (or classes). These two methods provide as well good diagnostic tools for the detection of biomarkers. However, one drawback may be their potential to "overfit" the data (i.e., to lead to statistical models based

more on the high number of variables available than the real robustness of individual contributors to the expected discrimination between samples). Each model built therefore has to be carefully validated before it is used for predictive modeling of unknown samples [19]. A recent review gives an overview of the various strategies that can be applied for metabolomics analysis [79].

Importantly, data pretreatment must be considered before any multivariate analysis. It corresponds to transformations that are applied to the dataset to convert data to a different scale. Subsequent data-processing steps may include centering, scaling, or other transformation of the original variables. Centering corresponds to spatial translation of the data from around the mean value to around zero (i.e., the barycenter of all individuals is set to zero). Transformations (log or square root) correspond to nonlinear conversions of the data aiming at reducing the influence of a few particularly intense signals (due to the huge dynamic range in terms of metabolite-concentration levels) that can strongly influence statistical analysis and subsequent interpretation. Scaling methods are data-pretreatment approaches that divide each variable by a given (namely scaling) factor. By scaling the data, the relative weight given to each variable whatever its abundance in the dataset can be adjusted, in order to avoid for instance a systematic overweight of highly abundant variables. The most used scaling methods are autoscaling (the SD of a variable is the scaling factor) and Pareto scaling (the square root of the SD is the scaling factor). One drawback of autoscaling may be the dilution of the analytical information of biological relevance. Indeed, an excessive/uncontrolled scaling may results in overrepresentation of small variables that can be just noise, which can have undesirable side effect on the statistical analysis. Pareto scaling finally appears as a good compromise. A summary of the various data pretreatment methods is given in a previous publication [80].

2.3.6 Structural elucidation

Structural elucidation of the MS signals of interest is an important step in metabolomics approaches. Indeed, using metabolomics exclusively for fingerprinting without identifying the metabolites that cause clustering of experimental groups will only deliver a classification tool but not directly contribute to biochemical knowledge and understanding of underlying mechanisms of action. Given the chemical diversity of most metabolomes and the character of most metabolomics data, metabolite identification is still a real challenge. When it comes to characterize metabolic profiles wider than in a defined list of substances (metabolic profiling), this step still remains extremely long and difficult, especially for LC-MS based methods as acknowledged by several articles in the field [81–83]. Consequently, a great deal of effort in metabolomics over the past decade has been focused on making metabolite identification better, faster, and cheaper [81, 84]. The first step in metabolite characterization is the identification of the MS molecular ion for database queries. As already mentioned, in GC-EI-MS, the highly energetic process of ionization makes it difficult to identify the molecular ion. However, EI leads to a very reproducible fragmentation in which advantage is to allow the establishment of universal EI mass spectra databanks (e.g., the NIST database—*National Institute of Standards and Technology* [85]), which can be queried to ease the identification step. The situation is different when working with API-related techniques since they exhibit high interinstrument variability regarding fragmentation, impeding the creation of universal libraries. A guide to identification of metabolites using LC-MS is provided in Watson et al. [84]. Briefly, since in API techniques the ionization process is very soft, it can be easier to identify the protonated or deprotonated molecule, $(M + H)^+$ or $(M - H)^-$, respectively. The monoisotopic mass observed can then be subjected to databases [86–89]. These databases allow a search on the basis of the compound monoisotopic mass. In this case, the knowledge of the accurate mass (thanks to high-resolution MS) is helpful, if not mandatory to shorten the list of possible candidates, but even though most of the time insufficient for complete identification of the compound of interest. Indeed, there are sometimes tens of possible structures even for a mass entered with three-digit precision. Algorithms for filtering molecular formulae from accurate mass determination and isotopic pattern (which can be exploited on an easier way compared to peptides or proteins since the charge state of metabolites is usually one or two) have been setup to allow restriction of the number of candidate compounds proposed by these databases [90]. Moreover, additional fragmentation techniques can be used if adequate sample amount is available for additional analysis [82, 91]. Indeed, MS is a powerful tool in metabolomics investigations since it enables the detection of low-abundance metabolites. However, identification of such metabolites remains tricky because of lack of necessary sensitivity to obtain useful fragment information and difficulties to interpret the generated fragments (which are in some extent less rules driven than peptides fragments). Other parameters can also assist in metabolite identification: for example, the polarity/volatility of the molecule. The retention time in LC/GC can give useful indications that can sometimes rule out candidate molecules. Nevertheless, in the end, the only conclusive and unambiguous method to identify a given metabolite is the comparative analysis of the corresponding authentic reference standards to verify the different identification criteria to respect [92]. To this end, internal databanks are starting to be built allowing automatic annotation (based on experimental masses and retention times) of a maximum of organic compounds present in complex metabolic fingerprints [93, 94]. A good review of the tools developed to improve the annotation of MS-based metabolomics datasets may be found in Junot et al. [60]. However, even if several tools are available today for metabolite identification, there is still incomplete identification in fingerprints, this step being the bottleneck of metabolomics investigations.

3 How the approach is used today in research

Approaches usually conducted in metabolomics involve the collection of fingerprints by comprehensive analytical techniques for the detection, by a differential approach, of changes in the metabolome of samples collected from a "control" population and a "test" population. The objectives can be descriptive and explicative, through the characterization and understanding of metabolic changes resulting from one or more given factors, or diagnostic, through the identification of biomarkers that can further be used for predictive modeling. The field of metabolomics continued to grow rapidly over the last decade and has been proven to be a powerful technology in predicting and explaining complex phenotypes in diverse biological systems. Metabolomics, originally developed for human biomedical applications [4] has been an application driven science with broad range of applications in various fields, including medical and environmental science, food quality and safety, and engineering of microbial systems [95].

In medical sciences, metabolomics enables the differential assessment of the levels of a broad range of endogenous and exogenous molecules and has been shown to have a great impact on the investigation of physiological status, diagnosis of diseases, and identification of perturbed pathways due to disease or treatment [96]. There are two major purposes for its use in medicine. The first is to acquire knowledge on the mechanisms of drug action [97] or the disease itself [98] by bringing new insights in disease etiology. Another is to explore biomarkers [99, 100]. Indeed, metabolomics may provide advantages that classical diagnostic approaches do not have, based on the discovery of a suite of clinically relevant biomarkers, whose levels are simultaneously affected by the disease. Metabolomics may also be of good help in toxicological science since it has the ability to evaluate/monitor potential toxicity [101]. Metabolomics can assist in predicting and classifying different modes of action of toxics and identifying novel biomarkers of toxicity, either biomarkers of exposure to a substance [102] or biomarkers of effect after exposure [103].

In environmental science, metabolomics is used for determination of the effect of biochemical or environmental stresses on plants, which include genetically modified plants [104]. Many studies have concentrated on the physiological development of plant tissues as well as on the stress responses involved in heat shock or treatment with stress-eliciting molecules. Plant–host interactions represent one of the most biochemically complex and challenging scenarios that are currently being assessed by metabolomic approaches [105]. Untargeted analyses have been used as well in the identification of possible fingerprints of biological phenomena such as plant diseases [106]. Beyond plants, metabolomics technology has made significant inroads into the environmental research community [107]. This approach has considerable potential for characterizing the responses of organisms to natural and anthropogenic stressors and to finally assess environmental quality.

In food science, metabolomics has recently risen as a potent tool [108]. It has many applications in different areas of food science and technology. It can be used to identify and classify food constituents giving the opportunity to understand the molecular details of what gives certain foods and drinks their unique taste, texture, aroma, or color. The capacity to assess food constituents can be used as well to assess both food adulteration and quality [109]. Metabolomics provides as well new screening approaches to prevent fraud in food-producing animals and therefore can help in ensuring food safety [110]. In nutritional science, metabolomics is expected to characterize effect of nutrients, food, or diet on the organism with precision [111]. As nutrition nowadays focuses on improving health of individuals through diet, metabolomics can be used to assess metabolic responses to deficiencies or excesses of nutrients and bioactive components [112].

Finally, metabolomics has great potential in metabolic engineering. Indeed, rational engineering of metabolism is important for bioproduction using microorganisms and information on how the cell is using its biochemical resources is a good way to inform strategies to engineer a cell to produce a target compound [113, 114].

4 Worked example

This illustrating example is taken from a published study related to the control of forbidden substances in breeding animals [115]. β-Adrenergic agonists are a class of sympathomimetic agents that act upon the β-adrenoreceptors. β2-Agonist compounds are used as bronchodilators, tocolytics, or heart tonics in human and veterinary medicine. During the past 20 years, several studies focused on the effects on growth rate and performances, when administered per os, mixed with feeding stuffs. Because of their ability to shift nutrients toward protein instead of lipid anabolism, such molecules were gathered under the generic name of "repartitioning agents." Because pharmacological residues found in slaughtered cattle were found to have caused acute intoxication in consumers, European Commission Directive 96/22 prohibited its use in food producing animals except for well-defined therapeutic purposes and under strict veterinary control. Nevertheless, β2-agonist compounds are still misused in food-producing animals for growth-promoting purposes. Efficient methods based on MS detection have then been setup for this class of anabolic agents to ensure their control; such methods rely on the direct measurement of drugs in a targeted mode, allowing the detection of a given number of known compounds. Nevertheless, there are many possible chemical structures for β2-agonists and along with the illegal use of some well-known such as clenbuterol, salbutamol, and bromobuterol, new compounds, which exhibit activity at the β2-adrenoreceptor have been progressively discovered. Therefore, a range of compounds, either of known chemical structures but not yet included in the methods, or of unknown

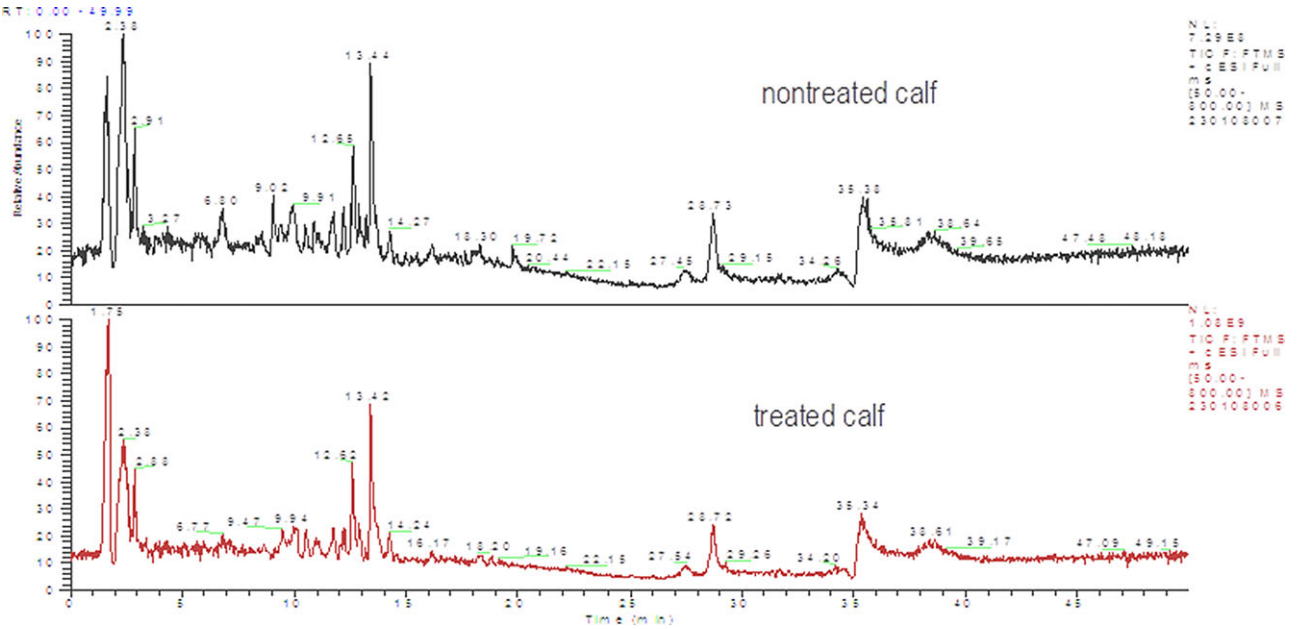

Figure 8. Typical total ion chromatograms (TIC) obtained in RP LC–positive electrospray–high-resolution MS for urine samples collected from a nontreated (up) versus a treated (down) calf with clenbuterol (taken from Courant et al. [115]). The beginning of the chromatograms corresponds to polar compounds (amino acids and hydrophilic small molecules). The region from 10 to 15 min corresponds to the elution of moderately polar metabolites, for example, organic acids and small peptides. The region from 15 to 35 min corresponds to the elution of apolar compounds, which are scarce in urine samples explaining the lack of chromatographic peaks in this region.

chemical structures, can be skipped during routine screening and confirmatory analysis. In addition, some practices consisting in the use of "cocktails" composed of mixtures of low amounts of several substances that exert a synergistic effect and exhibiting similar growth-promotion properties have been reported. The combination of unknown β2-agonists and low levels makes these illegal practices difficult to handle with. In this context, metabolomics may represent a new emerging strategy for investigating the global physiological effects associated to a family of substances (rather than monitoring the illegal molecules itself) and therefore, suspect the administration of β2-agonists (either "cocktails" or unknown compounds). As a demonstration of feasibility, metabolomics approach based on LC coupled to high-resolution MS was applied to investigate changes in the urinary metabolome of calves treated orally with clenbuterol for six consecutive days and compared with nontreated calves. All calves were male of 80-day old that were allowed to acclimatize for 1 week before treatment. The sample preparation procedure was the less selective as possible in order to reveal as much information as possible and consisted only in the filtration of the urine samples and freeze-drying for reconstitution in water at a defined dry matter of 30 mg/mL. Fingerprints were acquired through high-resolution MS in full-scan mode (m/z 50–800) at a resolving power R of 30 000. A typical total ion chromatograms is presented in Fig. 8 illustrating the high difficulty to distinguish any significant differences between samples collected in treated and nontreated calves on the basis of a simple visual examination of such global profiles. All the metabolic fingerprints acquired by LC-high resolution mass spectrometry were then processed with XCMS software. XCMS used a simple univariate t-test to identify metabolites presenting abundances significantly different between sample subgroups to be compared, and ranks these metabolites according to a statistical confidence level parameter (p-value) associated to this observed difference of abundance. Metabolomic changes occurring after clenbuterol treatment could be assessed through a plot (Fig. 9) on which each point corresponds to one ion [m/z; rt_i] plotted according to its p-value and its rank. Ions are then sorted out on the x-axis from the most (left) to the less (right) discriminant between control and treated animals. Thus, the higher the number of ion below the statistical p-value limit (0.05), the more important seems to be the metabolic modifications induced by the clenbuterol administration. Figure 9 represents the overlapped results obtained for metabolomic profiles collected 2, 3, 4, and 10 days after clenbuterol administration. During the successive three treatment days, a gradual change in the metabolome of treated animals can be observed. Nevertheless, the metabolome seemed to be the most affected by clenbuterol administration 4 days after treatment, and tends to return close to the control group 10 days after treatment. The end of the clenbuterol treatment occurring at day +6 could be explained by a feedback phenomenon. By means of multivariate statistics (i.e., orthogonal partial least squares-discriminant analysis), these metabolic differences were used to build predictive models able to suspect clenbuterol administration in calves. Biomarkers were highlighted, among which creatine had already been

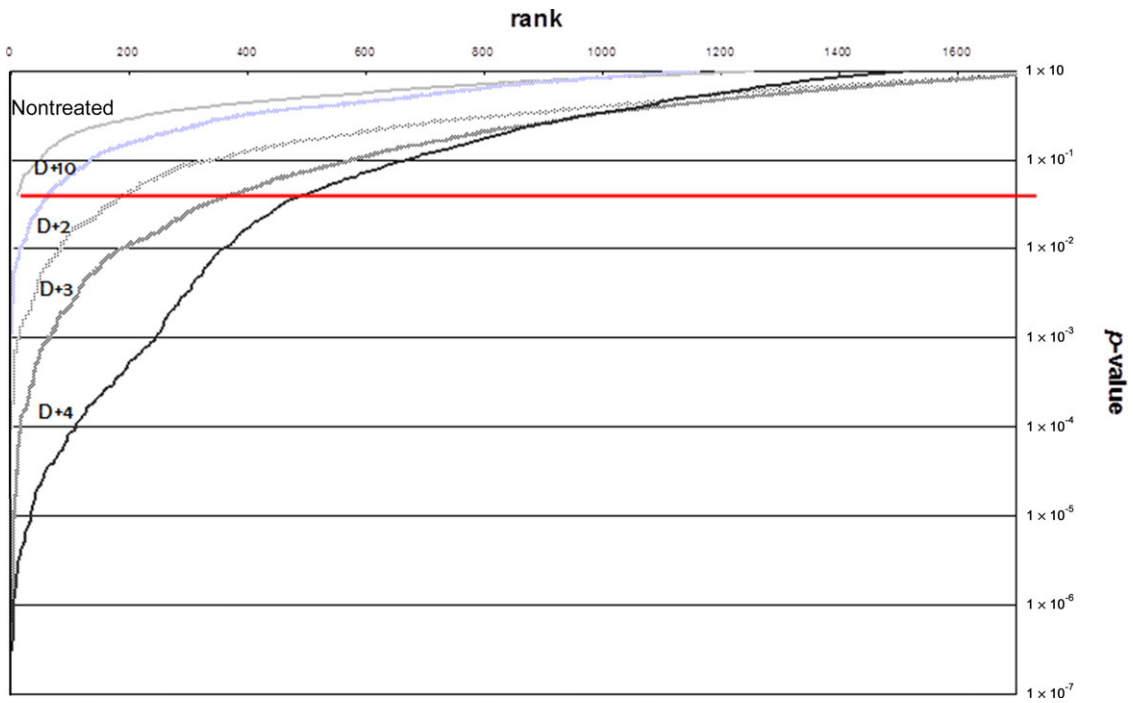

Figure 9. Overlapped results (each point corresponds to one ion plotted according to its *p*-value and its rank) obtained for metabolic fingerprints collected before clenbuterol treatment (day 0) and on days 2, 3, 4, and 10 (taken from Courant et al. [115]).

demonstrated as being downregulated after an anabolic treatment. Creatine being a nitrogen donor, its downregulation appeared consistent with the reinforcement of anabolism. In a second step, this study was completed by a parallel study involving several independent bovine experiments with different animals, different breeding conditions, and different parameters (doses, compound mixtures, etc.) to extract the β2-agonists characteristic changes. Biomarkers were highlighted (and confirmed the previous ones highlighted under a clenbuterol treatment) and a weighted combination of the selected biomarkers intensities allowed establishing an efficient and robust tool as a screening method to suspect illegal practices [116]. This example demonstrates how metabolomics approaches may be considered of valuable interest to overcome current limitations in the control of growth promoters' abuse, with promising perspectives as new effect-based screening approaches.

5 Current limitations and useful working limits

More and more original studies are emerging that indicate the promising capabilities of metabolomics, but that also demonstrate that there are many issues that must first be addressed if metabolomics is to reach its true potential. Indeed, metabolomics still retains several intrinsic limitations that have a great impact on its widespread implementation, limitations that lie in biological issues and experimental measurements. On a biological point of view, the metabolome is sensitive to various genetic and environmental stimuli. Therefore the execution of a metabolomic study requires the consideration of a number of factors so that confounders can be limited and information recovery optimized [117]. This issue is complicated when dealing with discrete signatures that can be blurred by multiple sources of variation, interindividual variations being sometimes larger than the studied factor. Sources of variation can result from various factors occurring at different stages of the experimental process. These include diversity among the selected sample population (gender, diet, medications, etc.), uniformity of sample treatment, and consistency in sample handling and analysis. Further research in advanced statistics is then needed to be able to separate interindividual variability caused by genetic, diet, and environmental factors from related metabolic effects, which are under study. Moreover, for studies based on biofluids, representativeness of a single sample collected at one moment, which represents a snapshot of the metabolite content of the analyzed specimen is another challenge. Therefore, in any experiment, standardization of protocols is a valuable practice for minimizing data variation. Standard operating procedures can be employed to overcome variation in sample handling and processing techniques. Another challenge is to be able to properly interpret results of metabolomics investigations in terms of biological meaning. Indeed, due to the incapacity to cover the whole metabolome and to structurally elucidate all the biomarkers, biological interpretation remains a tricky issue since we must compile the fragments of information

obtained from the experiment and try to explain the link between changes in metabolite concentrations and the factor being studied. Data visualization also remains a difficult task. The development of biochemical pathway analysis tools aiming at building associations between the MS signals collected and existing knowledge may help in that sense [118, 119].

On an analytical point of view, there is no universal technique covering the diverse range of analyte structures and polarities present in most biological samples. A substantial benefit could then be expected from crossing complementary analytical techniques in the scope of reaching a more comprehensive view of the metabolome. However, if this may be possible when working on a small number of samples, this is practically impossible on a large number of samples as the one encountered in epidemiology studies. Indeed, multiple platforms profiling directly impact the cost and duration of analytical treatment, and consequentially the real high-throughput capabilities of the approach. The fusion of data collected on different analytical platforms also remains an issue that has not been solved yet. Another analytical limitation arises from the semiquantitative nature of metabolomics. Indeed, the goal of absolute quantification, as performed for targeted analysis with the use of internal standards (compounds presenting a similar structure than the analytes to be quantified to allow normalization for recovery and matrix effect), is not reachable considering the large number and diversity of metabolites. Therefore, the systematic variability between LC-MS measurements must be kept under control to make sure that the differences observed between groups of samples are not spurious. The inclusion of quality controls (consisting in a pool of all the analyzed samples) injected at the beginning, end, and also randomly all along the analytical run, and allowing the characterization and control of the total variance associated with the MS instrument, is a prerequisite [120]. When large metabolomics studies divided over a series of analytical blocks cannot be avoided, normalization of the data can be considered. Dunn et al. suggest using quality control sample to allow for signal correction within and between analytical blocks [40]. However, such normalization protocols proposed to correct signal intensity (when a drift is acknowledged) do not correct for metabolites being no more detected when the ionization source is too dirty. The limitation is the still insufficient stability of the instrumentation used for generating the metabolic descriptors. Another analytical challenge is to be reproducible between analytical platforms (intra- and interlaboratory) to allow comparison of results obtained by different teams on the same subject and enhance the potential of metabolomics. However, there is no consensual strategy emerging from the scientific community regarding MS-based metabolomics, and a wide heterogeneity of analytical strategies and workflows exists today. There is an urgent need for standardization procedures. Finally, as discussed earlier, the identification of the large number of metabolites that are detected but whose chemical nature is unknown is a real challenge in metabolomics and constitute another limitation of the approach. Indeed, in best cases, only 30% of the total of MS signals detected in an experiment is structurally identified. The 70% left, which may represent information on changes in metabolic profiles, remains unknown and consequently lacks for the biological interpretation of the results.

6 Future developments

In order to overcome the limitations described above, research has been undertaken in different areas. Metabolomics organizations are currently working toward guidelines for commonality in metabolomics experiments as development of optimal methodologies and study designs are needed [40, 92, 121]. Reporting of standard metadata will provide a biological and empirical context for the data, facilitate experimental replication, and enable the reinterrogation and comparison of data by others [122]. Additionally, knowledge of the effects of normal physiological variation on metabolic profiles is essential for accurate interpretation of profile changes, particularly in human studies, because of diversity in lifestyle and environmental factors. The effects of factors such as ethnicity, gender, age, body composition, health, dietary intake, physical activity, gut microflora, and stress need to be further explored in order to advance the understanding of the human metabolome and therefore improve data interpretation [29]. Developments in advanced statistics could be extremely useful to help separating interindividual variability caused by genetic, diet, and environmental factors from related metabolic effects, which are under study. Additionally, as data structures become increasingly complex, particularly when considering the addition of supplementary dimensions associated to the generated metabolomic data, for example, a kinetic aspect for integrating the temporal evolution of the considered metabolic profiles (time series data representing multiway tensors of high-order or multiblock data tables), new solutions are being developed to cope with these new types of data. Recently, strong interest has developed in multiway (parallel factor analysis, N-way partial least squares) and multiblock methods (constrained principal component analysis, multi-block partial least squares) [79].

With regards to the coverage of the metabolome, advances in separation sciences will help in this wish [123]. Appropriate separation of complex samples is an important part of metabolomics analysis. The chemical information in the retention time may be used to identify a metabolite. Moreover, MS alone cannot distinguish isobaric compounds, whereas chromatography is usually powerful to separate isomers. In the next future, the democratization of fast chromatographic systems (fast-LC or fast-GC) is probably an evolution, which will yield substantial benefits for metabolomics. Narrower peak width and increased S/N will help getting more and more information from a sample analysis in a high-throughput fashion. The development of robust hydrophilic interaction chromatography represents a complementary normal-phase LC separation mode for resolving

polar/ionic metabolites and their isomers that are poorly retained by RP chromatography. Another future direction would be to combine efficiently two or more complementary stationary phase systems in order to achieve a satisfactory separation. One of the latest advancements in the GC system is the development of a 2D GC/MS system (GC × GC/MS) [124]. However, additional time resources are needed for data alignment, peak picking, and normalization that complicate data preprocessing and statistical analyses. Supercritical fluid chromatography (SFC), which combines some advantages of both GC and LC, is a high-resolution and high-throughput separation method using supercritical fluid (noncondensable fluid over liquid-gas at the critical point) as a mobile phase. Due to the high diffusion coefficient of the supercritical fluid, the separation capacity of SFC is much higher compared with LC. In addition, since the polarity of the supercritical fluid carbon dioxide is similar to hexane, SFC can also be employed for the analysis of hydrophobic metabolites [125]. Regarding mass analyzers, it is expected that future improvements in analytical sensitivity may reveal the presence of metabolites that are at the moment close to or under the detection limit of the current instrumental methods. Instruments with higher mass resolving power and higher scan speed may also allow a better characterization of the samples and ease the structural elucidation process. The development and generalization of ion-mobility spectrometry (IMS) is another future possibility [126]. The peak capacity of mass spectrometers can be increased by coupling IMS as a preseparation technique prior to MS analysis. IMS is used to separate ionized molecules in a gas phase based on their ion mobility in a carrier buffer gas.

Regarding metabolites identification, the use of databases has become well structured and more straightforward because of the initiatives allowing merging of several databases available on the Internet into the same software [74, 127]. However, as mentioned earlier there are sometimes hundreds of possible structures even for a mass entered with three-digit precision. The use of multidimensional mass spectra (i.e., multiple fragmentations) can aid in structure elucidation, particularly when dealing with isomers. The development of instruments and softwares that allow full-scan analysis along with MS/MS experiments in the same analytical run is another way forward [128]. For the data collected by different groups to be useful for the entire community, standardization of CID MS/MS spectra acquired with API is essential. It will allow to establish LC-MS identification databanks shared into the community as it is the case for GC-MS with the NIST. Initiatives are currently on-going on this topic in order to define normalization procedures [129].

In conclusion, the emerging field of metabolomics, in which a large number of small molecule metabolites are detected quantitatively, promises immense potential for different areas such as medicine, environmental sciences, agronomy, etc. Detection of crucial disturbances in the concentration of key biomarkers can be beneficial and bring new insights into innovative specific research questions. Provided that research continues in order to circumvent limitation discussed in this article, integrated applications with genomics, transcriptomics, proteomics, will provide greater understanding of global system biology.

The authors gratefully acknowledge Dr. Thierry Rabilloud (CEA, Grenoble, France) for reading the manuscript and for his useful suggestions.

The authors have declared no conflict of interest.

7 References

[1] Lederberg, J., McCray, A., One sweet omics: a genealogical treasury of words. *Scientist* 2001, *15*, 8–10.

[2] Wilkins, M., Pasquali, C., Appel, R., Ou, K. et al., From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Nat. Biotechnol.* 1996, *14*, 61–65.

[3] Oliver, S., Winson, M., Kell, D. F. B., Systematic functional analysis of the yeast genome. *Trends Biotechnol.* 1998, *16*, 373–378.

[4] Nicholson, J., Lindon, J., Holmes, E., 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivaraiate statistical analysis of biological NMR spectroscopic data. *Xenobiotica* 1999, *29*, 1181–1189.

[5] Ryan, D., Robarts, K., Metabolomics: the greatest omics of them all? *Anal. Chem.* 2006, *78*, 7954–7858.

[6] Nicholson, J., Connelly, J., Lindon, J., Holmes, E., Metabonomics: a platform for studying drug toxicity and gene function. *Nat. Rev. Drug Discov.* 2002, *1*, 153–161.

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

[8] Thompson, J. A., Markey, S. P., Quantitative metabolic profiling of urinary organic acids by gas chromatography-mass spectrometry: comparison of isolation methods. *Anal. Chem.* 1975, *47*, 1313–1321.

[9] Horning, E. C., Horning, M. G., Metabolic profiles: gas-phase methods for analysis of metabolites. *Clin. Chem.* 1971, *17*, 802–809.

[10] Lahner, B., Gong, J., Mahmoudian, M., Smith, E. et al., Genomic scale profiling of nutrient and trace elements in *Arabidopsis thaliana*. *Nat. Biotechnol.* 2003, *21*, 1215–1221.

[11] Holmes, E., Loo, R. L., Cloarec, O., Coen, M. et al., Detection of urinary drug metabolite (xenometabolome) signatures in molecular epidemiology studies via statistical total correlation (NMR) spectroscopy. *Anal. Chem.* 2007, *79*, 2629–2640.

[12] Dunn, W., Ellis, D., Metabolomics: current analytical platforms and methodologies. *Trends Anal. Chem.* 2005, *24*, 285–294.

[13] Förster, J., Famili, I., Fu, P., Palsson, B. et al., Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Res.* 2003, *13*, 244–253.

[14] Wishart, D., Jewison, T., Guo, A., Wilson, M. et al., HMDB 3.0—The Human Metabolome Database in 2013. *Nucleic Acids Res.* 2013, *41*, D801–D807.

[15] Psychogios, N., Hau, D., Peng, J., Guo, A. et al., The human serum metabolome. *PLoS One* 2011, *6*, e16957.

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

[17] Lutz, U., Lutz, R., Lutz, W., Metabolic profiling of glucuronides in human urine by LC-MS/MS and partial least-squares discriminant analysis for classification and prediction of gender. *Anal. Chem.* 2006, *78*, 4564–4571.

[18] Kieken, F., Pinel, G., Antignac, J.-P., Paris, A.-C. et al., Generation and processing of urinary and plasmatic metabolomic fingerprints to reveal an illegal administration of recombinant equine growth hormone from LC-HRMS measurements. *Metabolomics* 2011, *7*, 84–93.

[19] Antignac, J.-P., Courant, F., Pinel, G., Bichon, E. et al., Mass spectrometry-based metabolomics applied to the chemical safety of food. *Trends Anal. Chem.* 2011, *30*, 292–301.

[20] Suhre, K., Gieger, C., Genetic variation in metabolic phenotypes: study designs and applications. *Nat. Rev. Genet.* 2012, *13*, 759–769.

[21] Nicholson, G., Rantalainen, M., Maher, A. D., Li, J. V. et al., Human metabolic profiles are stably controlled by genetic and environmental variation. *Mol. Syst. Biol.* 2011, *7*, 525.

[22] Hodson, M., Dear, G., Roberts, A., Haylock, C. et al., A gender-specific discriminator in Sprague-Dawley rat urine: the deployment of a metabolic profiling strategy for biomarker discovery and identification. *Anal. Biochem.* 2007, *362*, 182–192.

[23] Lenz, E., Bright, J., Wilson, I., Hugues, A. et al., Metabonomics, dietary influences and cultural differences: a 1 H NMR-based study of urine samples obtained from healthy British and Swedish subjects. *J. Pharm. Biomed. Anal.* 2004, *36*, 841–849.

[24] Stella, C., Beckwith-Hall, B., Cloarec, O., Holmes, E. et al., Susceptibility of human metabolic phenotypes to dietary modulation. *J. Proteome Res.* 2006, *5*, 2780–2788.

[25] Van Ravenzwaay, B., Coelho-Palermo Cunha, G., Strauss, V., Wiemer, J. et al., The individual and combined metabolite profiles (metabolomics) of dibutylphthalate and di(2-ethylhexyl)phthalate following a 28-day dietary exposure in rats. *Toxicol. Lett.* 2010, *198*, 159–170.

[26] Pujos-Guillot, E., Hubert, J., Martin, J.-F., Lyan, B. et al., Mass spectrometry-based metabolomics for the discovery of biomarkers of fruit and vegetable intake: citrus fruit as a case study. *J. Proteome Res.* 2013, *12*, 1645–1659.

[27] Gavaghan, C. L., Wilson, I. D., Nicholson, J. K., Physiological variation in metabolic phenotyping and functional genomic studies: use of orthogonal signal correction and PLS-DA. *FEBS Lett.* 2002, *530*, 191–196.

[28] Llorach, R., Garrido, I., Monagas, M., Urpi-Sarda, M. et al., Metabolomics study of human urinary metabolome modifications after intake of almond (*Prunus dulcis* (Mill.) DA Webb) skin polyphenols. *J. Proteome Res.* 2010, *9*, 5859–5867.

[29] Walsh, M., Nugent, A., Brennan, L., Gibney, M., Understanding the metabolome-challenges for metabolomics. *Nutr. Bull.* 2008, *33*, 316–323.

[30] Kim, K., Mall, C., Taylor, S. L., Hitchcock, S. et al., Mealtime, temporal, and daily variability of the human urinary and plasma metabolomes in a tightly controlled environment. *PLoS One* 2014, *9*, e86223.

[31] Liu, L., Aa, J., Wang, G., Yan, B. et al., Differences in metabolite profile between blood plasma and serum. *Anal. Biochem.* 2010, *406*, 105–112.

[32] Bando, K., Kawahara, R., Kunimatsu, T., Sakai, J. et al., Influences of biofluid sample collection and handling procedures on GC-MS based metabolomic studies. *J. Biosci. Bioeng.* 2010, *110*, 491–499.

[33] Alvarez-Sanchez, B., Priego-Capote, F., Luque de Castro, M., Metabolomics analysis I. Selection of biological samples and practical aspects preceding sample preparation. *Trends Anal. Chem.* 2010, *29*, 111–119.

[34] Teahan, O., Gamble, S., Holmes, E., Waxman, J. et al., Impact of analytical bias in metabonomic studies of human blood serum and plasma. *Anal. Chem.* 2006, *78*, 4307–4318.

[35] Gika, H., Theodoridis, G., Wilson, I., Liquid chromatography and ultra-performance liquid chromatography- mass spectrometry fingerprinting of human urine sample stability under different handling and storage conditions for metabonomics studies. *J. Chromatogr. A* 2008, *1189*, 314–322.

[36] Lauridsen, M., Hansen, S. H., Jaroszewski, J. W., Cornett, C., Human urine as test material in 1 H NMR-based metabonomics: recommendations for sample preparation and storage. *Anal. Chem.* 2007, *79*, 1181–1186.

[37] Vuckovic, D., Current trends and challenges in sample preparation for global metabolomics using liquid chromatography–mass spectrometry. *Anal. Bioanal. Chem.* 2012, *403*, 1523–1548.

[38] Alvarez-Sanchez, B., Priego-Capote, F., Luque de Castro, M., Metabolomics analysis II. Preparation of biological samples prior to detection. *Trends Anal. Chem.* 2010, *29*, 120–127.

[39] Krastanov, A., Metabolomics: the state of art. *Biotechnol. Biotech. Eq.* 2010, *24*, 1537–1543.

[40] Dunn, W. B., Broadhurst, D., Begley, P., Zelena, E. et al., Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry. *Nat. Protoc.* 2011, *6*, 1060–1083.

[41] Raterink, R.-J., Lindenburg, P. W., Vreeken, R. J., Ramautar, R., Hankemeier, T., Recent developments in sample-pretreatment techniques for mass spectrometry-based metabolomics. *Trends Anal. Chem.* 2014, doi:10.1016/j.trac.2014.06.003.

[42] Gika, H. G., Macpherson, E., Theodoridis, G. A., Wilson, I. D., Evaluation of the repeatability of ultra-performance liquid chromatography-TOF-MS for global metabolic profiling of human urine samples. *J. Chromatogr. B* 2008, *871*, 299–303.

[43] Want, E. J., Wilson, I. D., Gika, H., Theodoridis, G., Plumb, R. S. et al., Global metabolic profiling procedures for urine using UPLC-MS. *Nat. Protoc.* 2010, *5*, 1005–1018.

[44] Legido-Quigley, C., Stella, C., Perez-Jimenez, F., Lopez-Miranda, J. et al., Liquid chromatography–mass

spectrometry methods for urinary biomarker detection in metabonomic studies with application to nutritional studies. *Biomed. Chromatogr.* 2010, *24*, 737–743.

[45] Antignac, J.-P., de Wasch, K., Monteau, F., de Brabander, H. et al., The ion suppression phenomenon in liquid chromatography–mass spectrometry and its consequences in the field of residue analysis. *Anal. Chim. Acta* 2005, *529*, 129–136.

[46] Dervilly-Pinel, G., Courant, F., Chéreau, S., Royer, A.-L. et al., Metabolomics in food analysis: application to the control of forbidden substances. *Drug Test. Anal.* 2012, *4*, 59–69.

[47] Jacob, C., Dervilly-Pinel, G., Biancotto, G., Le Bizec, B., Evaluation of specific gravity as normalization strategy for cattle urinary metabolome analysis. *Metabolomics* 2014, *10*, 627–637.

[48] Polson, C., Sarkar, P., Incledon, B., Raguvaran, V., Grant, R. Optimization of protein precipitation based upon effectiveness of protein removal and ionization effect in liquid chromatography–tandem mass spectrometry. *J. Chromatogr. B* 2003, *785*, 263–275.

[49] Want, E. J., O'Maille, G., Smith, C. A., Brandon, T. R. et al., Solvent-dependent metabolite distribution, clustering, and protein extraction for serum profiling with mass spectrometry. *Anal. Chem.* 2006, *78*, 743–752.

[50] Bruce, S. J., Tavazzi, I., Parisod, V., Rezzi, S. et al., Investigation of human blood plasma sample preparation for performing metabolomics using ultrahigh performance liquid chromatography/mass spectrometry. *Anal. Chem.* 2009, *81*, 3285–3296.

[51] Courant, F., Antignac, J., Monteau, F., Le Bizec, B., Metabolomics as a potential new approach for investigating human reproductive disorders. *J. Proteome Res.* 2013, *12*, 2914–2920.

[52] Anizan, S., Bichon, E., Di Nardo, D., Monteau, F. et al., Screening of 4-androstenedione misuse in cattle by LC–MS/MS profiling of glucuronide and sulfate steroids in urine. *Talanta* 2011, *86*, 186–194.

[53] t'Kindt, R., Morreel, K., Deforce, D., Boerjan, W., Van Bocxlaer, J. Joint GC–MS and LC–MS platforms for comprehensive plant metabolomics: repeatability and sample pre-treatment. *J. Chromatogr. B* 2009, *877*, 3572–3580.

[54] El Rammouz, R., Létisse, F., Durand, S., Portais, J. et al., Analysis of skeletal muscle metabolome: evaluation of extraction methods for targeted metabolite quantification using liquid chromatography tandem mass spectrometry. *Anal. Biochem.* 2010, *398*, 169–177.

[55] Fiehn, O., Extending the breadth of metabolite profiling by gas chromatography coupled to mass spectrometry. *Trends Anal. Chem.* 2008, *27*, 261–269.

[56] Pasikanti, K., Ho, P., Chan, E., Gas chromatography/mass spectrometry in metabolic profiling of biological fluids. *J. Chromatogr. B* 2008, *871*, 202–211.

[57] Spann, N., Aldridge, D. C., Griffin, J. L., Jones, O. A. H., Size-dependent effects of low level cadmium and zinc exposure on the metabolome of the Asian clam, *Corbicula fluminea*. *Aquat. Toxicol.* 2011, *105*, 589–599.

[58] Gao, X., Pujos-Guillot, E., Martin, J.-F., Galan, P. et al., Metabolite analysis of human fecal water by gas chromatography/mass spectrometry with ethyl chloroformate derivatization *Anal. Biochem.* 2009, *393*, 163–175.

[59] Dervilly-Pinel, G., Weigel, S., Lommen, A., Chereau, S. et al., Assessment of two complementary liquid chromatography coupled to high resolution mass spectrometry metabolomics strategies for the screening of anabolic steroid treatment in calves. *Anal. Chim. Acta* 2011, *700*, 144–154.

[60] Junot, C., Fenaille, F., Colsch, B., Bécher, F., High resolution mass spectrometry based techniques at the crossroads of metabolic pathways. *Mass Spectrom. Rev.* 2013, doi:10.1002/mas.21401

[61] Wagner, S., Scholz, K., Donegan, M., Burton, L. et al., Metabolomics and biomarker discovery: LC-MS metabolic profiling and constant neutral loss scanning combined with multivariate data analysis for mercapturic acid analysis. *Anal. Chem.* 2006, *78*, 1296–1305.

[62] Bajad, S. U., Lu, W. Y., Kimball, E. H., Yuan, J., Peterson, C. et al., Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. *J. Chromatogr. A* 2006, *1125*, 76–88.

[63] Madalinski, G., Godat, E., Alves, S., Lesage, D. et al., Direct introduction of biological samples into a LTQ-Orbitrap hybrid mass spectrometer as a tool for fast metabolome analysis. *Anal. Chem.* 2008, *80*, 3291–3303.

[64] Allwood, J. W., Parker, D., Beckmann, M., Draper, J., Goodacre, R., Fourier transform ion cyclotron resonance mass spectrometry for plant metabolite profiling and metabolite identification. *Methods Mol. Biol.* 2012, *860*, 157–176.

[65] Junot, C., Madalinski, G., Tabet, J.-C., Ezan, E. Fourier transform mass spectrometry for metabolome analysis. *Analyst* 2010, *135*, 2203–2219.

[66] Theodoridis, G., Gika, H., Wilson, I., LC-MS-based methodology for global metabolite profiling in metabonomics/metabolomics. *Trends Anal. Chem.* 2008, *27*, 251–260.

[67] Idborg, H., Zamani, L., Edlund, P.-O., Schuppe-Koistinen, I., Jacobsson, S., Metabolic fingerprinting of rat urine by LC/MS Part1. Analysis by hydrophilic interaction liquid chromatography-electrospray ionization mass spectrometry. *J. Chromatogr. B* 2005, *828*, 9–13.

[68] Nordström, A., Want, E., Northen, T., Lehtiö, J., Siuzdak, G., Multiple ionisation mass spectrometry strategy used to reveal the complexity of metabolomics. *Anal. Chem.* 2008, *80*, 421–429.

[69] Theodoridis, G., Gika, H., Want, E., Wilson, I., Liquid chromatography–mass spectrometry based global metabolite profiling: a review. *Anal. Chim. Acta* 2012, *711*, 7–16.

[70] Katajamaa, M., Oresic, M., Data processing for mass spectrometry-based metabolomics. *J. Chromatogr. A* 2007, *1158*, 318–328.

[71] Codrea, M., Jiménez, C., Heringa, J., Marchiori, E., Tools for computational processing of LC-MS datasets: a user's perspective. *Comput.Methods Programs Biomed.* 2007, *86*, 281–290.

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

[73] Kuhl, C., Tautenhahn, R., Bottcher, C., Larson, T. S N., CAMERA: an integrated strategy for compound spectra extraction and annotation of liquid chromatography/mass spectrometry data sets. *Anal. Chem.* 2012, *84*, 283–289.

[74] Brown, M., Wedge, D., Goodacre, R., Kell, D. et al., Automated workflows for accurate mass-based putative metabolite identification in LC/MS-derived metabolomic datasets. *BMC Bioinformatics* 2011, *27*, 1108–1112.

[75] Scheltema, R., Jankevics, A., Jansen, R., Swertz, M., Breitling, R., PeakML/mzMatch: a file format, Java library, R library, and tool-chain for mass spectrometry data analysis. *Anal. Chem.* 2011, *83*, 2786–2793.

[76] Guo, Y., Graber, A., McBurney, R., Balasubramanian, R., Sample size and statistical power considerations in high-dimensionality data settings: a comparative study of classification algorithms. *BMC Bioinformatics* 2010, *11*, 1–19.

[77] Brown, M., Dunn, W., Ellis, D., Goodacre, R. et al., A metabolome pipeline: from concept to data to knowledge. *Metabolomics* 2005, *1*, 39–51.

[78] Trygg, J., Holmes, E., Lundstedt, T., Chemometrics in metabonomics. *J. Proteome Res.* 2007, *6*, 469–479.

[79] Hendricks, M., van Eeuwijk, F., Jellema, R., Westerhuis, J. et al., Data-processing strategies for metabolomics studies. *Trends Anal. Chem.* 2011, *30*, 1685–1698.

[80] van den Berg, R., Hoefsloot, H., Westerhuis, J., Smilde, A., van der Werf, M., Centering, scaling, and transformations: improving the biological information content of metabolomics data. *BMC Genomics* 2006, *7*, 142.

[81] Wishart, D., Advances in metabolite identification. *Bioanalysis* 2011, *3*, 1769–1782.

[82] Werner, E., Heilier, J., Ducruix, C., Ezan, E. et al., Mass spectrometry for the identification of the discriminating signals from metabolomics: current status and future trends. *J. Chromatogr. B* 2008, *871*, 143–163.

[83] Bowen, B. P., Northen, T. R., Dealing with the unknown: metabolomics and metabolite atlases. *J. Am. Soc. Mass Spectrom.* 2010, *21*, 1471–1476.

[84] Watson, D., A rough guide to metabolite identification using high resolution liquid chromatography mass spectrometry in metabolomic profiling in metazoans. *Comput. Struct. Biotechnol. J.* 2013, *4*, e201301005.

[85] NIST, National Institute of Standards and Technology Database. Available at http://www.nist.gov/

[86] Wishart, D. S., Knox, C., Guo, A. C., Eisner, R. et al., HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res.* 2009, *37*, D603–D610.

[87] Smith, C., O'Maille, G., Want, E., Qin, C. et al., METLIN : a metabolite mass spectral database. *Ther. Drug Monit.* 2005, *27*, 747–751.

[88] Fahy, E., Sud, M., Cotter, D., Subramaniam, S., LIPID MAPS online tools for lipid research. *Nucleic Acids Res.* 2007, *35*, W606–W612.

[89] Kanehisa, M., Araki, M., Goto, S., Hattori, M. et al., KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 2008, *36*, D480–D484.

[90] Kind, T., Fiehn, O., Seven golden rules for heuristic filtering of molecumar formulas obtained by accurate mass spectrometry. *BMC Bioinformatics* 2007, *8*, 105.

[91] Xiao, J., Zhou, B., Ressom, H., Metabolite identification and quantitation in LC-MS/MS-based metabolomics. *Trends Anal. Chem.* 2012, *32*, 1–14.

[92] Sumner, L., Amberg, A., Barrett, D., Beale, M. et al., Proposed minimum reporting standards for chemical analysis. Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* 2007, *3*, 211–221.

[93] Roux, A., Xu, Y., Heilier, J.-F., Olivier, M.-F. et al., Annotation of the human adult urinary metabolome and metabolite identification using ultra high performance liquid chromatography coupled to a linear quadrupole ion trap-Orbitrap mass spectrometer. *Anal. Chem.* 2012, *84*, 6429–6437.

[94] Courant, F., Royer, A.-L., Chereau, S., Morvan, M.-L. et al., Implementation of a semi-automated strategy for the annotation of metabolomic fingerprints generated by liquid chromatography-high resolution mass spectrometry from biological samples. *Analyst* 2012, *137*, 4958–4967.

[95] Putri, S., Nakayama, Y., Matsuda, F., Uchikata, T. et al., Current metabolomics: practical applications. *J. Biosci.Bioeng.* 2013, *115*, 579–589.

[96] Wishart, D., Applications of metabolomics in drug discovery and development. *Drugs RD* 2008, *9*, 307–322.

[97] Beyoglu, D., Idle, J., Metabolomics and its potential in drug development. *Biochem. Pharmacol.* 2013, *85*, 12–20.

[98] Zhang, A., Sun, H., Wang, X., Serum metabolomics as a novel diagnostic approach for disease: a systematic review. *Anal. Bioanal. Chem.* 2012, *404*, 1239–1245.

[99] Lewis, G., Asnani, A., Gerszten, R., Application of metabolomics to cardiovascular biomarker and pathway discovery. *J. Am. Coll. Cardiol.* 2008, *52*, 117–123.

[100] Spratlin, J., Serkova, N., Eckhardt, S., Clinical applications of metabolomics in oncology: a review. *Clin. Cancer Res.* 2009, *15*, 431–440.

[101] Robertson, D. Metabonomics in toxicology: a review. *Toxicol. Sci.* 2005, *85*, 809–822.

[102] Werner, E., Croixmarie, V., Umbdenstock, T., Ezan, E. et al., Mass spectrometry-based metabolomics: accelerating the characterization of discriminating signals by combining statistical correlations and ultrahigh resolution. *Anal. Chem.* 2008, *80*, 4918–4932.

[103] Hao, D.-F., Xu, W., Wang, H., Du, L.-F. et al., Metabolomic analysis of the toxic effect of chronic low-dose exposure to acephate on rats using ultra-performance liquid chromatography/mass spectrometry. *Ecotoxicol. Environ. Saf.* 2012, *83*, 25–33.

[104] Okazaki, Y., Saito, K., Recent advances of metabolomics in plant biotechnology. *Plant Biotechnol. Rep.* 2012, *6*, 1–15.

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

[106] Duan, G., Christian, N., Schwachtje, J., Walther, D., Ebenhöh, O., The metabolic interplay between plants and phytopathogens. *Metabolites* 2013, *3*, 1–23.

[107] Viant, M. R. Metabolomics of aquatic organisms: the new 'omics' on the block. *Mar. Ecol. Prog. Ser.* 2007, *332*, 301–306.

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

[109] Castro-Puyana, M., Mendiola, J. A., Ibáñez, E., Herrero, M., in: Cifuentes, A. (Ed.), *Foodomics: Advanced Mass Spectrometry in Modern Food Science and Nutrition*, John Wiley & Sons, Inc., Hoboken, NJ, USA. 2013.

[110] Nebbia, C., Urbani, A., Carletti, M., Gardini, G. et al., Novel strategies for tracing the exposure of meat cattle to illegal growth-promoters. *Vet. J.* 2011, *189*, 34–42.

[111] Scalbert, A., Brennan, L., Fiehn, O., Hankemeier, T. et al., Mass-spectrometry-based metabolomics: limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics* 2009, *5*, 435–458.

[112] Rezzi, S., Ramadan, Z., Fay, L., Kochhar, S., Nutritional metabonomics: applications and perspectives. *J. Proteome Res.* 2007, *6*, 513–525.

[113] Courant, F., Martzolff, A., Rabin, G., Antignac, J.P. et al., How metabolomics can contribute to bio-processes: a proof of concept study for biomarkers discovery in the context of nitrogen-starved microalgae grown in photobioreactors. *Metabolomics* 2013, *9*, 1286–1300.

[114] Toya, Y., Shimizu, H., Flux analysis and metabolomics for systematic metabolic engineering of microorganisms. *Biotechnol. Adv.* 2013, *6*, 818–826.

[115] Courant, F., Pinel, G., Bichon, E., Monteau, F. et al., Development of a metabolomic approach based on liquid chromatography high resolution mass spectrometry to screen for clenbuterol abuse in calves. *Analyst* 2009, *134*, 1637–1646.

[116] Dervilly-Pinel, G., Chereau, S., Cesbron, N., Monteau, F., Le Bizec, B., LC-HRMS based metabolomics screening model to detect various β-agonists treatments in bovines. *Metabolomics* 2014, doi:10.1007/s11306-014-0705-3.

[117] Johnson, C., Gonzales, F., Challenges and opportunities of metabolomics. *J. Cell. Physiol.* 2012, *227*, 2975–2981.

[118] Cottret, L., Wildridge, D., Vinson, F., Barrett, M. P. et al., MetExplore: a web server to link metabolomic experiments and genome-scale metabolic networks. *Nucleic Acids Res.* 2010, *38*, W132–W137.

[119] Caspi, R., Foerster, H., Fulcher, C. A., Kaipa, P. et al., The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 2008, *36*, D623–D631.

[120] Sangster, T., Major, H., Plumb, R., Wilson, A., Wilson, I., A pragmatic and readily implemented quality control strategy for HPLC-MS and GC-MS-based metabonomic analysis. *Analyst* 2006, *131*, 1075–1078.

[121] Lindon, J., Nicholson, J., Holmes, E., Keun, H. et al., Summary recommendations for standardization and reporting of metabolic analyses. *Nat. Biotechnol.* 2005, *23*, 833–838.

[122] Haug, K., Salek, R. M., Conesa, P., Hastings, J. et al., MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res.* 2013, *41*, D781–D786.

[123] Kuehnbaum, N. L., Britz-McKibbin, P., New advances in separation science for metabolomics: resolving chemical diversity in a post-genomic era. *Chem. Rev.* 2013, *113*, 2437–2468.

[124] Ralston-Hooper, K., Hopf, A., Oh, C., Zhang, X. et al., Development of GCxGC/TOF-MS metabolomics for use in ecotoxicological studies with invertebrates. *Aquat. Toxicol.* 2008, *88*, 48–52.

[125] Putri, S., Yamamoto, S., Tsugawa, H., Fususaki, E., Current metabolomics: technological advances. *J. Biosci. Bioeng.* 2013, *116*, 9–16.

[126] Dwivedi, P., Schultz, A. J., Hill, H. H., Metabolic profiling of human blood by high resolution ion mobility mass spectrometry (IM-MS). *Int. J. Mass Spectrom.* 2010, *298*, 78–90.

[127] Creek, D., Jankevics, A., Burgess, K., Breitling, R., Barrett, M., IDEOM: an excel interface for analysis of LC-MS based metabolomics data. *BMC Bioinformatics* 2012, *28*, 1048–1049.

[128] Gallart-Ayala, H., Courant, F., Severe, S., Antignac, J. P. et al., Versatile lipid profiling by liquid chromatography-high resolution mass spectrometry using all ion fragmentation and polarity switching. Preliminary application for serum samples phenotyping related to canine mammary cancer. *Anal. Chim. Acta* 2013, *796*, 75–83.

[129] Champarnaud, E., Hopley, C., Evaluation of the comparability of spectra generated using a tuning point protocol on twelve electrospray ionisation tandem-in-space mass spectrometers. *Rapid Commun. Mass Spectrom.* 2011, *25*, 1001–1007.

