Journal of Chromatography B, 871 (2008) 143–163

Contents lists available at ScienceDirect

Journal of Chromatography B

journal homepage: www.elsevier.com/locate/chromb

Review

Mass spectrometry for the identification of the discriminating signals from metabolomics: Current status and future trends☆

Erwan Werner[a], Jean-François Heilier[a,b], Céline Ducruix[a,1], Eric Ezan[a], Christophe Junot[a,*], Jean-Claude Tabet[c,*]

[a] Commissariat à l'Energie Atomique, DSV-iBiTec-S-SPI, Laboratoire d'analyse du Métabolisme du Médicament, 91191 Gif-sur-Yvette, France
[b] Université catholique de Louvain, Faculty of Medicine, Industrial Toxicology and Occupational Medicine Unit, 1200 Brussels, Belgium
[c] Laboratoire de Chimie Structurale Organique et Biologique, CNRS UMR 7613, UPMC Univ. Paris 06, F-75005 Paris, France

ARTICLE INFO

Article history:
Received 25 February 2008
Accepted 1 July 2008
Available online 12 July 2008

Keywords:
Metabolomics
Metabonomics
Identification
Mass spectrometry

ABSTRACT

The metabolome is characterized by a large number of molecules exhibiting a high diversity of chemical structures and abundances, requiring complementary analytical platforms to reach its extensive coverage. Among them, atmospheric pressure ionization mass spectrometry (API-MS)-based technologies, and especially those using electrospray ionization are now very popular. In this context, this review deals with strengths, limitations and future trends in the identification of signals highlighted by API-MS-based metabolomics. It covers the identification process from the determination of the molecular mass and/or its elemental composition to the confirmation of structural hypotheses. Furthermore, some tools that were developed in order to address the MS signal redundancy and some approaches that could facilitate identification by improving the visualization and organization of complex data sets are also reported and discussed.

Contents

Abbreviations: ACS, American Chemical Society; APCI, atmospheric pressure chemical ionization; API, atmospheric pressure ionization; APPI, atmospheric pressure photo ionization; ARM, atomic reconstruction of metabolism; ArMet, architecture for metabolomics; CAS, chemical abstract service; CE, capillary electrophoresis; ChEBI, chemical entities of biological interest; CID, collision-induced dissociation; CID-SORI, collision-induced dissociation-sustained off-resonance irradiation; CSLS, chemical structure lookup service; C-trap, curve linear trap (shaped like the letter "C"); DART, direct analysis in real time; DRE, dynamic range enhancement; E_{com}, energy at the centre of mass; EI, electron ionization (*a.k.a.* electronic impact); EID, electron-induced dissociation; E_{lab}, energy in the laboratory framework; ERMS, energy resolved mass spectrometry; ESI, electrospray ionization; eV, electron volt; FTICR, Fourier transform ion cyclotron resonance; FT/MS, Fourier transform mass spectrometry; FWHM, full width at half maximum (peak height); GC, gas chromatography; H/D exchange, hydrogen/deuterium exchange; HMDB, Human Metabolome DataBase; IR, infra-red; IRMPD, infra-red multiphoton dissociation; ITMS, ion trap mass spectrometer; KEGG, Kyoto Encyclopaedia Of Genes And Genomes; LC, liquid chromatography; LMSD, The Lipid Maps Structure Database; LTQ, linear ion trap; LTQ-Orbitrap®, linear ion trap-Orbitrap®; m/z, mass-to-charge ratio; MALDI, matrix-assisted laser desorption ionization; MeMo, metabolomic modelling project; METPR, metabolite enrichment by tagging and proteolytic release; MIAMET, minimum information about a mettabolomics experiment; M_{iso}, monoisotopic mass; MoTo DB, the metabolome database for tomato; mRNA, messenger ribonucleic acid; MS, mass spectrometry; MS^E, untargeted mixed-mode (low- and high-energy) tandem mass spectrometry; MS^n, multistage (sequential) tandem mass spectrometry; MS/MS, tandem mass spectrometry; MVA, multivariate (data) analysis; MW, molecular weight; MySQL, My Structured Query Language; NCBI, National Center for Biotechnology Information; NCI, National Cancer Institute; NIST, National Institute of Standards and Technology; NMR, nuclear magnetic resonance; PHP, hypertext preprocessor; ppm, part per million; QqTOF, hybrid quadrupole time-of-flight mass spectrometer; Q TRAP®, hybrid quadrupole linear ion trap; RDBE, ring plus double bound equivalent; Rf-only, radiofrequence only; RIA, relative isotopic abundance measurement; SMB, supersonic molecular beam; SQL, Structured Query Language; TAIR, The Arabidopsis Information Resource; Th, Thompson; TOF, time of flight; TQ, triple quadrupole; u, atomic mass unit; UPLC, ultra performance liquid chromatography; UV, ultra-violet; UV–vis, ultra-violet–visible; XML, eXtensible Markup Language.

☆ This paper is part of a special volume entitled "Hyphenated Techniques for Global Metabolite Profiling", guest edited by Georgios Theodoridis and Ian D. Wilson.

* Corresponding authors.
E-mail addresses: christophe.junot@cea.fr (C. Junot), tabet@ccr.jussieu.fr (J.-C. Tabet).
[1] Present address: The Cancer Research UK Centre for Cancer Therapeutics, The Institute of Cancer Research, 15 Cotswold Road, Belmont, Sutton, Surrey, SM2 5NG, United Kingdom.

1570-0232/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2008.07.004

1. Introduction

Metabolomics deals with the comprehensive analysis of metabolites present in a biological sample by the combined use of analytical methods and multivariate statistical analyses (MVA) [1,2]. It emerged at the end of the 1990s as the third major path of functional genomics beside mRNA profiling (transcriptomics) and proteomics [3,4].

Metabolomics experiments start with the acquisition of metabolic fingerprints using various analytical platforms (RMN, GC/MS, LC/MS, FT/IR). When using MS-based methods, the resulting fingerprints are then pre-processed (*i.e.*, suppression of background chemical noise, variable alignment, peak picking) [5] before being eventually pre-treated (centering, scaling, etc.) [6], and finally processed. Processing generally relies on MVA, which facilitate data visualization through data dimensionality reduction and highlight relevant biological information [7,8]. Finally, identification of the discriminating signals is undertaken by combining mass spectrum analysis and database consultation.

Four particular issues should be addressed in the frame of metabolomics. First, the large diversity of metabolites in terms of chemical structures and concentrations prohibits the coverage of the whole metabolome with a single analytical method. Several analytical platforms are therefore used, relying either on nuclear magnetic resonance (NMR) [9], or on mass spectrometry with different ion sources and mass analyzers [10–19]. Each tool provides complementary but sometimes redundant information.

Secondly, processing the data prior to MVA may also be a source of errors. Artifactual signals could be generated due to erratic chemical background noise suppression or insufficient correction of retention time shifts when coupling with liquid chromatography (LC), and metabolites present in trace amounts or which poorly ionize may not be extracted due to low-intensity signals. In addition, the consistency of MVA results may be impacted not only by analytical and data pre-processing issues, but also by the scaling and normalization procedures [20–22].

Thirdly, another matter of concern is the high number of MS signals to be identified after MVA. A metabolite does not produce a single m/z peak since adduct and product ions can be generated during the desolvation step following the API process. Although informative, such a signal redundancy slows down the identification procedure by increasing the number of variables that have to be investigated and also by leading to unsuccessful database queries.

Fourthly, the identification of discriminating signals is perhaps the most laborious and time-consuming step of the metabolomics experiment. It is usually achieved by searching databases with the recorded metabolite molecular mass, the elemental composition, or the whole mass spectra. Rigorously, a panel of complementary spectroscopic methods such as NMR, UV, MS and IR detection is required to reach this goal [23,24].

Atmospheric pressure ionization mass spectrometry (API-MS)-based techniques have rapidly emerged as a popular and powerful tool for metabolomics. When combined with liquid chromatography, they exhibit a good sensitivity, high dynamic range and versatility but also provide soft ionization conditions giving access to the molecular mass of intact molecules from complex mixtures. Furthermore, the development of high and ultra-high resolution analyzers (TOF/MS and FT/MS, respectively) has yielded accurate mass measurements, whereas ion trap analyzers, among others, are used to perform MS^n experiments to get the additional structural information needed for metabolite identification [25,26]. Unfortunately, API-MS exhibits poor reproducibility and high inter-instrument variability in the generation of fragmentation patterns, and this has hampered the constitution of universal databases as done with electron ionization mass spectrometry [27] or with NMR [28].

For all these reasons, formal identification of metabolite structure is not easily achieved by using the API-MS experiments alone.

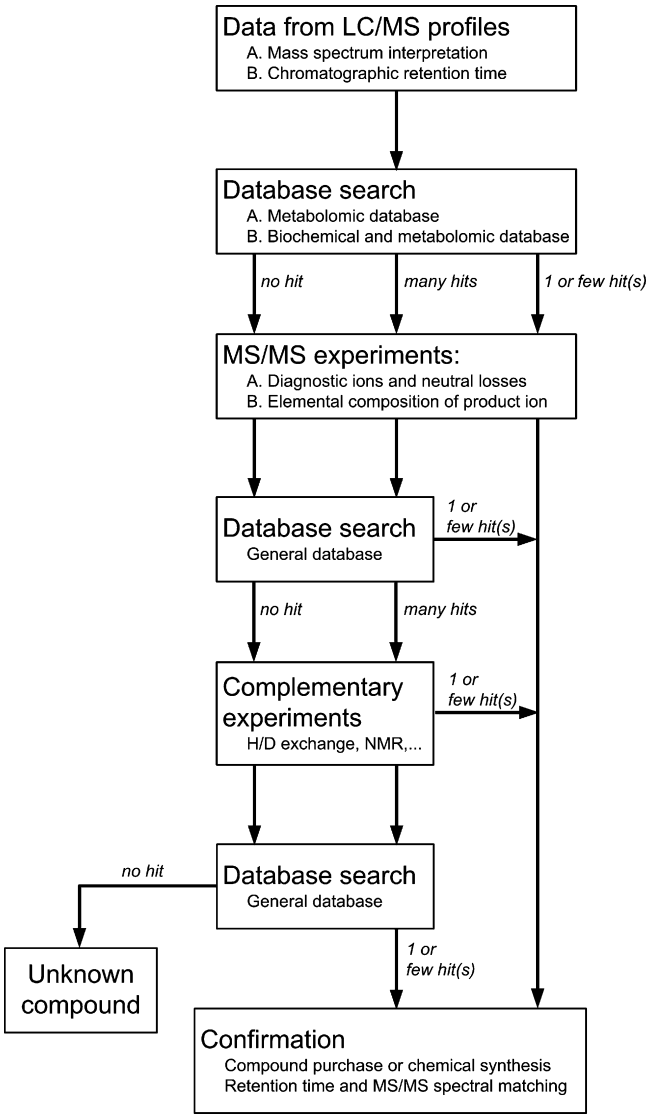


Fig. 1. Flowchart of a proposed strategy for the identification of potential biomarkers highlighted by MS-based metabolomics using complementary LC/API-MS(/MS) approaches and database queries.

As a result, most MS signals detected in metabolic fingerprints remain unidentified.

Recently, Chen et al. [29] reported an integrated approach for the identification of biomarkers detected in LC/MS fingerprints, based on LC/MS^2 experiments, micropreparative ultra-performance liquid chromatography (UPLC), Fourier transform ion cyclotron resonance mass spectrometry (FTICR/MS), gas chromatography (GC) retention time indices, database search and generation of stable-isotope-labeled standards. Based on the work of Chen et al. [29] and in our experience, a flow chart of the identification process of signals from metabolomics relying on complementary API-MS strategies is presented in Fig. 1. It starts with an interpretation of the mass spectra in order to ensure that the signal of interest corresponds to a monoisotopic ion and not to a natural isotopologue ion or an adduct ion. One or more relevant elemental compositions are deduced from accurate mass measurements if high or very high-resolution mass spectrometry is available, for further database queries. Collision-induced dissociation (CID) spectra are then acquired and interpreted in order to get information about the chemical structure. At this stage, chemical database queries may be refined and the highlighted compounds, if any, are kept for further consideration or ruled out based on chromatographic retention time and CID mass spectra information. Complementary experiments (*i.e.*, other sequential MS^n experiments or H/D exchanges) may be required to deduce a putative structure. Finally, formal identification is achieved when the metabolite to be characterized exhibits the same retention time and CID spectra as those of the reference molecule that has to be purchased or synthesized.

In this context, this review deals with advantages, limitations and future trends of API mass spectrometry for the identification of signals from LC/MS-based metabolomics. It aims to cover the identification process, from the determination of the molecular mass and/or elemental composition, to the confirmation of structural hypotheses. Furthermore, as the identification process cannot be reduced solely to the interpretation of mass and CID spectra, some complementary tools developed to address MS signal redundancy and some approaches that could facilitate identification by improving the visualization and organization of complex data sets will also be reported and commented upon.

2. Mass spectrometry-based tools

The coupling of liquid chromatography with mass spectrometry uses API sources to convert aqueous LC effluent containing analytes to gas-phase molecules. Ionization at atmospheric pressure differs from ionization under high vacuum conditions by the ability of the generated precursor ions to dissipate part of their acquired internal energy. In high-vacuum ionization conditions, the internal energy is not strictly controlled, ranging from 0 to 15 eV for electron ionization and being less than 3 eV for chemical ionization [30]. The situation is entirely different under API conditions for which the ion activation/cooling allows the production of more or less energetic charged molecular species [31]. By using gas phase ionization, *via* atmospheric pressure chemical ionization (APCI)[32], atmospheric pressure photo ionization (APPI) [33] or direct analysis in real time (DART) [34,35], it is possible to produce stable MH^+ (or $[M+NH_4]^+$) and $[M-H]^-$ (or $[M-X]^-$, X being an halogen anion or other deprotonated organic and inorganic acid) and to a lesser extent, $M^{\bullet+}$ or $M^{\bullet-}$ (mainly in APPI). With electrospray (ESI) [36] and desorption electrospray ionisation [37–39], although ions are produced by desorption, the desolvation occurring at the skimmer produces the same ions as those observed under gas phase ionization, except that ESI is known to produce multiply charged ions. Finally, under API conditions, ion aggregates are desolvated at the skimmer where the internal ion energy can be modulated through the controlled nozzle-skimmer potential difference in order to either analyze intact species or obtain in source CID spectra [40,41].

Among the high diversity of analyzer technologies that can be implemented with LC/ESI-MS systems, the analyst has the choice between two possible approaches: "in-space" instruments, and "in-time" mass spectrometers [42,43]. Two kinds of "in-space" instruments can be distinguished: (i) without field scanning devices (*i.e.*, TOF/MS-MS), and (ii) electrostatic or magnetic field scanning instruments (*i.e.*, quadrupole and sector analyzers, respectively).

"In-time" mass spectrometers are based upon ion storage in ion traps, enabling sequential MS/MS experiments. Different kinds of ion traps are commercially available: some of them rely on the scanning of quadrupolar or nonlinear fields, which is the case of 3D (Paul's trap) [44–46] and 2D (linear) ion traps [17,47], whereas others are based on ion trapping by using either an electromagnetic field (*i.e.*, the Penning trap used for ICR instruments) [48] or a quadro-logarithmic electrostatic field, as for the Kingdon cell (*i.e.*, the Orbitrap®) developed by Makarov and coworkers [49–52].

Table 1
Typical values for mass resolving power and mass accuracy of commercially available analyzers

Analyzers	Mass resolving power[a]	Mass accuracy
Triple quadrupoles and ion traps	~500	~0.3 to 1 u
TOF devices	8000–20,000	1–5 ppm[b]
LTQ-Orbitrap	Up to 100,000	<3 ppm
FTICR	Up to 1,000,000[c]	<1 ppm

[a] $m/\Delta m$, FWHM.
[b] When internal calibration or lock mass is used.
[c] Higher values can be achieved depending on the magnetic field intensity.

The Orbitrap® operates by radially trapping ions around a central spindle electrode and then generates a transient ion signal related to the axial ion motion rather than to the radial ion motion used with FTICR. Oscillation frequencies for stored ions with different m/z ratios are then obtained by using Fourier transform processes, resulting in an accurate reading of m/z ratios through the mass scale.

Multi-stage mass spectrometers (*i.e.*, tandem instruments) can be constructed by coupling the large variety of analyzers, leading to either homogenous tandem mass spectrometers such as TOF–TOF, triple quadrupole or multipoles, 2D and 3D (linear) ion traps, or hybrid instruments, such as QqTOF, Q TRAP®, and 2D trap coupled with ICR cell or Orbitrap®.

For identification purposes, the performance of the different kinds of analyzers will be discussed through their mass resolving power and mass accuracy (high and ultra-high resolution instruments), which are two major parameters involved in the determination of elemental compositions. In addition, their ability to provide structural information will be addressed through their capacity to perform two-, three- or multi-stage MS/MS (*i.e.*, MS^n) experiments. The issue of spectral database building with API data will also be discussed.

2.1. The analyzers: mass resolving power and accuracy

The mass resolving power is the capacity of a mass spectrometer to separate ions of adjacent but different m/z ratios. 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.*, $m/\Delta m$, full width at half maximum peak height, FWHM). The mass accuracy is the difference between the exact mass of an ion and its measured mass and is commonly expressed as parts per million (ppm) [43,53,54]. Mass resolving powers and mass accuracies of commercially available analyzers discussed below, are displayed in Table 1.

The mass resolving power of low-resolution analyzers such as ion trap and quadrupole devices does not exceed 500 at m/z 250 ($m/\Delta m$, FWHM) [55]. As a consequence, they permit the separation of adjacent peaks by 1 Th, but not that of isobaric ions at the same nominal m/z ratio. Higher resolution mass spectra can be obtained by reducing the scan rate (up to 5000 at m/z 250 using the Ultra ZoomScan® mode on a 2D ion trap), but the acquisition window is then limited to the 100 Th scale.

High-resolution analyzers based on TOF technology (TOF and hybrid Q-TOF mass spectrometers) provide accurate mass measurements with errors below 5 ppm when internal calibration or lock mass is used, and achieve isobaric ion separation within the limits of their resolving power ranging from 8000 to 20,000 ($m/\Delta m$, FWHM). However, TOF analyzers suffer from a modest dynamic range that is typically around two to three orders of magnitude. This has been improved to four orders of magnitude by using a new technology named Dynamic Range Enhancement (DRE®) [56,57].

By using FT-ICR devices, elemental compositions of low molecular weight analytes (*i.e.*, below 1000 Da) are provided by very accurate mass measurements with sub-ppm errors in mass spectra. However, FT-ICR devices are not popular because they utilize strong and maintenance-expensive magnetic fields. Accurate mass measurements are achieved at the expense of longer recording times (transient signal digitizing duration $\geq$ 1 s). This may limit their coupling with LC and especially ultra high pressure liquid chromatography (*i.e.*, UHPLC) [14]. The increase in superconducting magnetic fields up to 12 T allows better resolution for lower transient signal lifetimes.

More recently, a new technology was introduced with the release of the LTQ-Orbitrap® [52]. This new type of hybrid mass spectrometer consists of a linear ion trap coupled to an Orbitrap analyzer. The linear ion trap (LTQ) [58] is able to record its own full-scan mass spectra and sequential MS^n activation spectra from low-resolution precursor ion selection (mass window $\geq$ 0.3 u). Orbitrap® devices achieve accurate mass measurements with errors below 3 ppm and resolving powers up to 100,000 ($m/\Delta m$ at m/z 1000, FWHM). Their performances in terms of m/z ratio stability and dynamic range of mass accuracy have been evaluated by serial injections of five reference compounds [59]. The Orbitrap® provided accurate mass measurements (error < 2 ppm) for over 24 h post-calibration, whereas the dynamic range of mass accuracy was over three orders of magnitude. Similar values of mass accuracy were found in biological matrices in other studies [60,61].

High and ultra-high resolution analyzers are becoming increasingly popular in the field of metabolomics because they provide accurate mass measurements, which are useful for the determination of elemental compositions of metabolites, together with MS/MS or sequential MS^n experiments which provide useful additional structural information, especially when product ions are analyzed at high resolution. However, there is a lack of comparative studies about their analytical performances in biological matrices in terms of dynamic ranges of concentration and mass accuracy.

2.2. The determination of elemental compositions

When accurate mass measurements are available, the elemental composition determination of monoisotopic ions (mass noted as M_{iso}) is the first step of metabolite characterization from the mass spectrum because it provides a simple, efficient and automatable way to search chemical and metabolic databases.

It is often believed that mass accuracy is the most important parameter for the determination of elemental compositions. Nevertheless, the number of possible elemental compositions increases exponentially with increasing ion mass, even with ultra-high resolution instruments (*e.g.*, 15 different molecular formulas for mass > 600 Th within a 1-ppm window) [62]. Therefore, restrictive criteria based on physico-chemical rules and spectral information, as found in mass spectrometry textbooks [43,53] (*i.e.*, nitrogen rules, valence considerations, isotopic patterns), are required in order to remove irrelevant proposals.

The most popular of these chemical rules is the nitrogen rule that states that odd nominal molecular mass compounds contain an odd number of nitrogen atoms [54,63]. It is important to keep in mind that ESI produces almost exclusively even electron ions (during the ionization step) and therefore facilitates the determination of the integer nominal mass of the compound. Instrument software that calculate the elemental compositions can automatically take care of the nitrogen rule, which is also important for the product ions since it helps to determine the number of nitrogen atoms lost or if the product ion is a radical (*i.e.*, odd electron). Nevertheless, such a rule becomes unreliable for masses above 500 u because the large number of elements contributing to the mass may result in a mass defect above 0.5 u that leads to overestimation of the integer mass by 1 u.

The number of ring plus double bond equivalents (RDBE) in a molecule consisting of $C_aH_bN_cO_d$ can be evaluated by the formula: RDBE = $a-(1/2)b+(1/2)c+1$ [64]. Depending upon their valence, other elements should be counted as C, H or N. For example, in biochemical applications, halogen atoms that have a valence state of 1 should be counted as hydrogen atoms, whereas O, which has a valence of 2 has not to be included with C, H, and N since it has no impact on the number of H attached to a given C. RDBE for singly protonated ions should be above −0.5 for an elemental composition to be theoretically possible. Furthermore, the RDBE formula distinguishes between odd- and even-electron ions, since RDBE values are integers in the first case and non-integers with a reminder of 0.5 in the second case [54]. The RDBE formula only takes account of the lowest valence state of elements (*e.g.*, 3 and not 5 for phosphorus). Therefore, RDBE determinations of phosphorus-containing molecules may be of limited interest since more than one RDBE value should be obtained [65,66]. In addition, in a biological context, phosphorus is often present in a valence state of five (*i.e.*, as phosphate), which is not considered by the RDBE formula. The situation is far more complicated with sulfur that can have three different valences (*i.e.*, 6, 4 and 2 for sulfone, sulfoxide, and thiol or thioether, respectively). Moreover, if there are two or more sulfur atoms in a given molecule, they can be simultaneously present at different valence states. Advanced softwares calculate a range for RDBE but none to our knowledge uses every valence states of elements to return all possible results because it would require too much calculation time. This shows the limitation of the use of the RDBE formula which is not of great help for advanced structure elucidation when elements other than C, H, N, O are concerned.

Analysis of isotopic patterns is helpful in elemental composition determination. Kind and Fiehn [67] showed that interpretation of isotopic abundance patterns removes more than 95% of false candidate formulas for molecules above 500 u. Moreover, they concluded that instruments with 3 ppm mass accuracy and 2% relative error for isotopic abundance pattern outperformed those with less than 1 ppm accuracy that do not include isotope information in the calculation of molecular formulas [67].

Instruments providing ultra-high resolution (*i.e.*, Orbitrap® and FT-ICR) are of special interest in this context. Thanks to their resolving power, such instruments allow resolution of isotope peaks within isotope clusters. So, the isotopic contribution of each element may be isolated so as to give evidence of atoms with particular isotopic pattern or determine atom amounts.

In addition, the charge state could be determined based upon the fact that the smallest mass difference arising from natural isotopic variations in a molecule is close to 1 u. Thus, the spacing between the $^{12}C_n$ and $^{12}C_{n-1}$ $^{13}C_1$ (isotopologue ions) on the m/z scale is (1 u)/z, which allows the charge to be directly elucidated by measuring the m/z variation between the isotopologue ($^{12}C_n$, $^{12}C_{n-1}$ $^{13}C_1$, $^{12}C_{n-2}$ $^{13}C_2$, and so on) ions [63,68].

A typical example of the determination of a chemical formula from a LTQ-Orbitrap® mass spectrum is provided in Fig. 2. The ESI mass spectrum has been extracted from an LC/MS chromatogram of rat urine in the course of a metabolomics experiment. The discriminating signal to be identified is an ion at m/z 410.1025 in negative mode. The analysis of the mass spectrum before the calculation of possible formulas by the Qualbrowser/Xcalibur® post-processing software reduces the number of hypothetical structures. Analysis of the isotopic pattern reveals: (i) the presence of the first isotope cluster at (M_{iso} + 1), indicating that the ion is singly charged, (ii) the peak at (M_{iso} + 1.0034) that exhibits a relative abundance of 18% indicates a number of carbon atoms close to 16 (compared with ^{12}C, ^{13}C shows 1.1% relative abundance and a difference in mass of 1.00335 u), and (iii) the presence of a single sulfur-atom in the elemental composition (which simultaneously excludes the presence of Cl, Br or K), thanks to the separation of the ^{34}S (peak of 4.4% relative intensity at (M_{iso} + 1.9959)) from the $^{13}C_2$ isotopologue at the (M_{iso} + 2) isotope cluster.

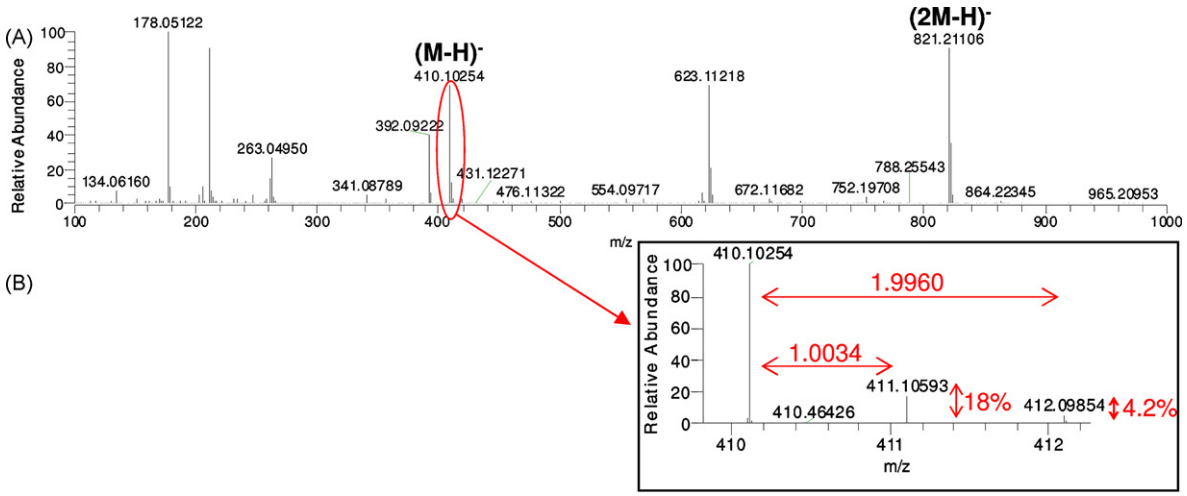
The Xcalibur® formula generator returned 41 possible elemental compositions within 3 ppm for the measured mass at m/z 410.1025, as shown in Fig. 2. If element settings are restricted with (i) at least one sulfur atom and (ii) 15–18 carbon atoms (corresponding to a 10% relative error in isotopic ion abundance measurement), then only four formulas remain. In addition, according to the classical nitrogen rules previously mentioned, the ion of interest should have an odd number of nitrogen atoms. Finally, a single formula remains: $C_{17}H_{20}N_3O_7S$. Simulation of the proposed elemental compositions under the recorded mass spectrum supports this elemental composition.

However, there are some limitations in the use of isotopic patterns for the selection of relevant elemental compositions in biological media. Isotope ratio measurements can be affected by interference and saturation effects due both to the matrix and the experimental conditions (that increase spectral chemical background noise or limit sensitivity). In addition, there is very limited information available in the literature about the precision of relative isotopic ion abundance measurements (RIA) because it depends upon both the matrix and the experimental conditions, and also because many signals cannot be formally identified. As a matter of fact, although it has been stated that isotopic ratios can be measured within a 2% signal intensity precision by using TOF devices [69], Grange et al. [70] reported relative isotopic abundance measurement errors in a hydro-organic medium within 20% of their predicted values for 35 compounds whose ion m/z ratios exceeded 140 Th by using a single MS stage orthogonal accelerated TOF mass spectrometer [70]. Other studies reported excellent agreement between experimental and theoretical distributions for selected ions [60]. However, no statistics about a significant number of measures on replicates and also different ions were provided.

In this context, it could be relevant to use stable isotope labeling when available (*e.g.*, for microorganism or plant metabolomics). The comparison of the monoisotopic masses from unlabeled and labeled biological extracts gives access to the number of both C and N atoms, decreasing the number of possible chemical formulas. A proof of concept has been performed on extracts from the model plant *Arabidopsis thaliana*: double isotopic labeling (^{13}C and ^{15}N) resulted in unique assignment of 87% of 5000 formulas versus 20% with a typical approach [71].

Another challenge of metabolomics is to deal with the huge number of masses to be investigated. It is then imperative to develop automatic procedures for the determination of relevant elemental formulas. In this context, pioneering work has already been performed in the analysis of petroleum oil and natural organic matter by using FTICR-MS [72]. Kujawinski and Behn [73] proposed an automated compound identification algorithm for the analysis of ultra-high resolution mass spectra of natural organic matter. Their approach looks for functional group relationships, such as CH_2, CO_2, NH or C_2H_2O, between all compounds within mass spectra. When applied to Suwannee River dissolved organic matter, such a procedure led to elemental composition assignments in 80–90% of the cases, whereas only 44% of assignments were obtained with bacterial dissolved organic matters. This example stresses the need for additional restrictive criteria when considering biological data in metabolomics.

Koch et al. also proposed some suitable techniques to be implemented in software for the analysis of ultra-high resolution mass spectra [62]. Formulas are generated without any restriction on hypothetical elements, but exclusion criteria, such as instrument error, the nitrogen rule, RDBE, and thresholds for molecular element ratios, are included. Two additional steps are added: (i) the



(C)

m/z	Theoritical Mass	Delta (ppm)	RDBE.	Composition			Reject causes	
410.1025	410.1025	-0.04	-0.5	$C_{11}H_{31}O_3N_3P_3S_2$		H/C	15<Nc<18	sulfur pattern
	410.1026	-0.15	6.5	$C_9H_{16}O_{10}N_9$			15<Nc<18	sulfur pattern
	410.1026	-0.16	1	$C_{10}H_{22}O_{15}N_2$	Nitrogen rule	H/C	15<Nc<18	sulfur pattern
	410.1026	-0.20	4.5	$C_{12}H_{25}ON_7PS_3$			15<Nc<18	sulfur pattern
	410.1024	0.33	14.5	$C_{19}H_{17}O_4N_5P$			15<Nc<18	sulfur pattern
	410.1027	-0.43	10	$C_{15}H_{20}O_4N_6P_2$	Nitrogen rule			sulfur pattern
	410.1023	0.56	9	$C_{16}H_{22}ON_6S_3$	Nitrogen rule			sulfur pattern
	410.1027	-0.58	15	$C_{16}H_{14}O_2N_{10}S$	Nitrogen rule			
	410.1027	-0.59	9.5	$C_{17}H_{20}O_7N_3S$				
	410.1022	0.72	4	$C_{15}H_{28}O_3N_2P_2S_2$	Nitrogen rule			sulfur pattern
	410.1029	-0.87	13	$C_{23}H_{24}OP_2S$	Nitrogen rule		15<Nc<18	
	410.1029	-0.91	2	$C_5H_{19}O_{10}N_{10}P$	Nitrogen rule	H/C	15<Nc<18	sulfur pattern
	410.1029	-0.95	0	$C_8H_{28}ON_8P_2S_3$	Nitrogen rule	H/C	15<Nc<18	sulfur pattern
	410.1021	1.00	0.5	$C_9H_{24}O_9N_5S_2$		H/C	15<Nc<18	sulfur pattern
	410.1021	1.09	19	$C_{23}H_{14}O_4N_4$	Nitrogen rule		15<Nc<18	sulfur pattern
	410.1030	-1.12	-0.5	$C_{10}H_{28}O_4N_5S_4$		H/C	15<Nc<18	sulfur pattern
	410.1020	1.16	1	$C_7H_{24}O_6N_8P_2S$	Nitrogen rule	H/C	15<Nc<18	
	410.1030	-1.18	5.5	$C_{11}H_{23}O_4N_7P_3$			15<Nc<18	sulfur pattern
	410.1030	-1.20	0	$C_{12}H_{29}O_9P_3$	Nitrogen rule	H/C	15<Nc<18	sulfur pattern
	410.1031	-1.35	5	$C_{13}H_{23}O_7N_4PS$	Nitrogen rule		15<Nc<18	
	410.1019	1.48	8.5	$C_{19}H_{25}O_3NPS_2$			15<Nc<18	sulfur pattern
	410.1032	-1.63	8.5	$C_{19}H_{27}ONP_3S$			15<Nc<18	
	410.1018	1.65	9	$C_{17}H_{25}N_4P_3S$	Nitrogen rule			
	410.1017	1.91	0	$C_{12}H_{27}O_{11}PS$	Nitrogen rule	H/C	15<Nc<18	
	410.1017	1.92	5.5	$C_{11}H_{21}O_6N_7PS$			15<Nc<18	
	410.1017	2.08	0.5	$C_{10}H_{27}O_8N_3P_3$		H/C	15<Nc<18	sulfur pattern
	410.1016	2.09	6	$C_9H_{21}O_3N_{10}P_3$	Nitrogen rule	H/C	15<Nc<18	sulfur pattern
	410.1034	-2.11	0.5	$C_9H_{26}O_7N_5P_2S$		H/C	15<Nc<18	
	410.1016	2.15	0	$C_8H_{26}O_3N_8S_4$	Nitrogen rule	H/C	15<Nc<18	sulfur pattern
	410.1034	-2.17	24	$C_{24}H_{10}N_8$	Nitrogen rule		15<Nc<18	sulfur pattern
	410.1034	-2.18	18.5	$C_{25}H_{16}O_5N$			15<Nc<18	sulfur pattern
	410.1016	2.23	13	$C_{23}H_{22}O_3S_2$	Nitrogen rule		15<Nc<18	sulfur pattern
	410.1034	-2.27	5.5	$C_{10}H_{20}O_5N_9S_2$			15<Nc<18	sulfur pattern
	410.1034	-2.28	0	$C_{11}H_{26}O_{10}N_2S_2$	Nitrogen rule	H/C	15<Nc<18	sulfur pattern
	410.1015	2.40	13.5	$C_{21}H_{22}N_3P_2S$			15<Nc<18	
	410.1014	2.68	10	$C_{15}H_{18}O_6N_6S$	Nitrogen rule			
	410.1036	-2.71	8.5	$C_{18}H_{24}O_2N_3S_3$				sulfur pattern
	410.1013	2.84	5	$C_{14}H_{24}O_8N_2P_2$	Nitrogen rule		15<Nc<18	sulfur pattern
	410.1013	2.85	10.5	$C_{13}H_{18}O_3N_9P_2$			15<Nc<18	sulfur pattern
	410.1037	-2.93	19.5	$C_{20}H_{13}N_9P$			15<Nc<18	sulfur pattern
	410.1037	-2.94	14	$C_{21}H_{19}O_5N_2P$	Nitrogen rule		15<Nc<18	sulfur pattern

Fig. 2. The determination of relevant elemental compositions. (A) Full-scan ESI spectrum at the retention time of the MS signal to be identified. Acquisition was performed on a LC/LTQ-Orbitrap® system scanning within the range 75–1000 Th at a mass resolving power of 60,000 ($m/\Delta m$, FWMH) in the negative-ion mode. The natural isotopic abundance distribution is given in an inset (B) and indicates an approximate number of carbon atoms equal to sixteen as well as the presence of a sulfur atom. (C) Table of the molecular formulas returned by the post-processing software based on the accurate mass measurement of m/z 410 (assumed mass accuracy better than 3 ppm), and summary of their reasons for rejection.

first one consists in sorting molecular formulas into homologous series (*e.g.*, molecules differing by CH_2 units) and (ii) the second one deals with the exploitation of mass spectral information such as isotope ratios and predicted carbon numbers for intense peaks. A similar approach based on compound classification into homologous series followed by formula extrapolation from lower to higher masses has also been successfully applied to crude petroleum oil by using Kendrick plots [74]. This will be addressed further in this article.

Finally, Kind and Fiehn described an algorithm for filtering molecular formulas which is derived from seven heuristic rules: (i) restriction on the number of elements, (ii) Lewis and Senior rules, which filter elemental compositions on the basis of valence considerations, (iii) isotopic pattern, (iv) H/C ratios, which are comprised between 0.125 and 2 in most cases, (v) heteroatom ratios, *i.e.* (NOPS)/C ratios, (vii) element probabilities and (vii) the presence of trimethylsilylated compounds (for GC/MS applications) [66]. All the generated formulas, which comply with the rules, are also checked for presence in a customizable list of chemical databases. These rules have been implemented in an automated Excel® script [66]. The adduct ion removal must be done manually before entering any values, but an adduct ion mass calculator is provided.

2.3. The different kinds of MS/MS experiments

LC/MS based on various ionization and scanning techniques can be considered as the first step of the identification process since it provides information about the metabolite molecular mass and elemental composition, when high-resolution analyzers are used. Although direct introduction of samples into the mass spectrometer is possible, it may suffer from lack of sensitivity in complex biofluids such as plasma or urine due to matrix effects, and it does not address the issue of isomeric compounds. However, LC/MS experiments only provide information restricted to retention times and molecular weights and therefore must be complemented by further LC coupled to tandem mass spectrometry experiments to access structural information. The following section deals with the different types of ion activation that are relevant in the field of metabolomics, *i.e.*, collision induced dissociations and infra-red multiphoton dissociation (IRMPD) [75], and also with several LC coupled to tandem mass spectrometry approaches that have been reported in the literature for metabolomics.

2.3.1. Fragmentation spectra: various ion activation modes

Ion dissociation of metabolites is related to the internal energy as well as the electron state (*i.e.*, open- or closed-shell species, which correspond to odd- and even-electron ions, respectively) of singly or multiply charged positive or negative ions. Odd-electron ions that are, for example, produced in electron ionization present high internal energy. Their gas phase stability is also lower than that of even-electron species, when they are formed under high-vacuum conditions in the ion source. This explains why many product ions are observed in EI mass spectra. Conversely, and in many cases (*i.e.*, ESI, APPI or APCI), the API process generates stable molecular species yielding limited "in-source fragmentations". Furthermore, these generated ions can be charged at various functionalized positions according to the experimental source conditions (solvent included), which results in an enhanced competitive dissociation of precursors. This yields a broad variety of fragmentations useful for ion structure elucidation.

There are two different modes of ion activation: (i) ergodic activation modes that deal with collisional to internal energy transfers, such as collision-induced dissociation (CID) and infra-red multiphoton dissociation (IRMPD) [76,77], and (ii) chemical activation modes, such as electron capture dissociation [78], which involve electron transfers.

2.3.2. Collision-induced dissociation

CID remains the most common ion activation technique used in current instruments. Collisions between a fast precursor ion and a neutral target gas lead to an increase in internal energy of the ion through the conversion of part of its translational energy into internal energy [79]. If the internal energy imparted is beyond the threshold for dissociation, precursor ion fragmentation may occur. Depending on instruments, collision energies can be in the keV range (high-energy collisions) or in the 1–200 eV range (low-energy collisions).

2.3.2.1. High-energy collisions. High energy collision conditions allow access to selected precursor ion excitation through electronic levels and indirectly through the vibrational levels of its electronic ground state [80,81]. The reproducibility of fragmentation patterns under variable collision conditions such as variations in high kinetic energy, target pressure, or initial internal energy is noteworthy and very useful for the building of product ion libraries from MS/MS experiments.

Such an excitation mode is mainly possible with TOF/TOF tandem instruments [82,83], although it has been used with multisector instruments for a long time [84]. However, TOF–TOF instruments [85] are only combined with MALDI and other laser desorption sources which are not suitable for LC coupling. Recently, a shotgun metabolomic approach has been developed using MALDI-TOF/TOF MS from mouse heart tissues [86]. From 285 peaks corresponding to negatively charged ions, 90 metabolites were identified and discussed. Further developments, enabling TOF–TOF operation with spray ionization sources could be very promising.

Among the most specific processes, charge remote processes involve homolytic pathways where the charge is not involved in the fragmentation [87–90]. Such dissociation takes place from deprotonated fatty acids (unsaturated, branched and oxidized side chain) or their cationized forms in positive ion mode. It occurs statistically along the hydrocarbon chain. Furthermore, using such an approach allows the localization of alkyl, double bond, hydroxyl, or ketone groups. It was applied to various compounds, including phospho- and glycolipids, acylglycerols and steroids which are relevant compounds for metabolomic applications [91].

Fragmentations promoted by charge occur also under high-energy collision conditions. However, distribution of ion species and their respective abundances are very different from those observed with charge remote fragmentation processes, as shown by a recent study of March et al. [92].

2.3.2.2. Low-energy collision. Low collision energy experiments (*i.e.*, E_{lab} lower than 200 eV, E_{lab} representing the ion kinetic energy in the laboratory reference frame) are the most popular activation methods. In low collision energy experiments, the selected precursor ions are exclusively vibrationally excited. In high energy CID, often using helium as target, less than 10–30% of precursor ions undergo dissociation, which results in a low product ion recovery efficiency (ion beam sputtering), whereas low energy CID can yield a total disappearance of precursor ions that is accompanied by a better product ion recovery. However, while the CID spectra recorded at several keV are reproducible, independently of the instrument, this is not so for low-energy CID spectra, which strongly depend upon (i) the instrument characteristics (*e.g.*, quadrupole versus ion trap or TOF mass spectrometers, for example), (ii) lens potentials, (iii) the collision energy, *i.e.*, E_{lab} values and (iv) the target gas pressure, a parameter that is difficult to control. A solution to this problem is to calibrate the collision energy, *i.e.*, to work in

constant experimental conditions in which the relative abundances of both the precursor and product ions are maintained constant. This can be achieved by using reference compounds for tuning the system in order to achieve a given ratio, and then by plotting the abundance of precursor and product ions as a function of E_{lab} (Energy Resolved Mass Spectrometry, ERMS). Only under these conditions, is it possible to build MS/MS libraries without large variations in relative ion abundances.

However, two essential differences characterize collisional activation occurring from ion beams in the Rf-only quadrupole collision cell and from the stored ions in 3D (and 2D) ion trap cells: (i) in the quadrupole collision cell, collisional activation takes place as a *fast heating* [93] from ion beam collision experiments in contrast to storage experiments in which *slow heating* [93] occurs due to the large number of low-energy collisions with collision gas, and (ii) a selective excitation of precursor ions occurs in ion traps, whereas both precursor and product ions can be activated in triple quadrupole instruments. These differences suggest that the number of consecutive dissociations must be reduced in the ion trap compared with that occurring in the quadrupole collision cell [44]. Furthermore, the detection of product ions depends upon the dissociation rate constant of precursor ions which must be consistent with the time-window for ion observation (*i.e.*, 10^{-5} s for triple quadrupoles and 10^{-2} s for IT/MS).

All these findings show why the building of database for CID in MS/MS experiments is difficult without many drastic precautions when CID spectra are recorded with various tandem instruments (*i.e.*, TOF, quadrupole, 3D–2D ion traps or FTICR analyzers). Indeed, all the previous experimental parameters must be controlled to be able to compare CID spectra.

2.3.3. Other activation modes

The IRMPD mode was first introduced by McLafferty and coworkers [77] and Hunt et al. [94] in FTICR devices. It is now conventionally available on FTICR [95] and as prototype on 3D- [96,97] and 2D-ion [47] traps. Following precursor ion fragmentation, product ions are trapped together with the selected precursor ions and can be activated during the next IRMPD process. The stored product ions being submitted to IR radiation can dissociate in turn, yielding consecutive cleavages not necessarily observed by CID in Rf-only multipole collision cells or during CID-sustained off-resonance irradiation (CID-SORI) processes in ICR cells. IRMPD was used by Hakansson and coworkers [75] for phosphate-containing metabolite investigation and was compared with CID processes. CID was shown to be more efficient for low-mass compounds (<600 u), but IRMPD appeared to provide complementary information since it was preferred for higher mass compounds (>600 u).

Recently, following the work of Freiser and coworker [98], McLafferty and coworkers [77], and Beauchamp and coworkers [76], Zubarev and coworkers [99] and Hakansson's group [75] used EID (electron-induced dissociation) for metabolite analysis with an FTICR instrument and demonstrated that both CID and EID processes give complementary fragmentation patterns.

2.3.4. Mass analyzers and MS/MS experiments

Triple quadrupole (TQ) instruments are very popular in the field of metabolomics [100,101]. They are dedicated to low-energy dissociation processes. The collision cell can be a Rf-only quadrupole or multipole (hexapolar or octapolar fields), both of them being suitable for ion transmission. Compared with the high keV ion beam tandem (TOF–TOF) mass spectrometers, TQ-instruments have a lower scanning speed and a limited resolution, which does not allow accurate mass measurements.

The advantage of the triple stage instruments is the possibility of simple screening by using different scanning MS/MS modes: (i) detection of the precursor ion of a selected product ion by maintaining the second mass filter on the m/z value of product species and by scanning the first mass filter which sequentially transmits all ions from the source to the collision cell (*i.e.*, precursor ion scans) and (ii) detection of the precursor ions able to release a common neutral after collision and excitation in the activation cell (*i.e.*, neutral loss scans). To this end, both the mass filters scan together with a mass shift corresponding to the released neutral. These different MS^2 scanning modes available with triple quadrupole instruments are relevant for targeted approaches by allowing the selective detection of the components of the same chemical class [102,103]. The neutral loss scanning combined with natural and stable-isotope labeled compounds allows high-throughput screening of metabolites based upon the MS signatures [104]. On the other hand, the constant precursor, product ion and neutral loss scanning modes were employed advantageously to identify metabolites in urine samples [105].

"In-time" analyzers, such as ion traps, can be used to perform low-mass resolution sequential MS^n experiments, which are of interest for structural elucidation [26,106]. The relevance of using LC/ITMS for complex product authentication and identification was shown by Kite et al. for the metabolomic analysis of a hundred saponins in crude plant extracts [107]. However, these experiments are limited by the intrinsic low mass cut-off (*i.e.*, the m/z ratios of product ions must be higher than 1/3 to 1/5 of that of the precursor ions). Linear ion traps are the last generation of ion trap mass devices and have several advantages over 3D-ion traps, such as a larger ion storage capacity, a higher trapping efficiency, a reduced low mass cut-off thanks to *pulsed-Q-dissociation* (PQD), which is a novel fragmentation mechanism developed for the LTQ [108,109], and the potential to use resonant radial ion activation with complementary advantages for product ion scanning [110]. Of note, linear ion traps have already been used for metabolomic purposes [111].

Triple quadrupole and ion trap analyzers provide complementary information. For example, a triple quadrupole operated in the neutral loss and precursor ion scanning modes was used to characterize sulfoconjugates in rat urine. Further MS^n experiments performed in an ion trap mass spectrometer, identified 4-ethylphenol sulfate and discriminated it from its two isomers (*i.e.*, 2- and 3-ethylphenol sulfates) by matching of chromatographic retention times, MS^2 and MS^3 spectra of the unknown to the synthesized compounds [102].

Such an approach is much more easily performed by using a hybrid Q TRAP® instrument [47]. Combining a triple quadrupole analyzer with linear ion trap technology in a single instrument retains the conventional triple quadrupole scan functions such as neutral loss, precursor and product ion scanning modes, and the selected product ion and multiple reaction monitoring modes, and provides access to sensitive MS scans over a large m/z ratio range as well as to MS^3 experiments [112]. Furthermore, "triple quadrupole-like" fragmentations without any low mass cut-off can be obtained because the precursor ion selection in the first quadrupole and the fragmentation in Rf-only cell are decoupled. Some applications have been published in the field of small molecule profiling [113], metabolomics [103] and metabolism studies [114,115].

A limitation of the previous instruments for identification purposes is their inability to provide mass measurements accurate enough to achieve elemental composition determination. Combining different analyzers such as quadrupole and TOF in QqTOF devices overcomes this limitation while retaining fragmentation capability. However, to maintain TOF properties, the quadrupole filter and the multipole collision cell are almost orthogonal to the TOF axis in order to limit the error due to the spatial ion beam distribution on the resolution of the TOF analyzer. The

use of a push–pull system transforms the continuous low kinetic ion beam coming from the ESI/quadrupole filter into a higher kinetic pulsed ion beam [116]. Such a device achieves high-resolution CID spectra of relevance for metabolite identification purposes [117].

QqTOF devices can be used for MS^2 experiments in the product ion mode and can be coupled to UPLC, as reported in a study of metabolomic maps for understanding the cell response to ionizing radiations [117]. A new approach called UPLC/MS^E has been developed to improve the collection of information from MS/MS experiments. Acquisitions are performed at low and high collision energies without any precursor ion selection, thus requiring additional software to generate information on both precursor and product ions [118]. Otherwise, 3D ion traps combined with TOF [119] and even better 2D ion traps/TOF [120,121] provide some interesting characteristics, which include MS^n scanning capability together with accurate mass measurement of product ions generated in the ion trap.

The use of TOF analyzer is also focused on accurate mass measurements for elemental composition determinations. However, greater accuracy is achieved using FT/MS instruments such as the LTQ-Orbitrap® which provides accurate mass measurements in multistage MS/MS experiments for metabolite profiling [122] and metabolism studies [59,60,123]. Interestingly, the C-trap, initially designed to store ions before their orthogonal ejection into the Orbitrap cell for high resolution analysis, can now be used as a collision chamber to enable triple quadrupole-like fragmentations, as recently shown for peptide sequencing [124]. This new CID approach is promising for metabolomic applications.

Finally, the best accuracy and mass resolving power are achieved with tandem QhFTICR instruments (h for hexapole) [125]. Furthermore, FTICR devices select precursor ions with high resolution and perform *in situ* (*i.e.*, in the ICR cell) ion activation such as IRMPD. This cannot be achieved with the Orbitrap® since until now isolation and fragmentation of precursor ions occur in the ion trap (maximum resolution for precursor ion selection: 0.3 Th). The selection of precursor ions at high resolution can be achieved in the ICR cell by injecting some pulsed collision gas. This decreases the resolution power of the device, with typical values ranging from 10,000 to 20,000 at FWHM (for an ion at m/z 600, JC Tabet Personal Communication). The performances of FTICR for proteomic and metabolomic applications have been reported in a study coupling 20 kpsi reverse-phase liquid chromatography to a 11.4 T-FTICR mass spectrometer: more than 5000 metabolites were detected in a bacterial cell extracts. However, complementary LC/MS/MS experiments were performed in the ion trap that is in front of the ICR cell, suggesting that such experiments (MS/MS) remain challenging in the FTICR cell [126].

2.4. Complementary approaches

2.4.1. H/D exchange

Hydrogen/deuterium exchange in solution is a widely used strategy for elucidation of mass fragmentation mechanisms [127–130], especially in the field of metabolism studies. Determination of the number of exchangeable hydrogen atoms (*i.e.*, attached to O, N, and S atoms) facilitates structural elucidation of metabolites. For example, from the formula $C_{10}H_{18}O_4$, several compounds are possible: a dicarboxylic acid, a monocarboxylic acid with two OH groups and an unsaturation (a ring or a double-bond), a monocarboxylic acid with a keto function and a hydroxyl group, and a compound with four OH groups and two unsaturations. These compounds will be partially discriminated by H/D exchange experiments since they bear 2, 3, 2 and 4 exchangeable hydrogen atoms, respectively. Furthermore, an increase in m/z ratio for a product ion during H/D exchange experiments will supply information on the position of a particular chemical function containing mobile protons in the charged molecule. Karlsson [131], one of the pioneers of H/D exchange application, introduced the use of deuterium oxide as a mobile phase for microcolumn liquid chromatography coupled to ESI/MS, which is now used extensively for structural elucidation of small molecules [132–135]. The application of deuterium oxide as the sheath liquid in CE/MS for structural elucidation purpose has also been reported [136]. In addition, gas phase H/D exchange experiments can be performed by introducing various deuterated agents (*e.g.*, CH_3OD, D_2O, ND_3) at the skimmer and in the collision cell. The extent and rate of H/D exchanges increase with the gas-phase basicity of the reagent, so it is possible to selectively label functional groups of the molecule. In addition, this sometimes permits the distinction of isomers [137]. However, the yield of deuteriation may not be optimal as compared with other classic H/D experiments. A relatively poor yield may lead to erroneous conclusions.

2.4.2. Derivatization strategies

Derivatization can be used to achieve numerous and different objectives, such as to improve the ionization efficiency of a family of compounds, highlight and localize a chemical function (*e.g.*, hydroxyl group, glucuronide), trap reactive metabolites, assist or induce ion dissociation, and modify polar compound retention. These latter points have been extensively reviewed [138,139] and although pharmaceutical and drug discovery issues were addressed, it is easily transposable to metabolomics.

A chemoselective tagging strategy termed "metabolite enrichment by tagging and proteolytic release" (METPR) has been introduced for enrichment and profiling of small molecules [140]. Briefly, metabolites are captured on a solid support by conjugation to resin-bound reactive groups that target different classes of metabolites. The trapped molecules are then released by a cleavage step promoted by a protease. This approach was used to profile highly polar compounds by derivatizing the metabolites with a hydrophobic *p*-Cl-phenylalanine residue, which improved their retention on reverse-phase columns. The chlorine of the tag also allowed discrimination of the tagged metabolites from background peaks [141].

3. Databases

A first objective of database inquiry is the annotation of MS signals. This requires queries by molecular mass or by elemental composition, depending on the instrument. Databases can as well be used for biological purposes. Indeed, it is essential to give biological sense to the acquired data, for example, to determine the biological functions of metabolites.

For these purposes, four kinds of tools can be distinguished: (i) general chemical databases (*e.g.*, PubChem) that encompass synthetic and/or natural compounds, (ii) metabolic databases that deal with annotated metabolic pathways, (iii) metabolomics databases that originate from metabolomic research project in a specific field (*e.g.*, LipidMaps in lipidomics) and (iv) mass spectral databases (*e.g.*, NIST).

Actually, some databases combine several of these functions. It is important to keep in mind that all databases are works in progress, with some going through a rapid growth phase with constant additions and corrections being made. Therefore, additional functions could be found in the near future. Table 2, which is certainly not exhaustive, summarizes some of the most widely used databases that are addressed in the following sections.

Table 2
Databases for the identification of metabolomic signals by mass spectrometry

Database			Thematic	Conception/URL	Reference
BiGG	●		Human	University of California (USA) (www.bigg.ucsd.edu/)	[197]
BioCyc (HumanCyc, MetaCyc)	●		Biochemical pathways	SRI International (USA) (www.biocyc.org/)	[198,148]
ChEBI	●		General	European Bioinformatics Inst. (UK)/European Molecular Biology Lab (www.ebi.ac.uk/chebi/)	[144]
ChemFinder	●		General	Cambridge Soft (USA) (www.chemfinder.CambridgeSoft.com)	[143]
CHEMnetBASE (Dict. Nat. Prod.)	○		General	Chapman & Hall/CRC (www.chemnetbase.com/)	
CSLS	●		General	CADD Lab. Med. Chem NCI, NIH (USA) (http://129.43.27.140/cgi-bin/lookup/search)	
Enhanced NCI Database Browser	●		General	CADD Lab. Med. Chem NCI, NIH (USA), U of Erlangen-Nuremberg (Germ.) (http://129.43.27.140/ncidb2/-)	[142]
Fiehn library	●		General	Fiehn Laboratory Univ California Davis: Genome center (http://fiehnlab.ucdavis.edu/Metabolite-Library-2007)	
Golm	●	*	Plant	Max Planck Institute for Molecular Plant Physiology (Germany) (www.csbdb.mpimp-golm.mpg.de)	[153]
HMDB	●	*	Human metabolites	Department of Computing Science, University of Alberta (Canada) (www.hmdb.ca/extrIndex.htm)	[149]
KEGG ligand database	●		General	Kyoto University Bioinformatics center (Japan) (www.genome.jp/kegg/ligand.html)	[199]
KNApSAcK	●		Natural products	RIKEN Plant Science Center (Japan) (http://kanaya.naist.jp/KNApSAcK/KNApSAcK.php)	[151]
LipidMaps	●		Lipidomics	LIPID MAPS Bioinformatics Core (USA) (www.lipidmaps.org/data/index.html)	[200]
LipidBank	●		Lipidomics	Japanese Conference on the Biochemistry of Lipids (Japan) (www.lipidbank.jp/)	[201]
Madison Metabolomics Consortium Database	●		General	National Magnetic Resonance Facility, University of Wisconsin-Madison (http://mmcd.nmrfam.wisc.edu/)	[28]
MassBank	●	*	General	Keio university, university of Tokyo, Kyoto university, RIKEN plant Science center (Japan) and others (www.massbank.jp)	[202]
Merck Index	○		General	Merck publishing	
Metlin	●	*	Human metabolites	Scripps Center for Mass Spectrometry (www.metlin.scripps.edu)	[150]
MoTo	●		Metabolome database for tomato	Wageningen University (www.appliedbioinformatics.wur.nl)	[152]
MSlib	●	*	Drugs, metabolites	University of Alberta (http://www.ualberta.ca/~gjones/mslib.htm)	
NIST	◉	*	General	National institute for standard and technology (USA) (www.nist.gov/srd/nist1a.htm)	
PubChem	●		General	National Center for Biotechnology Information (USA) (www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pccompound)	
SDBS	●	*	General	National Institute of Advanced Industrial Science and Technology (www.riodb01.ibase.aist.go.jp/sdbs/)	[203]
SciFinder	○		General	American Chemical Society (USA)	

(●) Free access, (◉) partially free access, (○) licenced and (*) spectral database.

3.1. General chemical databases

The American Chemical Society (ACS) provides several databases, including Scifinder (CAS: Chemical Abstract Service), that contain general information about millions of compounds. A subscription is required to access these databases. Searches can be performed by entering elemental compositions and substructures, which can also be used to refine queries.

Cactus (CADD group's Chemo informatics Tools and User Service) provides access to the Chemical Structure Lookup Service (CSLS) and Enhanced NCI Database Browser. CSLS allows searches of millions of indexed structures from 80 free and commercial chemical databases [142].

Chemfinder is a free chemistry database (CambridgeSoft Corporation, Cambridge, MA), but a subscription is required for its professional version ChemIndex [143]. The latter contains data on over 75,000 compounds including structures, names and synonyms.

PubChem is a database of chemical molecules managed by the National Center for Biotechnology Information (NCBI) which is part of the United States National Library of Medicine (NHL). The PubChem substance, the largest PubChem database, contains more than 38 millions of records describing chemical structures, molecular weight and formula. This database includes structures from chemical manufacturers (*e.g.*, Sigma–Aldrich) and from licensed databases (DiscoveryGate Elsevier, Thomson Pharma). The queries can be formulated as molecular weight (and MW range), molecular formula or using a chemical structure that may be sketched using the PubChem sketcher.

Chemical Entities of Biological Interest (ChEBI) is a freely available database of small chemical compounds. ChEBI combines information (on both synthetic and natural products) from the KEGG and IntEnz (the Integrated relational Enzyme database) databases. All entries have been reviewed in order to standardize biochemical terminology. In addition, ChEBI offers an ontological classification that gives comprehensive relationships between molecular entities [144].

The Merck Index is an encyclopedia of chemicals, drugs and biological compounds with over 10,000 monographs containing information such as CAS registry number, chemical formula, molecular weigh on single substances. It is available by subscription.

CHEMnetBASE is an online and CD-ROM chemical database from Chapman & Hall/CRC that includes "*The Handbook of Chemistry & Physics*" and some other combined chemical dictionaries as "*The dictionary of natural products*" for example. The latter provides a comprehensive source of chemical data on natural products.

3.2. Biochemical and metabolic databases

These databases are mostly used to give some biological sense to metabolomic data. They are also a good starting point for the iden-

tification of signals from cell extracts, plant biology or microbiology experiments.

The Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database of biological systems consisting of genetic building blocks of genes and proteins (KEGG GENES), chemical building blocks of both endogenous and exogenous metabolites (KEGG LIGAND, [145], molecular wiring diagrams of interaction and reaction networks (KEGG PATHWAY), and hierarchies and relationships of various biological objects (KEGG BRITE). It provides a reference knowledge base for linking genomes to biological systems [146].

Biocyc [147] is a collection of 371 pathways corresponding to various genomes. It describes the genome and pathways of a single organism. Two types of materials can be downloaded: data files and an executable program that combines the pathway software with the Biocyc database [148].

The Arabidopsis information resource (TAIR) maintains a database of genetic and molecular biology data for the model higher plant *Arabidopsis thaliana*. Data available include the complete genome sequence, gene product information, metabolism, gene expression, DNA seed stocks, genome maps, genetic and physical markers, publications and information about the *Arabidopsis* research community. The downloadable program Aracyc 4.1 contains 238 pathways, 85% of which are experimentally confirmed. Only queries by name among the 1908 compounds are available.

3.3. Metabolomic databases

Those databases are the most recent ones and in the near future should be the most comprehensive ones. They are of value for the first steps of identification because they contain spectral data about biochemically relevant metabolites in specific biological contexts such as human biofluids or plant extracts, for example.

The Human Metabolome Database (HMDB) is a comprehensive, Web-accessible metabolomic database that brings together quantitative chemical, physical clinical and biological data about thousands of endogenous human metabolites [149]. It contains records for more than 2180 endogenous metabolites and provides literature derived data, mass spectra, NMR-spectra and validated metabolite concentrations. The HMDB offers as well a relational data extraction tool and both MS and NMR spectra searching tools.

Metlin is a metabolite database containing over 15,000 entries. It provides a data management system designed to help the researcher to identify metabolites by offering public access to its repository of current and comprehensive mass spectral metabolite data. Metabolite data provide a list of known metabolites, their mass, chemical formula and structure, each of them being linked to outside resources such as KEGG. Metlin also contains FT/MS data consisting in FT/MS spectra of chromatographically separated human serum fractions. It gives access to LCMS profiles from various tissues and biofluids, which are provided with the experimental conditions[150].

Atomic Reconstruction of Metabolism (ARM) is a project of the Arita, Nishioka and Kanaya groups and represents the metabolism of more than 2700 reactions in about 1700 *Escherichia coli* subclasses. The official website (http://www.metabolome.jp) provides software tools and databases such as MassBank, flavonoid viewer, LipidBank and KNApSAcK. MassBank will be detailed in a further section. LipidBank contains biological activities, physico-chemical properties, literature information and spectral data on 7009 lipids. It is an open publicly free database of natural lipids including fatty acids, glycolipids, sphingolipids, steroids and various vitamins. KNApSAcK is a comprehensive metabolites-species relationship database. It includes 20,000 metabolites and their relative species information. It can be searchable by name, molecular weight, molecular formula and mass spectra [151].

The metabolome database for tomato (MoTo DB) is an open access metabolite database for LC/MS dedicated to tomato fruit [152]. It is based on literature information combined with experimental data derived from LC/MS-based metabolomic experiments. The assignment of mass signals relies on the combination of accurate mass (Q-TOF), retention time, UV–vis information and MS/MS fragmentation data. The accurate mass, together with a mass accuracy setting, is the main search entry for the database. Mass accuracy can be set from 1 to 1000 ppm, thus enabling the matching of data from detectors generating mass with either low or high accuracy. Links with PubChem and MedLine databases are also available.

3.4. Spectral databases

Mass spectral databases were initially constituted with data from GC/MS experiments due to high reproducibility between instruments. This is for example the case for the Golm metabolome database [153]. The NIST05 database offers a fully evaluated collection of electron ionization mass spectra obtained at standardized electron energy which also contains a MS/MS library (5191 compounds) and retention index data (25728 compounds) (http://www.nist.gov). LC/MS/MS using the particle beam interface [154] was expected to give EI spectra and be useful for constructing libraries, but was not sufficiently sensitive [155].

Ionization and fragmentation in LC/MS instruments could not be standardized even with similar instruments [54,156]. This observation highlights the need for the constitution of spectral libraries taking into account the information provided by different kinds of instruments. Despite this limitation, databases containing API CID spectra such as HMDB, Metlin, MassBank from metabolome.jp (www.metabolome.jp/), and LipidMaps are starting to be released.

The MS search function of HMDB allows access to 1200 MS/MS spectra acquired with a triple quadrupole at 3 different collision energies (*i.e.*, low, medium and high) for around 400 compounds.

Metlin incorporates 282 MS/MS spectra acquired in ESI+ and 140 MS/MS spectra acquired in ESI−. By specifying a precursor mass range on the MS/MS search form, a research can easily reference the MS/MS profile of known metabolites in Metlin against the MS/MS profile of an unknown compound [150].

The MassBank of metabolome.jp provides high-resolution mass spectra of 1749 metabolites. Experimental conditions of separation are also described whenever a separation technique (LC, GC, CE) is used. LC/MS spectra are recorded on QqTOF, TQ and ion trap instruments. Ten thousand spectral data are searchable through a spectral browser and a peak search function [157,158].

The LipidMaps structure database (LMSD) comprises structures and annotations of biologically relevant lipids. Lipid standards include MS/MS values with links to fragmentation spectra, including structures of principal product ions, as well as links to commercial catalogs and literature references. This online tool predicts possible structures based on MS data for precursor ions, mass tolerance, ion mode and head groups. The "structure drawing" function is able to create a fatty acyl structure from information about carbon chain and functional groups.

4. Endogenous metabolite identification: some case studies

Some metabolite identifications reported in metabolomic studies are typically not novel as they consist in assigning a signal to a compound that has already been rigorously described in the literature. Thus, these metabolites are often identified based upon the retrieval of properties such as retention time, accurate mass, fragmentation pattern that are common to endogenous metabolites and authentic reference compounds (standards). In high-resolution

mass spectrometry-based techniques, the identification of a compound typically begins with a database query based either directly on its experimental accurate mass measurement or on its deduced elemental composition.

Database queries here return one or more hits. If fragmentation experiments are available, they can supply information about chemical groups that are present in the compound by highlighting neutral losses or product ions, which are characteristic of a functional group and can serve to discriminate between database hits. In addition, spectral information from the authentic compound may be accessible via the database or the literature and lead to a confident structure assignment. As no consensus on what constitutes valid metabolite identification has been reached, Sumner et al. have reported four different levels of identification according to the information provided for the compound characterization within the Metabolomics Standards Initiative [159]:

(i) *Identified compounds*: A minimum of two independent and orthogonal types of data relative to an authentic compound analyzed under identical experimental conditions. In MS-based techniques this could include: retention time/index and mass spectrum, or accurate mass and tandem MS.
(ii) *Putatively annotated compounds*: Without chemical reference standards, based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries.
(iii) *Putatively characterized compound classes*: Based upon characteristic physicochemical properties of a chemical class of compounds, or by spectral similarity to known compounds of a chemical class.
(iv) *Unknown compounds*: Although unidentified or unclassified these metabolites can still be differentiated based upon spectral data, thus enabling relative quantification.

Two concrete cases are hereafter presented to illustrate the varying complexity of compound identification. These two examples are extracted from a metabolomic study and concern the identification of metabolites detected in rat urine by using an LC-LTQ-Orbitrap® system.

4.1. First situation: the hypothetic metabolite is described in biochemical or metabolomic databases

This first example deals with the identification of a discriminating signal at m/z 220.1180, recorded in positive electrospray ionization (Fig. 3A). Elemental composition determination led to the formula $C_9H_{17}NO_5$ (−1.1 ppm), which was used to query the human metabolome database. HMDB returned pantothenic acid as single hit.

From the CID spectrum, presented in Fig. 3A', it was possible to observe 2 losses of 18 (corresponding to water losses) and a loss of 28 (CO and not C_2H_4, as shown by accurate mass measurement), which indicated the probable presence of a hydroxyl and a carboxyl group. HMDB provided a CID spectrum of the authentic compound (pantothenic acid), acquired in positive ESI mode with a

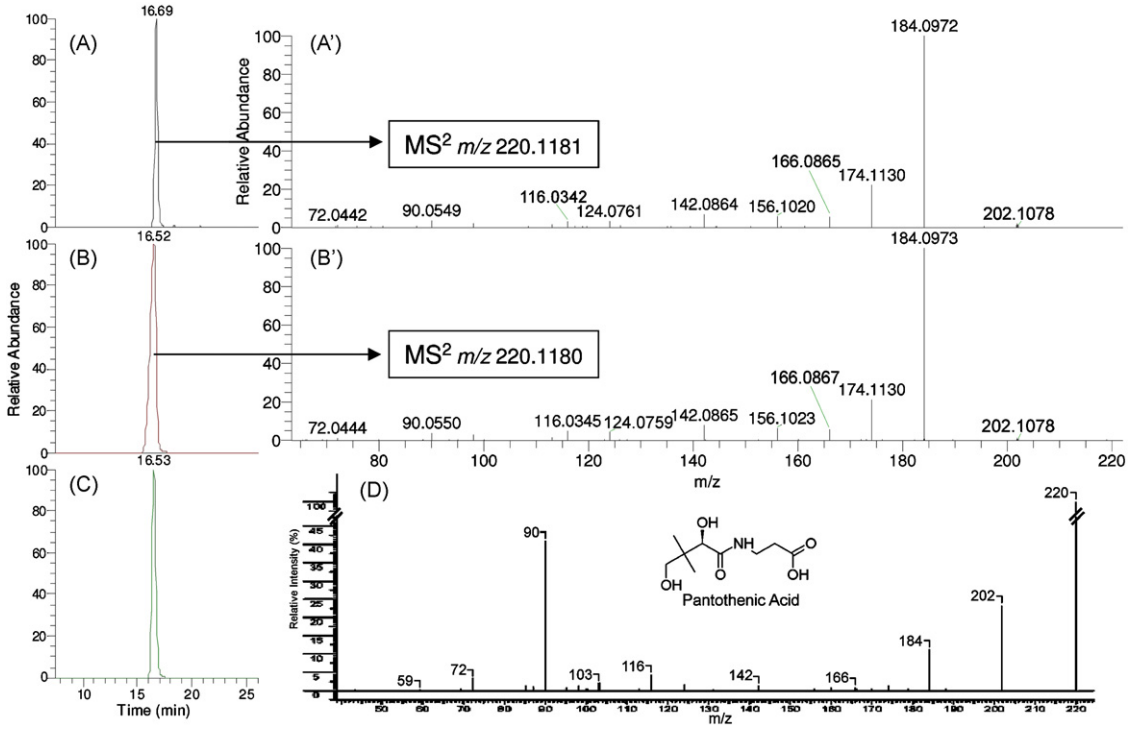


Fig. 3. Identification of pantothenic acid in rat urine. (A) Extracted ion chromatogram of the ion to be identified at m/z 220.1180 (±2 ppm). (A′) CID spectrum of the ion of interest at m/z 220.1180. (B) Extracted ion chromatogram from the authentic pantothenic acid (theoretical $[M+H]^+$ at m/z 220.11795 (±2 ppm)) and its CID mass spectrum (B′). (C) Extracted ion chromatogram of m/z 220.1180 from rat urine spiked with pantothenic acid. (D) CID mass spectrum of the authentic pantothenic acid provided by HMDB (acquired on a TQ instrument (Quatro Waters®)). *Conditions*: The chromatographic separation was performed on an XTerra MS C_{18} 5-μm, 2.1 mm × 150 mm column (Waters, Saint Quentin en Yvelines, France). The mobile phases were A: 100% water and B: 100% acetonitrile, both containing 0.1% formic acid. After an isocratic step of 5 min at 100% phase A, a linear gradient from 0 to 100% of B was run over the next 40 min with a mobile phase flow of 0.2 ml/min. Returning to 100% A over 1 min, the column was then allowed to equilibrate for 10 min leading to a total run time of 60 min. ESI-MS metabolic profiles were acquired using a LC/LTQ-Orbitrap® system within the range 75–1000 Th, successively in both positive and negative ion modes with a resolution set to 60,000 ($m/\Delta m$ at FWHM) in centroid mode. CID spectra were acquired using data-dependent scanning function. The scan event cycle comprised two data-dependent (MS^2 and MS^3) events acquired with a resolution set to 7500. Microscan count was set to unity and a repeat count for dynamic exclusion was set to 3. MS^2 acquisition parameters were an isolation width of 1 Th, normalized collision energy of 35%, and an activation time of 30 ms.

triple quadrupole instrument (Fig. 3D). Following the observation of strong similarities between the CID spectra of our compound of interest and those provided by HMDB, we injected the authentic pantothenic acid into our system for formal identification by comparing retention times (Fig. 3A–C, corresponding to unknown compound in rat urine, reference compound in buffer, and reference compound in spiked urine, respectively), accurate mass measurements, isotopic abundance and fragmentation patterns (Fig. 3A′ and B′).

4.2. Second situation: the hypothetical structure of the metabolite cannot be obviously deduced from databases

In this case, CID experiments may allow putative annotation or characterization of the metabolites. Careful interpretation of such CID spectra is required to generate structural hypotheses that need to be confirmed by further chemical syntheses.

The second example deals with an ion at *m*/*z* 201.1133 that was observed in LC/MS metabolic fingerprints from rat urine in negative electrospray ionization. This case is more complex because the extracted ion chromatogram of the ion of interest displayed three main peaks, even with a mass tolerance of 2 ppm (Fig. 4A). However, only the first eluted peak at 22 min was highlighted by MVA and therefore had to be identified. Isotope pattern analysis and elemental composition determination led to the same formula for the three peaks that is $C_{10}H_{18}O_4$ (RDBE = 2).

A query based on the elemental composition in the human metabolome database returned sebacic acid as a unique hit. The low-resolution CID spectrum of sebacic acid provided in the database (Fig. 4C) presented similarities (principal product ion at *m*/*z* 139, and several common product ions at *m*/*z* 183, 157, and 57) with the one obtained for the compound of interest (Fig. 4B, arrows mark the peaks of the compound of interest similar to those of the reference molecule). The injection of the authentic compound from Sigma resulted in the chromatogram and CID spectrum presented in Fig. 5B. From Fig. 5 and contrary to what was expected, it is clear that the endogenous metabolite of interest is not sebacic acid since it exhibits a different retention time.

CID and spiking experiments showed that sebacic acid corresponds to the third peak at 27 min. The three isomers presented in Fig. 4A can only be discriminated from their MS^3 fragmentation patterns, revealing the limitation of the sole use of high resolution for complex biological sample analysis (data not shown). Furthermore, although the product ion at *m*/*z* 183 had a high relative abundance in the CID spectrum of deprotonated sebacic acid from HMDB (recorded by using a triple quadrupole instrument) and in that of the unknown compound eluting at 22 min, it was produced neither for the authentic compound nor for the endogenous sebacic acid in our system (Fig. 5A and B). This highlights the difficulty of building multi-instrument databases with API-MS techniques.

As the sole hits from biologically oriented databases (HMDB, Metlin, KEGG) were sebacic acid and diethyl adipate, which were inconsistent with CID spectra, we had to move on to wider databases. However, querying CAS with $C_{10}H_{18}O_4$ returned 2334 hits. It was therefore necessary to conduct complementary experiments to restrict the search by using imposed chemical substructures. CID experiments (Fig. 4B) showed a neutral loss of 44 u (CO_2), yielding a peak at *m*/*z* 157 and so highlighted the presence of a carboxylic acid function that should be hydroxylated in the α position to explain a competitive loss of 46 u (H_2CO_2) in the negative ion mode [160]. Moreover H/D exchange experiments supported the previous result by revealing the presence of three exchangeable protons in the compound. In this way, it was possible to assume a dihydroxylated carboxylic acid with an additional unsaturation rather than a dicarboxylic acid.

Using this information, the search in CAS returned 19 hits that were discriminated based upon H/D exchange experiments. This led to a single candidate since only one out of the 19 compounds comprised a carboxylic acid, two hydroxyl groups and a single unsaturation (Fig. 6A). Based on this formula, we tried to explain the product ions observed in the multi-stage MS/MS experiments by proposing a mechanism of fragmentation. Unfortunately, in our opinion, the compound found in CAS is unlikely to generate the observed fragmentation pattern, especially the ion at *m*/*z* 113. Indeed, we would rather have expected product ions at *m*/*z* 99 and 69 according to the putative proposed mechanisms shown

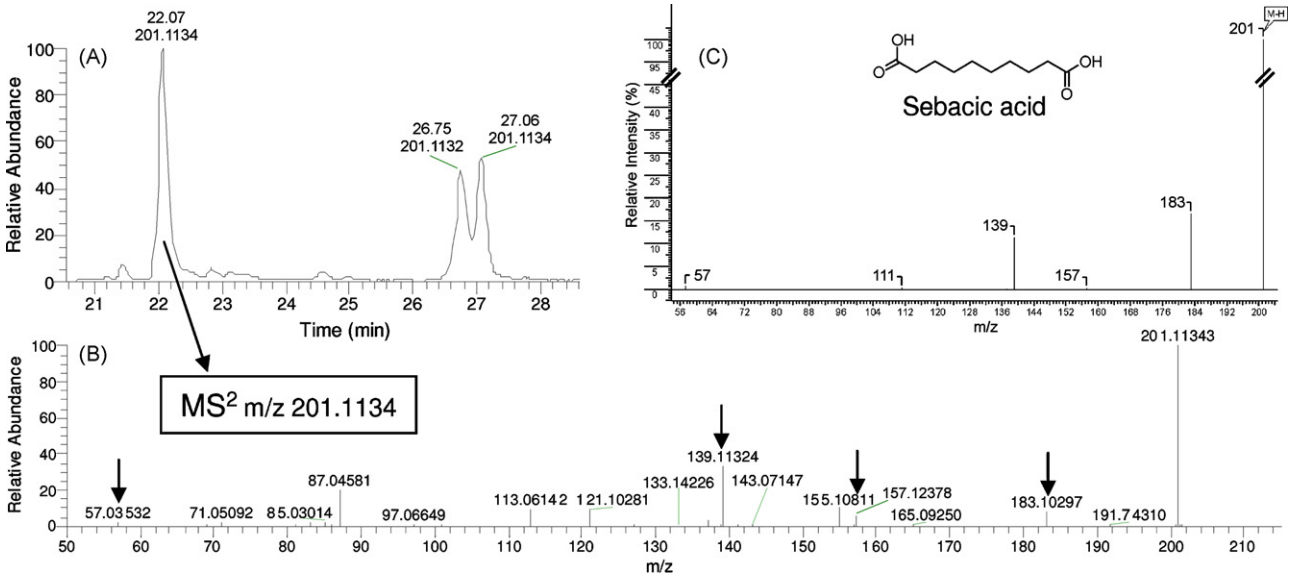


Fig. 4. Identification of an unknown compound in rat urine. (A) Extracted ion chromatogram at *m*/*z* 201.1133 (±2 ppm) from a LC/MS profile of rat urine acquired on an LC/LTQ-Orbitrap® system in negative ion mode. (B) MS^2 spectra of *m*/*z* 201.1134 at 22 min (with wide band activation function enabled). Arrows mark peaks similar to those observed at low-resolution on the TQ CID spectra provided by HMDB for sebacic acid. (C) MS/MS spectra of sebacic acid provided by HMDB (acquired with a TQ instrument Quattro Waters®). Same experimental conditions as those reported for Fig. 3.

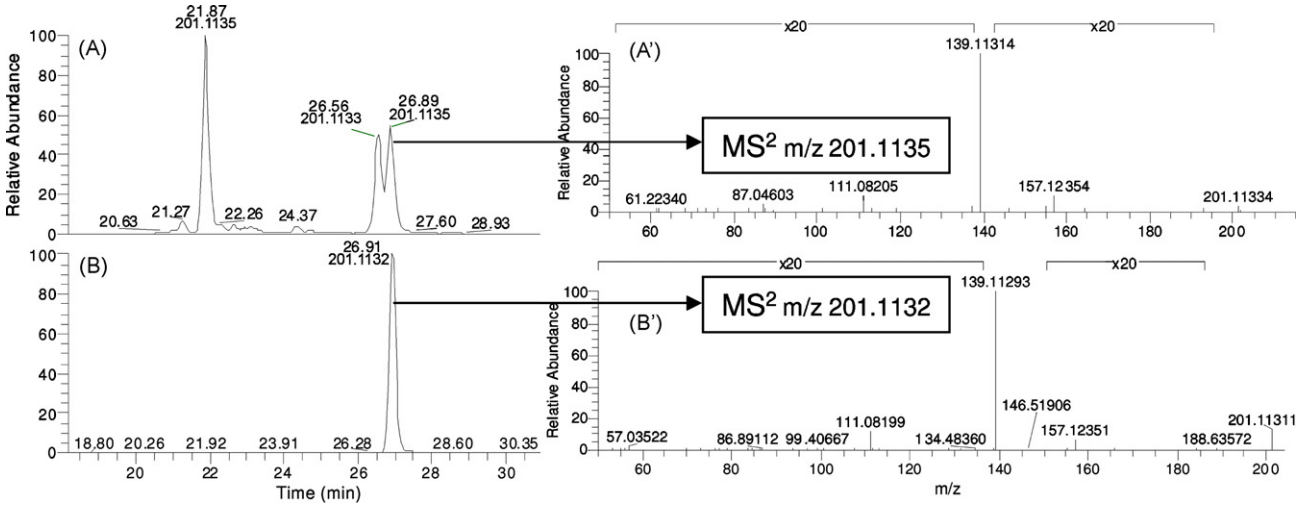


Fig. 5. Identification of sebacic acid in rat urine. (A) Extracted ion chromatogram at *m/z* 201.1133 from an LC/MS profile of rat urine acquired on a LC/LTQ-Orbitrap® system in negative ion mode. (B) Extracted ion chromatogram at *m/z* 201.1133 (±2 ppm) from an LC/MS profile of rat urine spiked with sebacic acid and acquired on an LC/LTQ-Orbitrap® system in negative-ion mode. Same experimental conditions as those reported for Fig. 3.

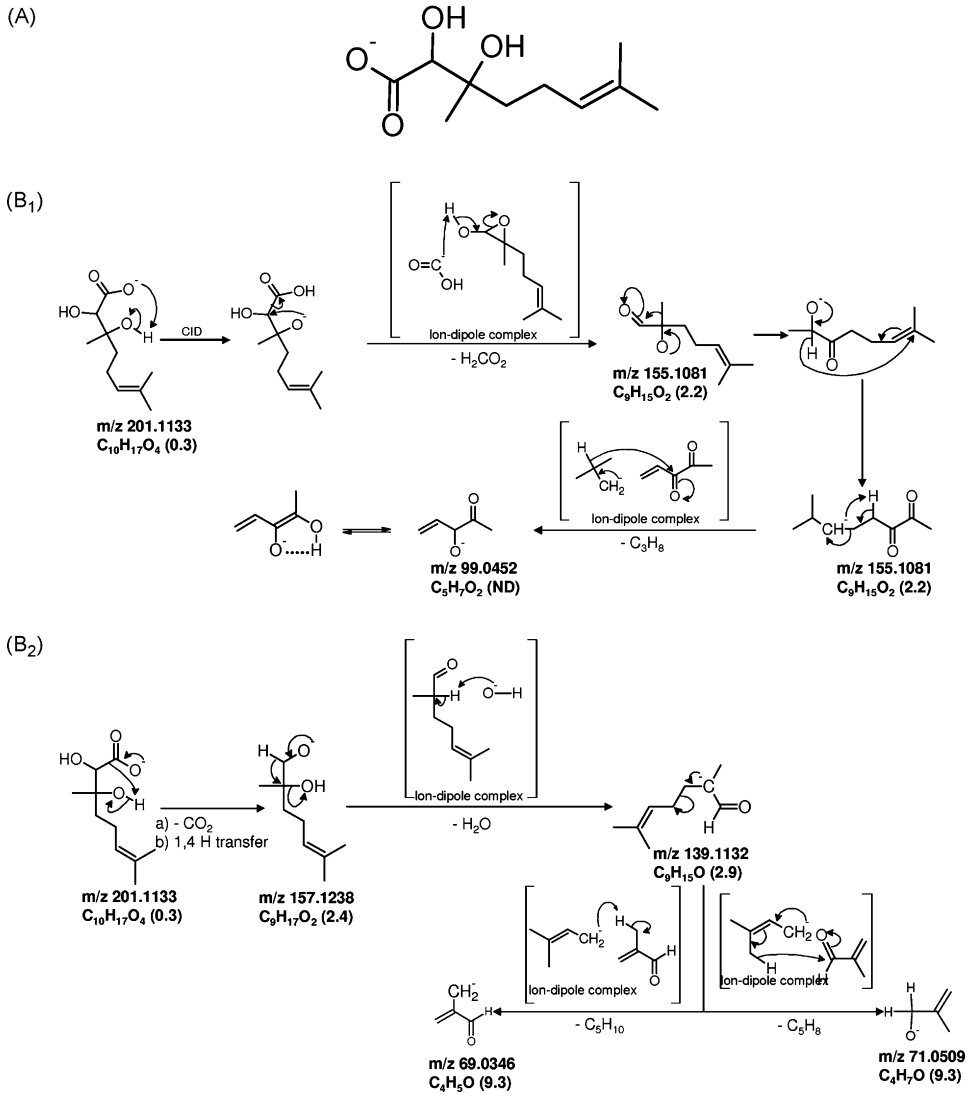


Fig. 6. (A) Structure proposed by CAS for the unknown metabolite at *m/z* 201.1133 ($C_{10}H_{18}O_4$) and putative predictive fragmentation mechanisms B1: starting with a neutral loss of H_2CO_2 and B2: starting with a neutral loss of CO_2. Exact masses and masses error on formula determination (ppm) are indicated under each product ion (ND = not detected).

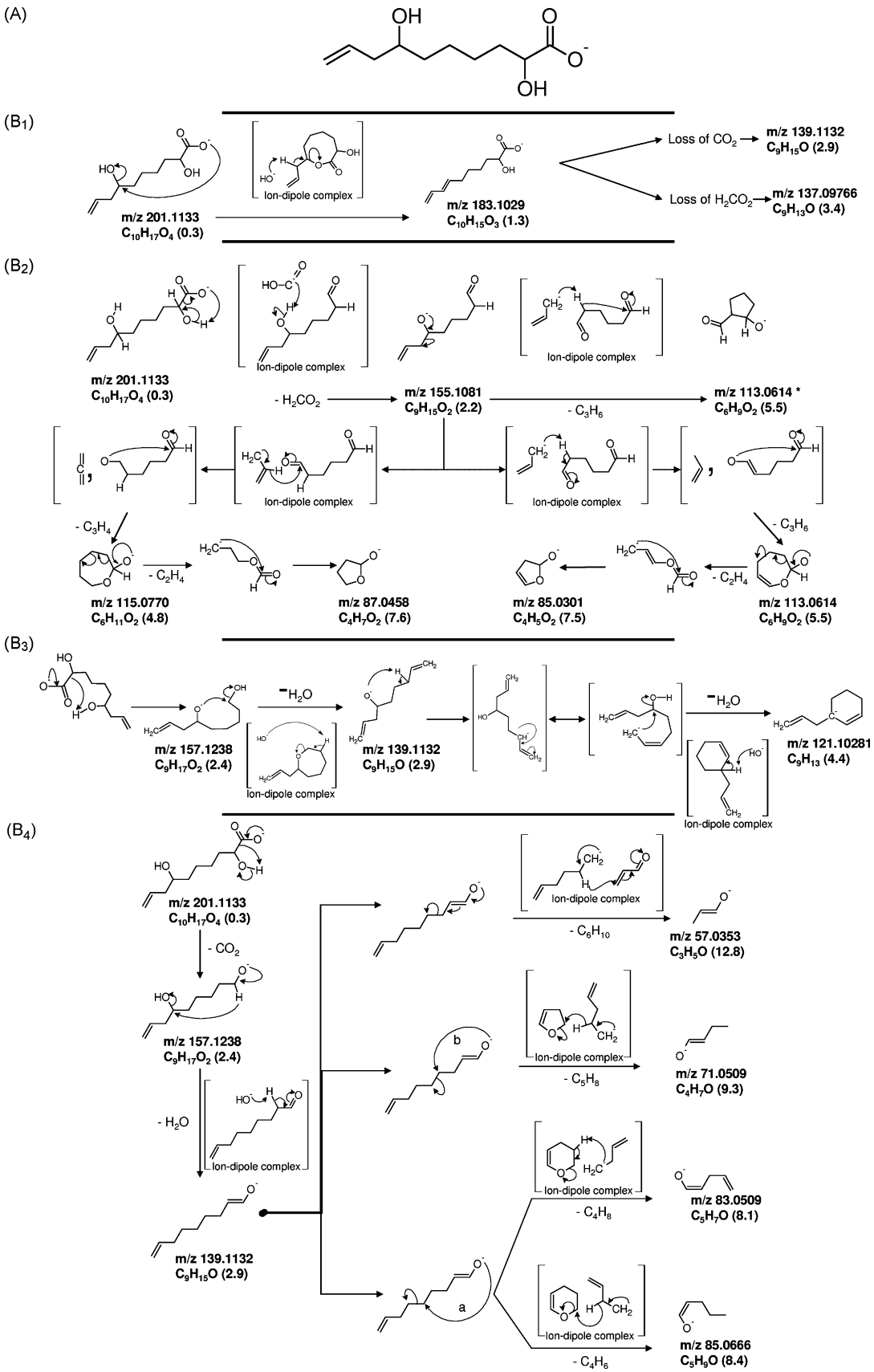


Fig. 7. (A) Proposed structure for the unknown metabolite at m/z 201.1133 ($C_{10}H_{18}O_4$) based on putative fragmentation mechanisms B1: starting with a neutral loss of H_2O yielding the ion at m/z 183.1029, B2: starting with a loss of H_2CO_2 yielding the ion at m/z 155.1081 and then to ions at m/z 113.0614, B3: starting with a loss of CO_2 and leading to the fragment ion at m/z 121.1028 and B4: starting with a loss of CO_2 but leading to fragment ions at m/z 85.0666, 83.0509, 71.0509, 57.0353. Exact masses and masses error on formula determination (ppm) are indicated under each product ion.

in Fig. 6B. However, such product ions were not retrieved in the experimental CID spectrum (Fig. 4A). As a consequence, we then undertook an interpretation of the CID spectrum of the endogenous compound of interest in order to propose a relevant structure that would fit with the observed product ions. A tentative structural elucidation of the product ions generated by the fragmentation of the metabolite of interest is shown in Fig. 7. The peak at m/z 113 can be considered as a diagnostic ion of the terminal position of the unsaturation. Indeed, it could be generated by an allylic cleavage that would be enhanced by the presence of a hydroxyl group at the homoallylic site. Of course, this putative structure must be confirmed by NMR experiments and further chemical synthesis for formal identification.

5. How to address MS signal redundancy?

With ESI-MS, a single compound may produce several signals corresponding to the formation of (i) adduct ions with cations in the positive ion mode (*e.g.*, Na^+, K^+, NH_4^+) or anions in the negative ion mode (*e.g.*, Cl^-), (ii) product ions formed by the spontaneous "in-source" CID of the precursor ion, and (iii) homo and heterodimeric ions (*e.g.*, $[2M+H]^+$, $[M+CH_3CN+H]^+$, $[2M-H]^-$). Such signal redundancy may be problematic for metabolomics because it increases time spent on identification and complicates database queries.

Using ESI ionization, it is not *a priori* obvious to determine whether the ion represents a quasi-molecular ion (*i.e.*, a protonated or deprotonated molecule) or an adduct ion. In positive ESI experiments, Na^+, K^+, NH_4^+ adduct ions can be observed in competition with MH^+. With ultra-high resolution instruments, the search for specific mass differences can help make this assignment. For example, looking for mass differences of 17.02655 ($M_{iso_{NH_4}} - M_{iso_H}$), 21.98195 ($M_{iso_{Na}} - M_{iso_H}$) and 37.95589 ($M_{iso_K} - M_{iso_H}$) traces the presence of NH_4^+ Na^+, K^+, adduct respectively. Product ions could be identified by the search for typical characteristic product ions and neutral losses.

A similar approach consists in scanning mass spectra to detect MS peaks differing by exact masses that correspond to known typical compositional changes occurring during the generation of adduct or product ions. At least two programs present this functionality: the ESI package and IntelliXtract®. The ESI package is a freely available package implemented in R [161]. It is able to annotate peaks such as isotope, adduct and product ions from a user-defined difference mass list that can be amended. IntelliXtract® (ACD Lab) is a commercial program compatible with most MS instrument raw data formats. The program performs component extraction and automatic ion assignment of molecular species (*i.e.*, protonated/deprotonated ions, cationized or anionized molecules). Peaks are flagged on mass spectra in which different colors discriminate related peaks in the case of co-eluting compounds.

Another approach is based on the fact that the ratios between the intensities of quasi-molecular and adduct or fragment ions are theoretically fixed, at least within a limited range of concentration of the parent molecule. It is therefore possible to assume that correlation coefficients between the intensities of pair of related ions should theoretically be equal to 1 either across several spectra within a sample or across all samples where the signals are observed. In the first case, the intensity correlation calculation is performed in the chromatographic domain and highlights the theoretical linear dependence between the extracted ion chromatograms of the molecule and its related compounds (*i.e.*, related ions should exhibit the same detection profile). In the second case, by assuming that the different ions from a single metabolite should exhibit the same retention time and the same change in response to a perturbation, and that instrumental conditions were kept identical from one sample to another, the relative intensities of such ions across all samples where they are observed should theoretically be fully correlated (correlation coefficient $r = 1$). Due to matrix effects and ionization suppression the theoretical value of r is rarely achieved and relation between signals is assumed for $r > 0.8$ [162]. Such an approach has been manually implemented by Chen et al. as the first step of the metabolite identification process [29]. Computational scripts relying on correlation calculations are also available in software such as the ESI package [161], openMS [163] or IntelliXtract®. Finally, autocorrelation matrices that were initially developed for NMR data processing [165] can also be applied to LC/MS data sets in order to highlight redundant information and also correlation or anti-correlation between ions eluted at different retention times with the aim of piecing together metabolic networks [166].

6. Mathematical tools to improve analysis of MS data

Multivariate statistical analyses constitute the most popular kind of visualization tool by emphasizing differences between samples. However, in some situations such as the characterization of unknown signals for database building or targeted approaches, it may be of value to sort the variables according to their chemical structures. In this context, ultra-high resolution mass spectrometry provides useful tools such as the Kendrick and van Krevelen representations for data set organization [63].

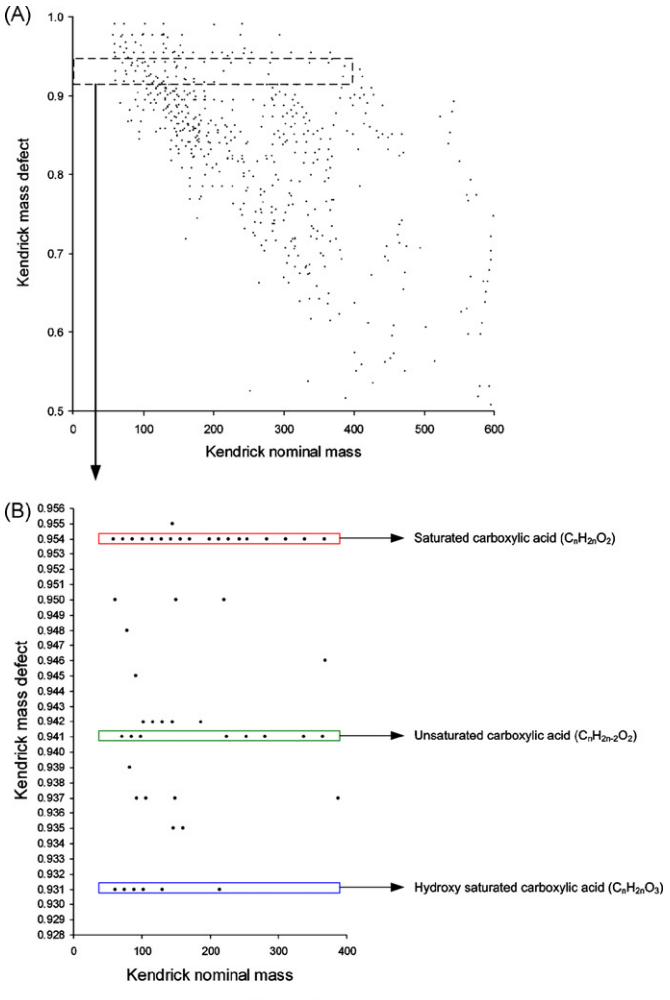
6.1. Kendrick plot

Kendrick proposed a mass scale based on the mass of CH_2 as an alternative to the IUPAC mass scale which is based on the mass of ^{12}C [167]. Such a scale converts the mass of CH_2 from 14.01565 to 14 u. Thus, homologous series (namely, compounds with the same constitution of heteroatoms and number of rings plus double bonds, but different numbers of CH_2 groups) will have an identical Kendrick mass defect [74,168]. The Kendrick representation plots the Kendrick mass defect as a function of the Kendrick nominal mass, thereby producing rows of data separated horizontally by numbers of alkyl groups and vertically according to degree of saturation and class.

The application of Kendrick plot improves the assignment of elemental formulas of high-molecular-weight compounds. As the elementary compositions of few low-molecular-weight compounds related to a given class have been identified, extension to higher masses with the same Kendrick mass defect (*i.e.*, belonging to the same "chemical class") allows for confident elemental composition assignment of ions [74].

The use of Kendrick mass defect analysis has recently been reported for lipidomics, in the course of a study aimed at evaluating changes in polar lipid levels in tumor cells in response to adenovirus therapy [169]. In this context, Kendrick plots provided a simple visual classification of different chemical classes of polar lipids such as phosphatidylinositols, gangliosides and sulfatides.

To illustrate the relevance of Kendrick indices in organizing biologically relevant compounds and their value for metabolomics, we provide here a Kendrick plot based on the exact masses of 7000 compounds extracted from the KEGG database. We limited Kendrick plot to compounds that exhibit even molecular weights within the range 50–1000 u and contain only C, H, N, O, P, and S atoms. As can be seen in Fig. 8A, it is possible to organize signals by sorting them into homologous series. Compounds of the same chemical class can be identified by calculating the Kendrick mass defect of one of them and then locating all the signals corresponding to this mass defect. This is illustrated by the isolation of carboxylic fatty acids (Fig. 8B and C). Furthermore, the Kendrick representa-



Saturated carboxylic acid ($C_nH_{2n}O_2$)

(C)

	Elementary composition	Identity	Exact mass	Kendrick mass	Kendrick nominal mass	Kendrick mass defect
	$C_1H_2O_2$	Formic acid	46.005	45.954	45	0.954
•	$C_2H_4O_2$	Acetic acid	60.021	59.954	59	0.954
•	$C_3H_6O_2$	Propionic acid	74.037	73.954	73	0.954
•	$C_4H_8O_2$	Butyric acid	88.052	87.954	87	0.954
•	$C_5H_{10}O_2$	Valeric acid	102.068	101.954	101	0.954
•	$C_6H_{12}O_2$	Caproic acid	116.084	115.954	115	0.954
•	$C_7H_{14}O_2$	Enanthic acid	130.099	129.954	129	0.954
•	$C_8H_{16}O_2$	Caprylic acid	144.115	143.954	143	0.954
•	$C_9H_{18}O_2$	Pelargonic acid	158.131	157.954	157	0.954
•	$C_{10}H_{20}O_2$	Capric acid	172.146	171.954	171	0.954
	$C_{11}H_{22}O_2$	Undecylic acid	186.162	185.954	185	0.954
•	$C_{12}H_{24}O_2$	Lauric acid	200.178	199.954	199	0.954
•	$C_{13}H_{26}O_2$	ridecylic acid	214.193	213.954	213	0.954
•	$C_{14}H_{28}O_2$	Myristic acid	228.209	227.954	227	0.954
	$C_{15}H_{30}O_2$	Pentadecylic acid	242.225	241.954	241	0.954
•	$C_{16}H_{32}O_2$	Palmitic acid	256.24	255.954	255	0.954
	$C_{17}H_{34}O_2$	Margaric acid	270.256	269.954	269	0.954
•	$C_{18}H_{36}O_2$	Stearic acid	284.272	283.954	283	0.954
	$C_{19}H_{38}O_2$	Nonadecylic acid	298.287	297.954	297	0.954
•	$C_{20}H_{40}O_2$	Arachidic acid	312.303	311.954	311	0.954
	$C_{21}H_{42}O_2$		326.318	325.954	325	0.954
•	$C_{22}H_{44}O_2$	Behenic acid	340.334	339.954	339	0.954
	$C_{23}H_{46}O_2$		354.35	353.954	353	0.954
•	$C_{24}H_{48}O_2$	Lignoceric acid	368.365	367.954	367	0.954

Fig. 8. (A) Kendrick plot based on the exact masses of 7000 compounds extracted from the KEGG database. The database query was constrained with the following criteria: (i) compound exhibiting even molecular weights within the range 50–1000 u and (ii) formulas containing C, H, N, O, P, and S atoms. (B) Zoom on the [0.928–0.956] Kendrick mass defect area showing the organization of biological compounds into homologous series that appear as ranks in the plot. (C) Table summarizing the carboxylic acids present in the database (marked by dots) and their associated IUPAC and Kendrick masses.

tion highlights some generic chemical modifications of compounds from homologous series. For example, hydroxy-carboxylic acid as well as mono-unsaturated carboxylic acids produce homologous series that can be distinguished from those related to saturated carboxylic acids, thus enabling rapid monitoring of biotransformation of a particular class of compounds (Fig. 8B). However, Kendrick plots do not succeed in discriminating isomers. As an example, the three compounds presented in Fig. 5 have the same Kendrick mass defect although they exhibit different chemical structures.

6.2. van Krevelen plot

Another way to organize data sets from accurate mass measurements is provided by the van Krevelen diagrams [170,171]. These diagrams are element–ratio plots, in which each dot represents the molar ratio of hydrogen to carbon (H/C) on the *y*-axis and molar ratio of oxygen to carbon (O/C) on the *x*-axis for one specific molecule. A prerequisite of such representations is to convert the accurate mass measurements of the data sets into elemental formulas. Then, the chemical classes exhibiting any characteristic H/C and/or O/C ratios cluster within specific regions of the diagram, as previously shown for biologically derived compounds, such as lipids, cellulose, lignins, proteins and condensed polyaromatic hydrocarbons [170–174].

In the van Krevelen plot, trends along the lines can be indicative of structural relationships among families of compounds linked by reactions that involve loss or gain of elements in a specific molar ratio. Lines from each reaction path have characteristic slopes or intercepts that can be easily demonstrated from mathematical calculations [170]. From these lines, a series of peaks, possibly products from various chemical reactions, can be visually identified. For example, a trend line representing methylation/demethylation reactions always intersects the ordinate at an H/C value of 2 and

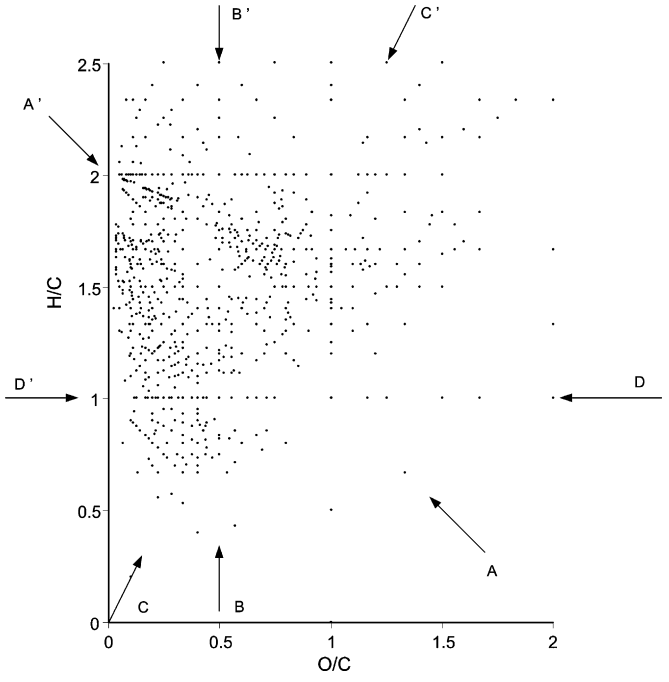


Fig. 9. Van Krevelen diagram based on the molecular formulas of 10,000 compounds extracted from the KEGG database. The database query was constrained with the following criteria: (i) compound exhibiting molecular weights within the range 50–1000 u and (ii) formulas containing C, H, N, O, P, and S atoms. The flagged lines in the plot denote the following chemical reactions: (AA′) methylation/demethylation, or alkyl chain elongation; (BB′) hydrogenation/dehydrogenation; (CC′) hydration/condensation; and (DD′) oxidation/reduction.

hydration/condensation reactions induce changes along a trend line with a slope of 2 [170,171].

Fig. 9 displays a van Krevelen diagram based on the exact masses of 10,000 compounds extracted from the KEGG database according to the following criteria: (i) molecular masses ranging from 50 to 1000; (ii) formulas containing C, H, N, O, P, and S atoms. The flagged lines in the plot denote the following chemical reactions: (AA′) methylation/demethylation, or alkyl chain elongation; (BB′) hydrogenation/dehydrogenation; (CC′) hydration/condensation; and (DD′) oxidation/reduction. Although the use of van Krevelen representations is still limited by the lack of software able to automatically generate and test elemental compositions, this example emphasizes the relevance of such a tool for metabolomic purposes. Alone or combined with Kendrick plots [168], they could be used to separate the signals previously highlighted by multivariate statistical analyses into different chemical families, thus providing another way to visualize the analytical information of biological relevance.

7. Conclusion and perspectives

The metabolome is characterized by a huge diversity of chemical structures requiring complementary analytical platform to reach its extensive coverage. Among these, atmospheric pressure ionization based technologies, and especially electrospray ionization are now very popular because they are adapted to the analysis of biological molecules and also because they can be combined with liquid chromatography. Furthermore, they give access to the molecular mass of uncharacterized metabolites, which facilitates the identification work. The aim of this review was to determine what can be expected from atmospheric pressure ionization mass spectrometry for the identification of signals highlighted by metabolomics. Mass spectrometry experiments alone are sufficient when the investigated metabolites are well described in databases and commercially available. In this case, identification is achieved by matching the retention time and CID spectra of the compound of interest to those of the putatively related synthetic reference molecule. In more complex situations, *i.e.*, when the metabolite is not reported in biochemical or metabolomic databases, additional analytical tools such as NMR will be used. However, in this context, atmospheric pressure ionization mass spectrometry can be used as a first approach for compound characterization. We have shown that careful and precise interpretation of CID spectra, combined with additional experiments such as H/D exchange, is useful for providing new structural hypotheses and in refining queries of general chemical databases. Collaboration with organic chemists is then necessary to synthesize the putative chemical structures, if possible, for further LC and CID spectral matching with the signal of interest. NMR experiments may also be required to confirm the structure and to address the issue of isomers.

In terms of perspectives, the identification process should be improved by developments in various fields reflecting the multidisciplinarity of the metabolomic-based approach: the development of tools for the standardization of metabolomic data, analytical platforms, relational databases, absolute quantification and statistical and mathematical tools. These different points are briefly addressed below.

7.1. Tools for the standardization of metabolomic data

Methodologies employed in metabolomics are diverse and evolving. Standardization of metabolomic data is required for the data to be re-used, shared and compared between laboratories. Several aspects have already been addressed or pointed out by the standard Metabolic Reporting Structure Working group [175] and also by working groups of the metabolomic society, including analytical chemistry [159], statistical analysis and experimental metadata [176,177].

One major issue remains the difficulty of comparing and evaluating the information generated by existing API-mass spectrometry platforms. To this end, reference mixtures of selected metabolites at appropriate concentrations or even perhaps synthetic biofluids have to be defined.

Absolute quantification could facilitate the comparison of results between laboratories. However, accurate and absolute quantification in LC/ESI-MS is not straightforward. Interestingly, some studies now report quantification of amine-containing molecules in biological matrices by using derivatization with a stable isotopic tag, thus suggesting that large-scale quantification of metabolite is possible [178,179].

Furthermore, we lack a mass spectral repository, *i.e.*, a central open source mass spectral database of CID spectra acquired with various instruments in different laboratories. Information such as the biological matrix, instrument and analytical conditions should be precisely documented, as previously suggested [159,175].

7.2. Analytical platforms

LC-NMR has already been applied to structural elucidation of drug and plant metabolites [180,181]. LC/MS and NMR can be coupled either on-line or off-line in order to improve structure determination [182,183]. However and despite the introduction of microflow NMR allowing investigations on the microgram scale [184], NMR still lacks sensitivity and requires appropriate chromatographic purification when applied to complex biofluids.

Ion mobility mass spectrometers can be coupled to conventional mass spectrometers. They allow separation of isomers or isobars by providing a third separation dimension if liquid chromatography is also used [185]. Ion mobility mass spectrometry has been evaluated for metabolomic applications: over 1000 metabolite peaks were observed from two-dimensional mobility mass spectra of cell extracts from *E. coli* cultures, with 42 isobaric pairs identified [186].

While API-MS techniques have proven efficient in the determination of ion monoisotopic masses, they also lack reproducibility in terms of fragmentation pattern, which has until now limited the constitution of LC/MS/MS spectral databases. Thus, an interesting improvement would be the interfacing of liquid chromatography with electron ionization mass spectrometry. Such a system has recently been reported: it exhibited minimal matrix effects, was fully compatible with non volatile buffers and salts present in mobile phases and allowed comparisons with EI spectra databases [187]. Although it lacks sensitivity, LC/EI-MS could be used to complement CID spectra obtained with API-mass spectrometers. Doing so, metabolite characterization could be improved through access to large EI spectral libraries and complementary fragmentation information.

Another promising approach consists in soft thermal vaporization followed by supersonic expansion and direct sample compound ionization, while in a supersonic molecular beam (SMB). This technique produces EI mass spectra with enhanced molecular ions that are used to eliminate false search candidates [188,189].

7.3. Toward relational databases

There is a crucial need to develop relational database covering the entire metabolomic experiment and allowing comparison

of data sets and interpretation of experimental results by various teams. To this end, associated metadata, *i.e.* biological data (treatment or status) or experimental protocols (sample origin, tissue and experimental conditions) should be stipulated, as has already been proposed for proteomics [190] and transcriptomics (microarray) [191]. Similar steps have been taken in the metabolomics field with MIAMET that provides a checklist of information that should be described in metabolomic publications [192].

A further improvement has been achieved with ArMet (*i.e.*, Architecture for Metabolomics) [193]. Its objective is to provide a formal description of plant metabolomic data sets, including associated metadata, for comparison, mining, data storage and exchange. Finally, standardization of relational databases remains a challenge that includes the choice of spectral data formats (instrument manufacturer format or rather open access file formats such as netCDF, tabular data format and storage format (*e.g.*, XML)). Among recent initiatives, MeMo is a hybrid SQL/XML based on ArMet structure that has been proposed for metabolomic data management (*i.e.*, data storage, organization and annotation) [194].

7.4. Statistical and mathematical tools

The performance of metabolome analysis could be improved by the development of multivariate statistical and chemometric methods that could link the information provided by complementary analytical tools on the same data set. Such approaches, based for example on correlation matrixes [195,196], could help to characterize unknown metabolites in analytical systems and could also be useful in gaining new insights into metabolic pathways.

Acknowledgments

EW receives financial support through a research contract funded by the CEA (Commissariat à l'Energie Atomique) and Technologie Servier. JFH is supported by a grant provided by the FSR (Fonds Spéciaux de Recherche, Université catholique de Louvain). The authors wish to thank the editor and anonymous referees for their useful comments and suggestions.

References

[1] O. Fiehn, S. Kloska, T. Altmann, Curr. Opin. Biotechnol. 12 (2001) 82.
[2] J.K. Nicholson, J. Connelly, J.C. Lindon, E. Holmes, Nat. Rev. Drug Discov. 1 (2002) 153.
[3] O. Fiehn, Plant Mol. Biol. 48 (2002) 155.
[4] L.W. Sumner, P. Mendes, R.A. Dixon, Phytochemistry 62 (2003) 817.
[5] M. Katajamaa, M. Oresic, J. Chromatogr. A 1158 (2007) 318.
[6] S. Bijlsma, I. Bobeldijk, E.R. Verheij, R. Ramaker, S. Kochhar, I.A. Macdonald, O.B. van, A.K. Smilde, Anal. Chem. 78 (2006) 567.
[7] E. Holmes, H. Antti, Analyst 127 (2002) 1549.
[8] J. Trygg, E. Holmes, T. Lundstedt, J. Proteome Res. 6 (2007) 469.
[9] E. Holmes, A.W. Nicholls, J.C. Lindon, S. Ramos, M. Spraul, P. Neidig, S.C. Connor, J. Connelly, S.J. Damment, J. Haselden, J.K. Nicholson, NMR Biomed. 11 (1998) 235.
[10] O. Fiehn, J. Kopka, R.N. Trethewey, L. Willmitzer, Anal. Chem. 72 (2000) 3573.
[11] A. Aharoni, C.H. Ric de Vos, H.A. Verhoeven, C.A. Maliepaard, G. Kruppa, R. Bino, D.B. Goodenowe, OMICS 6 (2002) 217.
[12] S.A. Trauger, E.P. Go, Z. Shen, J.V. Apon, B.J. Compton, E.S. Bouvier, M.G. Finn, G. Siuzdak, Anal. Chem. 76 (2004) 4484.
[13] S. Vaidyanathan, D. Jones, T. Jenkins, D.B. Kell, R. Goodacre, Abstracts of Papers of the 229th ACS National meeting, 2005, p. BIOT 414.
[14] I.D. Wilson, J.K. Nicholson, J. Castro-Perez, J.H. Granger, K.A. Johnson, B.W. Smith, R.S. Plumb, J. Proteome Res. 4 (2005) 591.
[15] S. Vaidyanathan, S. Gaskell, R. Goodacre, Rapid Commun. Mass Spectrom. 20 (2006) 1192.
[16] K. Dettmer, P.A. Aronov, B.D. Hammock, Mass Spectrom. Rev. 26 (2007) 51.
[17] G. Huang, H. Chen, X. Zhang, R.G. Cooks, Z. Ouyang, Anal. Chem. 79 (2007) 8327.
[18] L.W. Jia, C. Wang, S.M. Zhao, X. Lu, G.W. Xu, J. Chromatogr. B 860 (2007) 134.
[19] S. Vaidyanathan, D. Jones, J. Ellis, T. Jenkins, C. Chong, M. Anderson, R. Goodacre, Rapid Commun. Mass Spectrom. 21 (2007) 2157.
[20] A. Craig, O. Cloarec, E. Holmes, J.K. Nicholson, J.C. Lindon, Anal. Chem. 78 (2006) 2262.
[21] R.A. van den Berg, H.C. Hoefsloot, J.A. Westerhuis, A.K. Smilde, M.J. van der Werf, BMC Genomics 7 (2006) 142.
[22] C. Ducruix, D. Vailhen, E. Werner, J.B. Fievet, J. Bourgignon, J.C. Tabet, E. Ezan, C. Junot, Chemom. Intell. Lab. Syst. 91 (2008) 67.
[23] J.L. Wolfender, K. Ndjoko, K. Hostettmann, J. Chromatogr. A 1000 (2003) 437.
[24] R.J. Molyneux, P. Schieberle, J. Agric. Food Chem. 55 (2007) 4625.
[25] V.V. Tolstikov, O. Fiehn, Anal. Biochem. 301 (2002) 298.
[26] A. Lafaye, C. Junot, B. Ramounet-Le Gall, P. Fritsch, J.C. Tabet, E. Ezan, Rapid Commun. Mass Spectrom. 17 (2003) 2541.
[27] M.J. Bogusz, R.D. Maier, K.D. Kruger, K.S. Webb, J. Romeril, M.L. Miller, J. Chromatogr. A 844 (1999) 409.
[28] Q. Cui, I.A. Lewis, A.D. Hegeman, M.E. Anderson, J. Li, C.F. Schulte, W.M. Westler, H.R. Eghbalnia, M.R. Sussman, J.L. Markley, Nat. Biotechnol. 26 (2008) 162.
[29] J. Chen, X. Zhao, J. Fritsche, P. Yin, P. Schmitt-Kopplin, W. Wang, X. Lu, H.U. Haring, E.D. Schleicher, R. Lehmann, G. Xu, Anal. Chem. 80 (2008) 1280.
[30] K. Vekey, J. Mass Spectrom. 31 (1996) 445.
[31] V. Grill, J. Shen, C. Evans, R.G. Cooks, Rev. Sci. Instrum. 72 (2001) 3149.
[32] E. Rosenberg, J. Chromatogr. A 1000 (2003) 841.
[33] A. Raffaelli, A. Saba, Mass Spectrom. Rev. 22 (2003) 318.
[34] R.B. Cody, J.A. Laramee, H.D. Durst, Anal. Chem. 77 (2005) 2297.
[35] G. Morlock, Y. Ueda, J. Chromatogr. A 1143 (2007) 243.
[36] C.M. Whitehouse, R.N. Dreyer, M. Yamashita, J.B. Fenn, Anal. Chem. 57 (1985) 675.
[37] Z. Takats, J.M. Wiseman, B. Gologan, R.G. Cooks, Science 306 (2004) 471.
[38] H.W. Chen, Z.Z. Pan, N. Talaty, D. Raftery, R.G. Cooks, Rapid Commun. Mass Spectrom. 20 (2006) 1577.
[39] Z.Z. Pan, H.W. Gu, N. Talaty, H.W. Chen, N. Shanaiah, B.E. Hainline, R.G. Cooks, D. Raftery, Anal. Bioanal. Chem. 387 (2007) 539.
[40] B.B. Schneider, D.D.Y. Chen, Anal. Chem. 72 (2000) 791.
[41] B.B. Schneider, D.J. Douglas, D.D.Y. Chen, J. Am. Soc. Mass Spectrom. 12 (2001) 772.
[42] J.C. Schwartz, A.P. Wade, C.G. Enke, R.G. Cooks, Anal. Chem. 62 (1990) 1809.
[43] E. De Hoffman, V. Stroobant, Mass Spectrometry: Principles and Applications, 3rd ed., John Wiley and Sons Ltd., West Sussex, UK, 2007.
[44] R.E. March, J.F.J. Todd, Practical Aspects of Ion Trap Mass Spectrometry: Fundamentals of Ion Trap Mass Spectrometry, vol. I, CRC-Press, Boca Raton, FL, 1995.
[45] R.E. March, J.F.J. Todd, Practical Aspects of Ion Trap Mass Spectrometry: Ion Trap Instrumentation, vol. II, CRC-Press, Boca Raton, FL, 1995.
[46] R.E. March, J.F.J. Todd, Practical Aspects of Ion Trap Mass Spectrometry: Chemical, Environmental and Biomedical Applications, vol. III, CRC-Press, Boca Raton, FL, 1995.
[47] D.J. Douglas, A.J. Frank, D.M. Mao, Mass Spectrom. Rev. 24 (2005) 1.
[48] R.M.A. Heeren, A.J. Kleinnijenhuis, L.A. McDonnell, T.H. Mize, Anal. Bioanal. Chem. 378 (2004) 1048.
[49] M. Hardman, A.A. Makarov, Anal. Chem. 75 (2003) 1699.
[50] Q. Hu, R.J. Noll, H. Li, A. Makarov, M. Hardman, R.G. Cooks, J. Mass Spectrom. 40 (2005) 430.
[51] A. Makarov, E. Denisov, O. Lange, S. Horning, J. Am. Soc. Mass Spectrom. 17 (2006) 977.
[52] A. Makarov, Anal. Chem. 72 (2000) 1156.
[53] F.W. Mclafferty, Interpretation of Mass Spectra, 4th ed., Wiley University Science Book, New York, 1993.
[54] A.W.T. Bristow, K.S. Webb, A.T. Lubben, J. Halket, Rapid Commun. Mass Spectrom. 18 (2004) 1447.
[55] K.K. Murray, R.K. Boyd, M.N. Eberlin, G.J. Langley, L. Li, Y. Naito, J.C. Tabet, Abstracts of Papers of the 229th ACS National meeting, 2005, p. ANYL-212.
[56] J. Castro-Perez, M. McCullagh, A. Millar, Waters Technical Note, 720001170EN, 2005.
[57] M. McCullagh, J. Castro-Perez, L. Calton, Waters Technical Note, 720001596EN, 2006.
[58] A. Makarov, E. Denisov, A. Kholomeev, W. Balschun, O. Lange, K. Strupat, S. Horning, Anal. Chem. 78 (2006) 2113.
[59] H.K. Lim, J. Chen, C. Sensenhauser, K. Cook, V. Subrahmanyam, Rapid Commun. Mass Spectrom. 21 (2007) 1821.
[60] S.M. Peterman, N. Duczak Jr., A.S. Kalgutkar, M.E. Lame, J.R. Soglia, J. Am. Soc. Mass Spectrom. 17 (2006) 363.
[61] Q. Ruan, S. Peterman, M.A. Szewc, L. Ma, D. Cui, W.G. Humphreys, M. Zhu, J. Mass Spectrom. 43 (2008) 251.
[62] B.P. Koch, T. Dittmar, M. Witt, G. Kattner, Anal. Chem. 79 (2007) 1758.
[63] J. Meija, Anal. Bioanal. Chem. 385 (2006) 486.
[64] F.W. McLafferty, F. Turecek, Interpretation of Mass Spectra, 4th ed., Wiley University Science Book, New York, 1993.
[65] A.W. Bristow, Mass Spectrom. Rev. 25 (2006) 99.
[66] T. Kind, O. Fiehn, BMC Bioinform. 8 (2007) 105.
[67] T. Kind, O. Fiehn, BMC Bioinform. 7 (2006) 234.
[68] K.D. Henry, E.R. Williams, B.H. Wang, F.W. McLafferty, J. Shabanowitz, D.F. Hunt, Proc. Natl. Acad. Sci. U.S.A. 86 (1989) 9075.
[69] S. Ojanpera, A. Pelander, M. Pelzing, I. Krebs, E. Vuori, I. Ojanpera, Rapid Commun. Mass Spectrom. 20 (2006) 1161.
[70] A.H. Grange, M.C. Zumwalt, G.W. Sovocool, Rapid Commun. Mass Spectrom. 20 (2006) 89.

[71] A.D. Hegeman, C.F. Schulte, Q. Cui, I.A. Lewis, E.L. Huttlin, H. Eghbalnia, A.C. Harms, E.L. Ulrich, J.L. Markley, M.R. Sussman, Anal. Chem. 79 (2007) 6912.
[72] A.C. Stenson, A.G. Marshall, W.T. Cooper, Anal. Chem. 75 (2003) 1275.
[73] E.B. Kujawinski, M.D. Behn, Anal. Chem. 78 (2006) 4363.
[74] C.A. Hughey, C.L. Hendrickson, R.P. Rodgers, A.G. Marshall, K. Qian, Anal. Chem. 73 (2001) 4676.
[75] H.J. Yoo, H. Liu, K. Hakansson, Anal. Chem. 79 (2007) 7858.
[76] R.L. Woodin, D.S. Bomse, J.L. Beauchamp, J. Am. Soc. Mass Spectrom. 100 (1978) 3248.
[77] D.P. Little, J.P. Speir, M.W. Senko, P.B. O'Connor, F.W. McLafferty, Anal. Chem. 66 (1994) 2809.
[78] N.A. Kruger, R.A. Zubarev, D.M. Horn, F.W. McLafferty, Int. J. Mass Spectrom. 187 (1999) 787.
[79] K. Levsen, Fundamental Aspects of Organic Mass Spectrometry, Verlag chemie, Weinheim, 1978.
[80] R.G. Cooks, Collision spectroscopy, vol. xiv, Plenum Press, New York, 1978.
[81] K.L. Busch, G.L. Glish, S.A. McLuckey, Mass Spectrometry/Mass Spectrometry: Techniques and Applications of Tandem Mass Spectrometry, VCH Publishers, New York, 1988.
[82] E. Clayton, R.H. Bateman, Rapid Commun. Mass Spectrom. 6 (1992) 719.
[83] U. Lewandrowski, A. Resemann, A. Sickmann, Anal. Chem. 77 (2005) 3274.
[84] R.G. Cooks, J.H. Beynon, J.F. Litton, Org. Mass Spectrom. 10 (1975) 503.
[85] R.J. Cotter, W. Griffith, C. Jelinek, J. Chromatogr. B 855 (2007) 2.
[86] G. Sun, K. Yang, Z. Zhao, S. Guan, X. Han, R.W. Gross, Anal. Chem. 79 (2007) 6629.
[87] N.J. Jensen, K.B. Tomer, M.L. Gross, J. Am. Chem. Soc. 107 (1985) 1863.
[88] K.B. Tomer, F.W. Crow, M.L. Gross, J. Am. Chem. Soc. 105 (1983) 5487.
[89] J. Adams, Mass Spectrom. Rev. 9 (1990) 141.
[90] C.F. Cheng, M.L. Gross, Mass Spectrom. Rev. 19 (2000) 398.
[91] R.C. Murphy, J. Fiedler, J. Hevko, Chem. Rev. 101 (2001) 479.
[92] R.E. March, H.X. Li, O. Belgacem, D. Papanastasiou, Int. J. Mass Spectrom. 262 (2007) 51.
[93] S.A. McLuckey, D.E. Goeringer, J. Mass Spectrom. 32 (1997) 461.
[94] D.F. Hunt, J. Shabanowitz, J.R. Yates, J. Chem. Soc. Chem. Commun. 8 (1987) 548.
[95] Y.O. Tsybin, M. Witt, G. Baykut, F. Kjeldsen, P. Hakansson, Rapid Commun. Mass Spectrom. 17 (2003) 1759.
[96] Y. Hashimoto, H. Hasegawa, K. Yoshinari, I. Waki, Anal. Chem. 75 (2003) 420.
[97] J.J. Wilson, J.S. Brodbelt, Anal. Chem. 79 (2007) 2067.
[98] R.B. Cody, B.S. Freiser, Anal. Chem. 51 (1979) 547.
[99] B.A. Budnik, K.F. Haselmann, Y.N. Elkin, V.I. Gorbach, R.A. Zubarev, Anal. Chem. 75 (2003) 5994.
[100] H. Idborg-Bjorkman, P.O. Edlund, O.M. Kvalheim, I. Schuppe-Koistinen, S.P. Jacobsson, Anal. Chem. 75 (2003) 4784.
[101] S.U. Bajad, W.Y. Lu, E.H. Kimball, J. Yuan, C. Peterson, J.D. Rabinowitz, J. Chromatogr. A 1125 (2006) 76.
[102] A. Lafaye, C. Junot, B. Ramounet-Le Gall, P. Fritsch, E. Ezan, J.C. Tabet, J. Mass Spectrom. 39 (2004) 655.
[103] S. Wagner, K. Scholz, M. Donegan, L. Burton, J. Wingate, W. Volkel, Anal. Chem. 78 (2006) 1296.
[104] Z. Yan, G.W. Caldwell, Anal. Chem. 76 (2004) 6835.
[105] W. Bicker, M. Lammerhofer, D. Genser, H. Kiss, W. Lindner, Toxicol. Lett. 159 (2005) 235.
[106] L. Coulier, R. Bas, S. Jespersen, E. Verheij, M.J. van der Werf, T. Hankemeier, Anal. Chem. 78 (2006) 6573.
[107] G.C. Kite, M.J. Howes, M.S. Simmonds, Rapid Commun. Mass Spectrom. 18 (2004) 2859.
[108] M. Bantscheff, M. Boesche, D. Eberhard, T. Matthieson, G. Sweetman, B. Kuster, Mol. Cell Proteomics (2008).
[109] T.J. Griffin, H. Xie, S. Bandhakavi, J. Popko, A. Mohan, J.V. Carlis, L. Higgins, J. Proteome Res. 6 (2007) 4200.
[110] J.C. Schwartz, M.W. Senko, J.E. Syka, J. Am. Soc. Mass Spectrom. 13 (2002) 659.
[111] A. Koulman, B.A. Tapper, K. Fraser, M. Cao, G.A. Lane, S. Rasmussen, Rapid Commun. Mass Spectrom. 21 (2007) 421.
[112] J.W. Hager, J.C.Y. Le Blanc, Rapid Commun. Mass Spectrom. 17 (2003) 1056.
[113] Y.A. Hammel, R. Mohamed, E. Gremaud, M.H. Lebreton, P.A. Guy, J. Chromatogr. A 1177 (2008) 58.
[114] G. Hopfgartner, E. Varesio, V. Tschappat, C. Grivet, E. Bourgogne, L.A. Leuthold, J. Mass Spectrom. 39 (2004) 845.
[115] A.C. Li, M.A. Gohdes, W.Z. Shou, Rapid Commun. Mass Spectrom. 21 (2007) 1421.
[116] A.N. Krutchinsky, I.V. Chernushevich, V.L. Spicer, W. Ens, K.G. Standing, J. Am. Soc. Mass Spectrom. 9 (1998) 569.
[117] A.D. Patterson, H. Li, G.S. Eichler, K.W. Krausz, J.N. Weinstein, A.J. Fornace Jr., F.J. Gonzalez, J.R. Idle, Anal. Chem. 80 (2008) 665.
[118] R.S. Plumb, K.A. Johnson, P. Rainville, B.W. Smith, I.D. Wilson, J.M. Castro-Perez, J.K. Nicholson, Rapid Commun. Mass Spectrom. 20 (2006) 1989.
[119] K. Tanaka, T. Tamura, S. Fukuda, J. Batkhuu, C. Sanchir, K. Komatsu, Phytochemistry 69 (2008) 2081.
[120] J.M. Campbell, B.A. Collings, D.J. Douglas, Rapid Commun. Mass Spectrom. 12 (1998) 1463.
[121] T. Yokosuka, K. Yoshinari, K. Kobayashi, A. Ohtake, A. Hirabayashi, Y. Hashimoto, I. Waki, T. Takao, Rapid Commun. Mass Spectrom. 20 (2006) 2589.
[122] J. Ding, C.M. Sorensen, Q. Zhang, H. Jiang, N. Jaitly, E.A. Livesay, Y. Shen, R.D. Smith, T.O. Metz, Anal. Chem. 79 (2007) 6081.
[123] M. Thevis, A.A. Makarov, S. Horning, W. Schanzer, Rapid Commun. Mass Spectrom. 19 (2005) 3369.
[124] J.V. Olsen, B. Macek, O. Lange, A. Makarov, S. Horning, M. Mann, Nat. Methods 4 (2007) 709.
[125] S.C. Brown, G. Kruppa, J.L. Dasseux, Mass Spectrom. Rev. 24 (2005) 223.
[126] Y. Shen, R. Zhang, R.J. Moore, J. Kim, T.O. Metz, K.K. Hixson, R. Zhao, E.A. Livesay, H.R. Udseth, R.D. Smith, Anal. Chem. 77 (2005) 3090.
[127] M.M. Siegel, Anal. Chem. 60 (1988) 2090.
[128] H.T. Chi, J.K. Baker, Org. Mass Spectrom. 28 (1993) 12.
[129] M.D. Sierra, A. Furey, B. Hamilton, M. Lehane, K.J. James, J. Mass Spectrom. 38 (2003) 1178.
[130] A.M. Kamel, B. Munson, Eur. J. Mass Spectrom. 10 (2004) 239.
[131] K.E. Karlsson, J Chromatogr. 647 (1993) 31.
[132] D.Q. Liu, C.E.C.A. Hop, M.G. Beconi, A. Mao, S.H.L. Chiu, Rapid Commun. Mass Spectrom. 15 (2001) 1832.
[133] A.M. Kamel, H.G. Fouda, P.R. Brown, B. Munson, J. Am. Soc. Mass Spectrom. 13 (2002) 543.
[134] A.M. Kamel, K.S. Zandi, W.W. Massefski, J. Pharm. Biomed. Anal. 31 (2003) 1211.
[135] B. Zhou, Z.Y. Zhang, Methods 42 (2007) 227.
[136] M.E. Palmer, L.W. Tetler, I.D. Wilson, Rapid Commun. Mass Spectrom. 14 (2000) 808.
[137] S. Campbell, M.T. Rodgers, E.M. Marzluff, J.L. Beauchamp, J. Am. Chem. Soc. 117 (1995) 12840.
[138] D.Q. Liu, C.E.C.A. Hop, J. Pharm. Biomed. Anal. 37 (2005) 1.
[139] C. Prakash, C.L. Shaffer, A. Nedderman, Mass Spectrom. Rev. 26 (2007) 340.
[140] E.E. Carlson, B.F. Cravatt, Nat. Methods 4 (2007) 429.
[141] E.E. Carlson, B.F. Cravatt, J. Am. Chem. Soc. 129 (2007) 15780.
[142] M.C. Nicklaus, W.D. Ihlenfeldt, D. Filimonov, V.V. Poroikov, Abstracts of Papers, 222nd ACS National Meeting, Chicago, IL, United States, August 26–30, 2001, CINF-091.
[143] J.S. Brecher, Chimia 52 (1998) 658.
[144] K. Degtyarenko, M.P. de, M. Ennis, J. Hastings, M. Zbinden, A. McNaught, R. Alcantara, M. Darsow, M. Guedj, M. Ashburner, Nucleic Acids Res. 36 (2008) D344–D350.
[145] S. Goto, Y. Okuno, M. Hattori, T. Nishioka, M. Kanehisa, Nucleic Acids Res. 30 (2002) 402.
[146] M. Kanehisa, S. Goto, M. Hattori, K.F. oki-Kinoshita, M. Itoh, S. Kawashima, T. Katayama, M. Araki, M. Hirakawa, Nucleic Acids Res. 34 (2006) D354–D357.
[147] M. Krummenacker, S. Paley, L. Mueller, T. Yan, P.D. Karp, Bioinformatics 21 (2005) 3454.
[148] R. Caspi, H. Foerster, C.A. Fulcher, P. Kaipa, M. Krummenacker, M. Latendresse, S. Paley, S.Y. Rhee, A.G. Shearer, C. Tissier, T.C. Walk, P. Zhang, P.D. Karp, Nucleic Acids Res. 36 (2008) D623–D631.
[149] D.S. Wishart, D. Tzur, C. Knox, R. Eisner, A.C. Guo, N. Young, D. Cheng, K. Jewell, D. Arndt, S. Sawhney, C. Fung, L. Nikolai, M. Lewis, M.A. Coutouly, I. Forsythe, P. Tang, S. Shrivastava, K. Jeroncic, P. Stothard, G. Amegbey, D. Block, D.D. Hau, J. Wagner, J. Miniaci, M. Clements, M. Gebremedhin, N. Guo, Y. Zhang, G.E. Duggan, G.D. Macinnis, A.M. Weljie, R. Dowlatabadi, F. Bamforth, D. Clive, R. Greiner, L. Li, T. Marrie, B.D. Sykes, H.J. Vogel, L. Querengesser, Nucleic Acids Res. 35 (2007) D521–D526.
[150] C.A. Smith, G. O'Maille, E.J. Want, C. Qin, S.A. Trauger, T.R. Brandon, D.E. Custodio, R. Abagyan, G. Siuzdak, Ther. Drug Monit. 27 (2005) 747.
[151] Y. Shinbo, Y. Nakamura, M. Altaf-Ul-Amin, H. Asahi, K. Kurokawa, M. Arita, K. Saito, D. Ohta, D. Shibata, S. Kanaya, Biotechnol. Agric. For. 57 (2006) 165.
[152] S. Moco, R.J. Bino, O. Vorst, H.A. Verhoeven, G.J. de, T.A. van Beek, J. Vervoort, C.H. de Vos, Plant Physiol. 141 (2006) 1205.
[153] J. Kopka, N. Schauer, S. Krueger, C. Birkemeyer, B. Usadel, E. Bergmuller, P. Dormann, W. Weckwerth, Y. Gibon, M. Stitt, L. Willmitzer, A.R. Fernie, D. Steinhauser, Bioinformatics 21 (2005) 1635.
[154] R.C. Willoughby, R.F. Browner, Anal. Chem. 56 (1984) 2626.
[155] J.C. Wolff, P.N. Hawtin, S. Monte, M. Balogh, T. Jones, Rapid Commun. Mass Spectrom. 15 (2001) 265.
[156] R. Jansen, G. Lachatre, P. Marquet, Clin. Biochem. 38 (2005) 362.
[157] M. Arita, Genome Res. 13 (2003) 2455.
[158] M. Arita, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 1543.
[159] L.W. Sumner, A. Amberg, D. Barrett, M.H. Beale, R. Beger, C.A. Daykin, T.W.M. Fan, O. Fiehn, R. Goodacre, J.L. Griffin, T. Hankemeier, N. Hardy, J. Harnly, R. Higashi, J. Kopka, A.N. Lane, J.C. Lindon, P. Marriott, A.W. Nicholls, M.D. Reily, J.J. Thaden, M.R. Viant, Metabolomics 3 (2007) 211.
[160] M.L. Bandu, T. Grubbs, M. Kater, H. Desaire, Int. J. Mass Spectrom. 251 (2006) 40.
[161] R. Tautenhahn, C. Bottcher, S. Neumann, in: S. Hochreiter, R. Wagner (Eds.), Bioinformatics Research and Development, Springer, Berlin, 2007, p. 371.
[162] C.B. Clish, E. Davidov, M. Oresic, T.N. Plasterer, G. Lavine, T. Londo, M. Meys, P. Snell, W. Stochaj, A. Adourian, X. Zhang, N. Morel, E. Neumann, E. Verheij, J.T. Vogels, L.M. Havekes, N. Afeyan, F. Regnier, J. van der Greef, S. Naylor, OMICS 8 (2004) 3.
[163] E. Lange, C. Gropl, O. Schulz-Trieglaff, A. Leinenbach, C. Huber, K. Reinert, Bioinformatics 23 (2007) i273–i281.

[165] O. Cloarec, M.E. Dumas, A. Craig, R.H. Barton, J. Trygg, J. Hudson, C. Blancher, D. Gauguier, J.C. Lindon, E. Holmes, J. Nicholson, Anal. Chem. 77 (2005) 1282.
[166] E. Werner, V. Croixmarie, T. Umbdenstock, E. Ezan, P. Chaminade, J.C. Tabet, C. Junot, Anal. Chem. 80 (2008) 4918.
[167] E. Kendrick, Anal. Chem. 35 (1963) 2146.
[168] Z. Wu, R.P. Rodgers, A.G. Marshall, Anal. Chem. 76 (2004) 2511.
[169] H. He, C.A. Conrad, C.L. Nilsson, Y. Ji, T.M. Schaub, A.G. Marshall, M.R. Emmett, Anal. Chem. 79 (2007) 8423.
[170] D.W. van Krevelen, Fuel 29 (1950) 269.
[171] S. Kim, R.W. Kramer, P.G. Hatcher, Anal. Chem. 75 (2003) 5336.
[172] J.H. Reuter, E.M. Perdue, Mitteilungen des Geologisch-Paläontologischen Instituts der Universität Hamburg 56 (1984) 249.
[173] J.I. Hedges, Adv. Chem. Ser. 225 (1990) 111.
[174] R.L. Sleighter, P.G. Hatcher, J. Mass Spectrom. 42 (2007) 559.
[175] J.C. Lindon, J.K. Nicholson, E. Holmes, H.C. Keun, A. Craig, J.T. Pearce, S.J. Bruce, N. Hardy, S.A. Sansone, H. Antti, P. Jonsson, C. Daykin, M. Navarange, R.D. Beger, E.R. Verheij, A. Amberg, D. Baunsgaard, G.H. Cantor, L. Lehman-McKeeman, M. Earll, S. Wold, E. Johansson, J.N. Haselden, K. Kramer, C. Thomas, J. Lindberg, I. Schuppe-Koistinen, I.D. Wilson, M.D. Reily, D.G. Robertson, H. Senn, A. Krotzky, S. Kochhar, J. Powell, F. van der Ouderaa, R. Plumb, H. Schaefer, M. Spraul, Nat. Biotechnol. 23 (2005) 833.
[176] R. Goodacre, D. Broadhurst, A.K. Smilde, B.S. Kristal, J.D. Baker, R. Beger, C. Bessant, S. Connor, G. Calmani, A. Craig, T. Ebbels, D.B. Kell, C. Manetti, J. Newton, G. Paternostro, R. Somorjai, M. Sjostrom, J. Trygg, F. Wulfert, Metabolomics 3 (2007) 231.
[177] N.W. Hardy, C.F. Taylor, Metabolomics 3 (2007) 243.
[178] M.R. Shortreed, S.M. Lamos, B.L. Frey, M.F. Phillips, M. Patel, P.J. Belshaw, L.M. Smith, Anal. Chem. 78 (2006) 6398.
[179] K. Guo, C. Ji, L. Li, Anal. Chem. 79 (2007) 8631.
[180] J.C. Lindon, J.K. Nicholson, I.D. Wilson, J. Chromatogr. B 748 (2000) 233.
[181] J.L. Wolfender, K. Ndjoko, K. Hostettmann, Phytochem. Anal. 12 (2001) 2.
[182] Z. Yang, J. Pharm. Biomed. Anal. 40 (2006) 516.
[183] G. Glauser, D. Guillarme, E. Grata, J. Boccard, A. Thiocone, P.A. Carrupt, J.L. Veuthey, S. Rudaz, J.L. Wolfender, J. Chromatogr. A 1180 (2008) 90.
[184] D.L. Olson, J.A. Norcross, M. O'Neil-Johnson, P.F. Molitor, D.J. Detlefsen, A.G. Wilson, T.L. Peck, Anal. Chem. 76 (2004) 2966.
[185] A.B. Kanu, P. Dwivedi, M. Tam, L. Matz, H.H. Hill Jr., J. Mass Spectrom. 43 (2008) 1.
[186] P.P. Dwivedi, P. Wu, S.J. Klopsch, G.J. Puzon, L. Xun, H.H. Hill, Metabolomics 4 (2008) 63.
[187] A. Cappiello, G. Famiglini, E. Pierini, P. Palma, H. Trufelli, Anal. Chem. 79 (2007) 5364.
[188] A. Amirav, O. Granot, J. Am. Soc. Mass Spectrom. 11 (2000) 587.
[189] O. Granot, A. Amirav, Int. J. Mass Spectrom. 244 (2005) 15.
[190] S. Orchard, W. Zhu, R.K. Julian Jr., H. Hermjakob, R. Apweiler, Proteomics 3 (2003) 2065.
[191] A. Brazma, P. Hingamp, J. Quackenbush, G. Sherlock, P. Spellman, C. Stoeckert, J. Aach, W. Ansorge, C.A. Ball, H.C. Causton, T. Gaasterland, P. Glenisson, F.C. Holstege, I.F. Kim, V. Markowitz, J.C. Matese, H. Parkinson, A. Robinson, U. Sarkans, S. Schulze-Kremer, J. Stewart, R. Taylor, J. Vilo, M. Vingron, Nat. Genet. 29 (2001) 365.
[192] R.J. Bino, R.D. Hall, O. Fiehn, J. Kopka, K. Saito, J. Draper, B.J. Nikolau, P. Mendes, U. Roessner-Tunali, M.H. Beale, R.N. Trethewey, B.M. Lange, E.S. Wurtele, L.W. Sumner, Trends Plant Sci. 9 (2004) 418.
[193] H. Jenkins, N. Hardy, M. Beckmann, J. Draper, A.R. Smith, J. Taylor, O. Fiehn, R. Goodacre, R.J. Bino, R. Hall, J. Kopka, G.A. Lane, B.M. Lange, J.R. Liu, P. Mendes, B.J. Nikolau, S.G. Oliver, N.W. Paton, S. Rhee, U. Roessner-Tunali, K. Saito, J. Smedsgaard, L.W. Sumner, T. Wang, S. Walsh, E.S. Wurtele, D.B. Kell, Nat. Biotechnol. 22 (2004) 1601.
[194] I. Spasic, W.B. Dunn, G. Velarde, A. Tseng, H. Jenkins, N. Hardy, S.G. Oliver, D.B. Kell, BMC Bioinform. 7 (2006) 281.
[195] M.E. Dumas, C. Canlet, L. Debrauwer, P. Martin, A. Paris, J. Proteome Res. 4 (2005) 1485.
[196] D.J. Crockford, E. Holmes, J.C. Lindon, R.S. Plumb, S. Zirah, S.J. Bruce, P. Rainville, C.L. Stumpf, J.K. Nicholson, Anal. Chem. 78 (2006) 363.
[197] N.C. Duarte, S.A. Becker, N. Jamshidi, I. Thiele, M.L. Mo, T.D. Vo, R. Srivas, B.O. Palsson, Proc. Natl. Acad. Sci. U.S.A. 104 (2007) 1777.
[198] P. Romero, J. Wagg, M.L. Green, D. Kaiser, M. Krummenacker, P.D. Karp, Genome Biol. 6 (2005) R2.
[199] M. Kanehisa, S. Goto, Nucleic Acids Res. 28 (2000) 27.
[200] E. Fahy, M. Sud, D. Cotter, S. Subramaniam, Nucleic Acids Res. 35 (2007) W606–W612.
[201] K. Watanabe, E. Yasugi, M. Oshima, Trends Glycosci. Glycotechnol. 12 (2000) 175.
[202] H. Horai, K. Suwa, M. Arita, Y. Nihei, T. Nishioka, Abstracts of Papers of the 55th ASMS Conference on Mass Spectrometry and Allied Topics, Indianapolis, USA, 2007.
[203] O. Yamamoto, K. Someno, N. Wasada, J. Hiraishi, K. Hayamizu, K. Tanabe, T. Tamura, M. Yanagisawa, Anal. Sci. 4 (1988) 233.