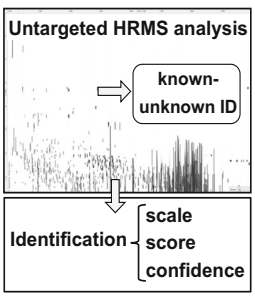
J. Am. Soc. Mass Spectrom. (2017) 28:709–723
DOI: 10.1007/s13361-016-1556-0

CRITICAL INSIGHT

Proposed Confidence Scale and ID Score in the Identification of Known-Unknown Compounds Using High Resolution MS Data

Bertrand Rochat

Centre Hospitalier Universitaire Vaudois (CHUV), University Hospital of Lausanne, 1011, Lausanne, Switzerland



Abstract. High-resolution (HR) MS instruments recording HR-full scan allow analysts to go further beyond pre-acquisition choices. Untargeted acquisition can reveal unexpected compounds or concentrations and can be performed for preliminary diagnosis attempt. Then, revealed compounds will have to be identified for interpretations. Whereas the need of reference standards is mandatory to *confirm* identification, the diverse information collected from HRMS allows identifying unknown compounds with relatively high degree of confidence without reference standards injected in the same analytical sequence. However, there is a necessity to evaluate the degree of confidence in putative identifications, possibly before further targeted analyses. This is why a confidence scale and a score in the identification of (non-peptidic) known-unknown, defined as compounds with entries in database, is proposed for (LC-) HRMS data. The scale is based on two representative documents edited by the European Commission (2007/657/EC) and the Metabolomics Standard Initiative (MSI), in an attempt to build a bridge between the communities of metabolomics and screening labs. With this confidence scale, an identification (ID) score is determined as [a number, a letter, and a number] (e.g., 2D3), from the following three criteria: I, a *General Identification Category* (1, *confirmed*, 2, *putatively identified*, 3, *annotated compounds/classes*, and 4, *unknown*); II, a *Chromatography Class* based on the relative retention time (from the narrowest tolerance, A, to no chromatographic references, D); and III, an *Identification Point Level* (1, very high, 2, high, and 3, normal level) based on the number of identification points collected. Three putative identification examples of known-unknown will be presented.

Keywords: High resolution mass spectrometry, Metabolomics, Screening, Known-unknown, Metabolites, Identification, Identification points

Received: 6 July 2016/Revised: 7 November 2016/Accepted: 11 November 2016/Published Online: 23 January 2017

Introduction

The appearance of *quantitative* and *robust* high-resolution (HR) MS instruments while recording HR-full scan, allow analysts to go further beyond pre-acquisition choices and targeted analysis [1–9]. With the same HRMS instrument, targeted and untargeted data, routine and research, quantitative or qualitative analyses can be performed and can reveal the presence, level, or fate of exogenous and endogenous compounds [10]. In routine labs using HRMS, global and untargeted data are easily acquired but data are usually treated in a targeted manner to quantify expected compounds [1–11]. However, when treated as untargeted, this large data can possibly reveal unpredicted compounds of interest in particular samples [10, 12–17].

Typically, global and untargeted (LC-)HRMS acquisition allows analysts to reveal biomarkers and toxic/illicit compounds in, respectively, metabolomics and screening labs (typically, labs involved in food safety, waste water, environmental, security, antidoping, forensic, and toxicology analyses). Even further, in clinical labs, untargeted analysis could also be applied to patients' blood samples in order to reveal *unexpected* compounds or *unexpected* levels for preliminary diagnosis attempt [18–21].

Metabolomics labs are used for untargeted workflow (see online resources, Figure S-1) [22]. Untargeted data treatment usually starts with the detection of all ions, called *features*. A feature is a *m/z* accurate value and its retention time.

Electronic supplementary material The online version of this article (doi:10.1007/s13361-016-1556-0) contains supplementary material, which is available to authorized users.

Correspondence to: Bertrand Rochat; *e-mail:* bertrand.rochat@chuv.ch

Frequently, metabolomics has to cope with unknown-unknown features that are compounds with no entries in databases. In this case, the determination of chemical composition with substructure identification (neutral losses and fragmentation trees) is the first step [23–26]. Here, the identifications of unknown-unknowns will not be considered. However, the degree of confidence in known-unknown identification will be treated carefully because it is the fundamental step prior biological interpretations.

Screening labs focus essentially on known-unknowns that are compounds with entries in databases such as *Chemspider*, *HMDB, MetLin, mzcloud* (https://www.mzcloud.org), etc. [23, 27–35]. In screening labs, reference (authentic, pure) standards are generally available and (semi)-targeted analysis are performed. But untargeted approach has already shown its potential for the discovery of unexpected xenobiotics [10–14] or to show the effects of toxins on endogenous metabolites (toxicometabolomics) [36]. From LC-HR-full scan acquisition, data mining can be time-consuming [12]. But, using a *very* stringent selection of compounds (e.g., *P*-values, fold differences in comparison to controls, etc.), a rapid and easy data treatment is possible (e.g., ≤4 h) when sample number is limited (e.g., ≤25/d) and if a limited number of unknown molecules are considered (e.g., ≤50).

In the past, total ion current (TIC) or base peak chromatograms could be used to reveal unexpected but intense ions. Today, dedicated software allows treating HRMS full scan data more in depth and comparing efficiently between- or within-samples or subjects (Figure 1 and Supplementary Materials, Figure S-1). Today, user-friendly softwares allow using such untargeted workflow routinely (Figure 1a) by lab heads or experienced technicians. This untargeted workflow would consist of a preliminary diagnosis attempt with the identification of unexpected metabolites or unexpected levels (relative quantification) and could be followed by further targeted, quantitative, and confirmative analysis.

In untargeted diagnostic analysis, the number of uncovered compounds that would have to be potentially identified are >1000 (endogenous compounds and xenobiotics). If all reference standards have to be systematically injected in the same analytical sequence, identification would be too time-consuming. In addition, some reference standards would probably be unavailable or too expensive to purchase.

In the identification of features/compounds, there are two main "visions" coming from metabolomics and screening lab communities. Whereas various recent articles can be taken into account, both visions are based on two central reference documents that have been edited in 2002 and 2007 by the European Communities, EC [37] and the Metabolomics Standard Initiative, MSI [38], respectively.

For LC-HRMS analysis, the document from the European Communities (2002/657/EC) [37] asks for (1) the relative retention time (RRT) of the analyte(s), set with a tolerance of ≤ ±2.5%, (2) relative product ion intensities (ideally, 4 MS^2 ion ratios), set with a tolerance of ±20 to ±50% with the reference MS^2 spectra, and (3) ≥4 identification points (IPs): 2 IPs for the precursor *m/z* and 2.5 IPs for a product ion determined accurately.

In the MSI document [38], there are four levels of metabolite identification: the first level (1) is a definitive identification performed with two orthogonal analyses compared with the reference standard(s) [38–43]. The second level (2) is the putative annotation (identification) of compounds and refers to the correspondence with MS data from databases or literature. The third and fourth levels (3 and 4) stand for putatively characterized compound classes and unknown compounds, respectively.

In light of HRMS data, key information and confidence level for known-unknown identification should be renewed as it has been mentioned already by some authors [5, 39, 40, 44]. In this article, we will focus on the degree of confidence in known-unknown identification using HRMS instruments [2, 45]. The vision of metabolomics and screening labs, sustained by the references mentioned above [37, 38], will be considered and merged. An identification confidence scale and an ID score will be proposed in an attempt to build a bridge between the metabolomics and screening lab communities.

The need to identify known-unknowns should grow with the increasing availability of HRMS in various routine labs, and the further implementation of global and untargeted approaches as preliminary diagnosis prior to possible further targeted analyses (Figure 1a). The aim of the proposed ID scale and score is to establish a base and a degree of confidence in known-unknown identification with an inter-laboratory and inter-community understanding. Confident identification is crucial for biological interpretations in order to stratify diseases, reveal known biomarkers, understand disease mechanisms, monitor therapeutic outcomes, and uncover acute or chronic intoxications, etc. Finally, using LC-HRMS data from plasma extracts, concrete examples of identification with ID score determination will be discussed.

Parameters to Consider for the Identification of Unknown Compounds Using HRMS Analysis

Before proposing a scale of confidence for the identification of non-peptidic known-unknown compounds (see next chapter), this is crucial to evaluate the probability of misidentification. There are various parameters influencing false positive identifications, defined here as a wrong match between a feature (*m/z*-RT item) and an attributed hit (entry) from a database. These parameters depend mainly on compound composition and isomeric possibilities, MW and mass defect, matrix and its clean-up, chromatography, and detection specificity including mass accuracy (MA) and mass resolution (R; R ≥ 20,000 at full-width-half-maximum) (see on-line resources, Figures S-2a to S-2g).

In the Supplementary Materials, Figure S-2e depicts the compound entries in *ChemSpider* per MW ranges [27]. The Gaussian bell curve shows that the majority of compounds are

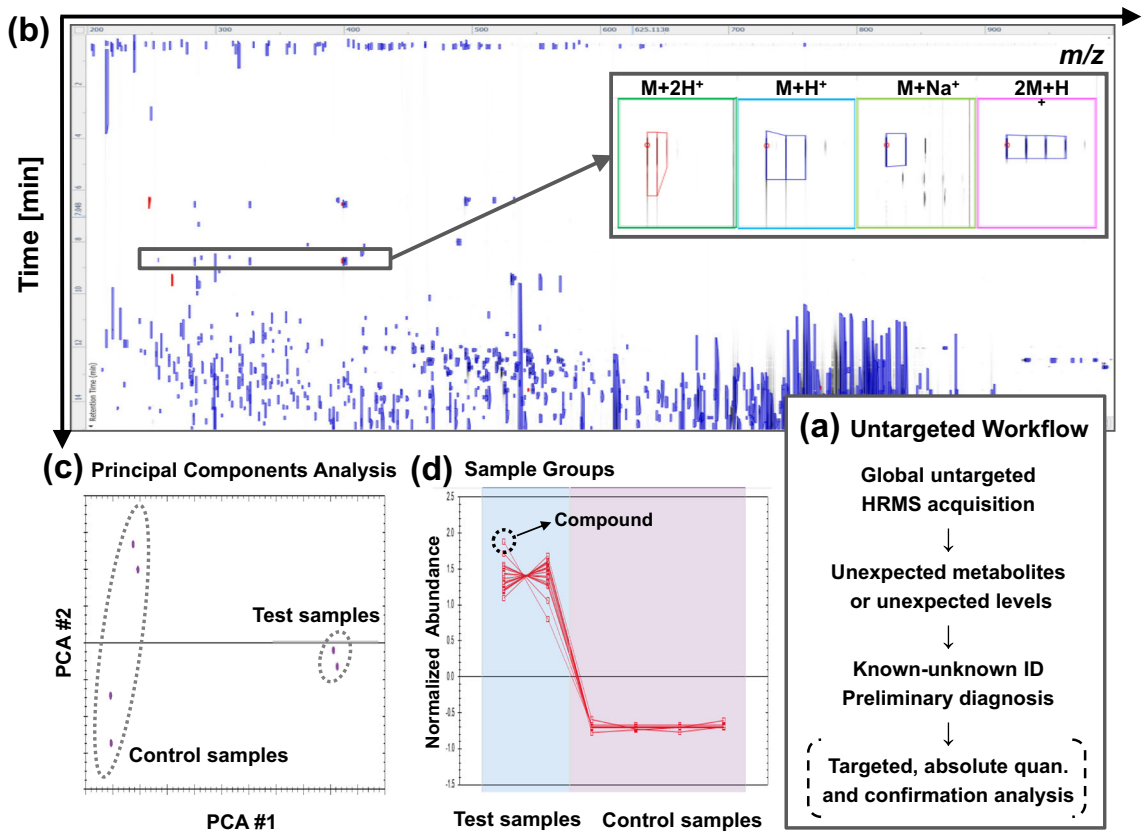


Figure 1. Data treatment and representation in untargeted analysis with HR-full scan acquisition. (**a**) Untargeted workflow in metabolomics/screening laboratories. (**b**) 2D-gel representation of the global HR-full scan (Progenesis QI software, Newcastle, UK); the software data treatment allows regrouping related adducts (right-hand box) for further processing. Principal component analysis showing differences between two groups and (**c**) main compounds involved in the group discrimination (**d**)

between 300 and 500 Da. Figure 2a shows, with a log scale, the number of hits found in *Chemspider*, with *Xcalibur* and in *HMDB* against MA tolerance (in ppm) for four $m/z_{monoisotopic}$ values, corresponding to four drugs from 150 to 500 Da. *Xcalibur* software (*Xcalibur* ver. 3.0.63; Thermo Fisher Scientific, San Diego, CA, USA), used as a chemical composition generator, was set with the following possibilities: $C_{0\text{-}30}$, $H_{0\text{-}60}$, $O_{0\text{-}15}$, $N_{0\text{-}10}$, $P_{0\text{-}3}$, $S_{0\text{-}3}$ with or without $Cl_{0\text{-}3}$, $F_{0\text{-}3}$, $Br_{0\text{-}3}$.

In general, there is a log difference in the number of hits found in the *HMDB*, a chemical composition generator, and *Chemspider* (Figure 2a). With the chemical composition generator, the number of compositions significantly increased with (1) higher MA tolerance, (2) higher $m/z_{monoisotopic}$ values, and (3) when Br, Cl, and F atoms were taken into account (e.g., synthetic xenobiotics) (Supplementary Materials, Figure S-2a and b).

Overall, these figures shows that the risk of false positive identification depends on the MW (*m/z*), MA, and the database chosen, especially if all possible isomers are considered. This risk has to consider two extremes: (1) "small size" databases (e.g., *HMDB*, *Metlin*, *KEGG*, etc. [28–31]) with $<0.5 \times 10^6$ entries, and (2) large databases (e.g., *ChemSpider*, *PubChem*, CAS Registry, etc. [33–35]) with $>50 \times 10^6$.

A low risk of false positive identification is obtained by searching in dedicated "small size" MS databases that propose molecules with a realistic probability to be observed (e.g., in a specific matrix). The high risk of false positive identification is obtained by searching in large databases from *m/z* values or postulated chemical compositions. Considering 30 different chemical compositions, the number of hits was 255 times larger in *Chemspider* in comparison to *HMDB* (Supplementary Materials, Figure S-2c). Usually, in *Chemspider*, the number of hits per chemical composition or *m/z* value is significant (>100 hits, Figure S-2c) even if most proposed molecules are unlikely in a real-world environment.

The number of most common compounds is estimated to be around 300,000 [46], which corresponds to about 0.5% of the total chemical entries in *Chemspider* (54 million in 2016). This is why the numerous possibilities found in *Chemspider* or other large databases can be drastically reduced when the number of articles or reports is used as a filter (e.g., number of links to *PubMed*). For instance, considering 30 chemical compositions in *Chemspider*, the mean number of hits per formula was 844, whereas when hits with at least one link in *PubMed* was kept (*Chemspider* display), the mean number was only 4 (Supplementary Materials, Figure S-2d). The identification of known-unknown has to cope with the right balance between these small and large databases and the realistic probability/likeliness to find the putatively identified compound in the sample extract (e.g., links to *PubMed* in *Chemspider*).

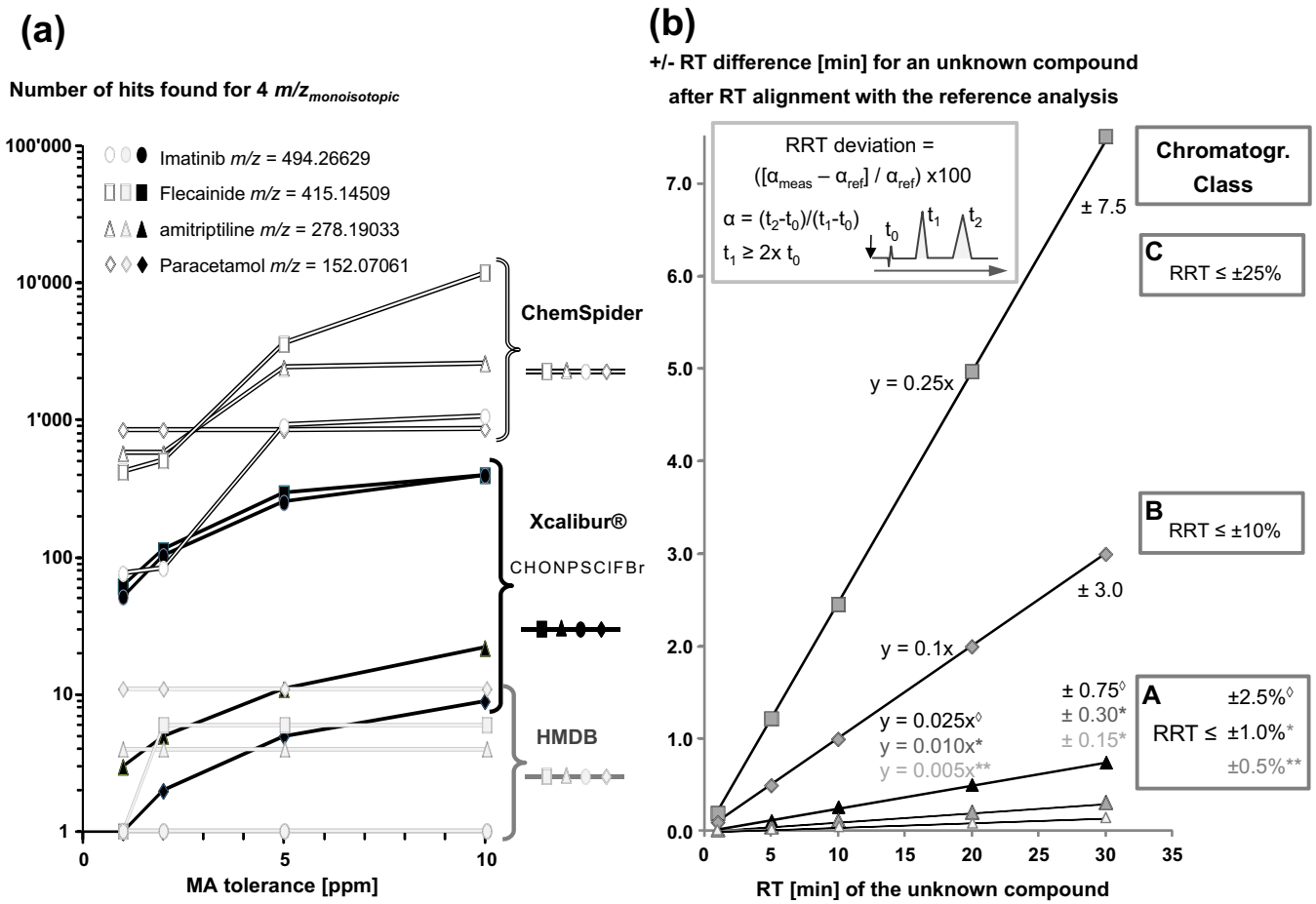


Figure 2. (**a**) Number of hits (entries) for 4 *m/z* (corresponding to four drugs) found in *Chemspider*, *Xcalibur* used as a chemical composition generator (here set with $C_{0\text{-}30}$, $H_{0\text{-}60}$, $O_{0\text{-}15}$, $N_{0\text{-}10}$, $P_{0\text{-}3}$, $S_{0\text{-}3}$, $Cl_{0\text{-}3}$, $F_{0\text{-}3}$, and $Br_{0\text{-}3}$) and *HMDB* against mass accuracy tolerance (in ppm). (**b**) Retention time difference (+/–; in min) for unknown compounds after retention time alignments (α_{meas}) with the reference analysis (α_{ref}) and for the three chromatography classes, A, B, and C, corresponding to three relative retention time deviations. Chromatography Class D has no chromatographic references. (◊), (*) and (**) are RRT tolerance values according to [37] and [51] with and without isotopic internal standards, respectively. RT of unknown compound should be >2 void volumes ($2xT_0$)

Chemical composition or accurate *m/z* determination, alone, is not sufficient for robust identification [47, 48] and this is why analysts have considered (relative) retention time, ion ratios, MS^2 ions among other parameters. One key question is about the need *or not* of reference standards for dependable identification. Whereas reference standards can be necessary for some laboratories to confirm the identification (e.g., forensic) or to discriminate isomer possibilities, reference standards are/should *not always* be necessary for reliable identification. This is especially true with the diverse information that can be collected from today's HRMS instruments. It makes sense to identify without the injection of reference standards in the same sequence, a work that can be tedious and time-consuming, especially when considering the phenotyping of 100s to 1000s of metabolites (Figure S-1) and the already ≈ 5000 commercially available compounds (food additives, drugs, endogenous or plant metabolites, and phytochemicals) [23]. The crucial question is to define the degree of confidence and this is why an identification scale and ID score considering HRMS data among other information can be useful.

Confidence Scale for Known-Unknown Compound Identification Using HRMS Analysis

In their review article [49], Lehotay and colleagues criticized the identification points system and defined it as "non-scientific." They argued that identification requires "careful deductive thinking" rather than "arbitrary requirements that need to be met by an unthinking analyst." One can agree to some extent to this point of view. Indeed, identification depends on *many* non-proportional parameters and is a case-by-case work. Thus, it appears simplistic to identify compounds with points attributed or not (1 or 0). However, a "perfect" confidence scale taking into account all parameters in small increments would be unworkable. Nevertheless, the great

advantage of a confidence scale, ID score, and IPs, is the common basis given to the wide community of analysts. This is why similar work about identification confidence has already been done in-depth for NMR experiments [50].

The proposed scale tries to be simple and opens a space for deductive comments with free attribution of some identification points (IPs) by the analyst. Today, it is all the more important to give some flexibility to analysts because unknown identification can depend on *specific* HRMS system characteristics or software packages [23, 51]. On the other hand, a common base appears necessary to reduce possible confusion originating from specific MS manufacturers' possibilities [23]. The proposed ID scale and score allow building a common basis in known-unknown identification for scientists in screening and metabolomics labs. Whereas the scale is built on the quantitative collection of IPs, the degree of flexibility given for the attribution of IPs, is rather qualitative. This flexibility opens ID scores to inter-bioanalyst variability but ID score difference should remain moderate.

Peak shape symmetry, blank injection, carry-over, background subtraction, chromatography performance and retention time alignment, signal intensity, mass accuracy calibration, signal/noise ratio, and limit of identification/detection/quantification will not be considered but can influence somehow the identification confidence. Eventually, putative identification relies on the analyst's responsibility and competence to guarantee the above parameters and the traceability of the data [49].

The proposed scale is based on the following three criteria corresponding to the first three columns in Table 1: Criterion I, Criterion II, and Criterion III.

I. **General Identification Category**: a *number* from confirmed (#1) to unknown (#4) (four *categories*: 1, 2, 3, and 4).

II. **Chromatography Class**: a *letter* from the narrowest chromatographic tolerance (A) to no chromatographic references (D) (four *classes*: A, B, C, and D)

III. **Identification Point Level**: a *number* from the maximum (#1; >15 IPs) to the minimum level depending on the number of IP attributed (#3; ≥5 IPs) (three *levels*: 1, 2, and 3).

For each of the three criteria, a grade is given allowing to determine an Identification Score (ID score) as [a number, a letter, and a number]. For instance, an unknown compound, identified in one LC-HRMS analysis, with no chromatographic reference but with 5 IPs attributed, would have an ID score defined as 2D3. For each identification score, Table 1 associates a qualitative word expression [46]. Below is the complementary information for each criterion.

I. General Identification Category

In Table 1, the first criterion/column (I) is based on the denomination given by the Metabolomics Standards Initiative (MSI) where four general levels of identification confidence have been proposed [38]. Here, these four levels are denoted as *General Identification Category*. According to the MSI and other references [38–43], these four categories stand for 1, confirmed compounds (definitively identified), 2, putatively identified compounds, 3, putatively identified classes or annotated compounds, and 4, unknown.

Category #1, *confirmed compounds*, is based on NMR analysis [50] or two orthogonal analyses with the reference standard(s) injected in the same analytical sequence and, in the proposed scale, with at least one analysis using HRMS (e.g., LC-, CE-, SFC- or GC-). Category #2, the large box of *putatively identified compounds*, is based on HRMS analysis with or without reference standards. Finally, Categories #3 and #4 concern *putatively characterized compound classes* or *annotated compounds* (possibly identified compounds but with a too low confidence), and *unknown compounds*, respectively.

Low to high identification confidence is established according to a Chromatography Class and a number of IPs attributed (column #2 and #3 in Table 1, respectively). Below are the four Chromatography Classes (II) and three IP-levels (III).

II. Chromatography Class

The second criterion is based on relative retention time (RRT) and has four classes of confidence, A to D (Table 2 and Figure 2b). Classes A, B, and C stand for identical, very similar, or close hydro/lipophilicity, respectively, whereas for Class D, there are no chromatographic references. Even if chromatographic references are not always mandatory for good identifications, it is usually inevitable when different isomers have to be discriminated. This point has to be documented by the analysts.

RRT is determined in the analytical run containing the unknown molecule(s) and in the reference run containing the reference standard(s). RRT definition is given in Figure 2b (box in the right-hand side) and determination of RRT deviation is shown in Supplementary Materials, Figure S-3. The two runs: #1, unknown(s) and chromatographic internal standard(s) (IS) and #2, authentic reference standard(s) and IS, are compared. The Class A corresponds to RRT deviation with the smallest tolerance and is based on the 2002/657/EC and WADA TD/2015/IDCR documents [37, 52]. According to 2002/657/EC, RRT tolerance should be ≤ ±2.5%, whereas for the WADA document [52], RRT tolerance is ≤ ±1% but ≤ ±0.5% when isotopic internal standards (IS) are used. The Chromatography Class A "imposes" reference standards to be injected in the same chromatographic conditions and generally within a few days. According to various documents [37, 52, 53], RT of the unknown compounds should be at least 2× the RT of the void volume, t_0 (LC system and column). When a second analysis is performed for confirmation, it should be recommended to have RT of the unknown ≥ 10× void volume RT.

Classes B and C correspond to RRT deviations (tolerance) ≤ ±10% and ≤ ±25%, respectively, which allow using previous

Table 1. Confidence Scale for Unknown Compound Identification and ID Score. The first 3 Columns Allow Determining the ID Score as [a Number, a Letter, and a Number] (e.g., 2D3) Corresponding, Respectively, to a I. *General Identification Category* (1 to 4), II. *Chromatography Class* (A to D; see Table 2) and III. *Identification Point (IPs) Level* (1 to 3). IP-Level is Attributed by Collecting a Certain Number of IPs (from 5 to 30 IPs; see Table 3)

I. General Identification Category (1 → 4)	II. Chromatography Class (A → D)	III. IPs level (1 → 3)	Std *	◊	Identification Confidence □
1 Confirmed compounds	/	1	no	(a)	utmost certainty
	A	1	yes	(b)	confirmed
		2			
2 Putatively identified compounds	A	1	yes	(c)	very strong
		2			very strong
		3			strong
	B	1	no	(c)	very strong
		2			strong
		3			good
	C	1	no	(c)	very strong
		2			strong
		3			good
	D	1	no	(c)	strong
		2			good
		3			fair
3 Annotated compounds / classes	D	/	no	(c)	tentative suspected presumptive
4 Unknown compounds	D	/	no	(c)	non-match

Examples of ID score	General Identification Category	ID Confidence
1A2	confirmed	confirmed
2C3 (see Results and Figure 4)	putitatively identified	good
2D2 (see Results and Figures 5 and 6)	putitatively identified	good

(*) Reference standards;
◊ (a) NMR: ^{1}H, ^{13}C, ^{1}H-^{1}H (COSY/NOESY), ^{13}C-^{1}H (HSQC/HMBC) [50];
(b) ≥ 2 orthogonal MS analyses, incl. ≥ 1 HRMS and ≥ 2 different chromatography;
(c), HRMS analysis (LC-, GC-, CE-, SFC-, etc.); (□), see [46]

analytical runs, in silico calculated RRT [54, 55], and chromatography databases. The lowest, Class, D, is attributed when there are no chromatographic references or RRT deviations > ±25%. Figure 2b depicts for Classes A, B, and C, the typical delta of time (min) that are accepted after alignments, with RRT tolerance ≤ ±0.5, 1, and 2.5% (A), ≤ ±10% (B), and ≤ ±25% (C) (Figure 2b and Supplementary Materials, Figure S-3a).

Table 2. Chromatography Classes, A to D, According to Relative Retention Time (RRT) Tolerance Values

Chromatography classes	RRT tolerance
A	≤ ±2.5%◊, 1%*, or 0.5%** relative to reference RRT
B	≤ ±10% relative to reference RRT
C	≤ ±25% relative to reference RRT
D	no RRT references

Narrowest tolerance is based on, respectively, (◊) [37], (*) [51] with non-isotopic internal standards, and (**) [51] with isotopic internal standards.
RRT parameters are determined with reference standard injected in the same analytical sequence (A), found in databases or calculated (B and C). Class D has no chromatographic references or a RRT tolerance >25%

Beside the fact that an accurate *m/z* or a specific chemical formula can meet many stereochemical possibilities, there is a probability associated with each isomer for its detection in a specific environment (occurrence in matrix, ubiquitous molecule or not, etc.). For instance, *Chemspider* can rank the isomers with their links to various sources, including *PubMed*. This allows giving an occurrence probability for a compound in a real situation. This information should be provided especially for Chromatography Classes B, C, and D. Moreover, this occurrence probability can be used by the analyst to allocate IPs.

III. Identification Point Levels and IPs attribution

There are three IP-levels from 1 to 3 for very high, high, and normal level, respectively. IP-levels are attributed based on the number of IPs collected for the unknown compound: 5 to 9.5,

Table 3. Attribution of Identification Points (IPs) from 6 Groups (*i* to *vi*) and IP-Levels (1 to 3) According to the Number of IPs Collected. There is a Maximum Number of IPs per Group and a Total of 30 IPs can be Collected

Group	Information	Max. IPs/group	IPs per information collected	
i	**Presursor ion** (in MS^1 or MS^2 acquisitions)	3	2	Precursor ion, if possible with identified adduct [(1)] with a MA ≤ 5 ppm and a mass resolution at *m/z* ≥ 25,000
			3	Precursor ion with identified adduct (e.g. $[M+H]^+$) [(1)] with a MA ≤ 3 ppm and mass resolution at *m/z* ≥ 50,000
ii	**MS^n Fragment ion** [(2)]	10	2.5	MS^2 product ion with precursor isolation width ≤ 5 Da, [(3)] with a MA ≤ 10 ppm and mass resolution at *m/z* ≥ 25,000
			2	MS^2 product ion with precursor isolation width ≤ 25 Da, [(4)] with a MA ≤ 10 ppm and mass resolution at *m/z* ≥ 25,000
			1	MS^2 product ion with precursor isolation width > 25 Da, [(5)] with a MA ≤ 10 ppm and mass resolution at *m/z* ≥ 25,000
			3	MS^{3-4} with precursor isolation width ≤ 5 Da, [(3)] with a MA ≤ 10 ppm and mass resolution at *m/z* ≥ 25,000
iii	**Relative ion intensity in MS^n spectra** [(6)] (ion ratio)	4	1	± 45% tolerance in *m/z* ratio in comparison to the reference MS^n spectra [(6)]
			0.5	± 90% tolerance in *m/z* ratio in comparison to the reference MS^n spectra ions considered should be ≥ 10% base peak ion
iv	**Relative Isotopic Abundance RIA**	2	1	M+1/M: RIA tolerance according to analyst's defined criteria
			1	M+1/M. M+2/M, M+n/M: RIA tolerance according to analyst's defined criteria
v	**Fine Isotopic Abundance FID**	4	1	Resolution of ^{15}N, possibly ^{33}S and ^{13}C isotopes in M+1 ions [(7)]
			1	Resolution of ^{34}S, possibly ^{18}O, and $^{13}C_2$ isotopes in M+2 ions [(7)]
			1-2	± 30% tolerance in fine isotopic ratio in comparison to theoretical ratios (fine RIA) [(7)]
vi	**Additional information**	7	1	Biological effect (macroscopic) of the unknown compound on the system
			1-2	Level correlation between the unknown and known metabolite(s) [(8)]
			1-3	Degradation/biotransformation product(s) of the unknown compound; max. 3 *m/z* [(8)]
			1	Occurrence probability of the unknown (e.g. number of links to *Pubmed*)
			1	Ratio of different MS^1 adducts relative to reference run [(9)]
			1-3	In source fragment ion in MS^1 spectra; max. 3 *m/z* [(8)]
			1-2	Information from the patient, medical staff or police/judge
			1-2	Collision cross section (ion mobility)
			1-3	Other; to be documented
	Sum = max. IPs collected	30		

IP-Levels	Number of IPs collected from at least 2 groups
1	5 to 9.5
2	10 to 15
3	> 15

[(1)] Additional adducts should be detected in order to identify the adduct considered (e.g., $[M + H^+]^+$). MA should be determined with a known compound in the analysis. MA up to 20 ppm can be tolerated for *m/z* ≤100.
[(2)] MS^n corresponds to any kind of induced fragmentation taking place after the MS entrance. MS^n fragment ions should co-elute with the precursor ion (determined in MS^n or MS^1 acquisitions). Unspecific loss should be discarded (e.g., H_2, CO_2, H_2O, and NH_3) or documented. The precursor ion in the MS^n can be used to get IPs with the rule of group *i*.
[(3)] Typically, product ion scan acquisition.
[(4)] Typically, SWATH-MS acquisition.
[(5)] Typically, MS^{ALL}, MS^{AIF} or MS^E acquisitions.
[(6)] ±45% Tolerance (relative) is based on [59].
[(7)] Typically, R should be ≥ 1.5x [(*m/z*)/(delta *m/z*)], between the 2 isotopes and the different isotopes should co-elute.
[(8)] Identified with ≥3 IPs (based on MA, MS^n ions, RIA etc.).
[(9)] In MS^1 spectra with ±20% tolerance to the reference run with pure standards.

10 to 15, and >15 IPs, respectively. Below 5 IPs, the unknown should be degraded to Category #3 as *annotated compounds*.

Table 3 lists the six different groups (from *i* to *vi*) where IPs can be collected from HRMS data. The six groups take into account (*i*) the precursor ion determination with MA, R at *m/z* ($R_{m/z}$), and if the adduct has been identified (e.g., $[M + H^+]^+$), (*ii*) the fragment ion(s) from different MS^n acquisitions, (*iii*) the relative ion intensity in MS^n spectra (ion ratio), (*iv*) the relative isotopic abundance (RIA), (*v*) the fine isotopic distribution (FID), and (*vi*) additional information. MS^n stands for any kind of induced dissociation/fragmentation taking place after the MS entrance (e.g., MS^2, MS^3, MS^E, MS^{ALL}, MS^{AIF}, SWATH-MS, vDIA, etc.). Additional information is given in Table 3 and its legend.

Each group has a maximum number of IPs that can be collected (column #3, Table 3) whereas the number of IPs per group could exceed these maxima (column #4, Table 3). For instance, the maximum number of IPs for the group (*ii*) [fragment ion(s)], is 10, whereas 5 MS^2 ions measured with a MA <5 ppm in a product scan acquisition (MS^2 with precursor isolation width ≤5 Da), could give 12.5 IPs. Figure 3 plots the maximum number of IPs per group and the number of IPs needed to obtain one of the 3 IP-levels. IPs have to be collected from at least two groups.

Details on the attribution of IPs are given in the 4th column of Table 3. Whereas the attribution of IPs is, to some extent, arbitrary, this is based on real data. For instance, in *HMDB*, there are significantly more hits corresponding to accurate *m/z* values when *all* adducts are considered rather than $[M + H^+]^+$ adduct only (mean = 8× in the positive mode; Supplementary Materials, Figure S-2g). In *HMDB*, there are 33 and 15 adducts listed for positive and negative mode, respectively. Thus, when

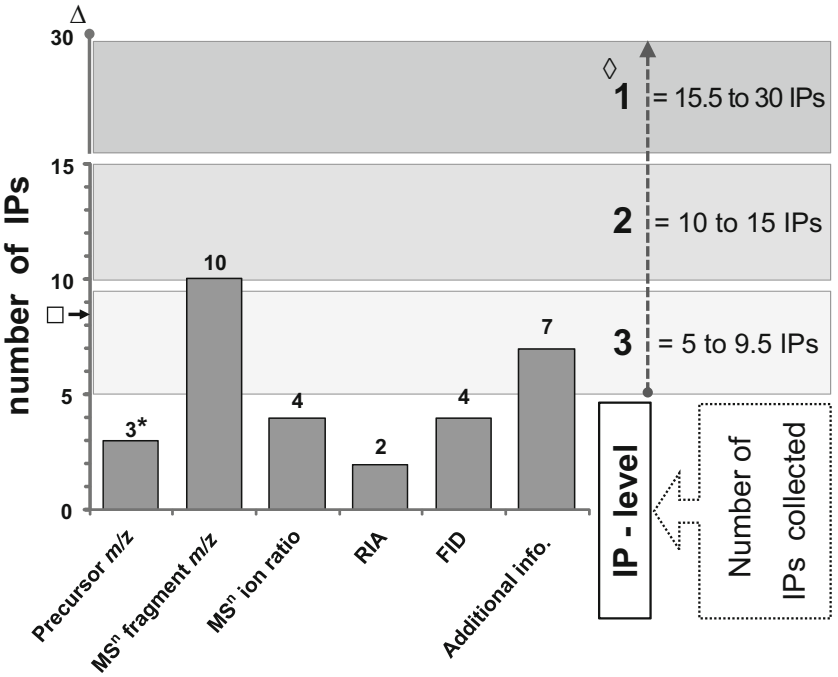


Figure 3. Number of identification points (IPs) for each of the six different groups where IPs can be collected in order to establish the IP-level. IP-level is 1, 2, or 3 (◊) for very high, high, and normal level corresponding to 5 to 9.5, 10 to 15, and >15 IPs collected, respectively. Each group has a maximum of IPs (*) and a maximum of 30 IPs (Δ) can be collected. (□) Corresponds to the minimum number of IPs collected for identification, according to 2002/657/EC recommendations (8.5 = 2 IPs for precursor *m/z*, ±2.5 IPs for one MS^2 fragment *m/z*, and 4 IPs for relative ion intensities of 4 *m/z*)

the adduct is identified (e.g., $[M + H^+]^+$) thanks to other adducts; see Figure 1b, [43]), MA ≤5 ppm and $R_{m/z}$ ≥50,000, an additional IP can be allocated.

In most HRMS analyses, the determination of precursor ion MA is a prerequisite (group *i*) and, if possible, identification of the adduct type by other observed adducts should be investigated (Figure 1b). Group *ii* can collect up to 10 IPs that is consistent with the fact that MS^n fragments (unless unspecific neutral losses) from a precursor give strong evidence in unknown identification [42]. Additional IPs must be collected from another group than *ii*. MS^n fragments should be compared with *measured* fragments found with reference standard fragmentation(s) (≠ in silico). IPs can be collected from in silico proposed MS^n fragments but in group *vi* (see below). MS^n fragments can be produced in (1) MS^2, MS^3, or MS^4, (2) SWATH-MS or vDIA [56, 57], and (3) MS^E, MS^{ALL}, MS^{AIF}[57, 58], which are all different denominations corresponding to, respectively, narrow (≤5 Da), medium (usually between 20 and 150 Da), or large (>200 Da) precursor isolation in the quadrupole. Thus, different IP numbers are attributed depending on precursor isolation width (≤5, ≤25, and >25 Da; Table 3, column #4) [37, 54]. The attribution of 2.5 IPs for a MS^2 product ion (precursor isolation width ≤5 Da) corresponds to 2002/657/EC recommendations [37].

Group *iii* consider MS^n ion ratios that are key information in GC-MS spectra with electron-impact fragmentation. In the 2002/657/EC document, tolerance (relative) for ion ratios is between ≤ ±10% and ≤ ±50%, depending on fragment ion intensities [37]. Unfortunately, in LC-HRMS, the MS^n ion ratio is very dependent on mobile phase, ion source design ESI parameters, collision cell design, fragmentation energy, etc., making of the tolerance proposed in 2002/657/EC unrealistic unless reference standards are injected with the exact same LC-HRMS conditions. More globally, and according to Mol et al. [59], ion ratio tolerance of ±45% is an excellent fit-for-purpose compromise in (LC-)HRMS analysis. Even wider tolerance can be useful to reduce false positive identification as far as additional MS^n ions are used [42]. This is why two wide tolerances in MS^n ion ratios between unknown and reference spectra have been proposed to collect IPs in (LC-)HRMS analysis: 1 IP and 0.5 IP for a ≤ ±45% and ≤ ±90% tolerance (relative). For instance, with a least/most ion ratio = 0.58, tolerance values of ±45% and ±90% would give ratio margins from 0.406 to 0.754 and 0.058 to 1.102, respectively.

Searching for the identity of a known-unknown compound, the *m/z* value is frequently used to match entries in databases. The number of hits proposed can be reduced if the chemical composition is established or at least if some chemical possibilities can be discarded. In this respect, relative isotopic abundance (RIA; group *iv*) as well as the fine isotopic distribution (FID; group *v*) give key information [47]. RIA and FID take into account M, M + 1, M + 2 etc. intensities and resolved isotopes (e.g., ^{15}N, ^{33}S, ^{34}S, ^{18}O and ^{13}C), respectively, and can be determined theoretically with appropriate software. Various RIA calculations have been proposed and results reliability can be dependent on ion intensities, HRMS technology used, $R_{m/z}$,

and the number of isotopes considered [60–66]. The use of FID is straightforward (even if isotopic abundance can vary slightly, (< ~4% [67, 68]) but relies on $R_{m/z}$ and the degree of overlap between isotopes. Therefore, in groups *iv* and *v* (RIA and FID, respectively), the number of IPs is attributed according to the analyst's documented interpretation in the capability to reduce the number of chemical compositions with these two parameters.

Figures S-3b and S-3c present typical information that can be obtained with RIA and FID. The RIA tolerance can be adapted to the number of carbons. Indeed, the discrimination between chemical compositions is more easily obtained for small molecules with <15 [C] than for compounds containing >30 [C] (Supplementary Materials, Figure S-3b). The interest of FID can be seen with the discriminations of ^{13}C, ^{2}H, ^{15}N, ^{18}O, ^{33}S, and ^{34}S isotopes at M + 1 and M + 2 of sulfamethoxazole ($C_{10}H_{11}N_3O_3S$, Supplementary Materials, Figure S-3c).

Some parameters have not been taken into account. For instance, as found in *mzcloud* database, the number of entries increases exponentially when MS^n*m/z* is <100 (Supplementary Materials, Figure S-2f). Similar results were found for GC-MS [49]. It suggests that identifications of unknown using product *m/z* ≤100 are less specific in comparison to *m/z* >100. No bonus/penalty system has been considered for the ID scale as some authors have proposed [5], but the analyst should also report contradicting information.

Finally, additional information (group *vi*, Table 3) can be used to collect up to 7 IPs. This maximum represents the second value behind group *ii* (up to 10 IPs attributed for MS^n fragment ions; Figure 3). For instance, non-analytical information from a policeman, judge, witness, clinician, or patient testimony could be considered to corroborate the analytical data. In source fragment ions [69], MS^1 adducts and their ratios, collision cross-section (measured from a drift time in ion mobility MS) [70–72], (bio)transformation products from the unknown [73], correlation between the levels of known and unknown metabolite(s) [48], and specific effects of the unknown compounds can be considered to collect IPs. In group *vi*, up to 3 IPs can be attributed in a subdivision denominated "other, to be documented" that gives some degree of flexibility to the analysts for using specific HRMS instrument characteristics, other chemical information, free or commercial tools, software and databases, etc. (e.g., collision cross-section library, in silico fragments, etc.).

From these six groups (*i*–*vi*, Table 3), a maximum of 30 IPs can be collected and three IP-levels can be determined as normal (3), high (2), and very high (1) for, respectively, 5 to 9.5, 10 to 15, and >15 IPs attributed. Below five IPs collected, compound identification should be considered as low and belong to the Category #3 (*annotated compounds*). Noteworthy, the use of intra-lab rather than inter-lab databases should be more efficient to obtain higher ID scores. Indeed, RRT, fragment ions, etc. are usually dependent on the LC-HRMS systems and conditions used.

In this proposed identification confidence scale, the lowest ID score for the identification of a known-unknown compound considering the 2002/657/EC document [37], would be 2A3 (*strong confidence*, Table 1). This relates to I, one LC-/GC-HRMS analysis (not confirmed by a second orthogonal analysis), II, Chromatography Class A for the narrowest RRT tolerance possible, and III, 8.5 IPs for the accurate *m/z* determinations of the precursor (2 IPs) and one MS^2 fragment ion (2 IPs) and for four relative ion MS^2 intensities (4 IPs).

Today, good or strong identification confidence of known-unknown compounds can be obtained from HRMS data without the need to have RRT ≤ ±2.5% and reference standard(s) injected in the same sequence (Chromatography Class A). Indeed, many IPs can be attributed allowing to accept a larger tolerance in chromatography constraints. The following Concrete examples will demonstrate this point.

Concrete Examples and Discussion

Theoretical *m/z* values, RIA, and FID have been calculated with *Xcalibur* software (Thermo Fisher Scientific). The *first example* is the putative identification of imatinib, an anticancer drug, in a human plasma extract by LC-HRMS analysis recording HR-full scan (no MS^n). This unique analysis (Criterion I, Category #2 or further) was compared with a previous analysis performed 2 years earlier with roughly the similar LC column (C18) and conditions (gradient with MeCN:H_2O with formic acid). The previous analysis was considered as the reference because pure imatinib standard was spiked and quantified in plasma.

RRT values (Criterion II) were calculated using an endogenous metabolite known to be abundant in plasma, LysoPC(18:2(9Z,12Z), which showed the same profile in both chromatograms (Figure 4a). RRT difference was –20.4% allowing to allocate the Chromatography Class C in the ID score. In this first example, 7 IPs were attributed (Criterion III, IP-level 3; Figure 4a-e). Table 3 and Figure 3 show how IP-level is attributed in relation to the number of IPs collected. In group *i*, 3 IPs were collected for the precursor ion identified as the $[M + H^+]^+$ because other co-eluting adducts were detected (Na^+, K^+, and $2H^+$; same RT) and because MA was ≤3 ppm and R ≥50,000 (Figure 4a). In group *iii,* 1 IP was obtained for RIA at M + 1 and M + 2 that matched imatinib RIA_{theor} with –8% and –17%, respectively (Figure 4b). In group *vi*, 3 IPs were collected as follows: 2 IPs for the detection of 2 known MS^2 fragment ions but detected in the MS^1 scan (in source fragmentation; MA <5 ppm, same RT as imatinib, Figure 4d) and 1 IP for S, Cl, and Si removals in proposed chemical formulae given by *Chemspider* (FID incompatibility) (Figure 4c) combined with the much higher number of references in *PubMed* for imatinib than further hits (according to *Chemspider*) (Figure 4e). Then an ID score of 2C3 (7 IPs) was established (*good evidence* in Table 1). More stringent acceptance on RIA errors could have been applied: RIA at A + 1 ≤ 5% (according to Knolhoff et al. [66]). This would have

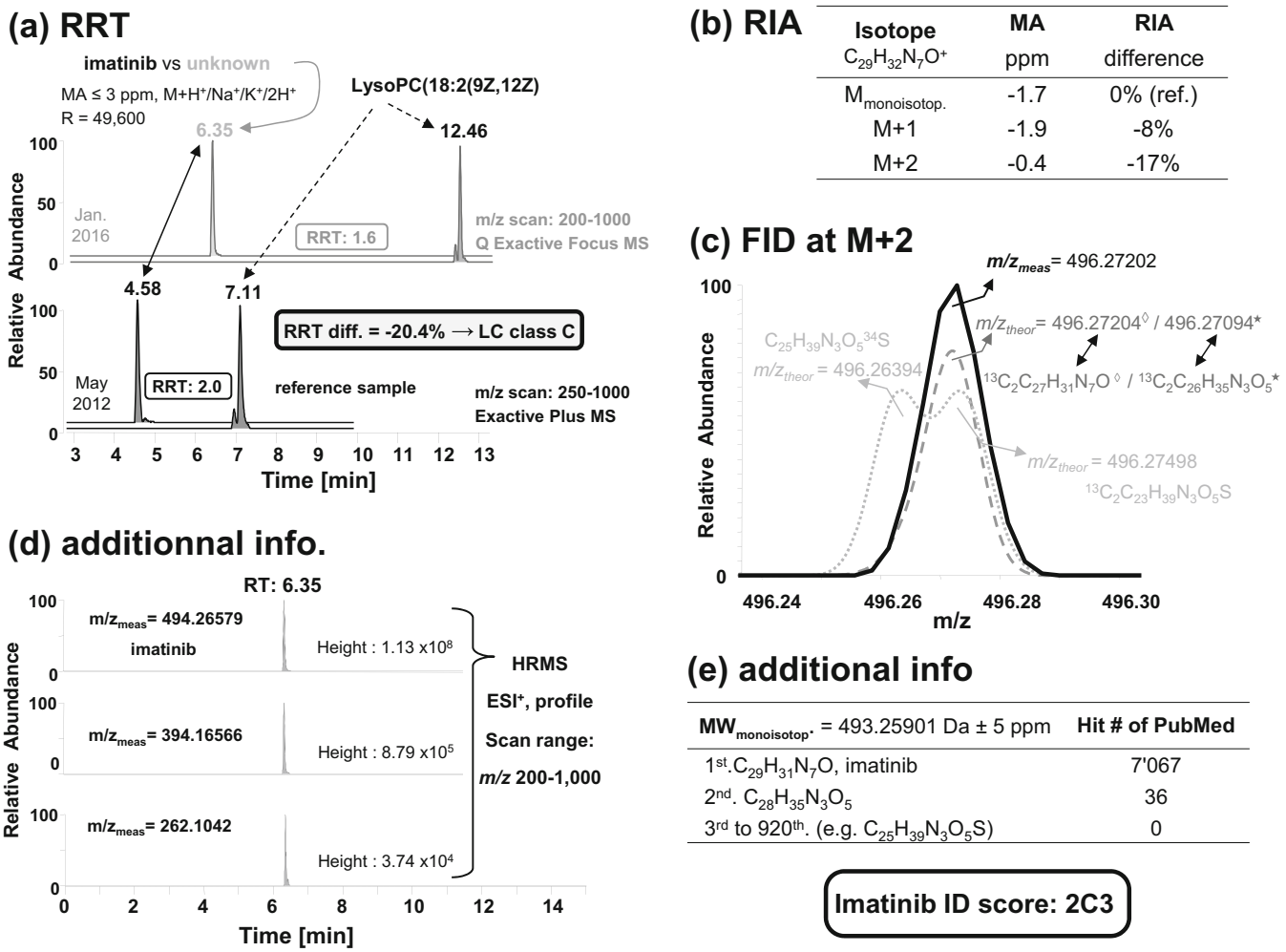
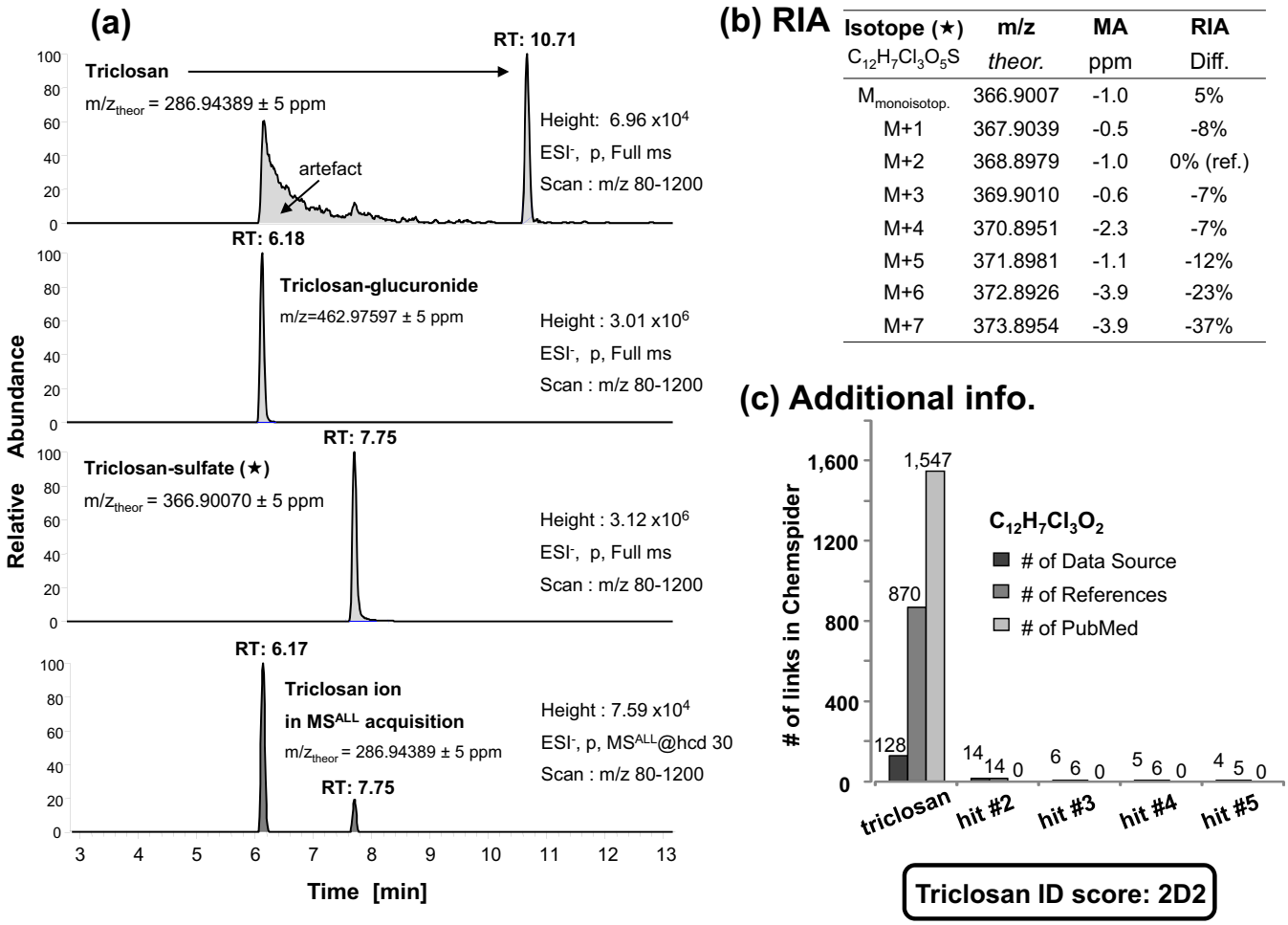


Figure 4. Putative identification of imatinib found in a human plasma extract by LC-HRMS analysis. Seven IPs were collected (IP-level 3; see main text) establishing the ID score for imatinib identification at 2C3 (*good evidence*). (**a**) After the comparison with a similar analysis performed 4 y before and containing imatinib pure standard, the chromatography class was determined to be C (RRT with reference sample was ≤25%; IS for RRT calculation was lysoPC(18:2(9Z,12Z), m/z_{theor} = 520.33977 extracted with ±5 ppm extraction window). Imatinib *m/z* was observed with H^+, Na^+, K^+, and $2H^+$ adducts with MA ≤3 ppm and at R ≥50,000 (3 IPs). (**b**) Relative isotopic abundance (RIA; 1 IP) and (**c**) fine isotopic distribution (FID) allow discarding a few chemical composition possibilities (e.g., C ≈ 30, no Si, no Cl, no Br, no S). (**d**) Extracted ion chromatograms from a LC-HR full scan acquisition, depict imatinib and two known imatinib MS^2 ions that were detected as in-source fragments (2 IPs). (**e**) In combination with FID, only $C_{29}H_{31}N_7O$ corresponding to imatinib is the most credible hit in large database (1 IP) because *Chemspider* shows >7000 links with *PubMed* whereas further possibilities show <40 hits

removed the attribution of 1 IP based on RIA even if the ID score had not changed. This underlines that the flexibility given to the analyst could possibly change the number of IP collected and the ID score but differences should remain moderate.

The *second example* shows the putative identification of triclosan, an antibacterial and antifungal agent, found in a human plasma extract. Triclosan was detected in only one individual, whereas 50 anonymous patients were analyzed by LC-HRMS performing two simultaneous acquisitions: HR-full scan and MS^{ALL} (MS/MS with no precursor selection). No additional analyses were performed (Criterion I, Category #2 or further) and no RRT comparison with a reference sample was possible (Criterion II, Chromatography Class D). From the HRMS data, 10 IPs were collected allowing determination of an ID score of 2D2 (*good evidence* in Table 1).

Figure 5 depicts this identification and the number of IPs collected. In Group *i* (see Table 3), 2 IPs were collected for the precursor ion (not formally identified as $[M - H^+]^-$ by other adducts). Whereas no IPs were attributed for MS^n fragments (no fragment ions found in the MS^{ALL} scans), RIA allowed the collection of 2 IPs (Group *ii*). In fact, the chemical composition of triclosan, $C_{12}H_7Cl_3O_2$, shows a specific isotopic pattern (see spectra and XIC in Figure S-3d), whereas the low levels detected gave relatively poor RIA difference (up to 37% for lowest abundance ions) between measured and theoretical RIA ($R_{m/z}$ =

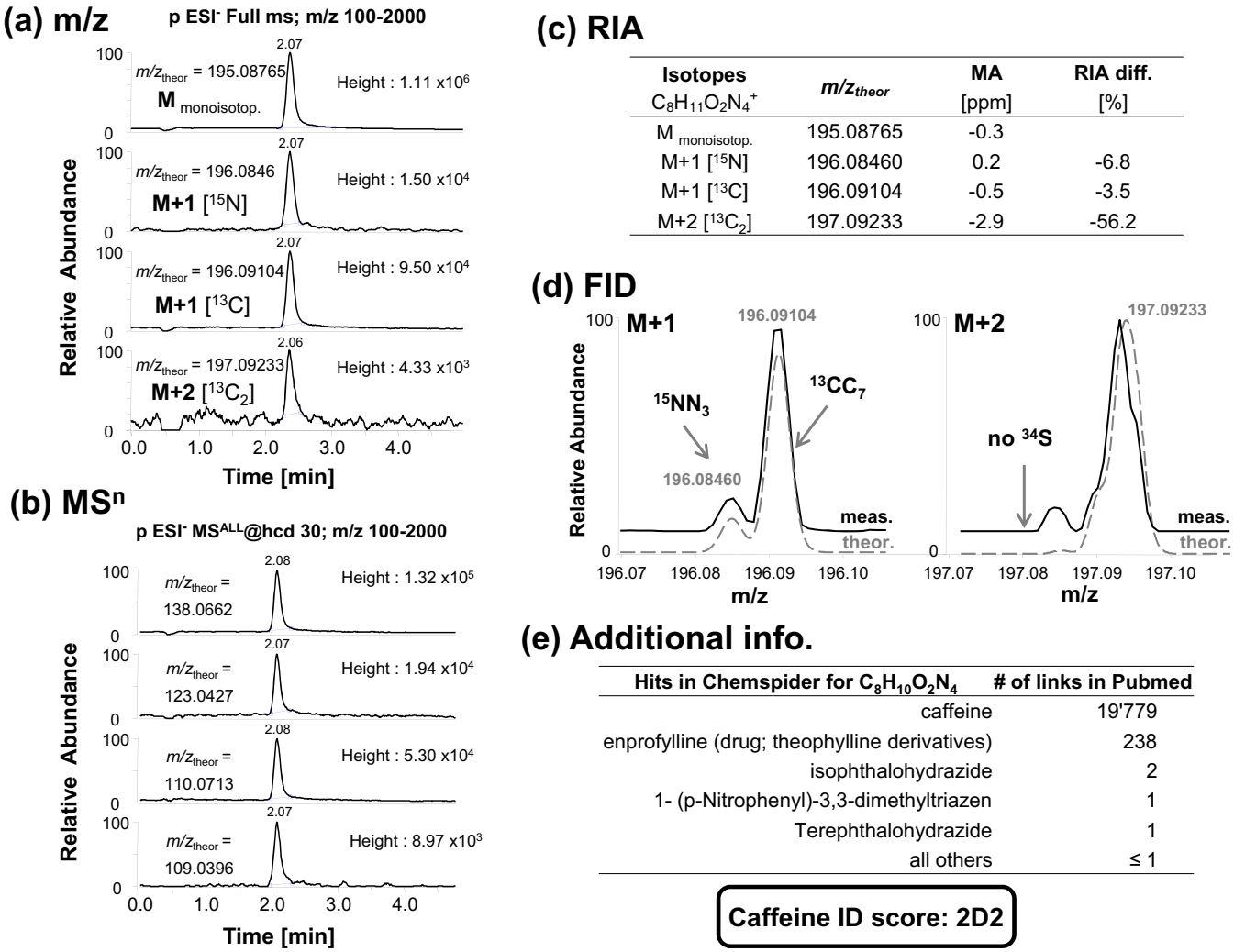


Isotope (★) C$_{12}$H$_7$Cl$_3$O$_5$S	m/z *theor.*	MA ppm	RIA Diff.
M$_{monoisotop.}$	366.9007	-1.0	5%
M+1	367.9039	-0.5	-8%
M+2	368.8979	-1.0	0% (ref.)
M+3	369.9010	-0.6	-7%
M+4	370.8951	-2.3	-7%
M+5	371.8981	-1.1	-12%
M+6	372.8926	-3.9	-23%
M+7	373.8954	-3.9	-37%

Figure 5. Putative identification of triclosan in a human plasma extract by LC-HRMS analysis. Ten IPs were collected and corresponding to IP-level 2 (see main text) establishing the ID score for triclosan identification at 2D2 (*good evidence*). (**a**) Extracted ion chromatograms (HR full scan) depict triclosan (two IPs) and two known triclosan metabolites (triclosan-glucuronide and triclosan-sulfate with 50× more peak intensities (2x1 IPs, $[M - H^+]^-$; MA <5 ppm; not found in other plasma extracts). In the bottom chromatogram, a MS^{ALL} fragment ion corresponding to triclosan *m/z* can be observed at the two metabolite RT, 1 IP). (**b**) Measured relative isotopic abundance (RIA) of triclosan-sulfate, detected here up to M + 7 (2 IPs; see spectra and XIC in Figure S-3d). MA and RIA differences between theoretical and measured values ($R_{m/z}$ = 54,000) are depicted and allow establishing the chemical composition, $C_{12}H_7Cl_3O_5S$; see spectrum in Figure S-3. Similar Cl_3-specific RIA patterns were found for triclosan-sulfate and triclosan-glucuronide and, therefore, two additional IPs were attributed. (**c**) In *Chemspider*, triclosan is the most credible compound for $C_{12}H_7Cl_3O_2$ because it has a lot more links to various Sources than the other hits (1 IP)

54,000). In Group *vi*, 6 IPs were obtained in the additional information as follows. Two IPs were collected from the detection of two well-known plasma metabolites of triclosan: triclosan-glucuronide, $[C_{18}H_{15}Cl_3O_8\text{-}H^+]^-$, and triclosan-sulfate, $[C_{12}H_7Cl_3O_5S\text{-}H^+]^-$. One IP was given for a MS^{ALL} fragment ion corresponding to triclosan *m/z* but detected at the RT of the two triclosan metabolites (see Figure 5a), 2 IP were attributed for the Cl_3 specific RIA of triclosan and its two metabolites (up to M + 7 for triclosan-sulfate, see Figure 5b and Supplementary Materials, Figure S-3d) and 1 IP was collected for the very low number of other credible hit(s) found in *Chemspider* and *HMDB* database. Searching with the unknown accurate *m/z*, only triclosan was returned from *HMDB* (all adducts in negative mode, MA ≤5 ppm) and considering $C_{12}H_7Cl_3O_2$, only triclosan shows references in *PubMed*, according to *Chemspider* (Figure 5c). This is in agreement with the ubiquitous use of triclosan in many consumer products in contrast to the other known isomers.

The *third and last example* shows the putative identification of caffeine in a human plasma extract. The ID score was 2D2 (*good evidence* in Table 1), which relates to one LC-HRMS analysis, performed with HR-full scan and MS^{ALL} acquisitions (Criterion I, Category #2), no chromatographic references (Criterion II, Chromatography Class D), and 10 IPs attributed (Criterion III, IP-levels 2). Two IPs were obtained from the precursor ion (MA <5 ppm; Figure 6a), 4 IPs were attributed from 4 known fragment ions of caffeine (see https://www.mzcloud.org or [74]), (1 IP per ion detected in MS^{ALL} scans; Figure 6b), 1 IP was attributed from RIA and 2 IPs were given



Isotopes $C_8H_{11}O_2N_4^+$	m/z_{theor}	MA [ppm]	RIA diff. [%]
M monoisotop.	195.08765	-0.3	
M+1 [^{15}N]	196.08460	0.2	-6.8
M+1 [^{13}C]	196.09104	-0.5	-3.5
M+2 [$^{13}C_2$]	197.09233	-2.9	-56.2

Hits in Chemspider for $C_8H_{10}O_2N_4$	# of links in Pubmed
caffeine	19'779
enprofylline (drug; theophylline derivatives)	238
isophthalohydrazide	2
1- (p-Nitrophenyl)-3,3-dimethyltriazen	1
Terephthalohydrazide	1
all others	≤ 1

Figure 6. Putative identification of caffeine ($[C_8H_{11}O_2N_4]^+$) in a human plasma extract by LC-HRMS analysis. Ten IPs were collected (IP-level 2; see main text) establishing the ID score for caffeine identification at 2D2 (*good evidence*). (**a**) Extracted ion chromatograms (m/z_{theor} ±5 ppm) depict the unknown compound (2 IPs; adduct not formally identified with other adducts) and three of its isotopes, Display (**b**) depicts 4 MS^{ALL} fragments known to be MS^2 ions of caffeine (4x1 IPs; see https://www.mzcloud.org or [72]). Display (**c**) shows MA (ppm) and RIA difference (%; 1 IP) of caffeine isotopes between measured and theoretical values ($R_{m/z}$ = 70,000). Spectra in (**d**) show the FID ([2] for consistency IPs) and the interest to resolve fine isotopes for chemical composition determination. Here, $[^{13}CC_7H_{11}O_2N_4]^+$ and $[C_8H_{11}O_2{}^{15}NN_3]^+$, corresponding to M + 1 isotopes, are resolved, whereas, at M + 2, no ^{34}S such as $[C_8H_{19}O^{32}S^{34}S]^+$ at m/z = 197.0829 is observed. The fine discrimination of m/z isotopes together with the ion intensity (RIA) allow establishing that the unknown composition has N_4, C_8, and S_0 and is eventually $C_8H_{10}O_2N_4$. (**e**) In *Chemspider*, caffeine is the most credible hit because the unknown was found in many plasma extracts and has >80× more links to *PubMed* in comparison to further hits (2 IPs)

by FID (Figure 6a and b). The last IP was attributed from the following analyst's deduction: from the spectra at M + 1 and M + 2 (FID and RIA), a high probability that the unknown chemical composition contains N_4, C_8, and S_0 was established (Figure 6). From *Xcalibur* used as formula generator, the only realistic hit for the accurate m/z measured was $C_8H_{11}O_2N_4^+$. In *Chemspider*, although more than 600 isomers have this formula, only caffeine and enprofylline have a significant number of links to *PubMed,* respectively, 19,779 and 239. The presence of the unknown compound in ≈ 30% of analyzed plasma adds evidence for the identification of caffeine rather than enprofylline or other isomers.

Concluding Remarks

With the increasing availability of HRMS in routine labs, new investigations will be realized using full scan or data-independent acquisition. The untargeted data treatment will reveal unexpected compounds of interest (pollutants, drugs, etc.) or unexpected concentrations. These revealed compounds

will have to be identified for pertinent biological interpretations [preliminary diagnosis] and possible further targeted analyses. The injection of reference standards in the same analytical sequence and possibly in two orthogonal analyses can be mandatory for identifications with utmost certainty. But this is too time-consuming to inject 100s to 1000s of reference standards in the same analytical sequence for known-unknown identification. This is why an ID scale and score has been proposed to establish a base and a degree of confidence in known-unknown identification. Indeed, fair, good, or strong confidence can be established without the injection of reference standards when various identification points (IPs) can be collected from HRMS data. The three presented examples of identification have shown this point.

Based on two key references for screening and metabolomics labs [37, 38], a confidence scale and an ID score have been proposed for the rating of (non-peptidic) known-unknown identification in (LC-)HRMS analysis (Tables 1, 2 and 3) and are based on three criteria: I, a general identification category (1 to 4); II, a chromatography class (A to D); and III, an identification point level (1 to 3) attributed from the number of IPs collected. Therefore, the ID score is based on these three criteria and given as [a number, a letter, and a number].

This scale and ID score are new proposals and logically could/should be modified and improved. The purpose of these proposals is also an attempt to build a bridge between the metabolomics and screening lab communities. It is a step forward in the need to rate the confidence of *putative* identifications of known-unknown compounds with or without the availability of reference standards. The need to identify known-unknowns should grow with the increasing availability of HRMS in various routine labs and the further implementation of global and untargeted approaches to reveal biomarkers, monitor therapeutic outcomes, uncover acute or chronic intoxications, etc. The proposed ID scale, score, and confidence provide a base to these untargeted analyses.

Acknowledgments

Although this article has been written by a single author, it is logically the result of interactive discussions with many colleagues and friends. They would be too numerous to mention all. I would like to express my profound gratitude to them. On the other hand, I also thank Drs. Anne Boddecs, Anne Cissencco, and Piotr Rouphail for their unwavering support.

References

1. Ramanathan, R., Jemal, M., Ramagiri, S., Humpreys, W.G., Olah, T., Korfmacher, W.A.: It is time for a paradigm shift in drug discovery bioanalysis: from SRM to HRMS. J. Mass Spectrom. **46**, 595–601 (2011)
2. Rochat, B.: From targeted quantification to untargeted metabolomics: why LC-high-resolution-MS will become a key instrument in clinical labs. Trends Anal. Chem. *Advance On-line publication* (2016)
3. Rochat, B., Kottelat, E., McMullen, J.: The future key role of LC-high-resolution-MS analyses in clinical laboratories: a focus on quantification. Bioanalysis **4**, 2939–2958 (2012)
4. Kaufmann, A., Butcher, P., Maden, K., Walker, S., Widmer, M.: Quantitative and confirmative performance of liquid chromatography coupled to high-resolution mass spectrometry compared to tandem mass spectrometry. Rapid Commun. Mass Spectrom. **25**, 979–992 (2011)
5. Kaufmann, A., Butcher, P., Maden, K., Walker, S., Widmer, M.: Reliability of veterinary drug residue confirmation: high resolution mass spectrometry versus tandem mass spectrometry. Anal. Chim. Acta **856**, 54–67 (2015)
6. Meyer, M.R., Helfer, A.G., Maurer, H.H.: Current position of high-resolution MS for drug quantification in clinical and forensic toxicology. Bioanalysis **6**, 2275–2284 (2014)
7. Henry, H., Sobhi, H.R., Scheibner, O., Bromirski, M., Nimkar, S.B., Rochat, B.: Comparison between a high-resolution single-stage Orbitrap and a triple quadrupole mass spectrometer for quantitative analyses of drugs. Rapid Commun. Mass Spectrom. **26**, 499–509 (2012)
8. Fedorova, G., Randak, T., Lindberg, R.H., Grabic, R.: Comparison of the quantitative performance of a Q-Exactive high-resolution mass spectrometer with that of a triple quadrupole tandem mass spectrometer for the analysis of illicit drugs in wastewater. Rapid Commun. Mass Spectrom. **27**, 1751–1762 (2013)
9. Herrero, P., Cortés-Francisco, N., Borrull, F., Caixach, J., Pocurull, E., Marcé, R.M.: Comparison of triple quadrupole mass spectrometry and Orbitrap high-resolution mass spectrometry in ultrahigh performance liquid chromatography for the determination of veterinary drugs in sewage: benefits and drawbacks. J. Mass Spectrom. **49**, 585–596 (2014)
10. Rochat, B., Bilat, J., Grund, B.: A close look at the fate of compounds we are exposed to. Chimia **68**, 818 (2014)
11. Bruce, S.J., Rochat, B., Béguin, A., Pesse, B., Guessous, I., Boulat, O., Henry, H.: Analysis and quantification of vitamin D metabolites in serum by ultra-performance liquid chromatography coupled to tandem mass spectrometry and high-resolution mass spectrometry—a method comparison and validation. Rapid Commun. Mass Spectrom. **27**, 200–206 (2013)
12. Schymanski, E.L., Singer, H.P., Slobodnik, J., Ipolyi, I.M., Oswald, P., Krauss, M., Schulze, T., Haglund, P., Letzel, T., Grosse, S., Thomaidis, N.S., Bletsou, A., Zwiener, C., Ibáñez, M., Portolés, T., de Boer, R., Reid, M.J., Onghena, M., Kunkel, U., Schulz, W., Guillon, A., Noyon, N., Leroy, G., Bados, P., Bogialli, S., Stipaničev, D., Rostkowski, P., Hollender, J.: Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis. Anal. Bioanal. Chem. **407**, 6237–6255 (2015)
13. Zomer, P., Mol, H.G.: Simultaneous quantitative determination, identification and qualitative screening of pesticides in fruits and vegetables using LC-Q-Orbitrap-MS. Food Addit. Contam. Part A **32**, 1628–1636 (2015)
14. Thoren, K.L., Colby, J.M., Shugarts, S.B., Wu, A.H., Lynch, K.L.: Comparison of information-dependent acquisition on a tandem quadrupole TOF vs a triple quadrupole linear ion trap mass spectrometer for broad-spectrum drug screening. Clin. Chem. **62**, 170–178 (2016)
15. Burish, M.J., Thoren, K.L., Madou, M., Toossi, S., Shah, M.: Hallucinogens causing seizures? A case report of the synthetic amphetamine 2,5-dimethoxy-4-chloroamphetamine. Neurohospitalist **5**, 32–34 (2015)
16. Aceña, J., Stampachiacchiere, S., Pérez, S., Barceló, D.: Advances in liquid chromatography-high-resolution mass spectrometry for quantitative and qualitative environmental analysis. Anal. Bioanal. Chem. **407**, 6289–6299 (2015)
17. Wu, A.H., Colby, J.: High-resolution mass spectrometry for untargeted drug screening. Methods Mol. Biol. **1383**, 153–166 (2016)
18. Rochat, B.: Is there a future for metabotyping in clinical laboratories? Bioanalysis **7**, 5–8 (2015)
19. Clish, C.B.: Metabolomics: an emerging but powerful tool for precision medicine. Cold Spring Harbor Mol. Case Study **1**, a000588 (2015)
20. Gertsman, I., Gangoiti, J.A., Barshop, B.A.: Validation of a dual LC-HRMS platform for clinical metabolic diagnosis in serum, bridging quantitative analysis and untargeted metabolomics. Metabolomics **10**, 312–323 (2014)
21. Wishart, D.S.: Emerging applications of metabolomics in drug discovery and precision medicine. Nat. Rev. Drug Discov. *Advance On-line publication* (2016)
22. Vuckovic, D.: Current trends and challenges in sample preparation for global metabolomics using liquid chromatography-mass spectrometry. Anal. Bioanal. Chem. **403**, 1523–1548 (2012)
23. Wishart, D.S.: Advances in metabolite identification. Bioanalysis **3**, 1769–1782 (2011)

24. Watson, D.G.: A rough guide to metabolite identification using high resolution liquid chromatography mass spectrometry in metabolomic profiling in metazoans. Comput. Struct. Biotechnol. J. **4**, e201301005 (2013)
25. Kind, T., Fiehn, O.: Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinformatics **8**, 105–125 (2007)
26. Kasper, P.T., Rojas-Chertó, M., Mistrik, R., Reijmers, T., Hankemeier, T., Vreeken, R.J.: Fragmentation trees for the structural characterization of metabolites. Rapid Commun. Mass Spectrom. **26**, 2275–2286 (2012)
27. Little, J.L., Williams, A.J., Pshenichnov, A., Tkachenko, V.: Identification of "known-unknowns" utilizing accurate mass data and ChemSpider. J. Am. Soc. Mass Spectrom. **23**, 179–185 (2012)
28. Wishart, D.S., Jewison, T., Guo, A.C., Wilson, M., Knox, C., Liu, Y., Djoumbou, Y., Mandal, R., Aziat, F., Dong, E., Bouatra, S., Sinelnikov, I., Arndt, D., Xia, J., Liu, P., Yallou, F., Bjorndahl, T., Perez-Pineiro, R., Eisner, R., Allen, F., Neveu, V., Greiner, R., Scalbert, A.: HMDB 3.0–The human metabolome database in 2013. Nucleic Acids Res. **41**, D801–D807 (2013).
29. Smith, C.A., O'Maille, G., Want, E.J., Qin, C., Trauger, S.A., Brandon, T.R., Custodio, D.E., Abagyan, R., Siuzdak, G.: METLIN: a metabolite mass spectral database. Ther. Drug Monit. **27**, 747–751 (2005).
30. Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K.F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., Hirakawa, M.: From genomics to chemical genomics: new developments in KEGG. Nucleic Acids Res. **34**, D354–D357 (2006).
31. Degtyarenko, K., de Matos, P., Ennis, M., Hastings, J., Zbinden, M., McNaught, A., Alcántara, R., Darsow, M., Guedj, M., Ashburner, M.: ChEBI: a database and ontology for chemical entities of biological interest. Nucleic Acids Res. **36**, D344–D350 (2008).
32. Fahy, E., Subramaniam, S., Murphy, R.C., Nishijima, M., Raetz, C.R., Shimizu, T., Spener, F., van Meer, G., Wakelam, M.J., Dennis, E.A.: Update of the LIPID MAPS comprehensive classification system for lipids. J. Lipid Res. **50**(Suppl), S9–S14 (2009).
33. Kelly, R., Kidd, R.: Editorial: ChemSpider—a tool for natural products research. Nat. Prod. Rep. **32**, 1163–1164 (2015).
34. Little, J.L., Cleven, C.D., Brown, S.D.: Identification of "known-unknowns" utilizing accurate mass data and chemical abstracts service databases. J. Am. Soc. Mass Spectrom. **22**, 348–359 (2011).
35. Kim, S., Thiessen, P.A., Bolton, E.E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B.A., Wang, J., Yu, B., Zhang, J., Bryant, S.H.: PubChem substance and compound databases. Nucleic Acids Res. **44**, D1202–D1213 (2016).
36. Bouhifd, M., Hartung, T., Hogberg, H.T., Kleensang, A., Zhao, L.: Review: toxicometabolomics. J. Appl. Toxicol. **33**, 1365–1383 (2013)
37. Commission Decision 2002/657/EC implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results. http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32002D0657&from=EN
38. Sumner, L.W., Amberg, A., Barrett, D., Beale, M.H., Beger, R., Daykin, C.A., Fan, T.W., Fiehn, O., Goodacre, R., Griffin, J.L., Hankemeier, T., Hardy, N., Harnly, J., Higashi, R., Kopka, J., Lane, A.N., Lindon, J.C., Marriott, P., Nicholls, A.W., Reily, M.D., Thaden, J.J., Viant, M.R.: Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). Metabolomics **3**, 211–221 (2007)
39. Schymanski, E.L., Jeon, J., Gulde, R., Fenner, K., Ruff, M., Singer, H.P., Hollender, J.: Identifying small molecules via high resolution mass spectrometry: communicating confidence. Environ. Sci. Technol. **48**, 2097–2098 (2014)
40. Nielen, M.W., van Engelen, M.C., Zuiderent, R., Ramaker, R.: Screening and confirmation criteria for hormone residue analysis using liquid chromatography accurate mass time-of-flight, fourier transform ion cyclotron resonance and Orbitrap mass spectrometry techniques. Anal. Chim. Acta **586**, 122–129 (2007)
41. Vanhaecke, L., Gowlk, P., Le Bizec, B., Van Ginkel, L., Bichon, E., Blokland, M., De Brabander, H.F.: European analytical criteria: past, present, and future. J. AOAC Int. **94**, 360–372 (2011)
42. Stein, S.E., Heller, D.N.: On the risk of false positive identification using multiple ion monitoring in qualitative mass spectrometry: large-scale intercomparisons with a comprehensive mass spectral library. J. Am. Soc. Mass Spectrom. **17**, 823–835 (2006)
43. Dunn, W.B., Erban, A., Weber, R.J.M., Creek, D.J., Brown, M., Breitling, R., Hankemeier, T., Goodacre, R., Neumann, S., Kopka, J., Viant, M.R.: Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics. Metabolomics **9**, 44–66 (2013)
44. Salek, R.M., Steinbeck, C., Viant, M.R., Goodacre, R., Dunn, W.B.: The role of reporting standards for metabolite annotation and identification in metabolomic studies. Gigascience **2**, 13 (2013)
45. Junot, C., Fenaille, F., Colsch, B., Bécher, F.: High resolution mass spectrometry based techniques at the crossroads of metabolic pathways. Mass Spectrom. Rev. **33**, 471–500 (2014)
46. Milman, B.L.: Identification of chemical compounds. Trends Analyt. Chem. **24**, 493–508 (2014)
47. Kind, T., Fiehn, O.: Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm. BMC Bioinformatics **7**, 234 (2006)
48. Weber, R.J.M., Viant, M.R.: MI-pack: increased confidence of metabolite identification in mass spectra by integrating accurate masses and metabolic pathways. Chemometrics Intelligent Lab. Syst. **104**, 75–82 (2010)
49. Lehotay, S.J., Mastovska, K., Amirav, A., Fialkov, A.B., Alon, T., Martos, P.A., de Kok, A., Fernández-Alba, A.R.: Identification and confirmation of chemical residues in food by chromatography-mass spectrometry and other techniques. Trends Anal. Chem. **11**, 1070–1090 (2008)
50. Dona, A.C., Kyriakides, M., Scott, F., Shephard, E.A., Varshavi, D., Veselkov, K., Everett, J.R.: A guide to the identification of metabolites in NMR-based metabonomics/metabolomics experiments. Comput. Struct. Biotechnol. J. **14**, 135–153 (2016)
51. LaMarche, B.L., Crowell, K.L., Jaitly, N., Petyuk, V.A., Shah, A.R., Polpitiya, A.D., Sandoval, J.D., Kiebel, G.R., Monroe, M.E., Callister, S.J., Metz, T.O., Anderson, G.A., Smith, R.D.: MultiAlign: a multiple LC-MS analysis tool for targeted omics analysis. BMC Bioinformatics **14**, 49 (2013)
52. Available at: https://wada-main-prod.s3.amazonaws.com/resources/files/wada_td2015idcr_minimum_criteria_chromato-mass_spectro_conf_en.pdf. Accessed: July (2016)
53. Available at: http://www.eurl-pesticides.eu/library/docs/allcrl/AqcGuidance_Sanco_2013_12571.pdf. Accessed: July (2016)
54. Stanstrup, J., Neumann, S., Vrhovšek, U.: PredRet—prediction of retention time by direct mapping between multiple chromatographic systems. Anal. Chem. **87**, 9421–9428 (2015)
55. Héberger, K.: Quantitative structure-(chromatographic) retention relationships. J. Chromatogr. A **1158**, 273–305 (2007)
56. Arnhard, K., Gottschall, A., Pitterl, F., Oberacher, H.: Applying 'Sequential Windowed Acquisition of All Theoretical Fragment Ion Mass Spectra' (SWATH) for systematic toxicological analysis with liquid chromatography-high-resolution tandem mass spectrometry. Anal. Bioanal. Chem. **407**, 405–414 (2015)
57. Zhu, X., Chen, Y., Subramanian, R.: Comparison of information-dependent acquisition, SWATH, and MS(All) techniques in metabolite identification study employing ultrahigh-performance liquid chromatography-quadrupole time-of-flight mass spectrometry. Anal. Chem. **86**, 1202–1209 (2014)
58. Berendsen, B.J., Wegh, R.S., Meijer, T., Nielen, M.W.: The assessment of selectivity in different quadrupole-Orbitrap mass spectrometry acquisition modes. J. Am. Soc. Mass Spectrom. **26**, 337–346 (2015)
59. Mol, H.G., Zomer, P., García López, M., Fussell, R.J., Scholten, J., de Kok, A., Wolheim, A., Anastassiades, M., Lozano, A., Fernandez Alba, A.: Identification in residue analysis based on liquid chromatography with tandem mass spectrometry: experimental evidence to update performance criteria. Anal. Chim. Acta **873**, 1–13 (2015)
60. Weber, R.J., Southam, A.D., Sommer, U., Viant, M.R.: Characterization of isotopic abundance measurements in high resolution FT-ICR and Orbitrap mass spectra for improved confidence of metabolite identification. Anal. Chem. **83**, 3737–3743 (2011)
61. Kaufmann, A., Walker, S.: Accuracy of relative isotopic abundance and mass measurements in a single-stage Orbitrap mass spectrometer. Rapid Commun. Mass Spectrom. **26**, 1081–1090 (2012)
62. Xu, Y., Heilier, J.F., Madalinski, G., Genin, E., Ezan, E., Tabet, J.C., Junot, C.: Evaluation of accurate mass and relative isotopic abundance measurements in the LTQ-Orbitrap mass spectrometer for further metabolomics database building. Anal. Chem. **82**, 5490–4501 (2010)
63. Grange, A.H., Sovocool, G.W.: Automated determination of precursor ion, product ion, and neutral loss compositions and deconvolution of composite mass spectra using ion correlation based on exact masses and

relative isotopic abundances. Rapid Commun. Mass Spectrom. **22**, 2375–2390 (2008)
64. Pelander, A., Decker, P., Baessmann, C., Ojanperä, I.: Evaluation of a high resolving power time-of-flight mass spectrometer for drug analysis in terms of resolving power and acquisition rate. J. Am. Soc. Mass Spectrom. **22**, 379–385 (2011)
65. Erve, J.C., Gu, M., Wang, Y., DeMaio, W., Talaat, R.E.: Spectral accuracy of molecular ions in an LTQ/Orbitrap mass spectrometer and implications for elemental composition determination. J. Am. Soc. Mass Spectrom. **20**, 2058–2069 (2009)
66. Knolhoff, A.M., Callahan, J.H., Croley, T.R.: Mass accuracy and isotopic abundance measurements for HR-MS instrumentation: capabilities for non-targeted analyses. J. Am. Soc. Mass Spectrom. **25**, 1285–1294 (2014)
67. Benson, S., Lennard, C., Maynard, P., Roux, C.: Forensic applications of isotope ratio mass spectrometry—a review. Forensic Sci. Int. **157**, 1–22 (2006)
68. De Laeter, J.R., Böhlke, J.K., De Bièvre, P., Hidaka, H., Peiser, H.S., Rosman, K.J.R., Taylor, P.D.P.: Atomic weights of the elements. Review 2000. Pure Appl. Chem. **75**, 683–800 (2003)
69. de Castro, A., Gergov, M., Ostman, P., Ojanperä, I., Pelander, A.: Combined drug screening and confirmation by liquid chromatography time-of-flight mass spectrometry with reverse database search. Anal. Bioanal. Chem. **403**, 1265–1278 (2012)
70. Lapthorn, C., Pullen, F., Chowdhry, B.Z.: Ion mobility spectrometry-mass spectrometry (IMS-MS) of small molecules: separating and assigning structures to ions. Mass Spectrom. Rev. **32**, 43–71 (2013)
71. Paglia, G., Williams, J.P., Menikarachchi, L., Thompson, J.W., Tyldesley-Worster, R., Halldórsson, S., Rolfsson, O., Moseley, A., Grant, D., Langridge, J., Palsson, B.O., Astarita, G.: Ion mobility derived collision cross sections to support metabolomics applications. Anal. Chem. **86**, 3985–3993 (2014)
72. Paglia, G., Kliman, M., Claude, E., Geromanos, S., Astarita, G.: Applications of ion-mobility mass spectrometry for lipid analysis. Anal. Bioanal. Chem. **407**, 4995–5007 (2015)
73. Schollée, J.E., Schymanski, E.L., Avak, S.E., Loos, M., Hollender, J.: Prioritizing unknown transformation products from biologically-treated wastewater using high-resolution mass spectrometry, multivariate statistics, and metabolic logic. Anal. Chem. **87**, 12121–12129 (2015)
74. Bier, D., Hartmann, R., Holschbach, M.: Collision-induced dissociation studies of caffeine in positive electrospray ionization mass spectrometry using six deuterated isotopomers and one N1-ethylated homologue. Rapid Commun. Mass Spectrom. **27**, 885–895 (2013)