Identification and confirmation of chemical residues in food by chromatography-mass spectrometry and other techniques

Steven J. Lehotay, Katerina Mastovska, Aviv Amirav, Alexander B. Fialkov, Tal Alon, Perry A. Martos, André de Kok, Amadeo R. Fernández-Alba

A quantitative answer cannot exist in analysis without a qualitative component to give enough confidence that the result meets the analytical needs (i.e. the result relates to the analyte and not something else). Just as a quantitative method must typically undergo an empirical validation process to demonstrate that it is fit for purpose, qualitative methods should also empirically demonstrate that they are suitable to meet the analytical needs. However, thorough qualitative method validation requires analysis of a great number of samples (possibly more than can be reasonably done), which is generally avoided due to the time and the effort involved.

Instead, mass spectrometry (MS) is generally assumed to be the gold standard for qualitative methods, and its results are typically unquestioned. For example, a system was developed by European regulators of veterinary drug residues in food animals (2002/657/EC), in which the number of identification points given in MS analyses depends on the general degree of selectivity of the MS technique used. This well-defined approach gives a definite answer for decision-makers, so it has grown in popularity.

However, the identification-points system is not scientific. The reality is that each situation requires information gathering and careful deductive thinking on the part of the analyst to make MS identifications. Rather than devise arbitrary requirements that need to be met by an unthinking analyst, we remind the analytical community that confirmation can be given only if two or more independent analyses are in agreement, preferably using orthogonally selective (independent) chemical mechanisms.

In this article, we discuss the proper use of terminology, highlight the identification power of various MS techniques, demonstrate how MS identifications can fail if precautions are not taken, and re-assert the value of basic confirmation practices, qualitative method validation, information checklists, routine quality-control procedures, and blind proficiency-test analyses.

Published by Elsevier Ltd.

Keywords: Confirmation; Food; Gas chromatography; Identification; Liquid chromatography; Mass spectrometry; Method validation; Pesticide residue; Qualitative analysis; Tandem mass spectrometry

Steven J. Lehotay*, Katerina Mastovska, U.S. Department of Agriculture, Agricultural Research Service, Eastern Regional Research Center; 600 East Mermaid Lane, Wyndmoor, Pennsylvania 19038, USA

Aviv Amirav, Alexander B. Fialkov, Tal Alon, School of Chemistry, Sackler Faculty of Exact Sciences, Tel Aviv University, Tel Aviv 69978, Israel

Perry A. Martos, University of Guelph, Laboratory Services, Chemistry Method Development, 95 Stone Rd. West, Guelph, Ontario, N1H 8J7, Canada

André de Kok, VWA - Food and Consumer Products Safety Authority, Chemistry Laboratory, Pesticides and Mycotoxins Analysis R&D Group, Hoogte Kadijk 401, 1018 BK Amsterdam, The Netherlands

Amadeo R. Fernández-Alba, Department of Hydrogeology and Analytical Chemistry, University of Almería, La Cañada de S. Urbano s/n, 04071 Almería, Spain

*Corresponding author.
Tel.: +1 (215) 233 6433; Fax: +1 (215) 233 6642;
E-mail: steven.lehotay@ars.usda.gov

0165-9936/$ - see front matter Published by Elsevier Ltd. doi:10.1016/j.trac.2008.10.004
1. Introduction

1.1. Definitions
The fundamental purpose of analytical chemistry is to meet needs for qualitative and quantitative analysis of samples.

In qualitative analysis, two questions can be posed:
1) “What is in the sample?” (general screening, structure elucidation, or component analysis); and,
2) “Is the analyte in the sample?” (targeted screening and analysis).

In quantitation, the central question is: “How much of the analyte is in the sample?” However, a quantitative answer should not be given without an acceptable degree of qualitative knowledge that the measured result relates to the analyte alone and not something else.

In this article, we mainly focus on this latter issue of analyte identification/confirmation-which is a major issue in many regulatory, forensic, clinical and other similar applications.

The first consideration to address is that many analytical chemists use the terms identification and confirmation interchangeably, as if they mean the same thing. However, they are defined differently in English dictionaries, and scientific usage should also reflect those differences. We wish to encourage widespread use of the terms below with the following definitions:

1) indication is a non-quantitative result from a general screening method (e.g., immunoassay), for which other factors may cause the result (i.e. “presumed” positive or negative);
2) determination is a quantitative result from a method that meets the acceptable performance criteria for the quantitative purpose of the analysis (e.g., chromatography with an element-selective detector);
3) identification is a qualitative result from a method capable of providing structural information (e.g., using mass spectrometric (MS) detection) that meets acceptable criteria for the purpose of the analysis; and,
4) confirmation is the combination of two or more analyses that are in agreement with each other (ideally, using methods of orthogonal selectivity, at least one of which meets identification criteria).

By definition, confirmation requires that one result must “confirm” the other, so at least two analyses are needed. As a result, a single analysis, no matter how selective it may be, is not confirmatory. The degree of selectivity required to satisfy “confirmation criteria” must also be fit for purpose, depending on the stakes involved [1], and, in most applications, the confirmatory methods should use orthogonally selective (independent) approaches based on different chemical mechanisms, such as liquid and gas chromatography (LC and GC) separations. Furthermore, confirmation of the analyte in the sample, which entails re-extraction of a duplicate sample, is preferable to confirmation of the analyte in the extract, and, when warranted by high enough stakes, multiple laboratories should be involved in the analyses.

Another important term for this article is limit of identification (LOI), which is defined as the lowest concentration for which the identification criteria are met. The identification criteria may be defined in different ways depending on fitness-for-purpose, and LOI should be determined empirically, much as the limit of detection (LOD) and limit of quantitation (LOQ) should be validated in quantitative methods. Alternatively, the concepts of lowest calibrated level (LCL) [2] or minimum required performance limit (MRPL) [3] can be adapted to qualitative analyses, which require demonstration of acceptable performance at a minimum concentration to suit the purpose for the analysis. In this latter model, if the analytical needs dictate that the LOI must be ≤ 10 ng/g, then defined identification criteria have to be met routinely for analyses of samples containing 10 ng/g of the analyte(s). System-suitability and quality-control (QC) tests should be conducted at this level before and during analysis of a batch of samples.

In the LOD model, if the identification criteria in an MS analysis entail that a minimum of three ions with signal-to-noise ratio (S/N) > 3 are needed to make an identification, then the LOI would be the concentration at which the least abundant ion gives S/N = 3 (i.e. the LOD of the least abundant ion is used as the quantitation ion). Typically, the LOD refers to the concentration at which S/N = 3 for the response used in quantitation (and LOQ has S/N = 10); however, the analyst may choose to make LOD (or LOQ) = LOI if the application is important enough that only acceptably identified analytes be reported as detected in a quantitative analysis. Such a decision would decrease the rate of false positives at the expense of increasing the rate of false negatives.

1.2. Sources of error
In measurement science, there are three sources of error:
1) random variability (precision);
2) systematic bias (trueness); and,
3) spurious or gross errors (mistakes).

The first two forms of error are taken into account during method validation to assess the quality of analytical results expected from a particular method. Analytical chemists and clients have devised minimal method-performance criteria for precision and trueness that must be met during validation, QC procedures, and proficiency testing (PT) to demonstrate that the method meets their needs.

However many analytical chemists tend to forget the occurrence and the impact of spurious forms of errors, even though mistakes cannot be eliminated in real-world analyses. In light of the high quality of modern analytical technology and instrumentation, human errors are undoubtedly the greatest source of error with respect to...
qualitative identifications and confirmations. The number of mistakes depends on the diligence and the intelligence of the people performing the work, but, even for the best analysts, the number of human errors will typically be the limitation in assessing qualitative MS identifications. For example, this has been found to be the case for DNA testing, which can have theoretical misidentification rates on the order of 1 chance per 100 billion people (except for being unable to distinguish identical twins), yet the rate of human errors in blind studies to evaluate the performance of analytical processes are measured in terms of percent [4].

Because human error will typically be the limiting factor during the analytical process, an accurate empirical measurement of the rates of false positives and false negatives arising solely from MS techniques is quite difficult, if not impossible, to determine. Statisticians, metrologists, or other scientists can go to great lengths to calculate the probabilities of false positives and false negatives from a theoretical basis of chemical structures, measurement errors, and the numbers of different molecules that exist, but no single approach will be valid for all situations. In any event, spurious forms of error (e.g., mislabeling, laboratory contamination and inadvertent spikes) will remain the most common reasons for misidentification.

Due to spurious sources of error, real-world qualitative analyses should entail not only confirmation of the presence (or absence) of the analyte in the extract, but also confirmation that the analyte originates from the sample. This requires re-analysis of a duplicate sample, ideally using another validated method involving different chemistries of isolation and/or detection. For many applications, such as regulatory enforcement actions, knowledge of the chemical form of the analyte in the sample is equally important (e.g., one metabolite of the original chemical may be legal, but another may be regulated, so the analytical method(s) must be able to distinguish them in order to take valid action).

In derivatization methods, the selectivity of the chemical reaction becomes part of the method, and typically provides a lower degree of selectivity than the MS detection, so MS cannot overcome the inherent limitation of the derivatization step(s) leading to the final results. In any analytical method, the selectivity of the entire sampling procedure, sample preparation and analysis must be considered holistically. The range of chemicals that can possibly be identified from the overall process is limited by the overlapping region among the subsets of chemicals that “pass through” each step in the process.

The topic of analyte identification and confirmation has been the subject of much discussion [1,5–14], especially recently as MS instruments have become more routinely available in many laboratories. This topic is very important in many applications because the stakes can be very high, even a matter of life and death, depending on the analytical results. We intend this article to review critically common approaches to MS identification, present real-world examples of pitfalls when identifying analytes, and remind the reader of basic traditional requirements and commonsense approaches to analytical identification and confirmation.

2. Historical perspective on MS identification

2.1. Dual-column and element-selective detectors

Prior to the widespread introduction of MS in routine residue laboratories, chemists relied on dual-column or separate methods of detection to confirm analytes [15]. This is still the only approach available to many laboratories in developing countries, and some regulatory guidelines officially permit this approach as a confirmatory method, but submission of such a technique by a pesticide or drug registrant to a developed nation would be questioned in this era of widespread availability of MS, tandem MS (MS²) and multi-stage MS (MSⁿ) instruments. The reason for this is demonstrated in Fig. 1 (A and B), which plots the similarities between the relative retention times (tR) of 263 (A) or 315 (B) pesticides analyzed using different GC columns and conditions [16]. Despite the different phases and methods, GC relies on the same physico-chemical mechanism for the separation, so these confirmation methods are not orthogonally selective. This type of dual-column GC approach may still be considered confirmatory (which could be the case even if the same method was used twice), but the value of such a confirmation is limited. The use of different sample-preparation methods and/or element-selective detectors adds to the value of the confirmations [15], but without extensive testing, it is questionable if the relatively low degree of selectivity in the confirmation suits the needs of high-stakes applications.

Although MS detection has not been extensively compared in side-by-side, real-world applications with dual-column GC confirmation, analytical chemists recognize the high degree of selectivity of MS detection in chromatography. It has increasingly been used to replace dual-column techniques for confirmation purposes, but this in part is also how terms “identification” and “confirmation” have become confused. MS was traditionally used to confirm analyte identity after determinations with non-MS methods, but, as with any single method, MS is not able to “confirm” the presence of analytes in the sample by itself. Nowadays, it may be more efficient to use MS to identify chemicals qualitatively in an initial screening method, and then use a traditional, non-MS method to make quantitative determinations and qualitative confirmations at the same time.
a particular molecule depends on a variety of factors, but
dation techniques. The uniqueness of a mass spectrum to
chemists tended to rely on traditional structure-eluci-
limited spectral-library information was available, so
very affordable and less used in routine monitoring labora-
circa 1980, MS instruments were less available, less
Administration scientist James Sphon [20,21].
In the U.S.A., the origins of MS-identification guidelines
2.2. MS-identification criteria
In the U.S.A., the origins of MS-identification guidelines
in regulatory decisions can be traced to Food and Drug
Administration scientist James Sphon [20,21]. Prior to
circa 1980, MS instruments were less available, less
affordable and less used in routine monitoring labora-
tories, and computer technology was in its infancy. Very
limited spectral-library information was available, so
chemists tended to rely on traditional structure-eluci-
dation techniques. The uniqueness of a mass spectrum to
a particular molecule depends on a variety of factors, but
a general guideline established through empirical dem-
onstrations asserted that three ions of the proper ratio
give enough selectivity to identify most compounds. The
experiences of many analytical chemists supported this
conclusion [1,22,23], but few would say that it should
be a “rule” because many exceptions can be found that
indicate a three-ion requirement is either too strict or not
strict enough.
For example, Sphon used an early mass-spectral
library to compare the spectra of ≈30,000 compounds
in the database with that of diethylstilbestrol (DES). He
found that isolating the three most intense ions in the
DES spectrum, with liberal relative-abundance criteria,
specifically distinguished DES from any other compound
in the library. Table 1 shows the results from this
demonstration, including an update from 1997 [22],
and a new set of results using NIST mass-spectral
libraries. As Table 1 shows, the use of the three most
appropriate ions, even with rather wide ion-ratio con-
straints, eliminated the possibility that DES would be
confused with any other compound in the libraries. In
one case, the use of only two ions with reasonable
constraints (m/z 268 and m/z 145 with ±10% permit-
ted ion-ratio variability) isolated DES from all other
107,885 compounds in NIST’98 (but two additional
compounds would be listed as possible hits among the
163,198 compounds in NIST’05). However, Table 1
also shows that if, in addition to m/z 239 as a qualifier
ion, the second most intense ion in the DES spectrum,
m/z 107, is chosen as the second qualifier ion, rather
than m/z 145 (the fourth strongest), three other com-
ounds overlap with DES in a search of the NIST
libraries. The choice of m/z 107 is not unreasonable
compared to m/z 145 in a selected-ion monitoring (SIM)
program, except in this case it may lead to more pos-
sible interferences. Moreover, if the molecular ion and
base peak, m/z 268, is excluded (e.g., due to an inter-
ference) and the three most intense fragment ions are
chosen instead, two different chemicals can give rise to
rather stringent ion ratios shown in Table 1 for the
three-ion “identification” of DES.
The above consideration refers only to the possible
number of compounds from NIST libraries that can be
confused with DES, while an unknown quantity of
compounds that are not in the library could also inter-
fere in the identification. For this general reason, Sphon
was careful not to provide anything other than “guide-
lines” for MS identification because no single set of rules
would necessarily apply to all situations.
Despite this, others in the U.S.A. began using what
became known as the “three-ion criterion” for mass-
spectral identification [1,22]. Simply put, at least three
ions of the correct m/z and relative-abundance ratio
±10% (absolute) are desired to make a mass-spectral
match. Other factors needed for identification generally
include proper tR (±2% error factor) and sufficient S/N

Recent documents about identification and confirmation
issues have focused on MS techniques, but confirma-
tory approaches should not exclude the use of other
methods. For example, the illegal usage of an insecticide
isofenphos-methyl was uncovered by nitrogen-phos-
phorus detection [17], and the elemental information
using pulsed flame-photometric detection can be very
valuable [18], particularly when it is employed simulta-
neously with MS [19].

### Table 1

<table>
<thead>
<tr>
<th>Compound</th>
<th>m/z 145</th>
<th>m/z 239</th>
<th>m/z 107</th>
</tr>
</thead>
<tbody>
<tr>
<td>DES</td>
<td>145</td>
<td>239</td>
<td>107</td>
</tr>
<tr>
<td>Isofenphos-methyl</td>
<td>145</td>
<td>239</td>
<td>107</td>
</tr>
</tbody>
</table>

**Figure 1.** Relative retention time (vs parathion) comparison for GC
analysis of pesticides using different 30 m, 0.25 mm i.d. columns
and oven-temperature programs by the Dutch Food Inspection Ser-
vice [16]: A) 263 pesticides comparing DB-1 and DB-5 phases; and,
B) 315 pesticides comparing DB-5 and DB-1701 phases.
(> 3) for the chromatographic peak of the least intense ion.

To provide greater stringency and more precaution against false positives, the European Union (EU) approach required four matching MS ions to make identification of banned substances in foods, rather than the three ions in the proper ratios recommended by Sphon. In time, this decision led to the establishment of the identification-point (IP) system as a “requirement” for identification of organic residues and contaminants in samples of animal origin in the EU regulatory system [5,24]. This constitutes a departure from the traditional approach in which only “guidelines” are provided, not “requirements.” For pesticide analysis in the EU, guidelines are still preferred [2].

The IP system has some practical benefits in that decisions can be made using clearly defined criteria, but, as in the case of essentially all identification guidelines to date, a critical drawback is that a rigorous assessment has not been conducted to determine the uncertainty of the approach(es). For example, what are the differences in the rates of false positives and false negatives by requiring four IPs for banned substances over three IPs for registered compounds? Why should a high-resolution ion always be worth two points in the IP system, and MS² ions always be worth 1.5, whereas the (pseudo)-molecular ion is only worth 1? What is defined as “high” resolution?

### 3. MS techniques and their relative selectivity and identification power

Without question, MS is currently the most powerful tool commonly available to analytical chemists for identification of organic compounds in a variety of matrices. We do not wish to disparage the well-known positive attributes of MS techniques but we want to remind analysts that MS is not a panacea, as some people would like to believe. Chemical identification is strongly affected by MS technology, which is constantly evolving and improving with time. In this section, we briefly discuss a few relevant MS techniques and compare their capabilities and limitations for identification purposes.

#### 3.1. Importance of the molecular ion

Some ions (in particular, the molecular ion) may be given more weight than others in an identification, but how can this weight be assessed? One possibility is to collect information about the potential number of interferences using data libraries and various algorithms. It is commonly accepted that the presence of the molecular ion in a mass spectrum enhances confidence in the sample identification for the following reasons:

a) it is the highest mass ion, so it tends to have the least amount of chemical or matrix interferences;

b) it safeguards against misidentification of homologous and degradation products; and,
c) it enables additional tools for identification purposes, either via isotope-abundance analysis (IAA) or via accurate-mass-related elucidation of the empirical formula.

Although atmospheric pressure ionization (API) techniques in LC-MS often yield the (pseudo)-molecular ion (e.g., [M+H]+), formation of adduct ions with sodium or ammonium ions also occurs frequently. In electron ionization (EI), the molecular ion is practically absent in about 30% of the mass spectra [25], and such a high probability of its absence introduces doubt that the highest mass ion present in a mass spectrum is indeed the molecular ion. In fact, the highest mass spectral peak could be a high mass fragment ion or emerge from an impurity. A unique way to enhance molecular ions in EI is by using supersonic molecular beam (SMB)-MS, which has been reviewed recently [26]. With cold EI (EI of vibrationally cold molecules in SMB-MS) the molecular ion is enhanced and is practically always observed [19,26] while cluster chemical ionization can be used to ascertain further the validity of the suspected mass-spectral peak as the molecular ion [27].

### 3.2. High-resolution MS and accurate mass

High-resolution MS (HRMS) is becoming more popular in laboratories, particularly in the form of time-of-flight (TOF) MS, while it can also be found in magnetic sector, Fourier transform (FT) MS, Orbitrap, and even quadrupole MS with software-calibration enhancements. For this discussion, we must distinguish between accurate mass and high resolution. For example, TOF-MS can have mass accuracy of 2 ppm while its full-width half-maximum resolving power is of the order of only 10,000. Thus, in comparison with FTMS or Orbitrap, which has similar mass accuracy but a higher resolving power of 100,000 [28], TOF-MS has less power to reduce matrix interferences than its capability to generate possible elemental formulas.

Accurate-mass TOF is more powerful in LC-API-MS (because the pseudo-molecular ion is nearly always present in the spectrum) than in GC-EI-MS because, as stated in Section 3.1, about 30% of the chemicals do not exhibit the molecular ion [25]. Also, common EI-MS library-search programs require unit-mass resolution to work properly (although that could be simulated with appropriate software).

Usually, the various vendors who sell high-resolution TOF instruments describe the power of high resolution in qualitative terms, which makes it hard to compare it to other options (e.g., MS²), so we wish to provide a more quantitative picture of the merit of HRMS. For this purpose, we used the NIST02 library, making the assumption that the library accurately represented the distribution of organic molecules in nature. For MW = 304 amu, the library contained 628 compounds, including our example-target analyte, diazinon. Fig. 2 shows how these 628 compounds give a MW range of 303.8–304.4 amu, with a distribution (full-width at half maximum, FWHM) of 0.166 amu. This does not include the numerous possibilities for fragment ions in this m/z region that can further complicate the situation, especially considering the typically low residue levels for the analyte among complex matrix components at high concentrations.

If the target analyte contains many elements with a large mass defect, such as Cl and Br, then HRMS truly excels in the suppression of matrix interference. Whereas hydrogen adds only +0.008 amu mass defect per H atom, each Cl atom has mass defect of −0.032 amu, and Br has an even greater mass defect of −0.083 amu. Other common elements in organic molecules have the following mass defects (amu): C = 0, N = 0.003, O = −0.005, F = −0.002, P = −0.026, and S = −0.028 [29]. The relatively large mass defects for Cl and Br cause a significant shift from the center of the mass distribution for a given MW (see Fig. 2), which improves selectivity of analysis for HRMS. Other MS approaches also give enhanced detection of halogenated compounds, such as negative-ion chemical ionization or even EI in full-scan mode with isotope-abundance analysis. However, the superior capabilities of HRMS are fully demonstrated in the analysis of dioxins, PCBs, and similarly multi-halogenated analytes [5,30].

Otherwise, most organic chemicals (analytes and matrix components) do not contain Cl or Br, and are located near the center of the histogram, as shown in Fig. 2 for diazinon. In this example with resolution of 10,000, about 72 other compounds (11.5% of those with MW = 304 amu) in the NIST02 library would have the same measured mass, so a nine-fold improve-

---

Figure 2. Histogram showing the exact-mass distribution of the 628 compounds with nominal molecular weight of 304 amu in the NIST02 mass-spectral library. The calculated average mass is 304.123 amu with a standard deviation of 0.083 amu. The exact mass of diazinon is marked, and the depicted region demonstrates how 72 other compounds in the library cannot be resolved from diazinon with MS resolving power of 10,000.
ment in selectivity over unit mass-resolution instruments is realized, which is good but not sufficient for full elimination of matrix interference.

However, the ability to determine accurate mass enables the analyst to have a table of possible elemental formulas listed in declining order of matching the experimental mass. Such a measurement provides a strong, independent (orthogonal) tool for analyte identification, which can serve to confirm other methods of identification. To a certain extent, accurate mass can also be obtained with “unit” resolution quadrupole MS, and commercial software algorithms can accurately locate the center of the mass-spectral peak. If a typical sample mass (e.g., diazinon at m/z 304) is measured with 5-ppm accuracy, this translates into 0.0017-amu accuracy, which reduces the number of possible compounds by an impressive factor of ≈100. However, the level of confidence in sample identification by HRMS cannot be greater than the level of confidence that the evaluated mass-spectral peak is indeed the molecular ion, and this should not be underestimated. If the analyzed ion is actually a fragment ion, unknown to the analyst, then the identification will be wrong.

### 3.3. Isotope-abundance analysis

The relative abundances of the various isotopomers (molecular ions with different isotopes) can provide accurate elemental formulas. Traditionally, the IAA approach has served to help elucidate molecular weights and chemical structures of synthesized organic compounds [31]. Recently, a unique IAA method and software were developed [32], among other things, to link with the NIST MS library and automatically support or reject the proposed library identification. In case of a rejection, the IAA software independently provides a list of elemental formulas with declining order of matching the experimental data, similar to accurate-mass measurements, but IAA does not require costly accurate-mass MS instrumentation.

Due to the low intensity of the isotopomer ions, key for use of the IAA approach is very low noise with few chemical and background interferences.

There is also demand for absence of protonation (due to chemical ionization or self-chemical ionization).

SMB-MS and SMB-MS² can provide high sensitivity with low background noise while their collision-free fly-through EI process excludes undesirable molecular ion protonation [33], which allows the use of IAA even at trace concentrations.

Although the IAA software was developed and tested mostly with GC-SMB-MS, IAA is also being used effectively to complement and to supplement accurate-mass data with high-resolution LC-MS. Again, as with accurate mass, the level of confidence in any sample identification, including combined IAA and NIST library search, cannot be greater than the level of confidence that the evaluated mass spectral peak is indeed the molecular ion.

### 3.4. Comprehensive two-dimensional GC×GC-MS

Comprehensive GC×GC-MS serves as another method for improved sample identification through improved GC separation and hence reduced matrix interference. Typically GC×GC has a theoretical gain in separation power by about 20 through having a second GC time window of 4 s with an average peak width of 0.2 s [34]. However, in reality, the GC×GC separation power is significantly lower because a proper GC×GC analysis requires that each first-dimension GC peak width will have to accommodate three to four GC×GC cycles. As a result, the GC×GC analysis time is generally longer, due to the need to generate broader GC peaks, so the GC×GC separation power should be compared with a one-dimensional GC separation with a longer column. Also, the second-dimension separation is often not fully orthogonal to that of the first dimension (see Section 2.1).

Despite the above criticism, GC×GC-MS is a powerful analytical tool that can particularly excel in the separation of polar samples from non-polar matrix interferences, and, as a result, improve library identification. Furthermore, the added separation power of GC×GC-MS may allow high-quality identifications using two ions (in reconstructed SIM) instead of following the traditional guideline of three ions. This would provide an even lower LOI since the third ion is typically the one with the least abundance and greatest chance for matrix interferences. The peak-focusing effect of the second-dimension column also increases S/N, which has the effect of further lowering LOI compared with one-dimensional GC separations.

### 3.5. MS² and MSⁿ

MS² is a powerful technique that uniquely combines improved sensitivity and selectivity; however, like SIM, it comes with the price of being a target-based method, which misses any compound that is not in its target list (so there is an inherent chance for many false negatives).

While there are many examples in the literature showing how MS² excels in reducing matrix interference and typically lowers the LOD, we are not aware of any publication that has explored how and when MS² fails and the factor by which it improves selectivity. In EI-MS², the inherent dissociation energies for any given molecule lead to the same fragments as it generates in EI mass spectra but with different relative intensities. Thus, the MS² product ions are not a matter of random distribution of all masses from m/z 1 to the precursor ion minus 1. If we analyze the NIST library of EI mass spectra, we conclude that, for a parent ion with m/z 250, only about eight masses are preferred in the mass range m/z 50–249. For example, hydrogen loss is usually unlikely and the next option is a loss of CH₃, which is a
difference of 15 amu apart with a void mass space of 14 amu. In LC-API-MS, the use of (pseudo)-molecular ions as MS² precursor ions further restricts the number of available possibilities for product ions.

Moreover, a small percentage of organic chemicals are not suitable for MS² analysis, particularly when EI is used, due to ion instabilities (or excessive stability to allow further fragmentation with typical instruments), lack of enough product ions, or low formation of high-mass ions in MS to allow MS² (e.g., in terms of identification power, MS² lacks the ability to exclude 150,000 other compounds in full-scan spectral libraries). However, in terms of targeted analysis, MS² has a second dimension of selectivity in its collision-induced dissociation voltage settings, which is part of the optimization process.

4. Limitations of current guidelines and rules

Milman has given an excellent overview of different MS identification criteria used by different organizations [10], and we do not repeat it here, but the information is helpful in order to improve understanding of the discussion in this section.

4.1. Targeted monitoring in MS and choice of ions

There are some major difficulties in implementing typical MS-identification guidelines and rules, especially with GC-quadrupole MS because its use in the analysis of chemical residues tends to require SIM mode to achieve the necessarily low LOD. It is quite common for analytes to yield only one or two ions of adequate intensity, even in EI, so it is impossible to identify many analytes at reasonably low concentrations using three or four ions.

Even when an analyte has ≥ 3 intense ions, some of the ions chosen for analysis over the range of targeted compounds frequently have an interference from the sample extracts at the tR of interest, so different ions must be chosen, depending on the analyte-matrix pair, to increase selectivity and minimize LOI. The choice of SIM ions can be difficult, especially in multi-residue analysis of complex matrices with scores of analytes in the method. A trade-off must be made in the number of analytes that can be included in the method vs time and the number of qualifier ions, all depending on background matrix interferences.

Furthermore, in targeted approaches (e.g., SIM and MS²), “false negatives” are guaranteed to occur for non-targeted compounds if they occur in the sample. Similarly, chemicals of interest or importance that are not recovered or detected by any particular method could be considered false negatives if they occur in the sample. This is why scope of analysis (or analytical range) is often the most important feature in a method, particularly in regulatory screening applications.

With respect to choice of ions and their relative selectivity, as Fig. 3 shows in the case of the NIST⁹⁸ database, ions of higher m/z have less chance of potential interferences than those with lower m/z. In a real application involving the GC-MS analysis of pesticides in very complicated spice extracts, the degree of matrix interference was shown to reduce exponentially by a factor of 20-fold per each 100-m/z increase [35]. This study also demonstrated how the presence of the molecular ion in the spectrum greatly enhanced the ability to isolate the analyte peak in a complicated matrix; however, this pattern is a general trend and any particular situation depends on the combination of analyte, concentration, matrix and method.

Some chemists have set policies for MS-identification decisions to account for the general trend shown in Fig. 3. For example, some criteria had dictated that ions with m/z <91 or many ions from a chlorine or bromine cluster should not be used for identification purposes [36]. The intention of such policies is to reduce the chances of false positives by analysts with poor judgment, but their arbitrariness can eliminate valuable information and preclude the use of new technologies that could result in improving identifications. For instance, chlorine-ion or bromine-ion clusters indicate both the presence and the number of those atoms in the molecule, and that eliminates a great number of other possible chemicals in the identification.

Another common predisposition is to avoid use of ions with <10% relative-ion abundances [5]. A very important aspect to consider when choosing quantitation and qualifier ions in MS is to maximize S/N and thereby minimize LOD or LOQ and LOI, but this does not mean that ions with relative abundance <10% will not yield higher S/N than ions with higher intensity. As shown in an example for permethrin [35], the presence of the molecular ion at the proper tR with adequate S/N, even if

![Figure 3. Number of spectra in the NIST⁹⁸ mass-spectral library plotted vs m/z of the base peak (relative abundance = 100%) and m/z of peaks with relative abundance >5%.

http://www.elsevier.com/locate/trac 1077
it was only 1% relative-ion abundance, is a powerful aid in the identification of analytes. IAA also works on the premise that small isotopomer ions may be used to assess molecular formulas of the chemical to help make identifications. Such powerful approaches should not be eliminated arbitrarily.

Fig. 4 demonstrates the effect of choosing different MS ions in GC-MS (SIM) for the same application in a different matrix is demonstrated in the analysis of pirimiphos-methyl in carrot and orange extracts vs added concentrations. Note how the relative intensities and variabilities of the ion ratios “change” for the same analyte in different matrices, due to chemical noise (matrix interferences). For carrots, m/z 125 should be avoided, but, for oranges, the base peak with m/z of 290 has a co-eluting interference. This is not unusual, and the analyst has to be aware of the different situations and take appropriate precautions. However, the wide diversity and large number of food, environmental, forensic, clinical, and other possible matrices makes it impractical to choose different qualifier ions depending on the interferences in each matrix [37]. Ultimately, the analyst must take good care and use sound judgment, especially when using SIM to make identifications.

4.2. Selectivity
The key to identification of chemicals is not necessarily to characterize the analyte to the greatest extent possible, but to exclude the possibility that any other molecule possesses the same measured trait(s). Current guidelines and rules narrow the focus (e.g., more ions, MS², tighter ion-ratio restrictions, and higher mass resolution), which tends to increase selectivity in general, but not necessarily in particular. An example is how MSⁿ of the same product ion generated from different compounds accumulates more information, but has no value in distinguishing between the different compounds. Fig. 5 demonstrates this situation in the case of MS² of ethion and terbufos using the same fragment, m/z 231, which is the base peak in their full-scan mass spectra. Another example is that higher mass resolution to obtain exact-mass measurements cannot distinguish between isomers with the same molecular formula. Some molecules (e.g., polychlorinated biphenyl or dioxin congeners) produce almost identical mass spectra, and a high-resolution chromatographic separation must be relied upon to help make the identification [30]. In that case, highly selective sample preparation followed by high-resolution GC (or GC×GC) with selective detection of chlorinated compounds would probably be more selective than a rapid GC-MS method using high-resolution TOF. In essence, the central failing of prescribed mass-spectral identification rules is that they do not apply in all situations [37].

The lack of a unified theory for the concept of selectivity in analytical chemistry is an underlying cause of the problem [12]. For example, it is impossible to provide the degree of selectivity in a method to such an extent that it demonstrates true specificity (the result can only originate from the analyte and no other factor). In science, a hypothesis, such as “this chemical is in the sample,” cannot be proved with 100% confidence, and experimental evidence can only lend further support to or dis-
prove the hypothesis, not prove it. Unlike quantitative analysis, qualitative analysis does not provide the actual degree of confidence that a result is accurately known because chemical selectivity does not follow a normal distribution pattern. We have already mentioned that spurious errors probably predominate and that mistakes also do not follow a normal distribution. In trace-residue analyses in complex matrices, perhaps qualitative answers with \( \approx 95\% \) confidence can be determined through a practical validation process [5], but statements of >99.9% or >99.99% confidence become problematic without a solid theoretical foundation for measuring selectivity and/or the analysis of many samples in carefully controlled and perfectly conducted experiments.

5. Real-world pitfalls in qualitative analysis

Everybody makes mistakes, and every analyst can give examples of circumstances that led to a mistake. Learning from the mistakes of others is better than learning from our own mistakes. In this section, we present some examples of real situations that have led the analyst either to make a misidentification or to take extra precautions in qualitative MS analyses. We provide these examples of real-world pitfalls to underscore the points made in this article, and perhaps help the reader recognize when a set of circumstances leading to a possible identification or confirmation is not as clear as it may seem.

5.1. Chemical degradation (e.g., GC-MS of carbamate pesticides)

In a presentation at a pesticide residue workshop, an MS expert showed how a targeted pesticide was detected in a difficult pepper sample using GC-MS in full-scan data-acquisition mode with mass-spectral deconvolution and contemporaneous mass-spectral library matching. The calibration curve was linear and no matrix interferences were observed. The deconvoluted mass spectrum (which was well isolated from the complex background) was essentially identical with the reference spectrum for the pesticide in his mass-spectral library of targeted analytes, which was generated by contemporaneous injection of certified pure reference standards using the same method on the same instrument. The spectrum obtained with EI possessed five intense ions of >15% relative abundance that gave a very similar pattern to the spectrum in the NIST mass-spectral library for the pesticide. The \( t_R \) was exactly the same, as was the peak shape, with respect to the reference standard for the analyte. The combination of all of these factors met the GC-MS confirmation criteria established for chemical residue analysis by several organizations [5,12,36], so, for all intents and purposes, the pesticide was “identified”.

Despite this, knowledgeable and experienced pesticide chemists in the audience knew that the finding was incorrect. The chemist claimed to have identified oxamyl in the sample by a GC method employing splitless injection, but oxamyl is thermally labile and cannot be analyzed directly at the GC conditions used. The method had actually detected oxamyl oxime, which was produced by thermal decomposition of oxamyl at the hot GC injector.

Typically, LC methods are used for the analysis of carbamate pesticides. Alternatively, as shown in Fig. 6, they can be analyzed directly by GC-SMB-MS [38], which can provide EI spectra at \(-70\) eV to show a prominent molecular ion, particularly for thermally-labile chemicals (e.g., oxamyl), and that is a major advantage of SMB over traditional GC-MS.

It is hard to fault the MS expert in misidentifying oxamyl in the conventional GC-MS analysis. The oxamyl-oxime metabolite appears in the NIST MS library, but it is listed as ethanimidothioic acid, 2-(dimethylamino)-N-hydroxy-2-oxo, methyl ester. The lower traces in Fig. 6 show the mass spectra of oxamyl and its oxime in the NIST’02 library (upper trace is the mass spectrum of cold oxamyl obtained using GC-SMB-MS [27]). Given the choice between the exotic sounding name given above and oxamyl, which was the injected compound after all, the chemist figured that oxamyl was the correct molecule and unwittingly entered the spectrum of oxamyl oxime as oxamyl into the MS data-processing software.

![Figure 6. Comparison (top to bottom) of cold EI mass spectrum of oxamyl using GC-SMB-MS [27] and the NIST library mass spectra of oxamyl and oxamyl oxime. Conventional GC-MS using EI does not yield the \( m/z \) 219 molecular ion, and hot splitless injection causes the conversion of oxamyl to its oxime in the inlet.](http://www.elsevier.com/locate/trac)
This type of mistake is not uncommon; the compound injected is not necessarily the compound detected or assigned in reference libraries (analysts have also uncovered misnamed spectra in commercial mass-spectral libraries), especially if the molecular ion is absent in the EI mass spectrum of that compound. There can be other sources of errors (e.g., mislabeling the reference standard, assigning the wrong chromatographic peak as the analyte in a mixture, or not performing background subtraction properly).

In another example, the MS expert had also included methomyl in the list of pesticide analytes in the GC-MS method. Like oxamyl, methomyl is thermally labile and readily converts to methomyl oxime in certain solutions (e.g., methanol) and/or during traditional GC analysis at both the injector and column. Again, the analyst obtained a consistent peak with strong ions of $m/z$ 105, 88, and 58, which looked much like the NIST library spectrum for methomyl. Unlike oxamyl and its oxime, methomyl oxime is not included in the library and the US tolerance definition for methomyl does not include its oxime [39]. To further complicate matters, the maximum regulatory limit for a separate pesticide, thiocarb, includes methomyl in its tolerance definition. If the pepper sample had contained methomyl rather than oxamyl and/or oxamyl oxime, then it would have appeared to the analyst that the presence of methomyl was identified in the sample with little doubt. In fact, the chemist would have unknowingly identified the presence of methomyl oxime in the sample, which has no regulatory bearing in the U.S.A., and which could have arisen from unquantifiable concentrations of methomyl, methomyl oxime, and/or thiocarb in the original sample.

In these cases, the presence of the molecular ion in the mass spectra of the parent molecules would have ensured that the oxime metabolites were not confused with the larger pesticide analytes. This example illustrates both importance of the presence of the molecular ion and also the selection of an appropriate analytical methodology that eliminates undesirable analyte conversion. In most cases, despite identification guidelines that typically require more ions for structure elucidation, it is actually the presence of distinct ions and other critical information (e.g., analyst knowledge and sample history) that tends to yield better results. Moreover, the use of a more appropriate type of approach, such as LC-MS$^2$, would have provided the evidence needed to find the errors. These examples also demonstrate the importance of using orthogonally selective (independent) approaches for the purposes of confirmation.

5.2. Similar analytes (e.g., LC-MS$^2$ of anthelmintic residues)

In another example involving LC-MS$^2$ method development for the residue analysis of veterinary drugs in cattle tissues, two analytes with the same nominal molecular weight [i.e. albendazole sulfone (MW = 297.328) and hydroxy mebendazole (MW = 297.313)] co-eluted using a C-18 column (15 cm x 3 mm i.d. with 5-µm particles) with a typical water/MelOH/MeCN buffered reversed-phase gradient [40]. Fig. 7 shows their structures and fragmentation pattern in electrospray positive MS$^2$. Without knowledge of this situation, the analyst would not be able to distinguish between the two chemicals in real samples using typical MS$^2$ instruments with unit-mass resolution. The situation is further complicated given that hydroxy mebendazole needs a much higher regulatory LOD than albendazole sulfone in the EU (there is no LOD for mebendazole and its important metabolites, whereas the maximum residue limits (MRLs) for alben- dadazole and its important metabolites are 100 ng/g in milk and 1 µg/g in bovine liver). In this case, a high-mass-resolution instrument could help resolve these compounds, but improving their chromatographic separation with a narrower column, smaller particles, and/or more selective stationary phase would also provide a good solution.

5.3. Identifying unknowns (e.g., isofenphos-methyl)

One of the most difficult jobs for an analyst is to find a completely unknown chemical in a sample. Unscrupulous athletes, racing-animal trainers, and food producers use illegal drugs or pesticides in an attempt to gain a competitive or monetary advantage, and they try to use chemicals that cannot or will not be detected. For example, farmers in Spain apparently obtained an unregistered pesticide, isofenphos-methyl, from a manufacturer in China and applied it to a crop of peppers [17]. Since there was no registration for the pesticide, the analytical laboratories had no knowledge of it and there was no reference standard for it. It is common practice in some chemical-residue laboratories to use GC-MS (SIM) and LC-MS$^2$, which can detect only targeted analytes. These methods are guaranteed to miss any chemical of possible interest that is not targeted in the method. Interestingly, the isofenphos-methyl...
residues were initially found in food extracts as a curious extraneous peak in GC using nitrogen-phosphorus detection. Full-scan GC-MS was used to make the identification and to obtain the multi-analysis confirmation after a reference standard of the illegally applied chemical had been synthesized.

Some analysts choose to use GC-MS in full-scan mode to provide the chance to find unknowns (and increase the potential number of analytes targeted). One advantage for those laboratories faced with the isofenphos-methyl situation was that they re-examined the total-ion chromatograms from earlier analyses to identify the previously non-targeted analyte. In the case of LC, full-scan MS with API is not so useful because typically only the pseudomolecular ion appears, and that is insufficient for making identifications even in high-resolution applications. The ability to assess the MW or the molecular formula of a chemical is helpful, but additional information is needed to make an identification in MS. Ferrer and Thurman [41] showed that it was possible to obtain in-source fragmentation in LC-TOF-MS to analyze 101 pesticides while maintaining acceptable sensitivity.

In full-scan EI-MS, there are typically so many matrix peaks (with many overlaps) in chromatograms that it is much too time consuming to evaluate the spectra for each peak [42,43]. Deconvolution software exists to help [44,45], but most chemicals in the chromatograms (e.g., the examples of isofenphos-methyl and methomyl oxime) do not appear in even the most extensive mass-spectral libraries. Furthermore, spectral matching with a contemporaneously analyzed reference standard is commonly required to make an identification, and simple matching with a library spectrum is not sufficient for a variety of reasons (e.g., mistakes in the library, differences between instruments and conditions, or incorrect tR). This topic has been debated in the literature [46], and experts in the American Society for Mass Spectrometry have devised guidelines that emphasize the need for contemporaneously analyzed reference standards [1].

5.4. Situations without a true blank (e.g., acrylamide in processed foods)
To ensure the quality of analysis, one of the most important factors, which is commonly absent from typical identification criteria, is that the analyte signal does not occur in the analysis of a blank matrix. A basic control experiment in scientific investigations is to isolate the parameter of interest (in this case, “that the signal occurs because the analyte is present in the sample, and not an artefact”). Analysis of reagent and matrix blanks is standard practice in QC procedures, which are designed to exclude certain factors (e.g., instrument-memory effects, carry over, laboratory or reagent contamination, or misidentification due to matrix interferents).

There are certain circumstances in which no blanks can be found because all samples tested contain the suspected chemical. A few examples of this are ultra-trace findings of certain ubiquitous persistent organic pollutants in environmental samples or semicarbazide in shrimp using very sensitive MS instruments. In these cases, the ultralow background levels are averaged, and a positive finding of interest only occurs when the calculated concentration exceeds an “action level” or the MRPL [3].

In other cases (e.g., the occurrence of acrylamide in certain types of processed food), the parameter of interest cannot be isolated in a blank without also generating the chemical that needs to be identified. In the case of acrylamide, no blanks of crackers, potato chips, French fries, and similar cooked products containing asparagine and reducing sugars [47] could be prepared because acrylamide is formed during the process used in making the food products, so, when acrylamide was found in these food products using LC-MS2, there was still a chance that there had been a false identification (or quantitative overestimation) because chemical interferents could not be excluded using a true matrix blank. Acrylamide is a small molecule (71 amu), and an interferent in potato chips (presumably the amino-acid valine that generates the m/z 72 immonium ion) was observed during sample-preparation experiments [48].

In this situation, confirmation by an alternate method, namely GC-MS2, was done to verify that the acrylamide had been found by both distinct methods at similar concentrations in the samples. But, even then, the analyst had to demonstrate that analytical artefacts had not occurred, because acrylamide can be formed from its starting materials in the extract during the hot injection process in the GC inlet [49] (or during any sample-preparation steps using enough heat). Incredibly, it is possible that the analyst can follow all standard practices for two independent MS identifications for orthogonally selective confirmation of acrylamide, and still be wrong if an interferent is misidentified in LC-MS2 and acrylamide is formed during sample preparation and/or during injection (split/splitless) in GC-MS2. In these instances, there is no substitute for an informed, careful chemist using well-validated methods.

5.5. Contaminated reference standards (e.g., erucamide and acetyl-deoxynivalenol)
One of the most insidious problems in developing targeted MS methods (e.g., SIM and MS2) is when a reference standard is mislabeled, has degraded, or contains a contaminant. Just because a bottle has a chemical listed with a given purity does not mean it is always correct, but this type of false information can be difficult to uncover.

In an example, reference standards for 3-acetyl-deoxynivalenol (DON) and 15-acetyl-DON were con-
taminated with erucamide. LC-MS\(^2\) conditions are typically optimized using infusion, and the largest peak is assumed to be the standard, since it typically has nearly 100% purity. The erucamide [M+H]\(^+\) was \(m/z\) 338, with its C-13 isotope at \(m/z\) 339, and, unfortunately, the [M+H]\(^+\) for 3- and 15-acetyl-DON was also \(m/z\) 339. With infusion, one could not therefore optimize the system for either of those compounds, since erucamide at \(m/z\) 339 created analyte-optimization problems, but an uninformed analyst would not recognize the issue.

In this example, the analyst was aware of the issue and chromatographically resolved erucamide from the target mycotoxins in LC-MS\(^2\) rather than using infusion. Column fractionation of the impurity and analysis by UV/VIS revealed that it gave no UV/VIS signal, yet fractionation of the standard showed a UV/VIS spectrum. Analysis of the contaminated standard by LC-UV/VIS-MS\(^2\) also indicated this. So, it was possible in this example to have an impurity in the standard that could be detected by LC-MS\(^2\). However, other detectors could assist the analyst in demonstrating the impurity chromatographically had no impact on the identification and quantitation of the target compound, but would have resulted in seriously incorrect optimization for the target compounds. Basically, information from an independent method (e.g., UV/VIS in this example) can be used in addition to MS techniques to aid in identifications as well as making confirmations.

5.6. Spurious errors (e.g., misteaks)
We mentioned that spurious forms of errors are probably the most common reasons for misidentifications in working laboratories. There are numerous examples of human mistakes that we have found, but the more troubling factor relates to the mistakes we have not found. In the following paragraphs, we give just a few examples that illuminate potential laboratory errors.

An analyst in a laboratory had been preparing a concentrated stock solution of many pesticides using disposable pipette tips. Even though the used tips were placed in a box clearly labeled as waste, they were accidentally confused as being new. A different analyst then used these tips to transfer sample extracts for routine monitoring to autosampler vials for analysis. Each extract in the sequence was found to contain a different pesticide at a rather high level. All identification criteria were met in each case, and the matrix and reagent blanks were clean, but the analyst questioned the results, and, fortunately, the cause of the problem was found.

In another example, high concentrations of endosulfans were found in river water [50]. Unknown to the analysts, the same pesticide had been applied in an agricultural field trial, and run-off water samples were prepared using the same filter apparatus in the laboratory. The other group of analysts thought they had rinsed the apparatus very well – but not well enough. It turned out. The pesticides were confirmed to be present in the extracts by independent GC-MS analyses, but the extracts were contaminated during the filtration steps, so the source of the pesticides in the samples had not been confirmed.

In a blind comparison study of samples shared between two laboratories to compare the different methods used, a student was asked to review the findings from the GC-MS (full scan) because the chemist did not trust the software to do the job automatically. Indeed, the instrument missed a large peak for a pesticide (phosmet, which gives only two strong ions), but the student was day-dreaming during the tedious review process and also missed the obvious peak. The chemist was embarrassed when the other laboratory identified the pesticide residue, and the chemist’s laboratory had not. The student’s excuse was: “I was part of the method, and if there was a false negative, then it is still the method’s fault.”

5.7. Analyte derivatization (e.g., nitrofurans and dithiocarbamates)
The use of the antibacterial agents nitrofurans is banned in many countries, due to their mutagenic and carcinogenic effects. In the analysis of nitrofurans in animal tissues, the common analytical method calls for overnight acid hydrolysis and derivatization of the sample followed by LC-MS\(^2\) analysis [51,52]. Nitrofurans are rapidly metabolized to smaller molecules, [e.g., furazolidone to 3-amino-2-oxazolidinone (AOZ) or nitrofurazone to semicarbazide (SEM)] that bind to proteins in the tissue. For nitrofurazone converting to SEM, Fig. 8 shows how there are different possibilities that can lead to the same detected derivatized analyte(s) in the method. Another possibility is that the chemicals can already be present in the sample, but originate from a source other than nitrofurans.

When the new nitrofuran method was first implemented, 30% of positive nitrofuran findings by an EU veterinary-drug reference laboratory in The Netherlands were for SEM, mainly in prawns and shrimp, at typical concentrations of \(\approx 1\) ng/g [53]. The other analyte commonly detected was AOZ, which like SEM (MW = 75 amu), is a rather small molecule (MW = 102 amu) comprising common elements C, H, N, and O. Many enforcement actions have been taken on nitrofuran findings in shrimp imported to the EU, which has led to large economic losses by exporters and producers [54]. In the meantime, SEM was shown to occur in packaged foods, most probably arising from plastic sealing rings and carrageenan [55]. However, we should note that initially no actions were taken until field investigations found containers of nitrofurans in the possession of shrimp producers.

This leads to the question of whether the EU requirements, which pertain to the number of IPs achieved in MS
The method detects and quantifies CS\textsubscript{2}, but there are common approach has been to analyze CS\textsubscript{2} liberation used illegally in fruit and vegetable production, the most which essentially set a limit of 1 ng/g for nitrofurans were used or not. To address this concern, at least in the case of chickens, investigations have been conducted to monitor for the parent drug directly in eyeballs, where they may accumulate. When SEM was found to be occurring naturally and as a contaminant in the samples, this created a conundrum for the regulators because their inflexible IP rules (EU/657/2002) required that regulatory laboratories take enforcement action when four identification points were met, even at extremely low concentrations of <0.1 ng/g. This led to revision of the EU legislation to devise MRPLs, which essentially set a limit of 1 ng/g for nitrofurans.

In a similar situation involved in the regulatory analysis to determine if dithiocarbamate fungicides were used illegally in fruit and vegetable production, the most common approach has been to analyze CS\textsubscript{2} liberation from the samples when treated with tin chloride. The method detects and quantifies CS\textsubscript{2}, but there are known sources of matrix components leading to CS\textsubscript{2} other than just dithiocarbamates, particularly in brassica-vegetable species, and that makes regulatory actions questionable. To address this problem, chemists have been investigating methods to analyze the dithiocarbamates in the food directly.

5.8. Multi-laboratory confirmation (e.g., aminopterin or folic acid, or melamine in pet food?) In this last example, sometimes the stakes are so high that many laboratories are needed to confirm an analytical result. An example of this was the large international investigation to find the cause of a rash of pet deaths in 2007, which had been associated with pet food. Before obtaining confirmation from other laboratories, one organization announced that aminopterin had been identified in the pet food and they believed it to be the cause of the deaths. Although a second laboratory did confirm the presence of aminopterin at the non-toxic level of <4 ng/g, other laboratories could not detect that chemical in the pet food. The known toxicological symptoms of aminopterin, and the low concentrations found in the pet food, did not match the pathologic causes of the pet deaths. It is possible that folic acid was confused with aminopterin, as they have similar structures, as shown in Fig. 9.

Later, a combination of high concentrations of melamine and cyanuric acid were found to be the cause of kidney failure of the pets that had died from eating the contaminated pet food.

This example illustrates the need for confirmation by a second laboratory in such a high-profile case, but, even within the same laboratory, use of a second method would have made it less likely that aminopterin would have been confirmed, and the press report would not have been issued.

6. Assuring data quality

The point of these true stories and real-world discussion is to emphasize that seemingly straightforward analyses and the analytical decisions that come from them can still be wrong, despite strict analytical requirements and laboratory accreditation.

Oxamyl and methomyl are just two among scores of pesticides that have “complicating details” in their analysis and regulatory control, and any method that analyzes something other than the true compound of interest, as in the nitrofuran example, should face increased scrutiny and lead to caution in interpretation of results. In today’s environment, greater pressure is placed on the laboratories to follow ISO 17025 practices and obtain accreditation for their analytical methods, but accreditation alone cannot guarantee data quality. In nearly all the examples given, a checklist, such as the one discussed in Section 6.2, would have helped the analysts avoid mistakes and make more accurate identifications.

The reliance on recipe-type instructions reduces the role of the thinking analyst and displaces the scientific burden of proof (“it wasn’t me, it was the method”). That is not to say that analytical procedures should not be followed precisely or that an organized system and proper documentation should not be in place, but, in any laboratory, the most critical factor that leads to high-quality analytical data comes from skilled, informed, responsible analysts, who have sufficient resources and time to do their job.

Independent of laboratory accreditation, a valuable tool for improving the performance of analysts comes from the feedback provided in PT-sample programs, as documented in many examples. Fig. 10 shows an example in the case of European Proficiency Testing (EUP) among approximately 130 accredited laboratories for analysis of pesticide residues in fruits and vegetables. The quantitative results improved by about 50% from PT 4 to PT 8 in terms of inter-laboratory reproducibility, and the qualitative performance also im-
proved in terms of fewer false negatives, as shown in Fig. 10 for Tests 6 and 8. In many cases, the laboratories had to expand the scope of analysis to reduce the number of “false negatives,” and, in other cases, the feedback obtained by participating in the programs helped analysts better identify the analytes in blind analyses.

Also, the number of false positives reported during that EUPT program has also reduced over time. In the first five EUPT sample sets (1–5), ≈20 false positives were reported among the laboratories in each test round, but, in the next five sample sets (6–10), this number decreased to ≤10 false positives per round among the ≈130 laboratories. This decrease is attributed to the wider implementation of MS systems and improved skills and experience among analysts in using the instruments and the techniques, especially LC-MS$^2$.

### 6.1. Qualitative method validation

Due to the lack of a solid theoretical basis for qualitative identifications, qualitative methods must typically be assessed empirically, just as quantitative methods are
validated. The real chances of false positive and false negative results in identifications, which constitute the most important qualitative performance factors, need to be determined with respect to analyte, concentration, matrix, and method parameters. Proposals have been made to do this during method development [24,61], and inter-laboratory trials have been performed to measure the confirmatory ability of MS techniques in blind analyses of real-world samples [62,63].

The empirical approach is daunting if the measurement is to be made to a high level of confidence (e.g., >99%), but a reasonable degree of validation can be performed to eliminate the use of poor methods (e.g., 95% level of confidence) [24].

A multivariate statistical model has been demonstrated to lower the rates of both false positives and false negatives in GC-MS identification [62,64] vs arbitrarily chosen criteria [5,24], but it may not be practical to analyze enough samples in a blind fashion to determine the optimized criteria.

Essentially, this practical difficulty leads back to the simpler concept of confirming analytes in the sample by repeating the analysis of a duplicate sample using an orthogonally selective, validated method (perhaps even in a second laboratory if the importance of the analysis warrants that).

6.2. Factors in making identifications
Section 5 presented several examples of how current approaches to identification can lead the analyst astray, not necessarily due to the identification criteria per se, but because not all the factors were considered in the overall method. In making qualitative decisions, the analyst (and/or software) should take into consideration a number of factors, many of which, but not all, appear in current guidelines [12]. These factors can be included in developing checklists for analysts to compile the information needed systematically to satisfy fit-for-purpose identification criteria:

(1) information about elemental composition gained from other detectors, isotopic patterns (e.g., Cl and Br), or the nitrogen rule (for molecules comprising C, H, N, O, P, S, Si, and halogens, an odd-numbered ion indicates that an odd number of N molecules occurs in the ion, and an even-numbered ion means that no or an even number of N molecules appears in the structure);

(2) correct isotopomer patterns for the assigned molecular formula;

(3) the choice of qualitative ions and the presence of the molecular ion in the spectrum, even if it is small;

(4) actual trends and variability of the chromatographic $t_R$ and appearance of peak shape of the analyte (e.g., peak shifts, tailing factors, and isomeric pattern);

(5) selectivity of the chromatographic separation;

(6) instrument performance and proper tuning;

(7) evidence concerning interferences in blank samples and that no analyte carryover or laboratory contamination has occurred;

(8) prior knowledge about the history of the sample in terms of likelihood that the analyte would be present, such as its potential use;

(9) whether the MS-fragmentation pattern can be explained;

(10) the detected chemical in the extract makes sense (e.g., analyte-stability issues, concentration, clean up done, and derivatizations);

(11) consistency of the analytical result with previous analyses of the sample(s);

(12) prior experience of the analyst; and,

(13) elimination of other possible compounds that could lead to the same result.

One must also recognize that more than a single compound is often present at any given time in a chromatogram, which can generate mixed spectra that can confuse the identification. For this reason, proper background subtraction, which may entail mass-spectral deconvolution of co-eluted peaks, is critical to aid and to improve the process.

6.3. Determination of ion ratios
A common flaw in typical MS-identification guidelines is the way used to set permissible variability in the relative ion-abundance ratios. The IP rules dictate that, for GC-MS in EI mode, relative ion abundances must be within ±10% (relative, not absolute value) of the ion intensity of the reference spectrum for ions >50% relative abundance, ±15% for ions <20–50%, ±20% for ions >10–20%, and ±50% for ions <10% [5]. In chemical ionization, API and other techniques, a wider acceptable degree of variability is permitted. Different confirmation criteria set different standards for acceptability, as shown in a previous report [1], but none are based on specific empirical measurements. In reality, some molecules, techniques and instruments yield variability in the mass spectra greater or less than others.

7. Proposed alternatives
7.1. Determining variability of ion ratios
During the validation of a quantitative method using MS detection, many analyses are done to evaluate recoveries, repeatability, reproducibility, linearity of calibration and other factors. The same data can be used to determine the variability of the ion ratios from day-to-day with respect to concentration in the sample matrix(es). Indeed, some chemists have chosen to measure these factors to help set the identification criteria for their methods [65].

http://www.elsevier.com/locate/trac
The error bars in Fig. 4 indicate the standard deviation (SD) of the measurements of pirimiphos-methyl with n = 4 at each concentration (taken from calibration standards in matrix analyzed on four different sequences). In carrot, the average %relative ion ratios vs m/z 290 for concentrations ≥ 38 ng/g were 99 ± 3 for m/z 276, 86 ± 2 for m/z 125, and 68 ± 2 for m/z 305; whereas, in orange, the same approach on the same instrument a few months later led to values of 110 ± 7, 108 ± 6, and 73 ± 4 for the same ions vs m/z 290, respectively. The change in mass spectra over time, including the base peak in this case, due to different tuning parameters or other factors, is not an uncommon occurrence and should be taken into account in the confirmatory procedures. Thus, reference spectra should be updated frequently, ideally after each tuning.

Assuming normal (Gaussian) distributions of ion intensities, the setting of variability criteria for relative ion intensity dictates the rate of false negatives that will occur. For example, at 10 ng/g in carrot, the limiting third ion for pirimiphos-methyl (m/z 305, the molecular ion) averaged 72 ± 8% relative abundance vs the m/z 290 base peak. This indicates that, if the ion-ratio window is set at ±10% (relative value, thus an acceptable range of 65–79% relative abundance) as the IP system requires, then the rate of false negatives at 10 ng/g in carrot would be ≈33% according to Gaussian distribution theory. A window of ±2 SD (which should be determined from ≥16 measurements) can be used instead, corresponding to ≈5% false negatives. Thus, a relative ion-abundance range of 56–88% for m/z 305 would yield an LOI of ≈10 ng/g for pirimiphos-methyl in carrot by the GC-EI-MS (SIM) method used. Furthermore, an ion-ratio window of ±3 SD relates to a mere 0.3% chance of false negatives, which would lead to a relative ion-abundance range of 48–96% in the pirimiphos-methyl example given.

In terms of potential false positives, comparison of these ion ranges in NiST’98 with the more narrow window defined by the IP system made no difference in being able to isolate the spectrum of pirimiphos-methyl from all others in the library using the three most intense ions of higher mass.

We suggest use of an empirical approach such as this: first, to determine the variability of the ion ratios in matrix extracts at the desired concentration; and, then an ion-ratio window can be defined to set the identification criteria for ion ratios at that concentration.

Instrument software already commonly allows the user to set qualifier - ion ratios or spectral matching factors in the method, and the software could be further devised to measure variability of mass spectra. By using real data rather than arbitrary criteria, the rate of false negatives should be reduced. Afterwards, a general assessment of false positives can be made using blanks, blind analyses, mass-spectral library searches, and other means to ensure that the criteria are reasonable and realistic. The multivariate statistical approach should work even better if enough samples can be evaluated to fit into the model [62].

The assessment of ion ratios is just one facet of the overall process, and the acceptability criteria should not be set so strictly that they act to increase unnecessarily the LOI without actually providing a greater degree of selectivity in the analysis. At this time, rather than defining criteria that are assumed to decrease the rate of false positives, such as defining an arbitrary points system and ion-abundance windows, the actual minimization of false positives is better addressed through checklists, many analyses of positive samples and blanks, analyses of blind PT samples, better analyst training, and knowledgeable judgment, taking into account all available information.

7.2. Matching factors
In library searches, ion ratios are calculated relative to the base peak, so more emphasis is placed on the base peak than other ions in calculating the spectral fit. The switch in base peak for compounds, such as pirimiphos-methyl in Fig. 4, can lead to lower match factors, despite the analyte’s presence. Endosulfan is an example where its spectrum contains dozens of ions, and its pattern is thereby easily recognizable by the analyst (provided it does not co-elute with chlordane), but the variability of the relative ion intensities is high, resulting in low matching factors. Identification criteria should not require high spectral match factors (e.g., >90%) that do not correspond to reality. Just as the ion ratios can be empirically measured during method validation, so can match factors vs reference spectra to determine the more realistic setting for the required match factor at the LOI needed. As discussed previously, contemporaneous comparisons with the standard spectra are critical, ideally from both solvent and matrix-matched standards at a concentration similar to that of the analyte in the sample.

In practice, matching thresholds for targeted pesticides are typically set very low (e.g., 40%) in order to minimize false negatives, so then the analyst can use judgment to assess whether the preliminary software finding is correct or not. Ideally, software would be trusted to identify any (or all) compound(s) in the chromatogram automatically, but this is seldom the case.

Another important, useful tool is the identification-probability factor provided by the NIST library and the listed order of hits. It is better to have a worse fit but better hit in the sense of having the suspected analyte come first in the identification-probability list and with a high level of confidence than necessarily with a high matching factor. A known example to illustrate this concept is the mass spectra of simple normal alkanes that provide very high matching factors to all, but they

1086 http://www.elsevier.com/locate/trac
7.3. Assessment of chromatographic factors

Another important factor in MS detection in combination with chromatography is the selectivity of the separation. Direct infusion of the sample into an MS instrument may be reasonable for screening methods, but, for complex samples, an analytical separation should be coupled to the MS detector for increased selectivity. Just as the MS-ion ratios can be assessed empirically, the assignment of the acceptable tR window and peak shape can also be validated empirically during the quantitative method-validation process. Measurement of tR is a central to chromatography and it is collected in each injection sequence with respect to system-suitability tests and calibration standards, and that provides a realistic range of values that can be measured. Moreover, the different ions used in an MS identification should co-elute with each other (within the scan speed and dwell times of the detection process).

Table 2 provides the average, SD and %relative SD (%RSD) in the tR of 16 pesticides from 24 injections of calibration standards in carrot extracts over the course of four sequences without performing instrument maintenance. One possibility is for the tR window in identification criteria to be established based on this kind of empirical measurement. For example, using a ±3 SD window for dimethoate would require that its peak must have tR of 11.841–12.045 min to be considered for identification purposes. Again, contemporaneous assessment of tR makes good analytical and practical sense.

Peak shape and isomeric patterns should also be taken into account. For example, the four isomers of cypermethrin that occur in the pesticide formulation present a specific pattern, which, when observed in the correct tR window, essentially guarantees that cypermethrin has been detected independent of MS information. Peak shapes, tR values, and asymmetry factors provide further information to help make an identification. The presence of a tailing peak at the tR for a compound that gives sharp peaks (or vice versa) indicates a possible false positive. Again, the use of contemporaneous injections of standards or standard addition to the injected sample (in matrix and at similar concentration level) would account for such chromatographic changes. In fact, it is common for standard operating procedures to require system suitability and/or QC tests be made before and during analytical sequences to ensure that instrument-performance criteria are being met (e.g., proper calibration, tR, peak shapes, and MS tuning factors).

7.4. Assessment of false positives and false negatives

An important possible measurement of method performance is to determine the percentage of sample analyses that give erroneous results using given identification criteria. One simple approach is to analyze in a blind fashion a number of samples fortified (or not) with analyte(s) of interest in matrices of interest. If only final extracts of blanks need to be fortified, then time-consuming sample-preparation steps can be minimized. MS specialists often believe such an approach would be a waste of time because they are confident in their instruments and techniques, but when blind analyses have been conducted in practice (e.g., PT samples), the number of erroneous results reported can be surprising [62,63]. This approach for perhaps 20 blind samples would certainly be useful to indicate a problem with the qualitative aspects of a method (or to verify that there is not a serious problem).

One of the authors of this report challenged application chemists from different MS manufacturers to provide results using their instruments on a series of fruit and vegetable extracts that have been fortified (or not) at relevant concentrations with 16 pesticides. Among five chemists who accepted the challenge, none has been able to perform the task well at the first attempt and all required feedback to help them learn from their mistakes.

Independent of instrumentation, qualitative analysis cannot be done properly without experience with the analytes and the matrices. The blind analysis of samples in a realistic situation is an excellent way to train analysts, and it is recommended that this practice be implemented among regulatory laboratories doing chemical-residue analysis. Perhaps in future, an automated system of analyte identification using a neural-network approach, in which the software is trained, much as an analyst is trained by experience, can be developed, but, until that time, analyst training is an essential part of qualitative analysis.

7.5. Spectra with too few ions

One of the main difficulties in GC-MS identification by traditional criteria (either EI or CI in full scan or SIM) relates to how many pesticides have mass spectra with only one intense ion. Notable examples include permethrin (m/z 183), fenithion (278), methoxychlor (227), phosmet (160), terbufos (231), pendimethalin (252), piperonyl butoxide (176), disulfoton (88) and pyrimethanil (198). Many other pesticides [e.g., methidathion (m/z 145 and m/z 85), give only two intense ions. To obtain three or more ions that meet typical MS-identification criteria leads to LOIs that are >100 ng/g.
especially in complicated matrices, which is too high for risk assessment and other common regulatory purposes. However, if the one or two strongest ions are used, then the LOI can be reduced greatly.

Independent of the number of ions, we propose that enforcement actions continue to entail a second injection of the original, re-extracted sample using a different analytical approach (e.g., LC vs GC), column phase (with orthogonal selectivity), detector (e.g., element-selective detector vs MS), and/or MS technique (e.g., CI vs EI, or MS² vs MS) plus inclusion of reagent and matrix blanks [2]. Due to the inability of non-MS methods to provide structural information, at least one of the analytical methods should use MS for detection, if the instrumentation is available to the laboratory.

For analyses that do not involve high stakes, such as non-violative findings for risk-assessment purposes, a single injection using MS detection should be demonstrated as fit-for-purpose. The method-validation process would provide a realistic estimate of the LOI, and blind analyses would have to be done to demonstrate that no false positives occurred in ≥ 10 analyses of different sample types using blanks and samples fortified at the LOI level. Again, non-MS techniques do not provide structural information, so a second analysis should be done using MS or an alternate approach if a non-MS detector is used in the first analysis.

8. Summary of proposed confirmation system for chemical residues

In the chemical-residue analysis of foods and environmental samples, the main purposes of analysis relate to:

1) surveillance monitoring and data collection for non-enforcement reasons; and,

2) enforcement applications to support legal actions.

Different degrees of confidence in identification and confirmation results are required for these two purposes. The current guidelines do not differentiate very well between these different needs, so they do not optimally suit either purpose of analysis.

The current criteria are not based on objective or empirical forms of measurement, but are based on arbitrary criteria using subjective assessments about the degree of selectivity provided by different MS techniques. The assessments are generally correct, but there are many exceptions, depending on the analyte-matrix pair, the concentration, the MS ions detected, the analytical technique, and the importance of the results.

The current rates of false negatives are thought to be too high in both types of application (enforcement and non-enforcement) because the identification criteria are too stringent. Furthermore, spurious errors are not typically addressed to reduce false positives, which are best addressed through confirmatory analysis.

As the word implies, “confirmation” requires the results from at least two analyses (which should ideally be orthogonal and at least one of them should involve MS detection after an analytical separation), which, in enforcement applications, must agree with each other in terms of analyte identity and concentration.

For non-enforcement purposes, a single analysis using MS detection should be satisfactory, or two analyses using non-MS methods, provided that there has been an empirical demonstration that acceptably low rates of false positives and false negatives occur for the analytes at an adequately low concentration in the targeted matrices. An MS qualitative screening approach may be the best option, followed by a different method to make the confirmation and to improve the quantitative determination.

We summarize the arguments and the points made in this article in the following proposed sequence for quantitative and qualitative method validation.

1) Define the need for the analysis (scope of analytes and matrices, concentration ranges, acceptable degree of accuracy, and tolerable rate of false negatives);

2) Define the primary analytical method(s) to meet the needs of the analysis most effectively and efficiently (e.g., GC-MS, LC-MS², GC with selective detection, LC-fluorescence, and others);

3) Conduct method validation of the primary method(s) for the targeted analyte(s) and matrix(es) (or representative analytes and matrices) to determine quantitative characteristics of the method(s) according to accepted practices;

4) At the same time as quantitative method validation, empirically determine MS criteria using samples of known concentrations on different days in fortified matrices to achieve the desired identification limit (with set criteria to minimize false negatives);

| Table 2. Reproducibility of t₁₈ in the GC-MS analysis of 16 pesticides in calibration standards prepared from blank carrot extracts (n = 24) |
|-----------------|----------------|----------------|----------------|
| Analyte         | Average (min)  | SD (min)       | %RSD           |
| o-Phenylphenol  | 8.935          | 0.020          | 0.22           |
| Diphenylamine   | 10.340         | 0.019          | 0.18           |
| Chlorpropham    | 10.630         | 0.020          | 0.19           |
| Triluralin      | 10.855         | 0.011          | 0.10           |
| Dimethoate      | 11.943         | 0.034          | 0.28           |
| Chlorothalonil  | 13.403         | 0.023          | 0.17           |
| Pirimiphos-methyl| 15.037        | 0.014          | 0.09           |
| Chlorpyrifos    | 15.655         | 0.014          | 0.09           |
| Methidathion    | 17.185         | 0.012          | 0.07           |
| Endosulfan I    | 17.404         | 0.011          | 0.06           |
| cis-Chlordane   | 17.453         | 0.011          | 0.06           |
| Endosulfan II   | 18.373         | 0.010          | 0.06           |
| Ethion          | 18.443         | 0.019          | 0.10           |
| Phosmet         | 19.584         | 0.013          | 0.06           |
| p,p’-Methoxychlor| 19.620        | 0.034          | 0.17           |
| cis-Permethrin  | 20.960         | 0.012          | 0.06           |
(5) Measure rates of false positives while also training analysts through analysis of blank fortified samples (≥ 10–20 blanks from different sources and a similar number of fortified samples at a variety of concentrations); [Note: this can be done over time during routine analyses, and such tests should be part of the QC requirements];

(6) Devise a second approach to be used for independent confirmation of suspected violative samples from the initial analysis, and conduct blind analyses of extracts to estimate rates of false positives and false negatives vs concentration, as in Step (5).

The proposed validation sequence covers only general factors, but, in future, we hope that more specific, practical procedures can be described, just as in the case of quantitative method validation for chemical-residue analysis.

Further scientific study of the degree of selectivity of MS techniques coupled with analytical separations is needed. Ideally, a systematic approach based on sound theory can be devised to assess accurately the degree of accuracy in qualitative MS analysis, but, until that time, we believe that the proposed general approach is more scientific and better fits the purposes of chemical-residue analysis.

Acknowledgements

We thank Lutz Alder, David Heller, Eugenia Soboleva, Charles Stafford and Stephen Stein for their participation in discussions. This research was supported in part by the Israel Science Foundation (Grant No. 1172/07). This research was further supported by the James Franck Center for Laser Matter Interaction Research, Chicago, Illinois, USA, and a grant from the Ministry of Science, Culture, Sport of the State of Israel, and Karlsruhe Research Center (FZK), Germany.

References


[60] http://www.crl-pesticides.eu