Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Chromatography B, 811 (2004) 3–12
JOURNAL OF CHROMATOGRAPHY B
www.elsevier.com/locate/chromb

Substance identification: the weak link in analytical toxicology☆

Rokus A. de Zeeuw*

Department of Analytical Chemistry and Toxicology, University Centre for Pharmacy, P.O. Box 196, 9700 AD Groningen, The Netherlands

Received 20 November 2003; accepted 23 July 2004
Available online 11 September 2004

Abstract

Although substance identification is a key factor in analytical toxicology, it is amazing that the subject is receiving very limited and often inappropriate attention. With regard to the latter, a "confirmation" approach is usually chosen, which does not yield unambiguous identification. Moreover, the criteria for establishing a "positive match" leave much to be desired. These observations are corroborated when comparing some recent guidelines for qualitative analysis (issued for various forensic areas by SOFT/AAFS, NCCLS, NLCP, WADA and EU). Apart from showing substantial differences between them on pivotal issues, the guidelines contain various elements that appear scientifically incorrect and/or legally untenable. Also, the guidelines focus primarily on mass spectrometry (MS) and pay little or no attention to other identification possibilities (such as chromatographic techniques, either in combination with MS or as stand-alone techniques. Moreover, they do not offer alternatives in situations where access to MS is not available. One must conclude, therefore, that substance identification is a neglected and misunderstood domain in analytical toxicology. Rapid and concerted actions are needed to: (1) improve the general knowledge; (2) to define uniform strategies in the analytical approach and in the interpretation of the results; and (3) to set up and maintain suitable banks of reference substances and computerized data bases to allow unambiguous identification.

Keyword: Substance identification

1. Introduction

The three major tasks in analytical toxicology are to detect, identify, and quantitate potentially harmful substances in biological or other relevant specimens. In the past, analytical toxicology was mainly applied in clinical and forensic toxicology, but in recent years, the horizon has broadened substantially, as can be seen from Table 1. The majority of the work now regards living subjects, the levels to be analysed are often at the ppb- or ppt-level, the range of relevant

Abbreviations: AAFS, American Association of Forensic Sciences; BDB, benzodioxolo-5-butanamine = l-(3,4-methylenedioxyphenyl)-2-butanamine; BOL, 17α,17β-boldenone; CI, chemical ionization; DES, diethylstilbestrol; EE, ethynylestradiol; EI, electron impact; EU, European Union; GC, gas chromatography; GC–MS, gas chromatography–mass spectrometry; HFBA, heptafluorobutyric anhydride; HPLC, high performance liquid chromatography; HRMS, high resolution mass spectrometry; IEC, International Engineering Consortium; ISO, International Organization for Standardization; LC, liquid chromatography; MB, methylboldenone; MDMA, methylenedioxymethylamphetamine; MS, mass spectrometry; MS–MS, tandem mass spectrometry; MT, methyltestosterone; *m*/*z*, mass-to-charge ratio; NCCLS,- National Clinical Chemistry Laboratory Standards (formerly); NIST, National Institute for Standardization and Technology; NLCP, National Laboratory Cerification Program; NOK, not OK; PMMA, *p*-methoxymethylamphetamine; ppb, parts-per-billion; ppt, parts-per-trillion; RI, relative intensity (of ions in mass spectrometry); RIVM, Rijksinstituut voor Volksgezondheid en Milieu; SIM, selected ion monitoring (in mass spectrometry); SOFT, Society of Forensic Toxicologists; TIAFT, The International Association of Forensic Toxicologists; TLC, thin layer chromatography; UV, ultraviolet; WADA, World Anti-Doping Agency

☆ From a keynote lecture presented at the 41st International Meeting of the International Association of Forensic Toxicologists (TIAFT) Melbourne, Australia, 16–20 November 2003.

* Tel.: + 31 50 3633336; fax: +31 50 3637582.
E-mail address: r.a.dezeeuw@wolmail.nl.

1570-0232/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jchromb.2004.07.043

Table 1
Major application areas of analytical toxicology

Area	Focus
Clinical toxicology	Intoxicated subjects
Forensic toxicology	Post-mortem cases
Urine drug testing	Drugs of abuse, doping in sports
Human performance testing	Workplace testing, drugs in traffic
Occupational toxicology	Exposure at work
Food toxicology	Residues in food products
Animal toxicology	Cattle, wildlife, fish, pets
Environmental toxicology	Environmental pollution, exposure

substances has broadened dramatically and the demand for rapid analysis is growing.

In all these areas, the task of unambiguous substance identification is the most challenging and crucial one. Yet, it is amazing that this subject is receiving very little and/or inappropriate attention in textbooks, journal articles, meeting presentations, and also in legal cases. Often, qualitative results are being summarized by short statements such as: (a) the identity was confirmed by GC–MS; (b) identification was achieved by monitoring three diagnostic ions in SIM-MS; and (c) the mass spectrum of the unknown matched that of the reference. Such one-liners apparently are considered satisfactory and give the impression that substance identification is rather simple and a matter of routine.

However, since qualitative results in analytical toxicology can have pivotal judicial, social, personal, economical and/or sports consequences, it is important to assess whether approaches and methodologies in this area are fit for purpose, scientifically sound and legally defensible. Therefore, in this paper, a critical review is presented on current practices in qualitative analytical toxicology, as exemplified by a number of recent guidelines in different areas, viz. forensic toxicology and human performance testing [1], confirmation of drugs [2], workplace testing [3,4], doping in sports [5] and residues in animals and animal products [6]. The focus will be on organic substances in biological specimens with molecular weights up to some 600 Da, because these are most frequently encountered in daily practice.

2. Common elements in the guidelines

When comparing the guidelines, a common set of confirmation criteria can be clearly distinguished:

- Confirmation should be based on mass spectrometry (MS). MS–MS is also allowed.
- The mass spectrum of the unknown should be compared with that of a contemporaneously analyzed reference standard.
- Selected ion monitoring (SIM) is preferable but full scan MS is allowed as well.
- Two to four diagnostic ions should be monitored, with signal-to-noise ratio's >3.
- Relative abundances must agree within maximum permitted tolerance windows.
- On-line chromatography (GC or LC) prior to MS analysis.

Since these elements are being emphasized in all guidelines, they may represent the core scientific opinion in this field. However, a number of pivotal shortcomings become apparent as well:

- The documents do not differentiate between confirmation and identification, focusing primarily on confirmation.
- The guidelines do not define what they consider a scientifically correct and/or legally defensible confirmation/identification.
- The guidelines differ substantially in their criteria for confirmation and in accepting or rejecting the matching of analytical results.

3. Confirmation versus identification

The term 'confirmation' has gained widespread acceptance in analytical toxicology after it appeared in the Mandatory Guidelines for Workplace Drug Testing in 1988 [7]. However, the following should be noted:

- Confirmation presumes the presence of a substance *Y* in a sample, based on initial tests or prior information. The presence of *Y* can then be 'confirmed' by further tests, such as MS.
- Identification does not make a priori presumptions based on initial tests or other information. Analytical results from a number of tests on the sample are randomly compared with reference data on all other substances that may come into consideration.

Moreover, it must be realized that a 'positive' confirmation test thus obtained is NOT an unambiguous identification of *Y*. It only indicates that the test result is not against the presumption; other substances may be able to give results that are the same or undistinguishable from those of *Y*. Unambiguous identification of *Y* requires that all other (relevant) substances can be excluded, so that *Y* remains as the ONLY possible candidate [8]. Utilizing this exclusion criterion may also be called the 'reverse angle approach'. For example, if methamphetamine is presumed to be present and the mass spectrum in the sample matches that of a reference sample within acceptable limits, it remains to be established that the test result in the sample cannot be due to any other substance likely to be encountered (e.g. other amphetamine-like substances, isomers, metabolites, endogenous compounds, omnipresent interferences, etc.). This is elucidated in Fig. 1 which depicts the EI mass spectrum of the suspect substance. The search in the NIST 98 data base resulted in a best match for methamphetamine with a match factor of 0.997, but also listed the substances in Table 2, all with match factors above 0.980.

The reverse angle approach requires that additional tests be carried out to exclude all other substances than

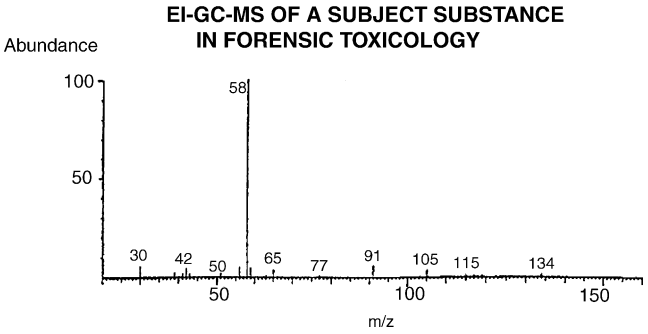

Fig. 1. Mass spectrum of a substance suspected to be methamphetamine. Sample analyzed by EI-GC–MS.

Table 2
Substances with match factor of 0.980 or higher for a spectrum of methamphetamine shown in Fig. 1

Phentermine	PMMA
Propoxyphene	3,4-MDMA
N-Methyl-4-thioamphetamine	6-Chloro-3,4-MDMA
2-OH-(3-OH-phenyl)-*N*-ethylethylamine	*N*,*N*-Dimethyltryptamine
Ephedrine	2,3-MDMA
N,*N*-Dimethyl-2-phenyl-propanolamine	4-Methoxy-*N*,*N*-dimethyl-amphetamine
Amitriptyline	3,4-BDB
Doxepin	2,3-BDB
Tramadol	Psilocine
	Phenyltoloxamine

methamphetamine, or—if this presumption was false—to determine which other substance comes into consideration. These additional tests may consist of any other technique that may provide appropriate information, including other MS techniques. Obviously, in the case of Fig. 1, in which EI-GC–MS was used, it would be logical to consider the GC retention index of the unknown, to narrow down the range of possible candidates. It should be noted, however, that this requires GC-data bases (and other chromatography data bases) which match the MS data bases in terms of entries and also that the chromatography system of the data base and that of the analyst be the same. Up till now, almost all MS data bases have much more entries than the available chromatography data bases.

It should also be emphasized that the term ‘relevant’ is important in the above context. With hundreds of thousands of substances known to society, it is clearly unfeasible to consider them all. Yet, even if one focuses only on those that have some relevance to the area one is involved in (e.g. forensic toxicology, doping, environmental pollution), reliable reference data on thousands of substances per area must be available.

4. How to compare mass spectra

Although a good quality full scan mass spectrum generally provides better qualitative information, the guidelines consider the SIM mode the method of choice because it is more sensitive and less affected by potential interferences. Yet, it should be realized that SIM always hinges on a presumption and that comparison should be made with a contemporaneously analyzed reference standard (although the guidelines leave it to the analyst to use a clean reference solution or a reference spiked in the matrix under consideration). Hence, each laboratory must have its own bank of reference substances. The latter may be feasible when the number of target compounds is limited. However, the consequence of the exclusion criterion requires large banks of thousands of precious compounds with a high turn-over rate.

Some guidelines allow computer-assisted library searching to match spectra, but—contrary to SIM comparisons—remain very vague in defining a satisfactory match, other than that a critical match factor should be exceeded. Obviously, the limited inter-laboratory reproducibility of mass spectra and the phenomenon that computerized MS data bases often contain a number of divergent mass spectra for the same substance (thus making it difficult to decide which spectrum to use as the reference) are rather complicating factors. On the other hand, if one follows the identification approach without presumptions (e.g. in general unknown cases, in cases where the initial presumption turns out to be wrong, or when searching for multi-drug cases), computer assisted library searching is the only viable option to the analyst.

5. Criteria for SIM

All guidelines are in agreement that:

- The diagnostic ions to be monitored in SIM must be sufficiently characteristic for the structure of the compound, should not originate from the same part of the molecule and should preferably include the molecular ion or the precursor ion.
- A minimum number of diagnostic ions is necessary.
- Relative abundances for all diagnostic ions must agree within certain maximum permitted tolerance windows.
- The chromatographic retention (GC or LC) and peak shape must be comparable to that of the reference.

However, there are marked differences when it comes to minimum number of ions and tolerance windows. This is depicted in Table 3. It is amazing to see, for example, that SOFT/AAFS requires the monitoring of only two ions (yielding only one ion ratio; the other automatically being 100%), whereas under EU guidelines, the identification of the same substance requires the monitoring of four ions with tighter windows at relative intensities >20% and much larger windows for relative intensities <10%. WADA even chooses to express tolerance windows in terms of absolute differences and relative differences: Thus, for a peak in EI-GC–MS with a relative intensity (RI) of 80% the tolerance windows in absolute terms are 70–90%, whereas for a peak of 40% the

Table 3
Minimum number of ions and tolerance windows for SIM

Guideline	Ion relative intensities (%)	EI-GC–MS Δ (%)	Other[a] Δ (%)	Ions to be[b] monitored
SOFT/ AAFS	All	±20 (rel)	±20–30 (rel)	2
NCLP	All	±20 (rel)		3
NCCLS	All	±20 (rel)		3–4
WADA	≥50–100	±10 (abs)	±15 (abs)	3
	≥25–49	±20 (rel)	±25 (rel)	
	<25	±5 (abs)	±10 (abs)	
EU	>50	±10 (rel)	±20 (rel)	4
	>20–50	±15 (rel)	±25 (rel)	
	>10–20	±20 (rel)	±30 (rel)	
	≤10	±50 (rel)	±50 (rel)	

[a] Includes CI-GC–MS, LC–MS. GC–MS–MS, LC–MS–MS, etc.
[b] Resulting in $n-1$ ion ratios.

window in relative terms will be 32–48%. Note also that the tolerance windows do not always fit properly when going from one RI class to another: For WADA in EI–GC–MS, a peak at 25% RI has a window of 20–30%, whereas the window for a peak at 24% RI would be 19–29%, providing a smooth transition from 10% absolute to 20% relative. Yet, for LC–MS, the respective windows would be 18.75–31.25 and 14–34%, respectively. Since the windows now overlap, this is not a smooth transition. Also note the big jump in the EU Guideline when going from 20 to 50% for ions with an RI ≤10% for EI-GC–MS.

What could be the scientific basis for the above criteria? As regards the tolerance windows, the general opinion is that the repeatability of ion ratio measurements decreases with lower RIs, but the guidelines do not provide substantiating references nor other indications how they arrived at the tolerance windows in Table 3. It seems that the latter are just based on arbitrary decisions. The reasons for switching between absolute and relative differences, as advocated by WAD A, also remains unexplained. On the other hand, Bethem et al. [9] suggested that it may not be fruitful to put exhaustive attention on the size of the tolerance windows, other than to use matching criteria that are not dramatically different than the ranges in Table 3. Yet, this is not scientifically sound: Too small windows may yield false-negatives, whereas too large windows may produce false-positives

The rationale for the number of ions to be measured will be addressed below.

6. Is the concept of tolerance windows correct?

Apart from the above observations, there is a more fundamental question as to whether the use of tolerance windows for RIs is acceptable from a conceptual point of view. Obviously, the idea is to let the windows compensate for small run-to-run variations in the measurements of the mass spectra. However, what is being overlooked here is that absolute ion intensities are recorded, which are then normalized against the base peak to give RIs.

Yet, as with any ion, the absolute intensities of the base peak may also vary from run to run but these variations are made invisible in the normalization process when the absolute intensities of the base peaks are automatically set at 100% RI. This can have crucial effects, as shown in Table 4, which uses tandem MS data from reference [10] and the corresponding tolerance windows of the EU [6].

In Table 4, middle part, the base peak in suspect sample 1 shows an acceptable variation in its absolute intensity as compared to the reference. All other ions in this sample have the same absolute intensities as the reference. When using absolute intensities, a positive match is obtained for all ions. However, when using relative intensities, three ions are outside the tolerance window, despite the fact that their absolute intensities were all OK. Thus, the conclusion based on relative intensities would lead to a false-negative result.

In the lower part of Table 4, the base peak in suspect sample 2 has an acceptable variation in absolute intensity of 15% (in the opposite direction as compared to sample 1), but the absolute intensities of ions 364 and 264 are now outside their respective tolerance windows, resulting in negative matches. Yet, if the relative intensities are compared, positive matches are obtained for all ions, yielding a false-positive result. One must conclude, therefore, that the normalization process against the base peak can lead to incorrect and arbitrary results, which do not allow equal justice.

7. Number of ions to be monitored

Though the number of ions to be monitored for spectral comparison in SIM varies from a minimum of two in the SOFT Guideline to a minimum of four in the EU Guideline, the scientific basis for these criteria lies in the early work of Sphon in 1978 [11], and an update of his approach at a workshop on the limits to confirmation, quantitation and detection in 1996 [12]. The recommendation by Sphon to monitor a minimum of three diagnostic is still the recognized standard for substance identification by the mass spectrometric community [9]. In brief, Sphon used diethylstilbestrol (DES) as model substance and an EI-GC–MS data base with some 30,000 spectra. Monitoring the RIs of three ions (viz. 268, 239 and 145) at discrete tolerance windows was sufficient to select DES as the only candidate in a computer-assisted library search. At the re-evaluation in 1996 [12], the data base contained some 270,000 spectra. Monitoring the same three ions, yet with tighter windows of 10% absolute, still showed DES to be the only candidate. However, a number of critical observations can be made:

- Although the approach by Sphon is valuable and scientifically correct in that it utilizes the exclusion criterion, it has so far been tested for only one model compound.

Table 4
Matching of SIM data from reference samples with those from suspect samples

Ion *m*/*z*	Absolute intensive arbitrary units	Relative intensity (%)	Tolerance window	Match with absolute intensity	Match with relative intensity
Reference sample					
440	10000	100	08000–12000[a]		
364	02700	27	02200–03200		
320	06800	68	05500–08100		
314	08300	83	06700–09900		
264	01200	12	00800–01600		
Suspect sample 1					
440	08000	100	80–120[b]	OK	OK
364	02700	34	22–32	OK	NOK
320	06800	85	55–81	OK	NOK
314	08300	105	67–99	OK	NOK
264	01200	15	08–16	OK	OK
Suspect sample 2					
440	11500	100	80–120[b]	OK	OK
364	03500	30	22–32	NOK	OK
320	06800	59	55–81	OK	OK
314	08300	72	67–99	OK	OK
264	01800	15	08–16	NOK	OK

NOK: not OK.
[a] Based on absolute intensities.
[b] Based on relative intensities.

- The mass spectrum of DES has all three diagnostic ions at relatively high masses, which provides much better selectivity than ions at lower *m*/*z* values.
- The suitability of the data bases to analytical toxicology remains uncertain (i.e. were toxicologically relevant substances adequately represented and were their spectra taken at conditions that are common in analytical toxicology?).
- To assess general applicability of the three-ion rule (or any other number of ions), tests with other model substances and classes are needed, especially where many structurally-related compounds can be encountered (e.g. amphetamines, opiates, hypnotics/sedatives, steroids, pesticides, antidepressants, etc.).
- The approach has not been checked for techniques other than EI-GC–MS.

Thus, the monitoring of three ions (not to speak of two, or even four) is unwarranted as a general rule at the present time. Follow-up studies with a much larger selection of relevant model substances and with other techniques than EI-GC–MS are urgently needed.

Recapitulating paragraphs 5–7, obviously the widely divergent criteria between the Guidelines for number of ions to be monitored and tolerance windows in SIM are scientifically unsound and legally untenable. It cannot be that one and the same test result may lead to a ‘positive’ identification when using Guideline A and a ‘negative’ identification when using Guideline B. However, even if the identification result remains the same under all guidelines, we cannot guarantee that the result is correct.

8. Techniques other than EI-GC–MS

Understandably, the primary focus in the guidelines is on EI-GC–MS because most of our presently available knowledge is in this domain. Yet, other techniques, such as CI-GC–MS, LC–MS, tandem MS and high-resolution MS are becoming more and more popular in analytical toxicology. It is good that some of the guidelines contain criteria for their use, although they suffer from the same shortcomings as discussed above. Moreover, it should be realized that our knowledge on their potentials and limitations, e.g. on items such as intra- and inter-laboratory reproducibility, tolerance windows, number of ions to be monitored, etc. is limited. Also, suitable data bases for these techniques are still in its infancy. A review on the utility of these techniques towards substance identification can be found in reference [13].

Undoubtedly, the availability of mass spectrometric techniques has been a great step forward in our ability to advance in qualitative analysis. However, other analytical techniques—when properly selected and interpreted—can also make useful and even decisive contributions to achieve unambiguous identification (immunoassays, TLC and colour reactions, GC with selective detectors, HPLC with UV or diode array detection, capillary electrophoresis, enantiochromatography, to name a few). Reviewing all these possibilities is beyond the scope of this article, but more information can be found elsewhere [14,15]. When using other techniques, it is important to note that they can only be of value when sufficiently large and reliable data bases are available as well. Also, for chromatographic techniques, the chromatographic system used by the analyst must be the same as the one used

to generate the data base. If these two conditions are not met, a suitable reverse angle identification approach is meaningless. Preferably, one should be able to interactively search the various data bases, including those with MS data [15].

Finally, there are still many laboratories involved in analytical toxicology without access to mass spectrometers. Can substance identification be done properly without the 'gold standard technique'? [16]. For all those financially strapped, the answer is encouraging: yes, it can be done but you should be aware of the potentials and limitations of the techniques available to you. It should not be forgotten that proper substance identification has been successfully attempted by many toxicologists in the past when MS was not even around. See reference [14] for approaches not using MS.

9. Miscellaneous issues

Although the guidelines are setting fairly firm and detailed requirements for substance identification, there remain a number of additional issues that are not, or only partly, addressed. Some of the more relevant ones are briefly discussed below:

(a) None of the documents contains a requirement to check peak purity. This should be mandatory for all peaks, not only those that are skewed but also for symmetrical peaks, and the test can be performed easily.
(b) How strict should the term 'contemporaneously analyzed' be interpreted? In the same batch/run, in a batch/run on the same instrument on the same day, in a batch/run on the same instrument on different days? Is it allowable that the instrument be used for other types of analysis between corresponding batches/runs? Recently, Parsons asked his colleagues on the TIAFT Internet site to which practice they adhered, but got a very limited and discouraging response [17].
(c) Regarding the recording of the reference spectrum, the guidelines leave it to the individual analyst to decide whether to dissolve the reference in a clean solution, to spike the reference in the appropriate matrix, or to use a biological specimen known to contain the reference substance. Obviously, since the matrix may play a substantial role by giving background signals or by inducing ion suppression or ion augmentation, the first option should be rejected. In addition, it should be realized that mass spectra may be dependent on the concentration of the analyte [18].
(d) It is unclear whether to construct a mass spectrum from a single run or from averaging a minimum number of runs.
(e) According to Laboratory Standard ISO/IEC 17025 [19], all testing laboratories must give an estimate of the uncertainty of measurement (the inverse is the probability of correctness). This applies to quantitative and qualitative results. Yet, the latter is often neglected and/or explicit validation requirements for this purpose are not given. This not only holds for the present guidelines but also for other official documents [20,21]. Clearly, confirmation approaches as advocated by the guidelines are unable to address the issue of qualitative uncertainty; only identification procedures utilizing the reverse angle concept will be able to do so [15]. However, the estimate will always depend on the size of the smallest data base used in the exclusion procedures. For example, if the identification is based on an LC retention index checked against a database of 5000 entries, a diode array UV absorption spectrum checked against a data base of 1000 spectra and a mass spectrum checked against a data base of 200,000 entries, the probability of correctness for the combined techniques can maximally be 99.9% (1 out of 1000). Moreover, it must be emphasized that this value only relates to the 1000 substances present in the UV data base.
(f) Most guidelines allow the use of professional judgement and expertise when interpreting the data. Although this appears logical, one should always be very careful that expertise does not turn into backbone routine. For example, seemingly unimportant deviations seen in what looks like the umpteenth MDMA-intoxication may just be the first hint that one is looking at a new party drug on the market. Moreover, professional judgement/expertise should always be applied with utmost objectivity. On the other hand, since the reliability of professional judgement/expertise cannot be expressed in numerical values, it cannot be taken into account when assessing the probability of correctness.

10. Examples of substance identifications in practice

After having discussed the fundamentals, it is good to examine how substance identification is being handled in practice and which pitfalls or problems may occur. It should be noted that the issues highlighted in the examples are not limited to the substances mentioned. The latter were chosen randomly from cases available to the author.

10.1. Diagnostic ions should be characteristic for the structure of the substance and should not originate from the same part of the molecule

The urinary metabolite 19-norandrosterone is indicative of nandrolone abuse in sports and its di-tms derivative is commonly analysed by EI-GC–MS, utilizing SIM of the ions 405 (base peak), 420 and 315. Occasionally, the ions 405, 420 and 73 are being monitored. These ions are indeed the most abundant in the MS, as shown in Fig. 2. However, they represent the following structures:

m/z 420 = molecular ion
m/z 405 = $[M–CH_3]^+$

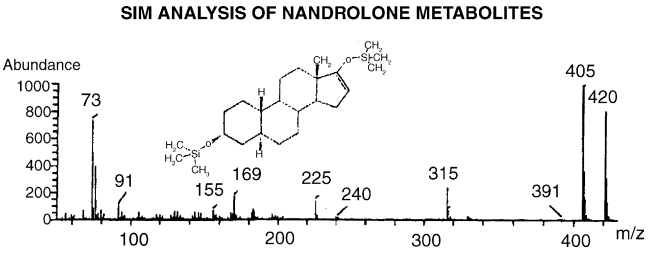

Fig. 2. SIM analysis of nandrolone metabolites.

m/z 315 = $[M–TMSiOH–CH_3]^+$
m/z 73 = trimethylsilyl ion

Obviously, apart from the molecular ion, none of the other ions provide much additional diagnostic information on the substance. The selection of the trimethylsilyl ion is even worse because it can be due to any silylated compound. Moreover, three of the ions originate from the same part of the molecule. One must conclude, therefore, that the identification criteria are by far not met. Unfortunately, the above phenomenon of not paying attention to the diagnostic value of the fragments chosen for substance identification is common in all areas of analytical toxicology, as can be seen from the literature and from case records.

Other options that can be used to increase the probability of correctness of the identification are: (1) high resolution MS; and (2) MS–MS. In HRMS, it is common to monitor the ions 405.2645, 420.2880 and 315.2144. Although the latter two ions do not provide much extra information, the presence of m/z 420.2880 strongly points to a molecular composition of $C_{18}H_{28}O_2$ for the suspect substance. Yet, it should be noted that, apart from 19-norandrosterone, there are at least 15 other substances commercially available with this molecular composition [22]. Application of MS–MS and monitoring ions 315, 225, 183, 169, 155143 would provide additional diagnostic value, but may also lead to a phenomenon that can be called “ion shopping”, i.e. to focus only on suitable ions: this is demonstrated in Table 5, which deals with a case analyzed under the 1998 criteria of the International Olympic Committee [23]. The data were interpreted as a positive confirmation because the RIs of three diagnostic ions matched with those of the reference within ±25% (rel). Here, the requirement was overlooked that all diagnostic ions must be taken into account. Instead, ion shopping was applied for those fragments

Table 5
SIM analysis of nandrolone metabolites by EI-GC–MS–MS (precursor ion 405)

Ion m/z	Sample relative intensity (%)	Reference relative intensity (%)	Δ Relative intensity (rel.)
315	100	100	–
225	70.7	65.9	7.2
183	6.83	8.15	16.2
169	8.35	5.84	30.1[a]
155	8.29	5.94	28.3[a]
143	7.78	7.49	3.7

[a] Outside the maximum permitted tolerance window.

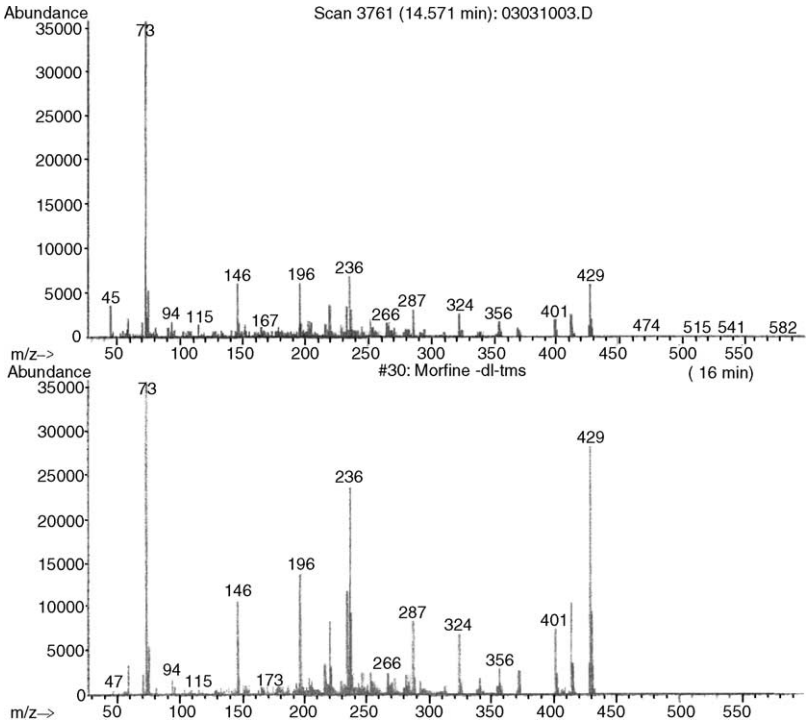

Fig. 3. Matrix impact on the mass spectrum of morphine. Top: morphine extracted from whole blood and analyzed by EI-GC–MS after silylation. Bottom: reference spectrum of morphine, analyzed under the same conditions.

Table 6
SIM analysis of morphine di-tms after extraction from whole blood and EI-GC–MS

Ion *m/z*	Sample relative intensity (%)	Reference relative intensity (%)	Δ Relative intensity
429	88	100	12 (absolute and relative)
236	100	84	16 (absolute), 19 (relative)
196	99	49	50 (absolute), 102 (relative)

that matched within the allowable tolerance windows, i.e. *m/z* 225, 183 and 143. A similar example of ion shopping can be found in Table 4 of reference [10].

10.2. Matrix effects on mass spectra

One should always be aware that extracts of (biological) samples, also after extensive clean-up, may contain compounds that will enter or be present in the ionization chamber of the mass spectrometer at the same time as the substance of interest. This may affect the resulting mass spectrum of the latter, e.g. through ion suppression, ion augmentation or ion clustering. Even when runs on a blank matrix do not indicate possible problems (e.g. the interference does not show up at the *m/z* values selected for the SIM process), interferences may still occur.

Therefore, if feasible, one should always strive to run the reference in the appropriate matrix. An example of the impact of the matrix is depicted in Fig. 3. The top scan is from morphine, extracted from whole blood, transformed to its di-tms derivative, and analyzed by EI-GC–MS. In the bottom panel, the reference spectrum is shown, taken from a similar concentration in a clean solution. Table 6 contains the results of the SIM analysis for the ions 429, 236 and 196.

Although the spectra look visually similar, in the SIM analysis, all ions in the sample are outside the tolerance windows permitted by the WADA and EU Guidelines, with ion 196 also being outside the windows permitted by SOFT/AAFS, NCLP and NCCLS. Thus, the SIM analysis would yield a negative result for morphine.

Another point to consider is that there is no such thing as a 'standard' matrix. They can vary from case to case (a 'fatty' liver versus a 'lean' one, 24 h urine versus morning urine, etc.). Other factors that may play a role are age of the individual, diet, ethnic background, standard medication, degree of putrefaction, etc. Obviously, one should strive to find a blank matrix that best resembles the one under investigation, but this is quite difficult in practice.

10.3. Repeatability/reproducibility of mass spectrometry

Obviously, as with any other analytical technique, the identification power of MS is ultimately dependent on the repeatability/reproducibility between experiments. With the guidelines focusing on SIM analysis, the key parameter to be known is the repeatability between runs on the same instrument that determines the maximum permitted tolerance windows. Surprisingly, none of the guidelines addresses this issue, and relevant literature does not seem to exist either. Therefore, it is likely that the windows in the guidelines have been set arbitrarily. Yet, a very recent and detailed study at the State Institute for Public Health and the Environment (RIVM) in The Netherlands has yielded some dramatic results regarding the repeatability of MS [24]. The authors validated the identification and quantitation of growth promoting substances in beef, pork, turkey and fish, against the EU Guidelines [6]. The target analytes were 17α,17β-boldenone (BOL), methylboldenone (MB), methyltestosterone (MT), 17α,17β-nortestosterone (NT) and ethynyl estradiol (EE) at 0.5–10 ppb. After enzymatic digestion of the spiked samples, the steroids were extracted with tert-methylbutylether and isolated by HPLC fractionation. After derivatization with HFBA the final analysis was by EI-GC–MS.

Table 7
Within-day repeatability of ion ratio's of anabolic steroids in standard solutions

Substance MT	Ion ratio's × 100		
	355/369	465/369	480/369
Std			
1	29.3	43.8	46.3
2	23.9+	34.3+	33.0+
3	25.4	34.1+	35.7+
4	28.9	46.0	48.4
5	33.5+	43.4	44.4
6	34.2+	49.2+	51.8+
7	26.1	37.5	39.5
8	27.9	39.2	42.5
9	25.7	37.9	38.2
10	28.3	42.2	43.9
11	36.3+	55.4+	54.6+
Average	29.0	42.1	43.6
Window	24.7–33.4	35.8–48.4	37.0–50.1

From [24]. +: outside window.

In Table 7, the within-day repeatability for standard solutions of MT (2–10 ppb) are given. It can be seen that more than 30% of the ion ratio's are outside the maximum allowable tolerance windows.

The day-to-day repeatability and the impact of the matrix (beef and pork) is exemplified in Table 8, with EE as target analyte. Apart from large variations in the ion ratio's, which are even more pronounced in the spiked samples, another disturbing phenomenon is that the base peak can change, i.e. from *m/z* 353 to 446. This was observed in two of the five standards and in all spiked samples. Base peak changes were also seen for BOL and for MT.

Therefore, it is not surprising that the number of correct identifications in blind, spiked samples in the above investigation was dramatically low. This is demonstrated in Table 9 for BOL in turkey and MT in beef at levels of 1 ppb.

From these extensive studies, a number of pivotal conclusions can be drawn:

- The repeatability of GC–MS needs to be proven, not assumed, and in many cases may not be as good as imagined. This also holds when analyzing standard solutions.

Table 8
Day-to-day repeatability and matrix impact on ion ratio's of anabolic steroids

Substance EE	Ion ratio's × 100		
	446/353	459/353	474/353
Std			
1	69.0	41.4	39.5
2	127b	97.6	95.7
3	104b	82.5	72.9
4	86.8	51.7	49.7
5	52.3	29.3	30.8
Range	52.3–127	29.3–82.5	30.8–95.7
Spiked			
1p	110b	97.3	92.0
2	129b	97.6	177b
3	146b	120b	108b
4	185b	151b	140b
Range	110–185	97.3–151	92–177

From [24]. b: base peak now at *m*/*z* 446; p: spiked in pork; other spiked samples in beef.

- The day-to-day repeatability of GC–MS with standards is significantly worse than within-day.
- The matrix reduces the repeatabilty of GC–MS even further.
- Changes in base peaks may occur due to matrix effects, but may also be seen with standards in clean solutions.
- The EU tolerance windows are not realistic.

It remains to be seen whether repeatability is feasible in practice with techniques like LC–MS and tandem MS. Also, studies with other toxicologically relevant substances are needed.

Table 9
Correct identifications of anabolic steroids in spiked samples

Substance	Ion ratio's × 100			Correctly identified
BOL, turkey	251/369	464/369	678/369	
Sample				
1	82.3	87.1c	31.7c	No
2	108	76.2	27.5c	No
3	69.2	66.8	23.1	No
4	104	56.6	85.6	No
5	72.2	86.1c	30.9c	No
Standard	56.4	85.6	28.6	No
Maximum tolerance, relative (%)	10	10	15	
Window	50.7–62.0	77.1–94.2	24.3–32.9	
MT, beef	355/369	465/369	480/369	
Sample				
1	79.5	103	132	No
2	102	124	147	No
3	55.6	75.8	90.5	No
4	85.5	100	130	No
5	68.8	107	114	No
Standard	29.0	42.1	43.6	
Maximum tolerance, relative (%)	15	15	15	
Window	24.7–33.4	35.8–48.4	37.0–50.1	

From [24]. c: ion ratio within tolerance window. For BOL 5 out of 15, for MT 0 out of 15.

11. Conclusions

The above review clearly indicates that, despite various technological advances in recent years, substance identification in analytical toxicology leaves much to be desired. This is due to a lack of understanding of the fundamental principles, in combination with practical shortcomings. They can be summarized as follows:

- The strategies and the interpretation criteria for substance confirmation/identification are not scientifically sound nor legally defensible.
- The utility of MS techniques forms a major bottleneck because of low repeatability/reproducibility.
- Suitable banks of reference substances and adequate computerized data bases are hardly available/accessible.
- Many laboratories around the world do not have access to MS techniques, so that there remains a need for less expensive identification approaches.

In view of the importance and complexity of these issues, rapid and concerted actions are urgently needed.

References

[1] Forensic Toxicology Laboratory Guidelines, Society of Forensic Toxicology/American Academy of Forensic Sciences, Colorado Springs, CO, 2002.
[2] Gas Chromatography/Mass Spectrometry (GC/MS) Confirmation of Drugs: Approved Guideline, NCCLS, Wayne, PA, 2002.
[3] Mandatory Guidelines for Federal Workplace Testing Programs, Substance Abuse and Mental Health Services Administration, US Department of Health and Human Services, Federal Register 66, 2001, pp. 43876–43882 (analytical criteria in [4]).
[4] Guidance Document for Laboratories and Inspectors, National Laboratory Certification Program, Research Triangle Park, NC, 2002.
[5] Laboratory Accreditation Requirements and Operating Standards, Version 1.0, World Anti Doping Agency, Montreal, 2002.
[6] European Union Decision 2002/657/EC 17.08.2002, Commission decision laying down performance criteria for the analytical methods to be used for certain substances and residues thereof in live animals and animal products, Off. J. Eur. Commun. 221 (2002) 8.
[7] Mandatory Guidelines for Federal Workplace Drug Testing Programs, Substance Abuse and Mental Health Service Administration, US Department of Health and Human Services, Federal Register 53 (1988). nr 69, Di-D12.
[8] R.A. de Zeeuw, J. Forensic Sci. 37 (1992) 1437.
[9] R. Bethem, J. Boison, J. Gale, D. Heller, S. Lehotay, J. Loo, S. Musser, P. Price, S. Stein, J. Am. Soc. Mass Spectrom. 14 (2003) 528.
[10] F. André, K.K.G. De Wasch, H. De Brabander, S.R. Impens, L.A.M. Stolker, L. van Ginkel, R.W. Stephany, R. Schilt, D. Courtheyn, Y. Bonnaire, P. Fuerst, P. Gowik, G. Kennedy, T. Kuhn, J.-P. Moretain, M. Sauer, Trends Anal. Chem. 8 (2001) 435.
[11] J.A. Sphon, J. Off. Anal. Chem. 61 (1978) 1247.
[12] R. Baldwin, R.A. Bethem, R.K. Boyd, W.L. Budde, T. Cairns, R.D.S. Gibbons, J.D. Henion, M.A. Kaiser, D.L. Lewis, J.A. Matusik, J.A. Sphon, R. Stephany, R.K. Trubey, J. Am. Soc. Mass Spectrom. 8 (1997) 1180.
[13] L. Rivier, Anal. Chim. Acta 492 (2003) 69.
[14] R.A. de Zeeuw, J.P. Franke, in: M.J. Bogusz (Ed.), Handbook of Analytical Separations, vol. 2, Elsevier, Amsterdam, 2000, p. 567.

[15] J. Hartstra, J.P. Franke, R.A. de Zeeuw, J. Chromatogr. B 739 (2000) 125.
[16] I. Sunshine, J. Anal. Toxicol. 23 (1999) 229.
[17] R. Parsons, Personal communication through mailing list of the International Association of Forensic Toxicologists, April 2003.
[18] S. Impens, K. De Wasch, M. Cornelis, H.F. De Brabander, J. Chromatogr. A 970 (2002) 235.
[19] ISO/IEC 17025, General Requirements for the Competence of Testing and Calibration Laboratories, International Organization for Standardization, Geneva, 1999.
[20] ILAC Document G17: Introducing the Concept of Uncertainty of Measurement in Testing in Association with the Application of Standard ISO/IEC 17025, The International Laboratory Accreditation Cooperation, Rhodes, NSW, 2002.
[21] ILAC Document G19: Guidelines for Forensic Laboratories, The International Laboratory Accreditation Cooperation, Rhodes, NSW, 2002.
[22] Steraloids Inc., Newport, R.I. Sales catalogue 2003 (http://www.steraloids.com/).
[23] Criteria for Reporting Low Concentrations of Anabolic Steroids, International 13 Olympic Committee, Lausanne, August 1998.
[24] S.H.M.A. Linders, A.A.M. Stolker, L.A. van Ginkel, Validation of the analysis of anabolic steroids in beef, pork, turkey and fish, Report 573003013/2003, RIVM, Bilthoven, 2003.