American Pharmaceutical Review™

Nelson Labs®
A Sotera Health company

Good Identification Practices for Organic Extractables and Leachables Via Mass Spectrometry

By Dr. Dennis Jenke, Dr. Piet Christiaens, Dr. Jan Baeten, Dr. Philippe Verlinde, Dr. Jean-Marie Beusen, Dr. Ward D'Autry, and James Mullis

Contents

Introduction

The function of a pharmaceutical drug product or a medical device is to provide the patient with a desired therapeutic benefit. If the drug product or the medical device were pure (meaning without impurities), ideally the therapeutic benefit would be largely realized with minimal adverse patient effects. However, since the practical reality is that drug products and medical devices contain impurities, a patient is exposed to these impurities during treatment and these impurities could potentially trigger an increase of adverse patient effects.

An important class of drug product or medical device impurities is leachables. Leachables are foreign impurities that are present in a drug product because of its chemical interaction (leaching) with its packaging and/or manufacturing systems. When patients are dosed with the drug product, they are exposed to the drug product's impurities, including leachables. In a medical device, leachables are substances -that are leached out of the device and into a patient, either directly or indirectly, during the device's clinical use. In either case, drug product or medical device, the leachables could have an adverse effect on patient health and safety.

As it is typically applied, the word "communication" signifies the exchange and/or transfer of information or news between the bearer of the information and the receiver of the information. In a manner of speaking, the packaging system or medical device "communicates" with a patient as an exchange or transfer of substances occurs between the packaging system or medical device (the bearer) and the patient (the receiver). This communication may be direct; for example, a medical device that is implanted in the body "communicates" with the patient via the transfer of leachables directly from the device to the patient. Alternatively, communication may be indirect; that is, the transfer of substances from packaging system or medical device to a patient occurs through an intermediary. For example, a medical device such as an IV administration set "communicates" with the patient indirectly via the drug product that is administered through the set. In the same way, a drug product's packaging system "communicates" with a patient indirectly through the administered drug product. Although the packaging system and the patient are never in direct contact, the patient is still exposed to packaging system-related leachables via the drug product.

The effect that a leachable will have on a patient's health depends on the route of exposure, the frequency and duration of exposure, the fate of the leachable in the human body, the toxicology of the leachable, and the exposure dose of the leachable. Of these parameters, establishing toxicity and dose is an exercise shared by an analytical chemist and a toxicologist. The role of the analytical chemist is to screen the drug product or medical device for leachables, establishing the leachables' identity and concentration. Armed with this information, the toxicologist establishes the leachable's toxicity (enabled through its identity as the link to the leachable's toxicology) and the patient's exposure to the leachable (enabled through its concentration). Thus, an important aspect of leachables screening of either

a drug product or a medical device is the identification of all leachables present in either the drug product or in the medical device at levels above what is essentially a toxicologically established "no adverse effect" level or threshold.

For various reasons, it may be relevant and/or necessary to address leachables abstractly, more as a possibility and less as an occurrence. For example, rather than screening a drug product for packaging-related leachables, the packaging system can be screened for extractables, which are substances that can be extracted from the packaging system under laboratory conditions. Screening of packaging systems and/or medical devices for extractables is relevant as extractables are either potential leachables, predictive of leachables, or precursors to leachables. As such, the effect of leachables on a patient can be estimated by considering the patient effect of extractables as if they were leachables. Like leachables, extractables must therefore be identified and quantified, thereby enabling toxicological safety assessment.

It is generally accepted standard good practice that organic extractables and leachables are discovered by complementary and orthogonal chromatographic techniques whose separation method provides the necessary selectivity and resolution and whose detection method provides the necessary information from which an identity can be inferred and a concentration can be estimated. Almost exclusively, mass spectrometry is the chosen method for identification, as the mass spectral properties of discovered extractables or leachables can provide multiple avenues for identification. Thus, the typical orthogonal and complementary hyphenated chromatographic techniques used in the screening of test articles for organic extractables or leachables are Headspace GC/MS (volatile organic compounds), GC/MS (semi-volatile organic compounds) and LC/MS (non-volatile organic compounds).

Securing a substance's accurate concentration and correct identity is essential in establishing the substances potential impact as a leachable. This Appendix focus on the aspect of identification and, more specifically, on the process by which mass spectral data and other supporting evidences are used to secure, judge and justify complete and correct identities for all relevant extractables or leachables surfaced by the chromatographic screening analyses.

Part 1

Identification Classes, Processes, and Practices

1.1 Identification

Once a chromatogram is obtained by analyzing a drug product or an extract, the chromatogram is reviewed to establish every chromatographic peak which (a) is absent from the associated blank (and thus is a legitimate analyte) and (b) whose response is greater than the response equivalent to a justified threshold (e.g., the Analytical Evaluation Threshold, AET, for patient safety assessment).

Once a peak has been established to be relevant (meaning its response is greater than the AET), the compound responsible for the peak must be identified and quantified, as it is the combination of the identity and concentration that allows the extractable's potential patient impact to be assessed. As noted in Figure 1, a mass spectrum, secured for each peak in the chromatogram, provides the basis for identification.

Practically speaking, a compound has been identified if it can be assigned a proper chemical name, an appropriate identifying number (for example, a CAS registry number), and a chemical structure with an acceptably high degree of confidence. A compound can be described (or partially identified) if it can be assigned to a class of compounds that share

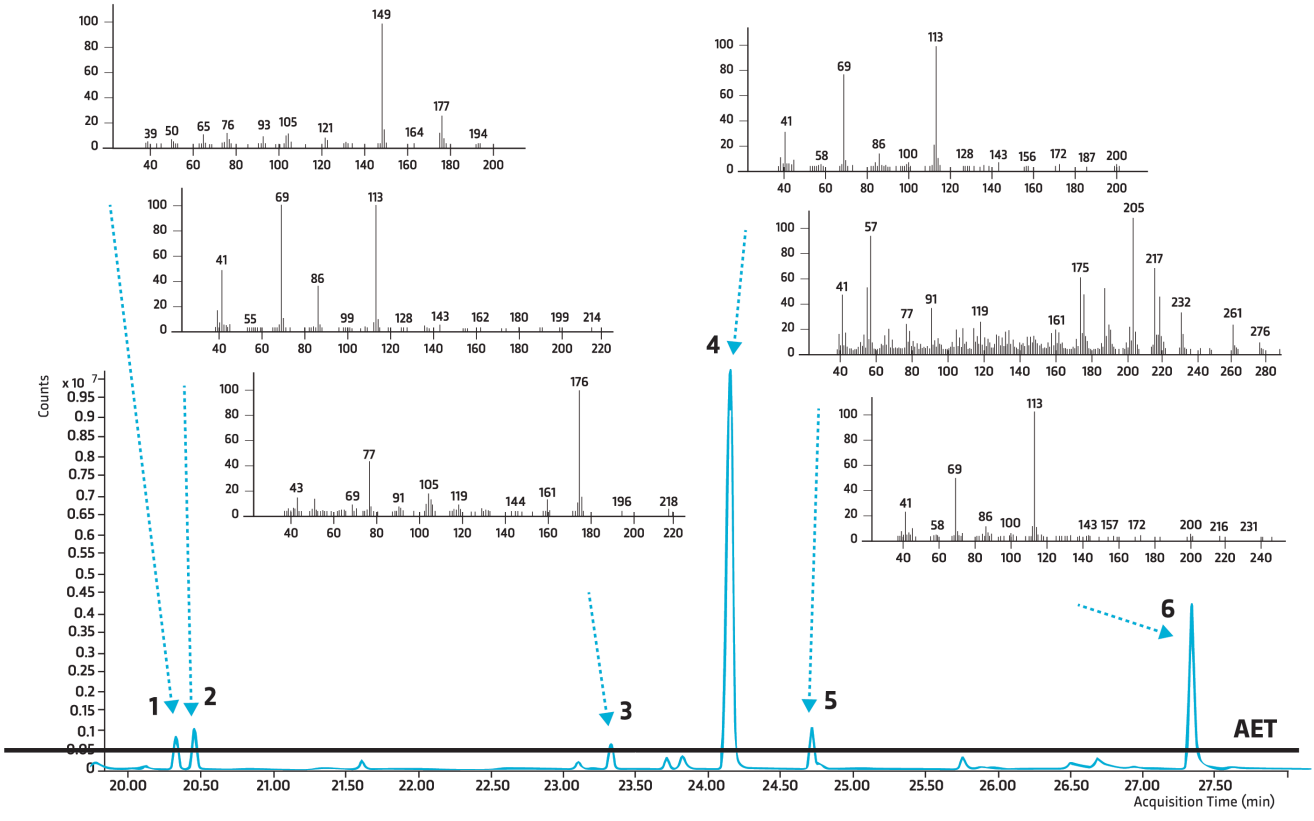


Figure 1. Typical outcome of a GC/MS analysis: A Total Ion Current (TIC) chromatogram where all peaks above the AET need to be identified. Each peak in the chromatogram has its own mass spectrum and based upon the tools that are available (e.g. mass spectra matching or an internally developed database), the skills of the mass spectrometrist, or other supporting evidence, a mass spectrum be used to establish a compound's identity with different levels of confidence.

a common functional characteristic. For example, an analyte can be identified as hexane or can be described as an aliphatic hydrocarbon.

It cannot be emphasized enough that a meaningful, rigorous, proper and accurate risk assessment of an extractable or a leachable can only be obtained when the extractable (or leachable) has been correctly identified with a high degree of specificity and confidence. If a compound remains unidentified, if the wrong identity has been assigned to the compound, or if there is little confidence that the proposed identity is correct, then the impact assessment (if it can even be completed), will likely be flawed. This is because (a) an unidentified compound's impact cannot be inferred, (b) a mis-identified compound's assessment will be erroneous (as it is based on the properties of the wrong compound), and (c) a low-confidence proposed identity is likely to be incorrect—triggering an erroneous assessment. As there is no way to fix an error in identification (other than securing the correct identity), an identification error is a fatal error which unavoidably undermines the overall impact assessment of a medical device or pharmaceutical container/closure component or system.

This discussion is important because the identity is used to infer an outcome, in many cases, the patient health impact of the extractable or leachable. The likelihood that the inferred outcome is, in fact, the observed outcome depends, in large part, on the certainty of the information upon which the conclusion is based. For example, if there is great

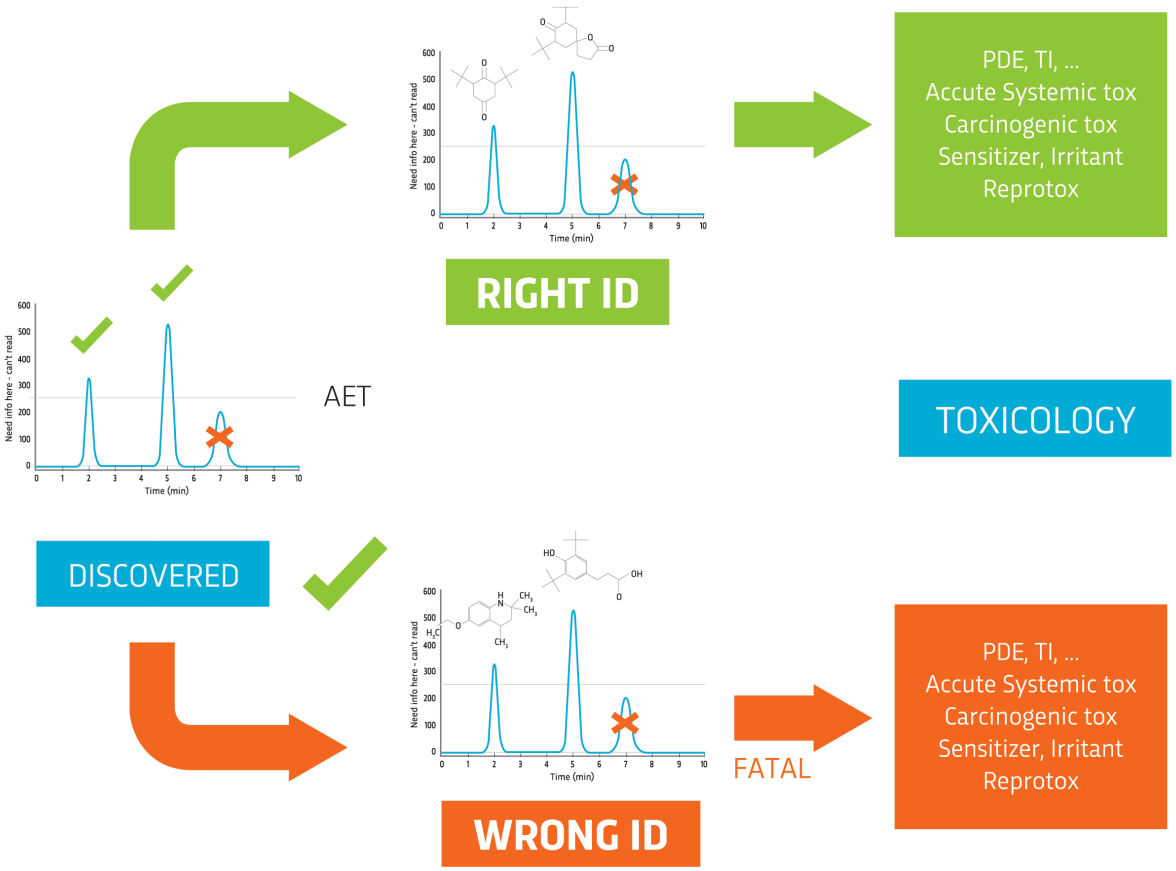


Figure 2. Illustration of the criticality of correct identifications. The "right ID" pathway will lead to the correct toxicological information of the respective compounds, while the "wrong ID" pathway will lead to toxicological information that is largely irrelevant.

uncertainty (and therefore lack of confidence) in the identity of a leachable, then there is corresponding great uncertainty (and therefore lack of confidence) in the subsequent safety assessment. It is for this reason that:

- Identities must be secured with the highest possible certainty
- Identities must be reported with an indication of their uncertainty

1.2 Identifying Information and Its Use

It must be understood that "identification" is not a direct product or outcome of screening (that is, the test methods used for screening do not produce the identity as a direct outcome of the test), and rather that securing the identity for a leachable or extractable involves processing and analysis of the generated test data. Various means for securing a compound's identity with mass spectral and supporting data are discussed as follows.

1.2.1 Mass Spectral Matching

Once a peak's mass spectrum has been obtained, perhaps the most commonly employed and efficient means of securing a possible identity for the compound responsible for the peak, solely based upon the merits of its mass spectrum, is a process termed mass spectral matching (Figure 3). Mass spectral matching is based on the premise that a library of reference mass spectra for relevant organic compounds exists and involves the ranking of the library spectra in terms their similarity with the mass spectrum of the compound of interest. The following are crucial questions to ask when using mass spectral matching with an external library as the sole means of identifying a compound:

- How comprehensive and relevant is the library?
- What constitutes a good and acceptable similarity or match between an analytical and a reference mass spectrum?
- If several "good" matches exist, which match is the "best" and thus establishes the compound's identity?

Answers to these questions, and the concepts behind them, are contained in Part 2.

These concepts notwithstanding, it is noted that there are no universally applied and accepted mass spectral match "quality criteria" that unequivocally establish that a matched identity is, in fact, the true identity of a compound. Therefore, it is recommended, "good practice" to have all match-based identities reviewed by a mass spectrometrist to substantiate the tentative identity, regardless of the mass spectral match quality. The poorer the mass spectral match between the mass spectrum of the compound to be identified and the reference mass spectrum, the greater is the need for a visual inspection and mass spectral interpretation by a mass spectrometrist to substantiate the matched identity.

1.2.2 Manual Mass Spectral Interpretation – Structural Elucidation

When mass spectral matching is not possible, or when the outcome of mass spectral matching is equivocal, "manual" mass spectral interpretation (structure elucidation) can be pursued to secure the compound's identity. The types of information that could assist a qualified mass spectrometrist in securing such an identification include (but are not limited to):

- The molecular formula (e.g., accurate mass measurements)
- The presence of specific elements (isotopic data)
- Substructural evidence (fragment interpretation)

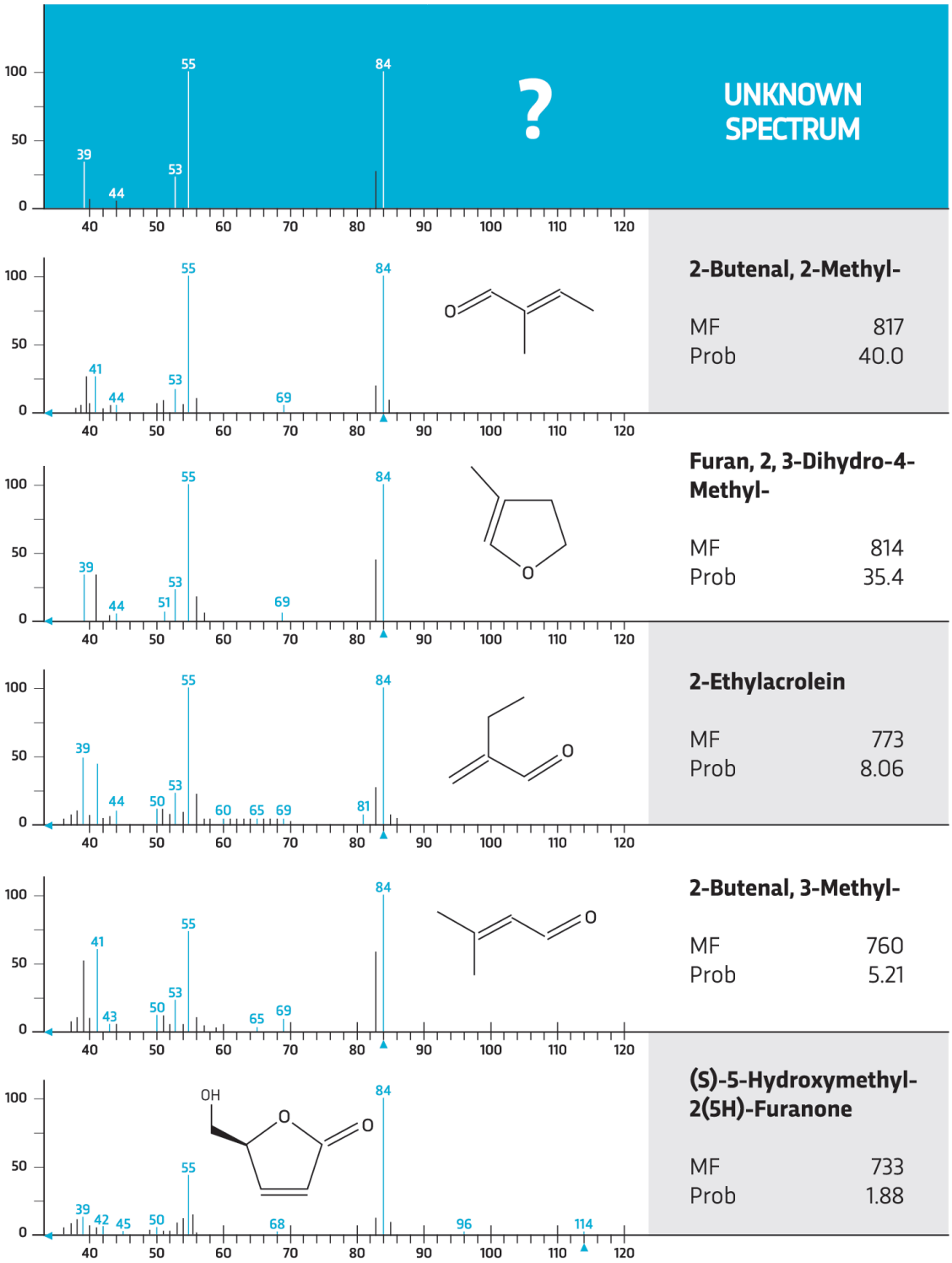


Figure 3. Typical example of using mass spectral matching information to assist in the identification of a peak, detected in GC/MS, using commercial MS libraries (in this case NIST match and probability scores). The compounds are typically ranked in decreasing mass spectral match quality, which often leads to the flawed conclusion that the highest ranked compound has the correct identity.

- "*De novo*" structural elucidation or a provable mass spectrum similarity to a compound that is identified as confirmed, confident or tentative

It should be emphasized that postulating a chemical structure, solely based upon mass spectral information and interpretation is not an easy task. The mass spectrometrist should be aware of the importance and consequences of postulating a defined chemical structure, as this information will be used to link the chemical compounds to its toxicological information and will be the basis for a subsequent toxicological evaluation of the compound. Cases where the wrong identity for the compound is assumed will inevitably bias the overall safety evaluation of the material, device, or container/closure component or system.

1.2.3 Additional Evidences – Securing the Identity with the Highest Confidence

Additional evidences that can increase the confidence that the matched or elucidated identity is in fact the compound's correct identity include (but are not limited to):

- A provable similarity between the identified compound and a second compound with a confirmed identity
- A provable relationship between the identified compound and the known composition of the test article
- A provable relationship between the identified compound and a compound with a confirmed identity secured by another orthogonal technique
- A good experimental retention index match between the identified compound and a database of retention indices
- Interpretation of a compound's accurate mass spectrum or MS/MS interpretation.

Part 4 will further elaborate on the various identification evidences that may assist in establishing and augmenting the level of identification of the compound of interest.

1.3 Classification Systems

To provide insight into how identities with the highest possible certainty can be reproducibly established, various identification classification systems have been proposed. Such systems accomplish three objectives:

1. They establish classes or levels of identification.
2. They rank the classes or levels in terms of degree of certainty and confidence.
3. They provide a high-level means for placing an identity in one of the classes based on the type, quantity, and quality of information used to secure the identity.

For example, an identification classification system for packaging-related extractables has been established by the USP[1] and consists of four classification levels:

- Unknown
- Tentative
- Confident
- Confirmed

More recently, an expanded system has been proposed to provide greater resolution than the USP classification[2] and consists of five levels, two with sub-levels (Figure 4):

- Unidentified
- Partial
- Tentative (matching or interpretive)
- Confident
- Confirmed (standard based or data based)

There is a lack of meaningful discussion of the mechanics and good practices required to produce and

properly interpret the requisite data to secure identities with the greatest certainty. This exists despite the authoritative works in mass spectrometry interpretation[11,12,13] and despite the fact that the aforementioned USP chapter and other publications[14,15,16] establish a hierarchy of identification classes and talk in generalities about "what it takes" to reach a certain level. The discussion of how collaborating data can be used to increase the certainty in identifications (that is, place the identity in its highest possible level) is also lacking.

1.4 The Multi-Dimensional Identification

1.4.1 "One-Dimensional" Identification – One Piece of Evidence: Partial or Tentative Identification

Based on either mass spectral matching (supported by expert review) or structure elucidation, three identification outcomes are possible:

1. In the worst-case, where the mass spectral information is inconclusive and the compound cannot be identified, the compound is reported as "unidentified." If such a compound is critical in an overall safety assessment of a medical device, a pharmaceutical container/closure component or system, or a material of construct, it would be necessary to further examine its chemical structure using high-end analytical instrumentation. Equipment such as accurate mass instrumentation (GC(Q-ToF), LC/(Q-ToF), LC/(Q-Exactive Orbitrap), NMR, or other techniques could provide additional, more definitive structural information that could lead to an identification.
2. In a better case, while neither mass spectral matching nor structure elucidation leads to a confidently proposed chemical structure,

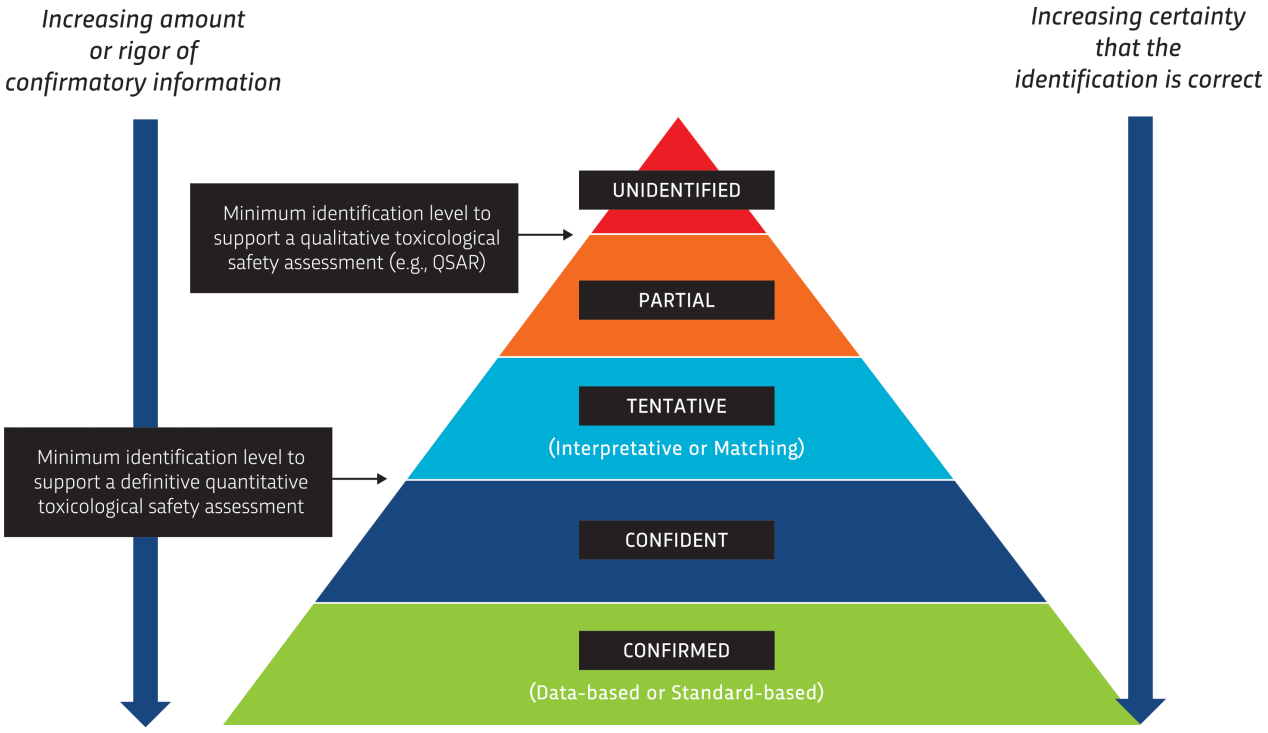


Figure 4. The identification hierarchy. The identity secured for an analyte is classified according to the hierarchy, based on the amount and rigor of available supporting information. Reprinted with permission from USP.

it may be that the mass spectral information is sufficiently rigorous that it suggests the compound of interest has distinct structural features or substructures. For example, mass spectra for a family of structurally related compounds (for example, phthalates, aliphatic hydrocarbons, or siloxanes) may contain mass spectral properties that are reproducible across the members of a family and which can be recognized by a trained and experienced mass spectrometrist. Thus, while the data may not be sufficiently robust to identify the compound of interest as a specific member of a family (for example, di-ethyl phthalate as a tentative identification), the interpreted mass spectral information can produce a "partial" identification (or description) of the compound's family (e.g., "it is a phthalate").

3. In the best case, the MS information is sufficiently convincing that an identity can be postulated for the compound of interest with a minimally acceptable level of confidence. In that case, the identification level for the compound becomes "tentative."

In a tentative identification (either through an acceptably good mass spectral match or via data-based MS interpretation), the identification was obtained in a one-dimensional way. This means that a tentative identification is based on a single action (either matching or interpretation) of a single piece of information, which is the mass spectrum of the compound. This level of identification may be sufficient in cases where the safety risk assessment establishes that the compound's effect on patient health and safety is likely to be small and negligible. In such situations, the tentative identification may be the end point of the identification effort for the compound. However, if the outcome of the safety assessment is equivocal, the identity of the compound will need to be secured at a higher level (i.e., with greater confidence). In such situations, additional supporting information about the identity of the compound must be secured to increase the confidence in the tentative identity (or to propose a different, more correct, identity).

Although there will be cases where the available information is sufficient to support only a tentative identification, it should be the goal of any identification activity to secure as confident an identity as is possible. Confidence in identity leads to confidence in the impact assessment. For example, a tentative identity can be confirmed (or invalidated) via the analysis of a reference standard of the identified compound. Should the analytical mass spectrum and retention time match those of the reference standard, then the tentative identity is confirmed, and one has the utmost confidence that the proposed identity is the true identity. If there is a disconnect between the analytical data and the reference standard data, then the tentative identity is invalidated, and it is back to the drawing board in terms of securing the correct identity.

1.4.2 "Three-Dimensional" Identification – Three Pieces of Evidence: Confirmed Identification

A confirmed identity is an example of a three-dimensional identity as it is based on three corroborating items: mass spectrum, retention time, and authentic reference compound.

Although a confirmed identity secured via matching to an authentic reference standard is generally and universally considered to be the "gold standard," authentic reference standards are not always readily available. In such a circumstance, there are alternate means of securing a confirmed identity, where in this case the term confirmed is generally applied to a proposed identity that has sufficient corroborating information that it is almost certain to be the true identity. As was noted previously, a confirmed identity secured by a match to an authentic reference standard is three-dimensional in the sense that the

identity is support by three independent pieces of corroborating information. More generally, then, a -confirmed identity is an identity supported by three (or more) independent and corroborating pieces of information. For example, consider a tentative identification that has been secured by spectral matching. If the same tentative identity can be secured independently via known information about the composition of the material (e.g., based on vendor information or information publicly known in the open literature), then the mass spectral match and the compositional information are collaborating information. If a third piece of collaborating information can be secured for the compound of interest (for example, its chemical formula obtained via accurate mass high resolution mass spectrometry), then the analyst has three independent items that corroborate the identification and the identification is established to be confirmed.

1.4.3 "Two-Dimensional" Identification – Two Pieces of Evidence: Confident Identification

As a tentative identity is a one-dimensional identification and a confirmed identity is a three-dimensional identification, an intermediate identity has a two-dimensional identification. As one has greater confidence in an identity based on two independent pieces of information versus an identity that is based on one piece of information (tentative), but has less confidence in the same identity versus one that is based on three pieces of information (confirmed), a two-dimensional identity is termed a confident identity. For example, consider the case where the same tentative identity is independently secured in two ways: spectral matching and known information about the composition of the material. As both pieces of information independently support the same tentative identification, taken together the two pieces of information support a confident identity.

1.5 The Identification Process

Considering the identification process as it relates to classes or categories of identities, identification starts at the point at which information about a peak's response (for example, its mass spectrum) has been obtained. At the time that the information is obtained, the compound responsible for the peak is unidentified. It is the job of the analytical chemist to take the available peak information and use that information to secure the compound's identity with the highest possible confidence.

Figure 5 proposes an identification "flow chart" diagram, largely based on the procedures and day-to-day practices at Nelson Labs. The diagram describes the actions, decisions, and documentation that is necessary to establish the right level of classification for the compound with respect to its identification status. However, it should be noted that the original nomenclature of classifications, used at Nelson Labs, may have been different from the nomenclature that is proposed in this document.

The flow chart specifically addresses the following actions, decisions, and types of documentation:

- Proper mass spectral matching practices as a means of securing a credible match-based tentative ID
- Proper structure elucidation practices as a means of either securing a credible inference-based tentative ID or elevating a match-based tentative identity to a higher classification level
- The proper role and use of accurate mass data
- Types and uses on non-MS correlating data
- The use of collaborating data to "move upward" through the classification hierarchy, thereby achieving the highest possible confidence (certainty) in all identifications

1.6 Practical Considerations in Identification

A topic of considerable debate in the E&L world is the level of identification one should achieve to support a rigorous toxicological safety assessment. It is possible to perform an impact assessment based on a tentative identification, as the tentative identification provides the essential information (name, structure and identifier such as the CAS number) required to secure impact-indicating information. Nevertheless, if there is a lack of confidence in the certainty of a tentative identity, then there will also be a lack of confidence in the certainty of the impact assessment. While such a low level of confidence might be acceptable when the outcome of the impact assessment is an emphatic "no impact", such a low level of confidence might not be acceptable in cases where the potential impact is "too close to tell." When the impact assessment approaches a "too close to tell" outcome, it is clear that the higher the level of confidence in the identification is, the more willing the impact assessor will be to accept the outcome of the assessment.

For example, consider the case of a toxicological safety risk assessment where a patient, during therapeutic use of a drug product, is also being dosed with a substance identified as a leachable during testing. The assessor is faced with the challenge of establishing the adverse effect this substance might have on the patient's health. An assessment measure that is commonly employed in such a toxicological safety risk assessment is the margin of safety (MoS). The MoS is the ratio of a leachable's permissible daily exposure (PDE), which is based on toxicological information linked to the leachable's identity, versus its total daily intake (TDI), which is based on the leachable's concentration and the drug product's prescribed dosing. When the PDE > TDI (MoS > 1), then the patient intake is less than the maximum permitted exposure and the leachable is said to pose a negligible risk of an adverse patient safety effect. When the PDE < TDI (MoS < 1), then the patient intake is greater than the maximum permitted exposure and the leachable could potentially produce an adverse patient safety effect.

In the case of an MoS > 10 (as an example of a situation where the safety risk assessment is an emphatic "no safety risk exists"), a PDE based on a tentative identification might be acceptable as there is a sufficiently large "margin of error". That is, the tentative identity is likely enough to be correct (at least in terms of the compound type) that even if the identity is wrong in terms of the specific compound, it is likely right in terms of the compound class. Having a "margin of error" of an order of magnitude (calculated MoS = 10 versus the minimal acceptable MoS = 1) should provide a sufficient "safety cushion" that the tentative identity, even if it is not completely correct, is adequately protective of a patient's safety.

Alternatively, consider the case where the resultant MoS = 2. Here, the "margin of error" is only a factor of 2 and it is obvious that basing the PDE on a tentative identity is not adequately protective of the patient. That is, if the tentative identity turns out to be incorrect, it is possible that the correctly identified compound would have a PDE that is sufficiently lower than the PDE based on the incorrect identity that patient safety could be compromised.

It is for this reason that it is generally accepted that a rigorous and acceptable toxicological safety risk assessment should be based on either a confident, or, preferably, a confirmed identity.

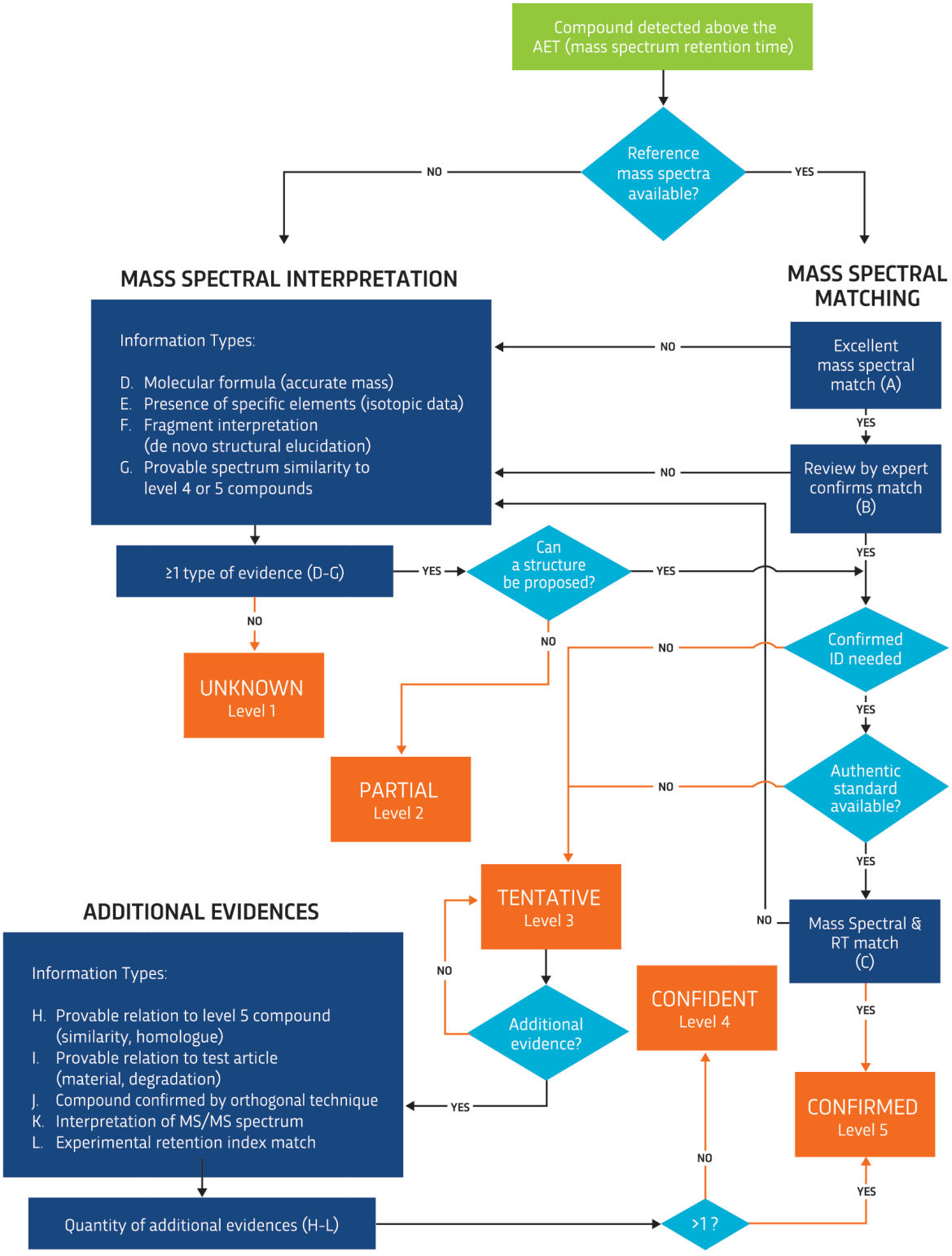


Figure 5. Decision tree for the determination of the level of confidence of identifications by chromatographic screening analysis in E/L screening.

Part 2

Identification via Mass Spectral Matching

2.1 Introduction to Mass Spectral Matching

The process of mass spectral matching is exactly what once would expect from its name. Screening chromatographic analysis of either an extract, a drug substance or a drug product produces chromatographic peaks associated with extractables or leachables. When mass spectrometry is employed as the chromatographic detection method, each peak has an associated mass spectrum, characteristic of the analyte responsible for the peak. This test spectrum can be compared to a compiled library of reference mass spectra generated via the analysis of authentic reference standards. An acceptable feature-by-feature match between test and reference spectra would suggest that the analyte that produced the test spectrum and the reference compound that produced the reference spectrum are on and the same. In this way, the analyte has been identified via mass spectral matching.

When reliable and current libraries containing mass spectra of relevant organic compounds are available, mass spectral matching is the most commonly employed and efficient means of establishing a tentative identity for the compound. General discussions of mass spectral matching and interpretation are contained in numerous publications and references.

Mass spectral matching is, in essence, the search for those library spectra that are similar to the test spectrum of the compound of interest. The primary outcome of library matching is a list of "hits", those library spectra that have some level of similarity to the test spectrum. In most commercially available libraries, "hits" are further delineated with the hits' critical identifying information, names, chemical structures, CAS numbers and possibility other information. Each "hit" is typically accompanied with a numerical value (e.g. match factors, probability scores, etc.) that is obtained by an established matching algorithm and which indicates the degree to which the hit's mass spectrum corresponds to the test spectrum. The simplest interpretation of the "hit list" is that the hit with the highest match score is taken as the identity of the compound of interest. Even if the top hit is not taken as the compound's identity, it is generally "assumed" that the proper identity for the compound of interest is among the highest ranked reference compounds.

Although the concept is intuitive and straightforward, it is not without its challenges. For example, it is evident that this identification strategy can only be successful if the mass spectrum of the compound of interest is actually present in the library. If the mass spectrum of the compound of interest is missing in the spectral library, it is obvious that the identification strategy via mass spectral matching cannot possibly lead to the proper identification of the compound of interest.

If the compound's mass spectrum is present in the library, the best-case outcome of matching will be that the library spectrum with the highest ranked hit will establish the true identity of the compound. While it would be desirable if this were always the case, there is absolutely no guarantee that this will always be the case in all circumstances. Whether the right match is the best hit depends upon the selected mass spectral match criteria and the similarity between the analytical conditions used to produce the test and reference spectra, among other factors.

Because the match score alone does not reliably produce the right identity every time, there is a certain level of uncertainty in a tentative identity secured in this way. Greater certainty in the identification can be secured if an experienced mass spectrometrist reviews the available spectral information and is able to substantiate the identification.

2.2 Relevance of External Mass Spectral Libraries for Identification

An ideal mass spectral library for the identification of extractables and leachables would:

1. Contain as many of the possible extractables and leachables as possible (increasing the likelihood of securing a match).
2. Contain no substances that are not extractables or leachables (decreasing the possibility of an incorrect match or false positive).
3. Be well controlled and maintained.
4. Be constantly and routinely updated (but in a controlled manner).
5. Be peer-reviewed.
6. Contain spectra that are secured under standardized analytical conditions.
7. Have scientifically validated search algorithms and scientifically vetted means of establishing the match factor.
8. Have a user-friendly output that supports review, interpretation, and assessment.
9. Be compatible with all data platforms used in all generally available commercial instrumentation.
10. Be universally available, accessible, and employed.

Considering point #1, it is likely that many extractables and leachables will be included in the most commonly used commercial spectral libraries (e.g. NIST, Wiley) because these extractables and leachables are themselves commonly encountered organic compounds such as residual solvents, monomers, processing aids, and commonly used additives (anti-oxidants, plasticizers, slip agents, acid scavengers, nucleating agents, curing agents, and others). However, there is also a large population of extractables and leachables whose mass spectra will not likely be included in the most commonly available commercial mass spectral libraries, as these compounds are infrequently encountered degradation and reaction products formed either in the production process of the test item or during the period of time that the test item is in contact with the "communicating" entity (for example, a drug product stored in its container closure system over shelf-life). Such infrequently encountered substances arise by a variety of mechanisms; for example, via oxidation reactions of the polymer or its additives or impurities, hydrolysis reactions, sterilization degradation (cleavage, cyclization, etc.), interactions between polymer additives/impurities during the formation process, oligomer formation and reactions, etc. Although a number of these degradation compounds may be known to the industry – and their mass spectra may be represented in internal, closely-held mass spectral libraries – they are less likely to have been reported and incorporated into the commercial databases.

Moreover, those extractables and leachables that are in the commercial databases are the "low hanging fruit" as they have probably already been encountered in numerous studies and are therefore relatively well-known. It is the fruit that is more difficult to reach, the rarely- or never-before-encountered compounds with complex and largely unrecognizable structures (and thus spectra) that will be most

challenging compounds to identify. If these compounds are not in the commercial libraries, another means will be required for securing their identities.

The two most commonly used hyphenated chromatographic methods for the screening of organic extractables or leachables are gas chromatography – mass spectrometry (GC/MS) and liquid chromatography – mass spectrometry (LC/MS). Mass spectral matching is a particularly powerful tool for securing identities in GC/MS for the simple reason that the operating conditions for the mass spectrometer in GC/MS have been standardized (e.g., electron impact (EI) mass spectra that were recorded at an ionization energy of 70 eV). Because of the highly standardized mass spectrometric data acquisition parameters, EI mass spectra are very reproducible across different GC/MS platforms and reproducible across test systems and commercial libraries. Well-known and well-controlled large commercial reference libraries such as the NIST/EPA/NIH Mass Spectral Library and the Wiley Registry of Mass Spectra are widely used tools to facilitate spectral matching of GC/MS data. Such databases fulfil most of the ideal requirements listed previously with the notable exception that the databases contain many more compounds that are not extractables or leachables than they contain extractables and leachables, thus increasing the likelihood that false and generally impossible identifications are secured.

While external mass spectral libraries are routinely used for mass spectral matching as a first step in identifying a compound in GC/MS, this is not the case for LC/MS. The fact that either in-source fragmentation spectra or multi-stage MS (MS^n) fragmentation spectra are necessary for mass spectral matching in LC/MS, adds to the complexity of the identification process. As there are no standard ionization and fragmentation settings for LC/MS^n detection and because the ionization can be strongly influenced by the chromatographic conditions, no commercial libraries are available that can readily and reliably be used to perform a useful first pass mass spectral matching for each and every LC/MS instrument platform. Although NIST and some instrument vendors have started to create collision-induced dissociation (CID) based MS^n spectral libraries that can serve as a resource to scientists who seek to establish a compound's identity in LC/MS, one should be careful in using these data as the mass spectra were acquired with a certain and specific combination of instrumental detector settings. Wrong selection of the precursor ion or deviations between the experimental conditions or instrument type used to collect the test and library spectrum may produce aberrations between the experimental mass spectrum and the library mass spectrum, complicating the identification process and leading to lower confidence in its outcome. As a consequence, mass spectral matching is not widely applied as a routine practice for securing a tentative identity via LC/MS. Rather, LC/MS matches should be considered to be supporting information for identifications made by other methods, for example, *de novo* structure elucidation using mass spectral interpretation by experienced mass spectrometrists.

2.3 Detection and Discrimination of Analyte Signals (Spectra) for Identification

The process of discovering, identifying, and quantifying organic extractables in extracts (or organic leachables in drug products) involves the analysis of the extract using compatible and orthogonal hyphenated chromatographic techniques, typically gas chromatography/mass spectrometry (GC/MS) and high-performance liquid chromatography/mass spectrometry (LC/MS). These hyphenated techniques yield at least two-dimensional compound specific information for analytes present in

an extract in the form of chromatograms which can be used to make inferences regarding the analyte's chemical structure based on chromatographic retention behaviour and mass spectral data.

Chromatographic selectivity is one of the determining parameters in terms of the reliability of the discovery and identification processes. When chromatographic selectivity increases, peak co-elution decreases and the discriminating power of the analysis method improves. However, peak co-elution cannot generally be avoided and it becomes a challenge to resolve peak responses sufficiently that useful, uncompromised responses can be obtained for the coeluting analytes. It is obvious that unreliable identifications arise when the determined mass spectra of the detected analytes are compromised due to spectral contamination. Typical sources of spectral contamination include ion signals from coeluting compounds, column bleed, solvent tailing or even electronic noise.

As visual inspection of complex chromatograms is an ineffective means of resolving chromatographic peaks and their associated mass spectra, application of data processing techniques that can do a better job than the naked eye are necessary. When determining the mass spectrum that will serve as the basis for identification either through library matching and/or mass spectral interpretation, background subtraction should be the absolute minimum approach. To support a higher quality of identification, however, application of a scrutinous deconvolution-based approach is more preferable as it delivers better quality mass spectra (i.e. free of interferences) and thus reduces the risk of misidentification.

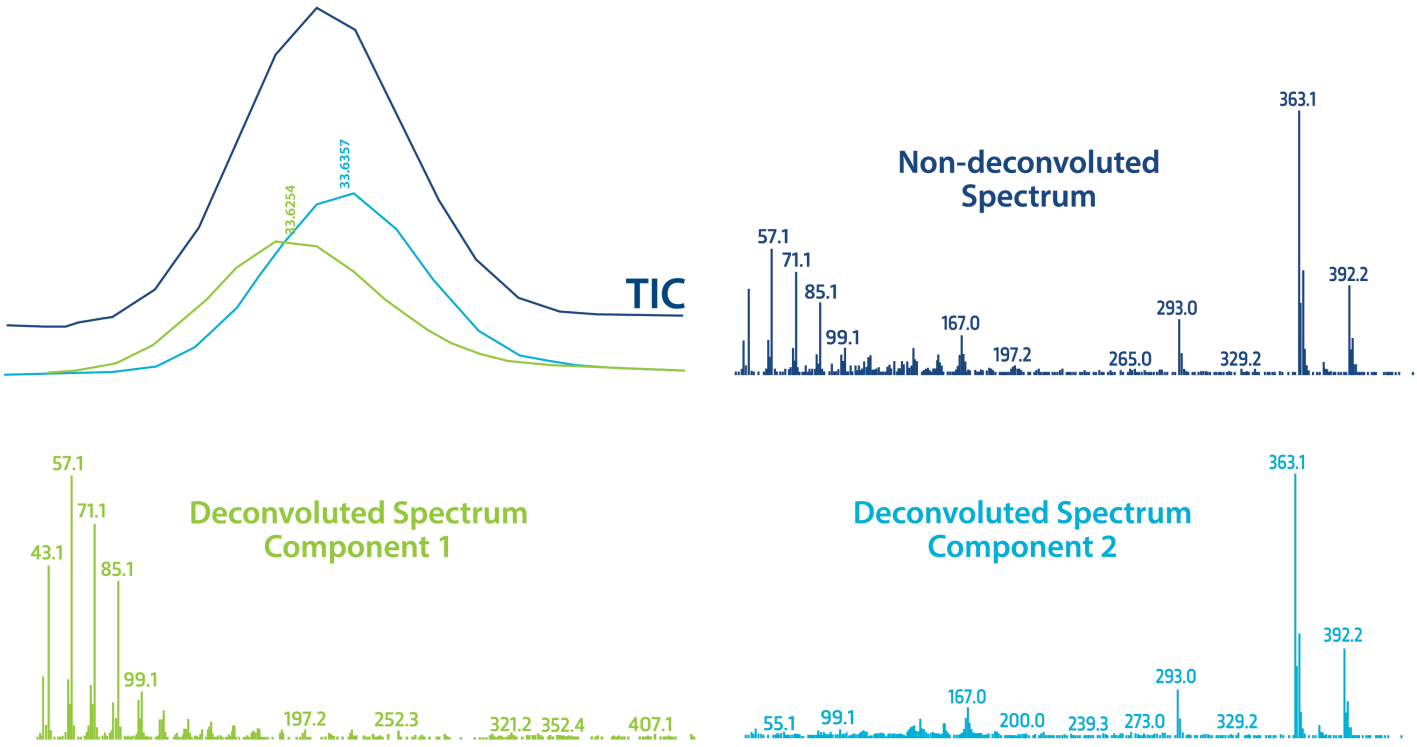


Figure 6. Deconvolution example of a single signal peak detected in a Total Ion Current (TIC) chromatogram that is composed of two closely eluting components.

Deconvolution is the process of computationally extracting analyte signals from a complex mass chromatogram, resulting in the elimination of background noise and the spectral separation of co-eluting compounds. An example is given in Figure 6.

The deconvolution process involves multiple steps such as noise analysis, peak shape analysis (ions belonging to the same peak should have the same apex and peak shape) and the assembly of a deconvoluted spectrum.[3] It should be noted that deconvolution algorithms and related parameters can differ among various software platforms and therefore it is possible to generate slightly different deconvoluted spectra for the same raw data.

To facilitate identification, most instrument software platforms enable the combination of deconvolution with a mass spectral library search, i.e. comparison of the resulting spectra against collection(s) of reference mass spectra included in either public, commercial or user defined libraries using matching algorithms. The result is an (indexed) hitlist containing the most similar spectra present in the searched libraries. Spectral library search programs are very powerful supporting tools for the identification of analytes in non-targeted analysis and are a well-established strategy that has long been applied in the reporting and identification of detected (non-target) analytes in environmental analysis.[4] Relying solely on mass spectral matching to secure an identity, however, is an inappropriate interpretation practice as more incorrect results are probably reported because of sole reliance on mass spectral library search programs than have been reported due to all other types of errors that can occur in mass spectral data.[5] A high correlation (high match index or probability of match) between an unknown spectrum and a library spectrum does not necessarily mean that the unknown has been identified unequivocally. Additionally, the match with the highest match score is not always the correct identity. The criteria needed to identify an unknown by chromatography mass spectrometry must include a visual comparison of the unknown and library spectrum as documented by an expert mass spectrometrist and may demand that the identification hypothesis is corroborated by additional information such as its expected retention behaviour, etc.[6]

2.4 Evaluation of Mass Spectral Matching Results

As noted previously, mass spectral matching can result in the tentative identification of detected compounds, producing the minimum level of information suitable for subsequent toxicological assessment. However, the use of undisciplined or unsubstantiated mass spectral matching is problematic. This is the case because the decision that the match is the correct identity is often based solely on the calculated similarity values (i.e. match factors) without subsequent critical review of the spectra. A high correlation between an experimental spectrum and a library spectrum does not necessarily mean that the identification is unequivocal. Moreover, unilaterally choosing the highest ranked hit as the reported identity is problematic at best and has been established to generate false positives in many cases.[3]

Due to the complexities of mass spectral interpretation, specifying "objective" criteria for a mass spectral matching identification strategy is not straightforward. Match factors (MF) are calculated values which indicate the similarity of two spectra by comparing individual m/z values and their corresponding intensities. In practice however, there is no "universal" match factor (MF) threshold value that exclusively by itself establishes that the corresponding match-based identity represents the true identity of a compound. The underlying reason for this is the varying degree of spectral uniqueness among the universe

of chemical compounds. Certain compounds may have a rather unique spectrum and are thus more likely to be correctly identified, while others may have a spectrum that very closely resembles the spectra of many other compounds. Nevertheless, the "goodness" of a mass spectral match factor can be correlated with the probability that the match factor has suggested the right compound. The lower a mass spectral match factor, the lower the quality of the fit and the more mass spectral interpretation efforts are necessary to justify an identification decision that is solely based on mass spectral matching. These efforts may include, for instance, inspection of a mirror image of both experimental and library spectrum to reveal the presence of additional or missing m/z values in either spectrum. Additionally, when using the NIST MS search software, the probability score and In Lib score can be evaluated. The former represents the relative probability that any matching spectrum in the hit list is correct, while the latter is a measure of the probability that the spectrum of the compound being searched is contained in the library. Hits with an MF below 700 are generally associated with a very low probability that the identification is correct. Never-the-less, identifications based on an MF below 700 have been reported by testing laboratories, particularly in the case when a MF below 700 is the highest ranked or the only match.

The uncertain nature of identification by mass spectral matching leads to the conclusion that mass spectral matching based identifications always require a close examination by a mass spectrometrist and that such identifications remain tentative unless they are corroborated by additional evidence, such as information obtained through analysing the authentic standard (mass spectrum and retention time) or additional supporting documentation.

Various approaches can be used to review mass spectral matching results and the intensity of the review could vary depending on the quality and the number of the returned match results from the search as illustrated in the examples that follow.

Various maximum MF values are used by different software vendors. In the examples that follow, MFs are expressed relative to a maximum value of 999.

When using the NIST Identity Search, spectrum search results in a hit list are summarized using four numeric descriptors: the Match Factor value, the Reverse Match Factor value, the Probability value and the InLib probability value. Maximal values for the descriptors represent a perfect match and are 1000, 100 or 1 depending on the data system.

The Match Factor for the unknown and the library spectrum assumes that the former originates from a single compound and uses all peaks in both spectra for spectral similarity determination, in other words, it is a direct match of peak m/z values and relative intensities (pure spectrum match factor).

The Reverse Match Factor for the unknown and the library spectrum assumes that the former spectrum can be contaminated by "impurities." In its calculation, peaks in the unknown spectrum that are missing in the library spectrum are disregarded (impure spectrum match factor). The Reverse Match Factor consequently enables the identification of multiple compounds represented by a single spectrum. The closeness of the Match and Reverse Match factor values should consequently be considered as a measure of the 'purity' of the similarity.

The Probability value describes the likelihood of the unknown and reference spectrum being from the same compound based on all the matches found during the search. It is derived if the compound is represented by a spectrum in the libraries and uses

the differences between adjacent hits in the hitlist to determine the relative probability that any hit in the list is correct. This value is derived from an analysis of the results of searching the NIST/EPA/NIH Main Library with a set of replicate spectra (given in the Replicates Library). The relative probability of each of the hits requires only the difference values because the total probability of the compound being in the searched libraries is assumed to be one.[5] When the best hit has a high Match Factor value (>900) and the next hit has a much lower value (e.g. 800 or less), the Probability value means that the probability of the compound being correctly identified is very large and that the probability of the compound being in the searched library is also large. When the Probability is high, it means that – apart from the hit – there are no other good matching mass spectra in the library which makes the hit 'unique' and obviously increases the likelihood of a correct tentative identification. When the Probability is low, it means that there are other good matching mass spectra present in the library, which makes it difficult to pick the best hit. This is typical when isomers exist (e.g. xylenes). Caveat: this descriptor assumes that the target molecule is present in the library, which is in reality a false hypothesis to start from!

As its name suggests, the In Lib probability value indicates the probability that the searched compound is present in the searched libraries and is meant as a guidepost. Generally, any positive value is acceptable. Values greater than approximately 300 usually mean that the spectrum is nearly unique. Negative values below 200 are generally a warning that the spectrum is not identified. Note that negative values will occur when there are many compounds with similar spectra. In these cases, the difference between the Match Factors for different spectra is very small, and the search cannot be assured of providing the correct unique answer. Especially when Match Factors are high, the In Lib value will provide very good guidance on the structure of the molecule.[8]

2.5 Examples of Identification by Mass Spectral Matching

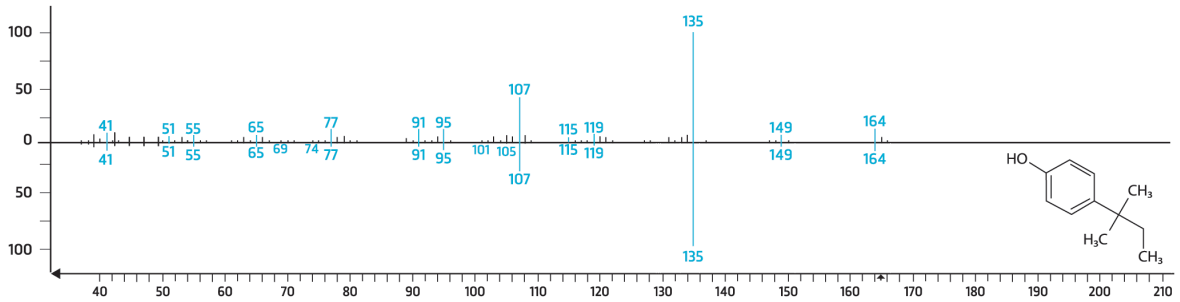
2.5.1 Example 1: Correct Identification for Best Hit (MF > 900); GC/MS

The top-3 hit list for a compound of interest is presented in Table 1 with the respective MS spectra being displayed in Figure 7. All 3 hits have relatively high match factors. The highest ranked candidate has a high match factor of 931 and visual review of the mass spectra shows an almost perfect mirror

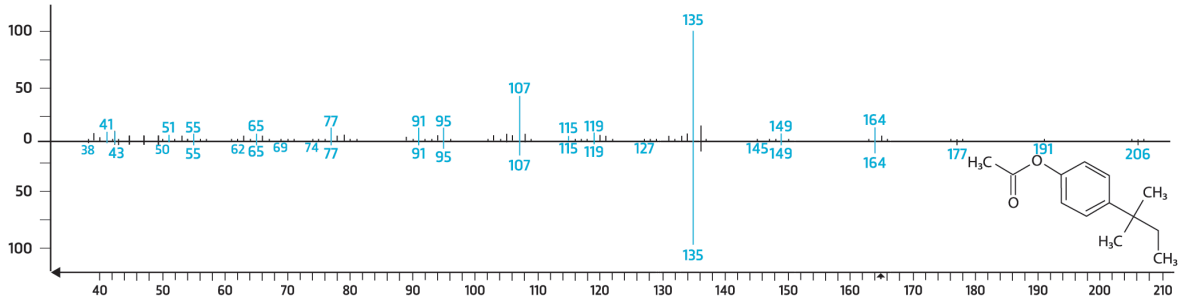
Table 1. Top 3 ranked hit list for a mass spectral matching example with an excellent (> 900) match score and a high probability of securing a correct tentative identity for a compound, matched using NIST MS Search v2.3. Experimental RI = 1394.

Rank	Candidate	Match	Reverse Match	Probability (%)	Retention Index (RI)[1]
1	p-tert-pentylphenol	931	931	81.2	1400 ± 4
2	P-tert-pentylphenol acetate	825	825	5.45	1502
3	p-tert-butylphenol	816	852	3.96	1295 ± 3

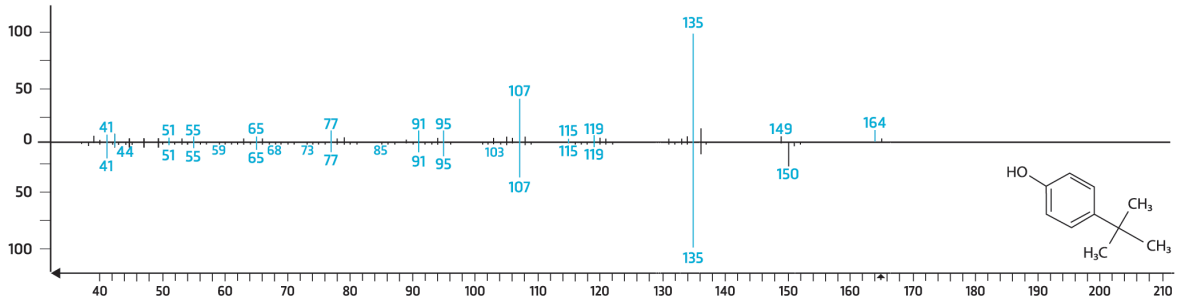
Notes: [1]Experimental retention index, mean ± standard deviation if there are multiple entries.
In Lib Score = 312.



A: Rank 1 candidate spectrum



B: Rank 2 candidate spectrum



C: Rank 3 candidate spectrum

Figure 7. Mirror mass spectrum plots of an unknown (top) and its match factor based top three ranked identification candidates where the Rank 1 candidate represents the correct identification

image match, without missing characteristic ions, between the experimental or library mass spectrum (spectrum A in Figure 7) and the test spectrum. This is in contrast to the two lower ranked hits, where image match is not nearly perfect. The second ranked spectrum contains additional peaks at m/z 177, 191 206 which are not detected in the experimental spectrum (spectrum B in Figure 7). The third ranked spectrum contains an additional peak at m/z 150 and lacks peaks at m/z 149 and 164 compared to the experimental spectrum (spectrum C in Figure 7). These observations lead to the conclusion that top

hit, p-tert-pentylphenol, has a high probability of being the correct tentative identity for the compound. In addition, comparison of the unknown's retention index (1394) to the experimental retention indices present in the library further corroborates the identification and typically would support upgrading the identification to the confident level.

2.5.2 Example 2: Correct Identification for Best Hit (800 < MF < 900); HS-GC/MS

This example shows a top-5 hit list for a second compound of interest (Table 2) where only the best hit has a match factor above 800. Visual inspection of the spectra (Figure 8) shows a good mirror plot for the best ranked hit; the only marked difference is the relative intensity of the peak clusters at m/z 249 and m/z 265. The lower ranked hits, on the contrary, show clearly deviating features such as additional or missing m/z values and very different relative intensities. In addition, Table 2 shows that the probability score of the best hit is very high and markedly different from the lower ranked hits. Therefore, the compound can be tentatively identified as octamethylcyclotetrasiloxane with a high degree of confidence.

2.5.3 Example 3: Incorrect Identification for Best Hit (800 < MF < 900); GC/MS

Table 3 shows another example for a third compound of interest where the highest MF value is approximately 800. Upon cursory examination, this top hit seems to provide a good and acceptable match. Mirror plots for the five best hits (Figure 9), however, show that none of the library spectra are very good matches to the experimental spectrum. Either there

Table 2. Top 5 ranked hit list for a mass spectral matching example with a good (800-900) match score and a high probability of securing a correct tentative identity for a compound, matched using NIST MS Search v2.3.

Rank	Candidate	Match	Reverse Match	Probability (%)	Retention Index (RI)[1]
1	Octamethylcyclotetrasiloxane	828	842	92.7	N/A
2	1,1,3,3,5,5,7,7-Octamethyl-7-(2-methylpropoxy)tetrasiloxan-1-ol	757	757	4.04	N/A
3	3-Ethoxy-1,1,1,5,5,5-hexamethyl-3-(trimethylsiloxy) trisiloxane	730	738	1.18	N/A
4	3-Butoxy-1,1,1,5,5,5-hexamethyl-3-(trimethylsiloxy) trisiloxane	726	739	1.00	N/A
5	2,6-Dihydroxyacetophenone, 2TMS derivative	697	697	0.28	N/A

Notes: [1]RI values not available for the stationary phase used
In Lib Score = 237

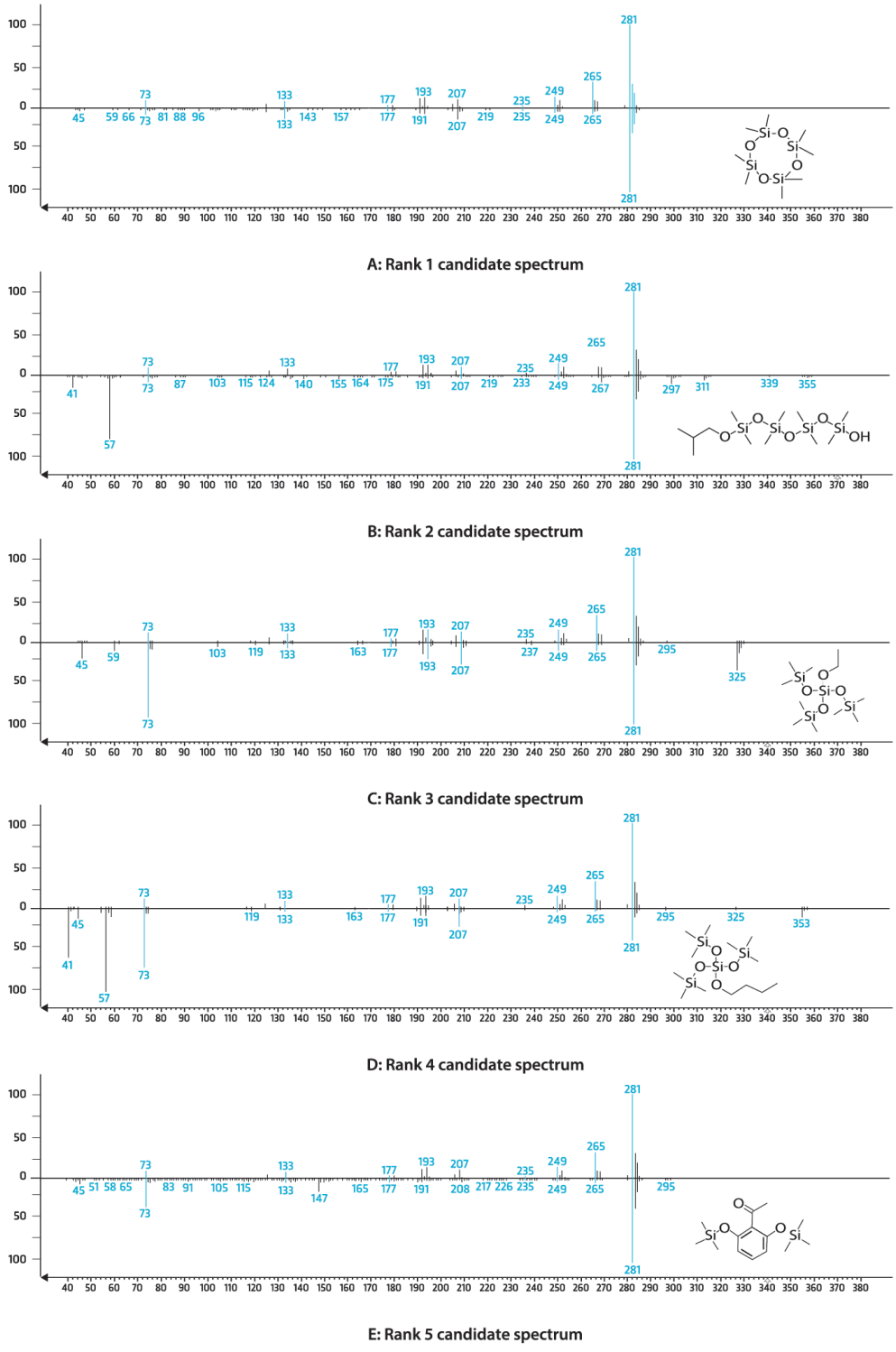


Figure 8. Mirror mass spectrum plots of an unknown (top) and its match factor based top five ranked identification candidates where the Rank 1 candidate represents the correct identification (Table 2).

Table 3. Top 5 ranked hit list for a mass spectral matching example with moderate and low match scores and a low probability of securing a correct tentative identity for a compound, matched using NIST MS Search v2.3. Experimental RI = 916.

Rank	Candidate	Match	Reverse Match	Probability (%)	Retention Index (RI)[1]
1	Methoxy-phenyl-oxime	804	822	85.3	1301 ± 382
2	Cyclopentyl-4-ethylbenzoate	698	710	5.73	1713 ± 201
3	Sec-butyl-4-ethylbenzoate	671	677	1.68	1507 ± 201
4	Cyclohexyl-4-ethylbenzoate	668	679	1.49	1883 ± 201
5	Isobutyl-4-ethylbenzoate	652	678	0.85	1507 ± 201

Notes: [1]Estimated value ± 95% confidence interval.
In Lib Score = -135.

are a number of characteristic ions which are present in the experimental spectrum and not in the library spectrum, or vice versa. For instance, peaks at m/z 42, 55 and 73 in the best hit library spectrum (spectrum A in Figure 9) are missing in the experimental spectrum which indicates that this identity is incorrect. The same decision can be made for the other hits (sectra B-D in Figure 9) because they either lack characteristic ions in the test spectrum or have deviating relative intensities.

Moreover, the In Lib value of the mass spectral search (In lib Score = -135) suggests that the compound being searched is not present in the mass spectral library and the unknown's retention index (916) does not correspond to any retention index in the hit list. Consequently, further mass spectral interpretation efforts by an expert are necessary to identify the compound of interest .

2.5.4. Example 4: False Positive Identification for Best Hit (700 < MF < 800)

Identification based on mass spectral matching becomes even more difficult when the quality of the match factors deteriorates further, as reflected in even lower MF values. An example of this is shown in Table 4 and the associated Figure 10 where the MF values are between 750 and 700 for the five best ranked hits. A visual inspection, performed by an experienced mass spectrometrist, would reveal than none of the candidate library spectra fit the experimental spectrum of the compound of interest. Consequently, additional efforts in mass spectral interpretation are essential to secure the correct identity of this compound.

Once a mass spectrum has been obtained, the process of identification begins in one of two ways, spectral matching, or structure elucidation. Considering spectral matching, it is noted that this technique is most productive when it is applied to GC/MS data as large, well-maintained and standardized commercial libraries of spectra exist. Nevertheless, even in this circumstance, spectral matching is not an exact science and the following recommendations are made to assist in securing the right identity:

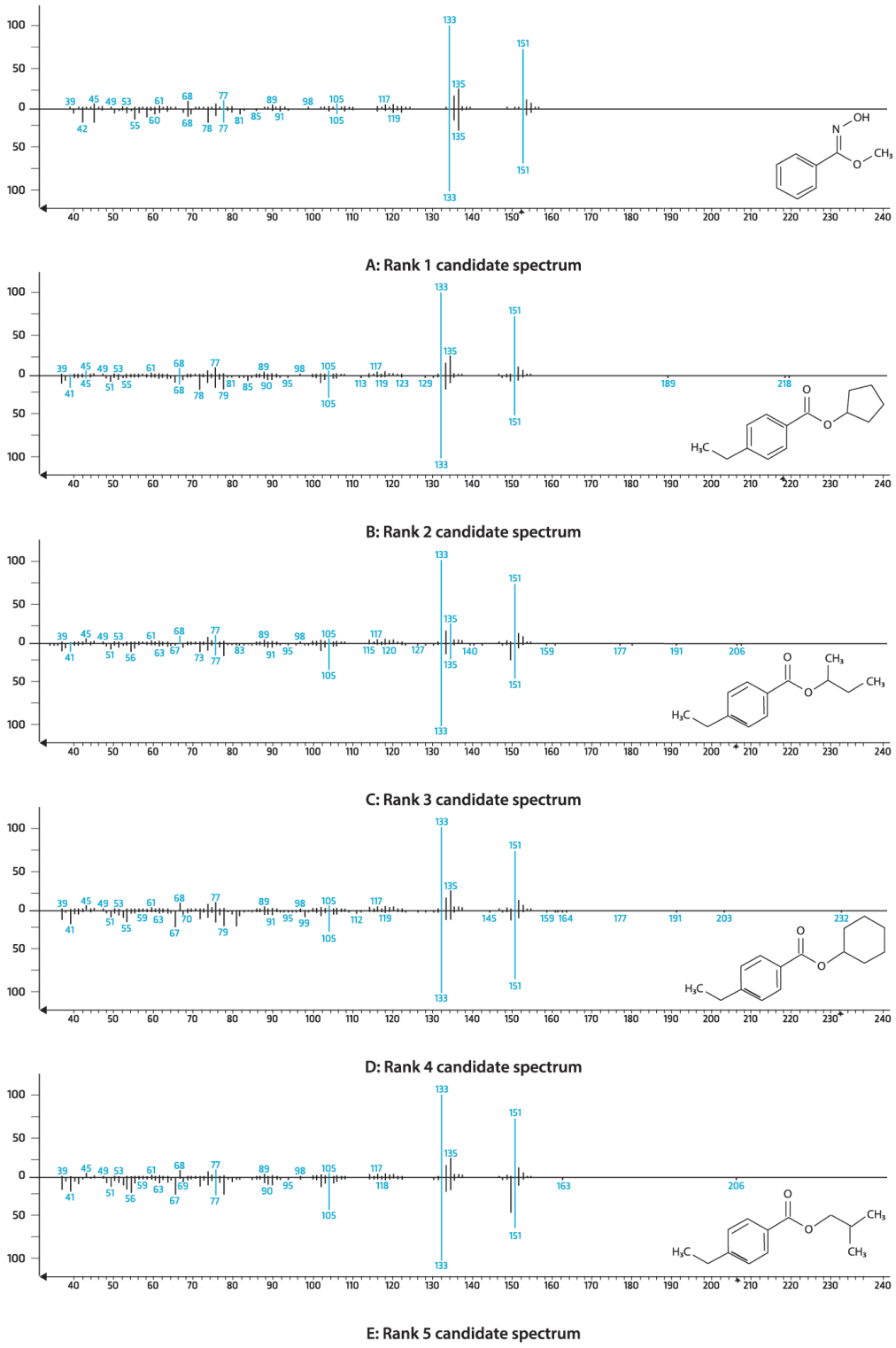


Figure 9. Mirror mass spectrum plots of an unknown (top) and its match factor based top five ranked identification candidates where none of the identification candidates is correct (Table 3).

Table 4. Top 5 ranked hit list for a mass spectral matching example with moderate and low match scores and a low probability of securing a correct tentative identity for a compound, matched using NIST MS Search v2.3. Experimental RI = 1178.

Rank	Candidate	Match	Reverse Match	Probability (%)	Retention Index (RI)[1]
1	4a,8a-Dimethyloctahydro-2(1H)-napthalenone	743	743	17.5	N/A
2	1-Methyl-1-cyclododecane	723	723	7.93	1387 ± 5
3	Tetrahydroionyl acetate	713	713	5.62	N/A
4	Neophytadiene	710	775	4.97	1837 ± 5
5	(2,2-Dimethylcyclopentyl) cyclohexane	709	712	4.77	N/A

Notes: [1]Estimated value ± 95% standard deviation. N/A = RI not available
In Lib Score = -733.

- If the mass spectrum of the compound of interest is missing in the spectral library, it is obvious that the identification strategy via mass spectral matching cannot possibly lead to the proper identification of the compound of interest.
- It should be stressed that there is no value for the mass spectral match factor (MF) that unequivocally guarantees that the correct identity of a compound has been determined, based upon the MF alone. It is clear that the exclusive reliance on mass spectral match factors without any expert review cannot robustly and routinely lead to correct identifications. Reporting the highest ranked hit as an analyte's tentative identity is an all-to-common error and matching based on absolute MF thresholds can still lead to false positive identifications. Therefore, the practice of reporting the identity of an analyte of interest as the compound with the highest match score by default is strongly discouraged and any proposed identity should be verified an expert mass spectrometrist, which may require that the identification hypothesis is corroborated by additional interpretation efforts or information.
- An expert should always visually evaluate the spectra of match candidates in the mass spectral matching's hit list, regardless of the quality of the MF. This evaluation serves as a means to compare the target mass spectrum with the library spectrum in order to verify the resemblance of all mass fragments in both mass spectra.
- It is consequently considered as good and necessary practice that tentative identifications based on mass spectral matching are always substantiated by comparative spectrum plots such as mirror plots, as such comparative data allows one to visually confirm the quality of, and increase the confidence in, the fit of the matched spectra.

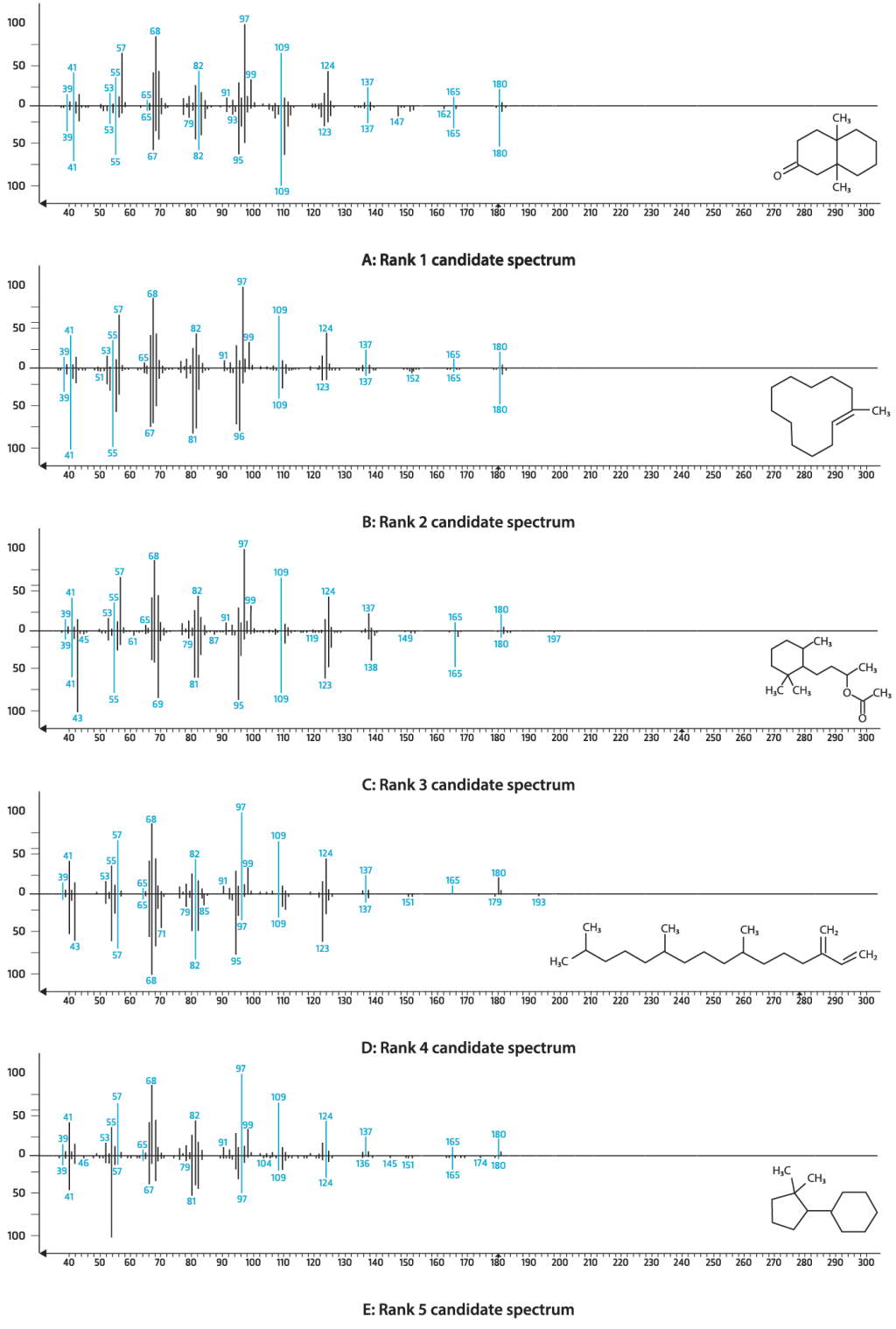


Figure 10. Mirror mass spectrum plots of an unknown (top) and its match factor based top five ranked identification candidates where none are the correct (Table 4).

- The lower the MF, the more intense the mass spectral interpretation exercise will need to be to secure the identity of the compound based solely upon the merits of its mass spectrum.
- While MF's above 80% (or 800 depending on the scoring scale) may lead, after a careful mass spectral evaluation, to an unequivocal identification, the probability of securing the right identity via mass spectral matching decreases quickly below these values. When the MF's are further deteriorating, e.g. below 70% (or 700), the probability of correctly identifying a compound primarily based upon MF is extremely low. In that case the most likely outcome will be that the compound remains unidentified, although it is possible that the match is sufficient to substantiate and support a partial identification, based on the principles outlined in Part 3: Identification by Mass Spectral Interpretation.
- Identifications of organic compounds, solely based upon the practice of mass spectral matching, should be considered as tentative identifications, as the identification is a "one" dimensional identification, where the one piece of evidence is its mass spectrum. The Identification Class can be augmented by acquiring information obtained through analyzing the authentic standard (mass spectrum and retention time: confirmed identity) or through additional supporting documentation.
- Despite the availability of large collections of reference spectra in commercially available MS-libraries, mass spectral matching is most effective and definitive when the reference spectra have been produced on the specific instruments and with the specific methods that will be used for the analysis of actual samples (extracts or drug products). It is therefore recommended that E&L laboratories create and maintain their own mass spectral library of spectra using reference standards of the analytes of interest to maximize the correctness of identifications. Moreover, it is likely that such custom libraries will be populated with highly relevant compounds, avoiding issues such as false identifications linked to irrelevant compounds (for example, identifying an extractable as an insecticide) and lack of identification because a compound unique to extractables and leachables is not present in the commercially available library. This latter point is of particular concern, given the unique chemical nature of extractables and leachables, the continuous evolution of materials used in pharmaceutical packaging and medical devices (translating to new and previously unencountered extractables and leachables), and the proprietary nature of such materials, where undisclosed compositions means that ingredients are not added to the commercial databases because they are unknown.

Part 3

Identification by Mass Spectral Interpretation

Mass Spectral Interpretation is the process of securing a compound's identity solely by expert interpretation of the information that is made available through the compound's mass spectrum. This is an identification strategy that is often applied to the mass spectral information generated in an LC/MS experiment, as there are no universal commercial databases available that can provide identities based solely on mass spectral matching. However, it can also be necessary to follow this type of approach for GC/MS when mass spectral matching does not lead to a reliable identity for the detected compound.

3.1 Introduction to Mass Spectral Interpretation

Although tentatively identifying a compound of interest via GC/MS can often be accomplished by mass spectral matching (or MSM), unequivocal identification based solely on MSM is not possible. Clearly, there will be situations where a compound's mass spectrum cannot be effectively matched to a library spectrum, meaning that MSM fails to provide even a tentative identity for the compound of interest. Moreover, even if MSM produces a credible match, the resulting identification is only tentative as it is based on only one dimension of supporting information (the spectral match itself).

In either case, securing an identity when MSM fails or elevating a tentative identification secured by MSM, an alternative identification strategy, which is also the basis for Mass Spectral Identifications in LC/MS, involves the expert interpretation of the spectrum's individual features (mass values and their relative abundances). While it is not the intent of this document to provide a comprehensive and detailed discussion of all the fundamentals of mass spectral interpretation (MSI), essential principles and practices are discussed and illustrated.

In general, the MSI identification consists of three consecutive steps:

1. Determining which peak in the mass spectrum corresponds to the molecular weight of the molecule. In case the spectrum is acquired with an accurate mass high resolution instrument, the mass-to-charge ratio (m/z) of that ion can be used to generate a candidate molecular formula.
2. Establishing whether the compound of interest contains certain elements, such as chlorine or bromine, which have specific isotope patterns which translate into recognizable spectral features, namely specific relative abundances of monoisotopic masses.
3. Performing *de novo* structural elucidation. All peaks in a mass spectrum with an m/z value below that of the molecular ion are formed during the ionization (or MS/MS fragmentation) of the compound of interest and relate in one way or another to substructures or functional groups of that substance. An expert in fragmentation chemistry can relate this information to the molecular ion and potentially propose a tentative structure via a process that is generally referred to as "*de novo* structural elucidation". Although algorithms have been developed to assist in certain aspects of such an elucidation, it should be emphasized that structural elucidation

is always a subjective interpretation performed by an expert mass spectrometrist and therefore that any identity secured by structural elucidation is classified as being a tentative identification until additional collaborating data allows for an upgrade of the identification level.

3.2 Identification of the Molecular Formula

Identifying the elemental composition (molecular formula) of an unknown from its mass spectrum typically starts with determining which ion peak in the mass spectrum represents the m/z value of the intact, ionized molecule, termed the molecular ion. In many cases, this determination is not as simple as just picking the highest m/z value in the spectrum. The ability and strategies to ascertain the (pseudo) molecular ion depend heavily on the type of ionization technique used.

GC/MS spectra are generally acquired with electron ionization (EI) which produces a radical molecular ion M+• with highly variable intensities. The intensity of the molecular ion depends on its propensity to decompose into several smaller fragments which in turn is dictated by the stability of the ion under the applied ionization conditions. For instance, the molecular ion is usually very intense for compounds which are highly stable under these ionization conditions, such as (*poly*)aromatics (example shown in spectrum A of Figure 11), while it is often not detected in spectra of largely unstable compounds such as

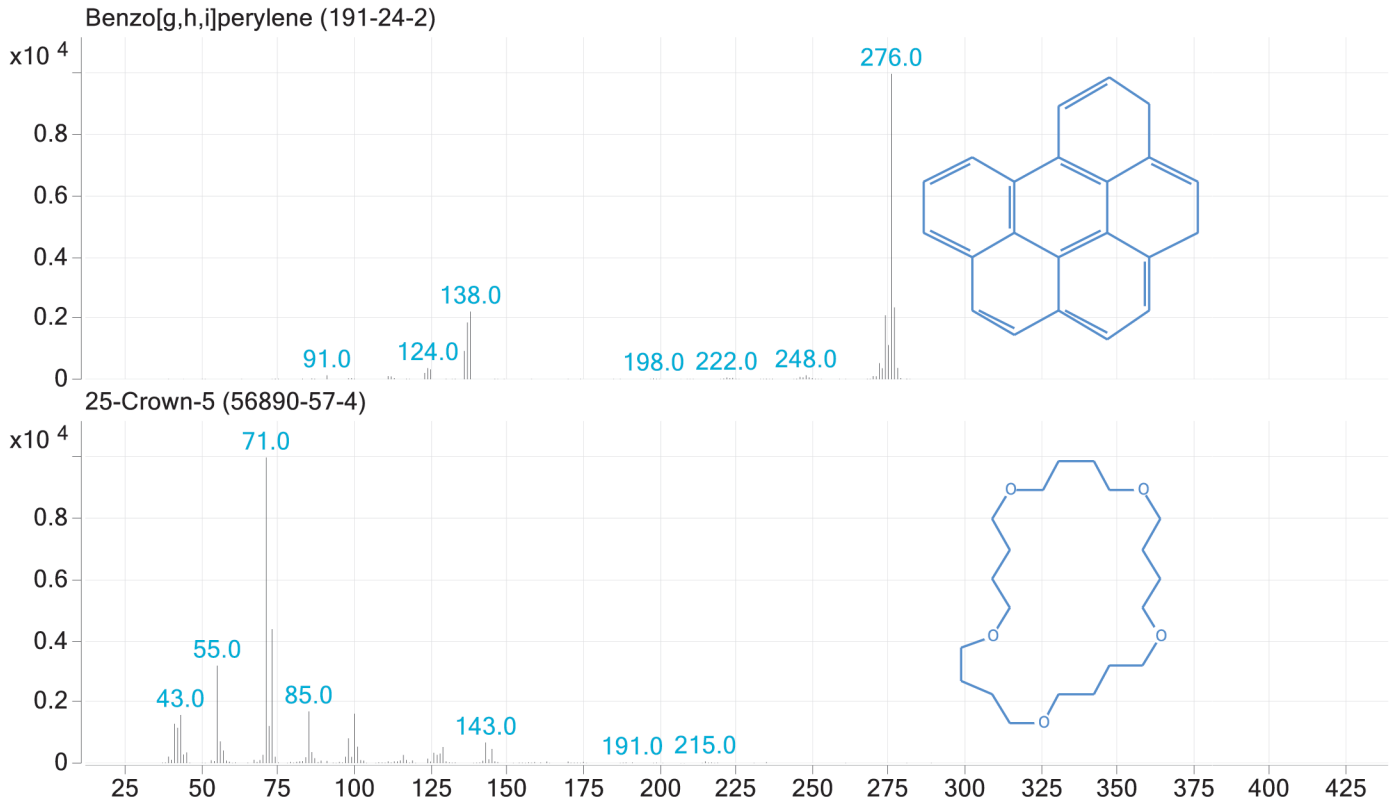
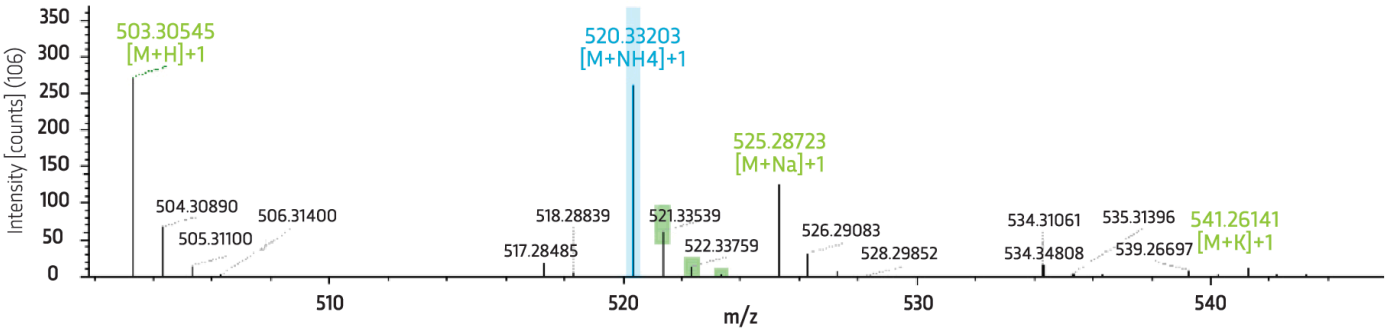


Figure 11. Variability in intensity of the molecular ion in EI mass spectra demonstrated for a polyaromatic aromatic hydrocarbon, Benzo[g,h,i]perylene (spectrum A), and a crown ether, 25-Crown-5 (spectrum B). Benzo[g,h,i]perylene ($C_{22}H_{12}$) shows a clear molecular ion at m/z 276, whereas the molecular ion expected for 25-Crown-5 ($C_{20}H_{40}O_5$) at m/z 360 is not detected.



igure 12. Example of a positive mode ESI high resolution accurate mass spectrum of a compound illustrating formation of the pseudo molecular ion ($[M+H]^+$) and concurrent adduct formation with alkali ($[M+K]^{+1}$; $[M+Na]^{+1}$) and ammonium ($[M+NH_4]^{+1}$) salts. The Ion at m/z 503,30545 will be used for further evaluation in Table 5.

aliphatic alcohols, highly branched compounds and polyether glycols (example shown in spectrum B of Figure 11). Therefore, an independent assignment of the molecular ion for EI spectra, while not impossible, can be prone to the error of incorrectly picking an m/z value which is, in fact, associated with a fragment of the molecular ion.

More intense molecular ions are usually produced by "soft" ionization techniques, such as atmospheric pressure chemical ionization (APCI) or electrospray ionization (ESI) for LC/MS methods and chemical ionization (*CI*) for GC/MS methods. Ionization of the molecule can result in protonated $[M+H]^+$ or deprotonated $[M-H]^-$ ions depending on the polarity

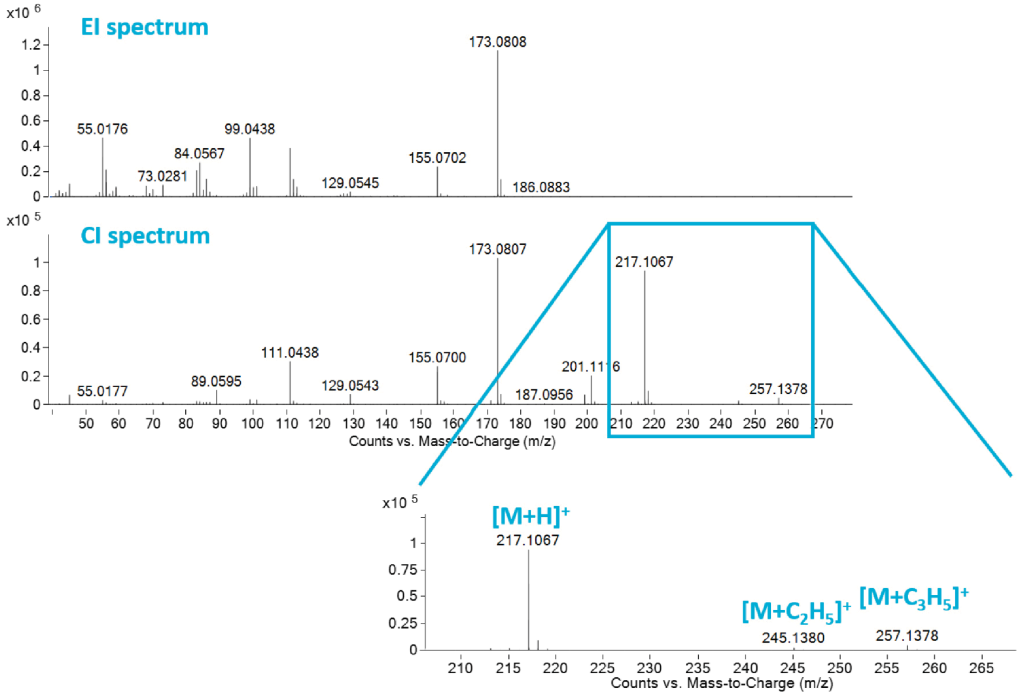


Figure 13. Identification of the molecular ion of 1,4,7-Trioxacyclotridecane-8,13-dione ($C_{10}H_{16}O_5$) based on CI mass spectra in GC/MS. While the EI spectrum (top) shows no clear molecular ion, the CI spectrum (bottom) shows a clear protonated molecular ion $[M+H]^+$ at m/z 217.1067. CI data were acquired with methane as reagent gas which favors the formation of typical methane adducts. These were detected at m/z 245.1380 ($[M+C_2H_5]^+$) and m/z 257.1378 ($[M+C_3H_5]^+$) and thus reinforce the identification of the molecular ion.

of ionization. In addition, adducts may be formed during ionization by reaction or clustering of the molecule with chemical entities present in the sample or mobile phase or due to reaction with a reagent gas. For instance, adducts with alkali or ammonium salts (e.g. Na^+, K^+, $NH4^+$ in positive mode, Cl^- in negative mode) are frequently observed in APCI or ESI spectra. An example of this general phenomenon is given in Figure 12.

In case of CI spectra (e.g. in GC/MS), the protonated molecule often co-occurs with adducts between the molecule and the ionized reagent gas. An example of this is given in Figure 13. In addition to adducts, soft ionization may also be associated with in-source fragmentation depending on the ionization conditions, the stability of the (pseudo) molecular ion and, in case of CI, on the proton affinity of the molecule. In general, a thorough evaluation of adducts and in-source fragments is necessary to confirm the molecular ion.

Additionally, dimeric ions or even higher clusters can also be formed in case of APCI or ESI.

For both hard and soft ionization technologies, it should be emphasized that identifying the molecular ion in a mass spectrum is a subjective interpretation performed by a mass spectrometry expert; thus, there is a degree of uncertainty in the interpretation. In many cases, this uncertainty will be greater for electron impact ionization than it is for soft ionization techniques. Unfortunately, the degree of uncertainty cannot easily be expressed as a mathematical number such as the probability score used in mass spectral matching.

Although establishing the (pseudo) molecular ion with unit mass resolution is a significant step in compound identification, such information in itself is rarely adequate to secure even a tentative identification. However, if the molecular weight of the ion could be established with a high degree of resolution, the exact (or accurate) mass so secured could be used to generate a short list of candidate compounds whose molecular formulas have molecular weights equal to the determined accurate mass. An example of such a table where "candidate" elemental formulas are ranked, based upon the deviation of their calculated m/z compared to the measured m/z, can be found in Table 5. This exact mass information can be obtained with a high resolution - accurate mass spectrometer (HRAMS) such as time-of-flight, orbitrap or ion cyclotron resonance mass spectrometers. The selectivity and mass accuracy of HRAMS instrumentation facilitates the distinction between candidate elemental composition formulas that would be undistinguishable on unit mass instrumentation obtained from quadrupole or ion trap-based mass spectrometers. On such instrumentation, ions that only slightly differ in m/z value would be detected as isobaric signals. For example, diethyl fumarate and 2-fluorobiphenyl have the same unit mass of 172 Da and thus would be indistinguishable on this basis alone. However, their accurate masses (172.0730 Da and 172.0683 Da, resp.) are sufficiently different that they would be readily distinguished on the basis of the elemental compositions obtained using HRAMS since such mass measurements enable the determination of the ions elemental composition by considering the sum of the exact masses of various nuclides ($C_8H_{12}O_4$ and $C_{12}H_9F$, respectively).

The generation of molecular formulas from accurate mass information is usually assisted by software algorithms using user-defined search constraints. Search criteria include the species and quantity of allowed elements, allowed mass accuracy (depends on the resolution of the mass spectrometer), the charge state (e.g. singly or multiply charged) and the allowed electron state, which refers to the

Table 5. Predicted elemental formulas and corresponding number of rings plus double bonds (RDB) values for an even electron ion with m/z 503.30545 considering C, H, O and N as allowed elements with a m/z deviation tolerance of 10 ppm.

Candidate #	Elemental Formula	Theoretical m/z	RDB	m/z Deviation (ppm)
1	C24 H39 O4 N8	503.30888	9.5	-6.811
2	C23 H43 O8 N4	503.30754	4.5	-4/154
3	C22 H47 O12	503.30620	-0.5	-1.497
4	C19 H39 O6 N10	503.30486	5.5	1.181
5	C18 H43 O10 N6	503.30352	0.5	3.839
6	C30 H39 O3 N4	503.30167	13.5	7.515

number of electrons (even or odd) and depends on the ionization technique. Soft ionization techniques normally produce ions with an even electron state. EI spectra, on the other hand, generate a radical molecular ion with an odd number of electrons while fragments of the molecular ion can either have an odd or even number of electrons. For the ESI-HRAMS example in Figure 12, several even electron elemental formulas can be predicted for the determined pseudo molecular ion detected as [M+H]+ at m/z 503.30545, Table 5.

In addition to establishing molecular formulas, accurate mass information can even give structural information based on the number of ring and double bond equivalents rule that is a conventional measure of the degree of the unsaturation of an organic molecule corresponding with the lowest formal valence state of the elements present in its elemental formula.

Depending on the mass resolution and mass accuracy of an HRAMS measurement and the structure of the compound of interest, it may or may not be possible to choose among the multiple candidate elemental formulas that are a reasonable match to the accurate mass established as the m/z peak of the (pseudo) molecular ion. In that case, the correct molecular formula can be established by evaluation of the isotopic data.

3.3 Interpretation of Isotopic Data

Most elements appear naturally as a mixture of isotopes of which the stable isotopes are of prime importance for identification purposes. For example, natural carbon is a mixture of 98.9 % of isotope ^{12}C and 1.1 % of isotope ^{13}C. The natural isotopic composition of a molecule is reflected in the mass spectrum by the presence of isotopic clusters. Such a cluster is composed of distinct monoisotopic masses with relative abundances that reflect their distinct isotopic compositions. By consensus, the first peak in the cluster of peaks corresponding to the most abundant isotopes of a given ion is designated as X. The isotopic regions in a mass spectrum that corresponds to 1 or 2 (or more) mass units further away from X are designated as X+1, X+2, etc. regions. The common elements such as C, H, N, and O – which have a diagnostic isotopic pattern with relatively

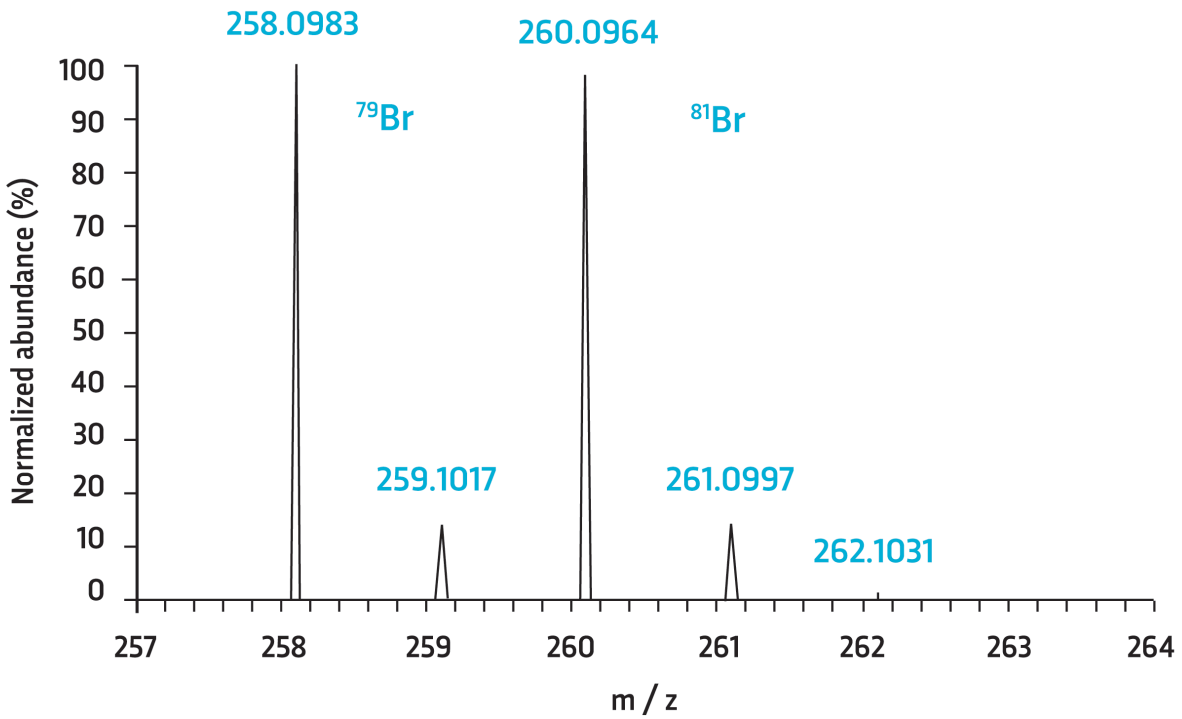


Figure 14. Example of a diagnostic bromine isotope pattern for elemental formula $C_{13}H_{23}Br$ rubber oligomer. The almost equal size of the two chromatographic peaks represent the almost equal natural abundances of these isotopes (50.7 % ^{79}Br and 49.3 % ^{81}Br).

low abundance of their first isotope at X+1 – do not show very intense and obvious isotopic compositions when measured as unit mass. On the other hand, certain elements such as Cl, Br, S, K and Si have very characteristic and intense isotopic distributions up to the X+2 region. For instance, bromine isotopes have a natural composition of 50.7 % ^{79}Br and 49.3 % ^{81}Br. Therefore, a molecular formula of $C_{13}H_{23}Br$ has an average molecular weight of 259.231 which in the mass spectrum will be observed as distinct peaks at 258.098 Da and 260.096 Da with almost equal intensities as shown in Figure 14.

These diagnostic isotopic clusters are readily recognizable by a mass spectrometry expert and can be used to reveal the presence of specific elements. In addition, algorithms have been developed to predict the presence of certain elements such as chlorine and bromine (e.g., NIST/EPA/NIH MS Search software). Accurate mass data are not required to infer the presence of these elements from their isotope patterns, although such data would certainly reinforce the isotopic evidence for the presence of certain elements in other cases.

Take for example a (pseudo) molecular ion that is detected at m/z 177.0944 ± 0.0004 using a HRAMS measurement. Considering the mass accuracy of this HRAMS measurement, three candidate formulas are possible: $C_7H_{17}O_3Si$ (m/z 177.0942), $C_8H_{17}O_2S$ (m/z 177.0944) and $C_6H_{11}N_4F_2$ (m/z 177.0946). In this case the correct formula can only be established by interpretation of the X+1 region's related isotopic pattern, i.e. around m/z 178.09. The theoretical isotopic clusters are shown in Figure 15 for $C_6H_{11}N_4F_2$ (spectrum A), $C_8H_{17}O_2S$ (spectrum B) and for $C_7H_{17}O_3Si$ (spectrum C) and clearly illustrate that the expected differences between these isotopic patterns in the X+1 region can be used to select the correct formula by comparison with the X+1 region in the

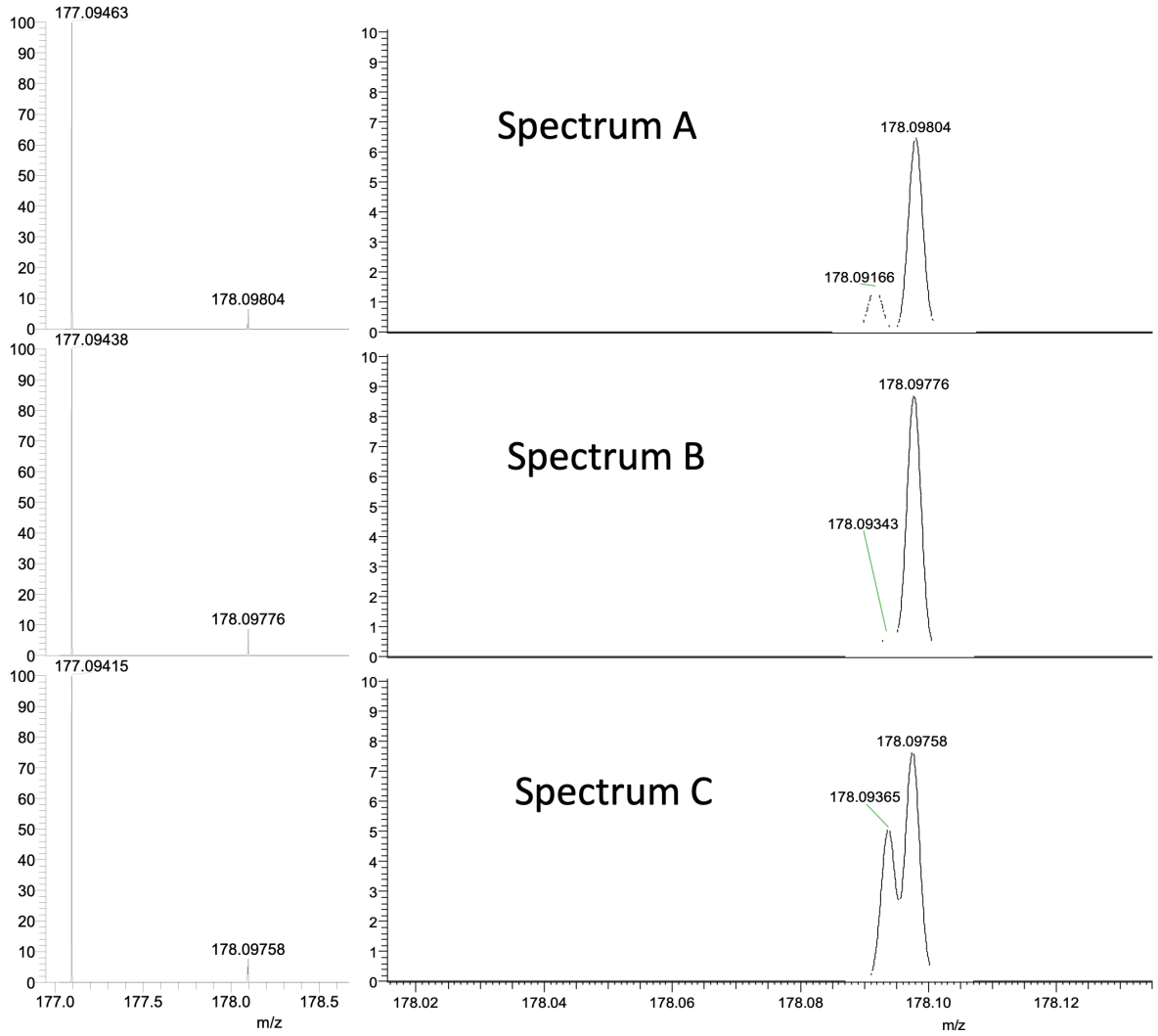


Figure 15. Left: Theoretical high-resolution accurate mass spectra for ions with an elemental formula of $C_7H_{17}O_3Si$ (spectrum A), $C_8H_{17}O_2S$ (spectrum B) and C6H11N4F2 (spectrum C) for m/z range 177 – 178.5 and Right: the corresponding zoomed X+1 region around m/z 178.09 at 70000 resolving power.

experimental spectrum for the compound of interest (if the measurement's mass resolution and mass accuracy is sufficient).

3.4 Interpretation of Mass Fragments: *De Novo* Structural Elucidation

The ion peaks present in a mass spectrum can be interpreted to establish the presence of functional groups or substructures in the compound of interest, to place the compound of interest into certain structure-based classes (e.g., alcohols, or phthalates) or even to propose a tentative molecular structure. This is based on the fact that the ions, formed during ionization, represent either the ionized molecule or ionized fragments thereof. Fragmentation of a molecule principally occurs in a predictable and reproducible way within the boundaries of the applied instrumental parameters. The general mechanisms for such fragmentation reactions have been extensively described in authoritative reference works on mass spectral interpretation.[4,5,9,10] For example, cleavages resulting in the loss of neutral molecules

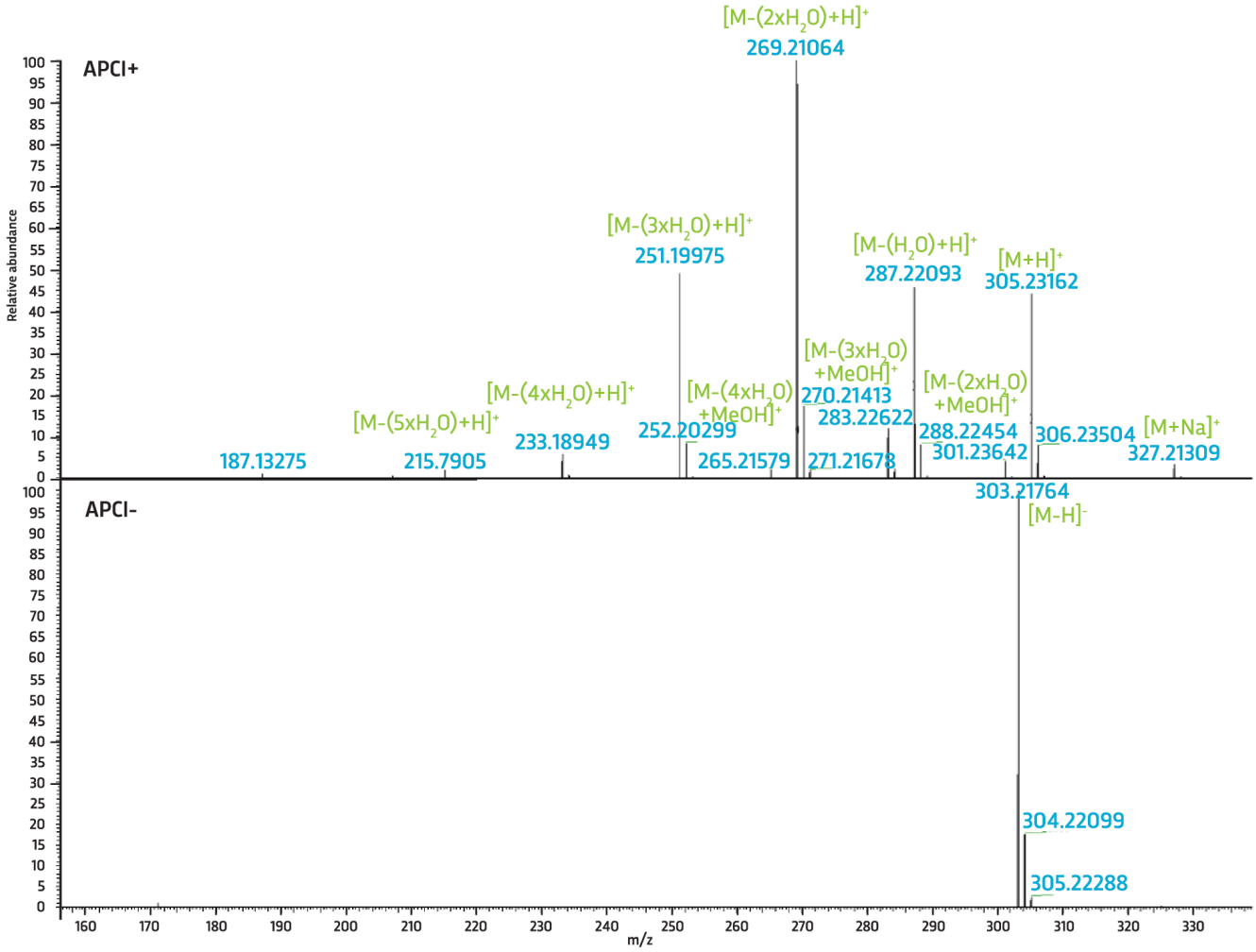


Figure 16. Example of an annotated positive(top) and negative mode APCI high resolution accurate mass spectrum of a compound (aleuritic acid) illustrating formation of the pseudo molecular ion ($[M+H]^+$, $[M-H]^-$) concurrent adduct formation and loss of neutral fragments by in-source fragmentation.

(water, carbon monoxide, methanol, etc.) are usually produced by structural rearrangements or proton shifts of the ionized molecule. An example of such a cleavage can be found in Figure 16.

Another type of cleavage involves the loss of a radical fragment (e.g., methyl radical), which is almost exclusively observed in EI spectra. A cleavage generates two types of informative spectral values, namely the mass value of the formed fragment and mass differences of that fragment with a heavier fragment or the molecular ion, commonly referred to as "losses".

An experienced mass spectrometrist who has an in-depth knowledge of fragmentation rules can relate observed mass values and losses to specific fragment structures which are linked together to establish a logical fragmentation pathway. Moreover, an extensive chemistry background is imperative to assess whether or not a proposed structure is viable, that is, that it is thermodynamically stable and whether it is likely to be detected with the applied technique.

The ultimate goal of structural elucidation is to elucidate as many fragments as possible and to link the

fragments together via a rational pathway, as so doing limits the number of possible structures to the smallest number of candidates. In general, the more fragments that can be fit into a defendable fragmentation pattern for a proposed structure, the greater the likelihood that the identity established by elucidation is, in fact, the correct identity.

In the initial stages of elucidation, it often happens that numerous structure candidates can be proposed which fit the generated molecular formula or observed fragments to varying degrees. As a general rule, the relative percentage of peaks that can be rationalized by a fragmentation pathway for a given structure is directly related to the likelihood that the spectrum indeed corresponds to that structure; that is, the higher the percentage of rationalized, peaks, the greater the likelihood that the elucidated identity is the correct identity. A complicating factor in structural proposal, however, is that not all structures have unique mass spectra. This is often the case for compounds with very similar structures. For instance, the degree and position of branching of hydrocarbon chains or the exact stereochemistry of a molecule often cannot be inferred from a spectrum. Therefore, the confidence level of identifications which are solely based on structural elucidation is limited to tentative identification at best. This is the case even for compelling elucidations, as the identification is still based only on one dimension of information. A higher level of confidence can be achieved by gathering additional data such as retention time, MS/MS spectra, or spectra recorded with a different type of ionization. In addition, other corroborating data, such as the result of an identification found in another orthogonal and complementary technique (such as GC/MS identification results for LC/MS compound identifications), disclosed compositional data of the material of construction or other analytical techniques that can assist in the confirmation of the elucidated structure (e.g. NMR on the isolated compound), can assist in upgrading the elucidated structure from a tentative to a confident or confirmed identity. It goes without saying that the highest level of identification is obtained by confirming the mass spectrum (and associated retention time) of the tentatively identified compound with its authentic standard and/or providing sufficient corroborating information so that the chances of an incorrect identification are small.

3.5 Case Studies

The ability to propose an initial tentative molecular structure and the strategy used to secure that structure depend largely on the ability to identify the molecular ion and molecular formula and the availability of reference mass spectra that are similar to the spectrum of the compound of interest. The following three cases demonstrate these strategies.

3.5.1 Case 1: Molecular Ion Not Identified

As mentioned previously, the likelihood of detecting and identifying the molecular ion depends on the ionization technology (EI, CI, APCI, ESI, etc.) and on the stability of the ionized molecule. If the molecular ion cannot be identified, potentially all peaks in the spectrum are fragments of a larger molecular structure. In that case, any proposal of a molecular structure would be highly speculative. At best, the presence of functional groups, substructures or general compound classes could be inferred based on the similarity of spectral features with available reference mass spectra. The underlying principle is that spectra from molecules with very similar structures also have similar spectral features. This is particularly relevant for sample spectra which are not present in a library of reference spectra. The similarity between spectra is not limited to the circumstance that the sample and reference spectra contain equal mass

values or relative abundances, but may also include equal losses.

Because a structure cannot be established, identifications secured in this manner are classified as partial identifications. Some examples of these partial identities include:

- EI spectra of phthalate esters contain an intense m/z 149 ion which is often the only major peak. The molecular ion is often not detected
- EI spectra of aliphatic hydrocarbons are characterized by a typical pattern of m/z 43, 57, 71, 85 etc.

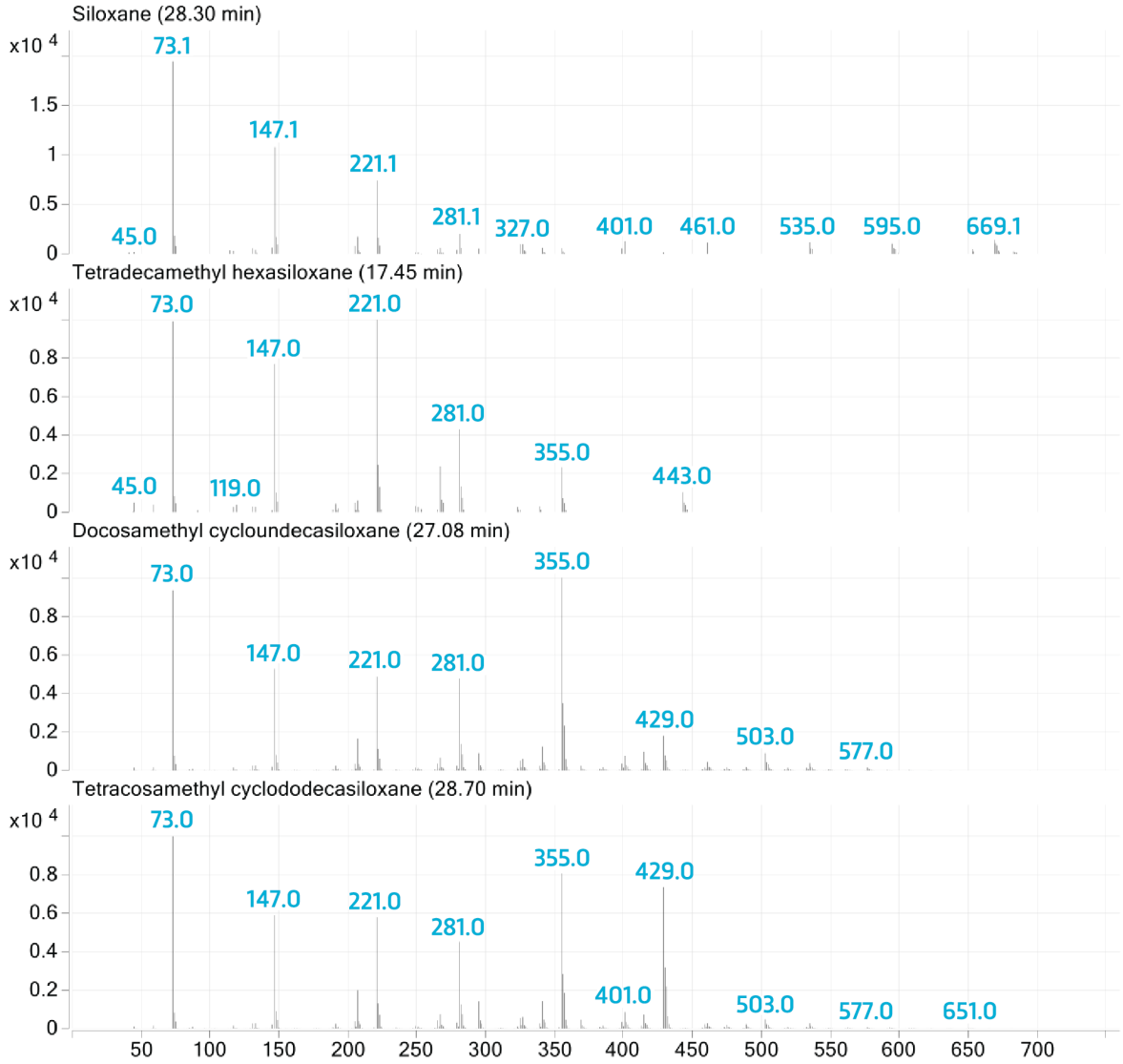


Figure 17. Example of a partial identification in which the mass spectrum of the compound of interest (top spectrum) contains typical fragments with m/z values 73, 147, 221 and 281. The mass spectra of siloxanes, confirmed with authentic standards, are shown for comparison and have the same characteristic pattern. This justifies a partial identification for the compound of interest (it is a siloxane). However, because neither the spectrum nor the retention time of the peak associated with the compound of interest is an acceptable match with the same data for any of the confirmed siloxanes, the identification cannot be elevated to tentative status.

- The presence of ions m/z 77 and 91 in an EI spectra is diagnostic for the presence of phenyl and benzyl substructures, respectively
- CI Spectra that contain a mass difference of 18 Da indicates the loss of water which is typically observed in alcohols or acids but not in ketones

Another example of a partial identification for a siloxane compound is shown in Figure 17. Experienced mass spectrometrists will recognize such spectral similarities more readily than will less-experienced analysts. Alternatively, software tools have been developed to assist with such substructure identifications. For example, the substructure analysis tool in the NIST/EPA/NIH MS Search software analyses the presence of substructure signatures in the hit list of a particular unknown spectrum and the match of the different hits to the unknown.[1] This analysis is then translated to a list of probabilities of substructures being present or absent in the spectrum of the compound of interest.

3.5.2 Case 2: Molecular Ion Is Identified (Unit Mass)

All strategies described previously can also be used when the molecular ion can be identified but the molecular formula cannot be established (e.g., low-resolution GC/MS). In this case however, knowledge of the molecular ion adds the possibility of relating all evident substructures to a certain molecular weight. In addition, a good but not perfect mass spectral match with a spectrum from an external reference library could be a reference to assist in the

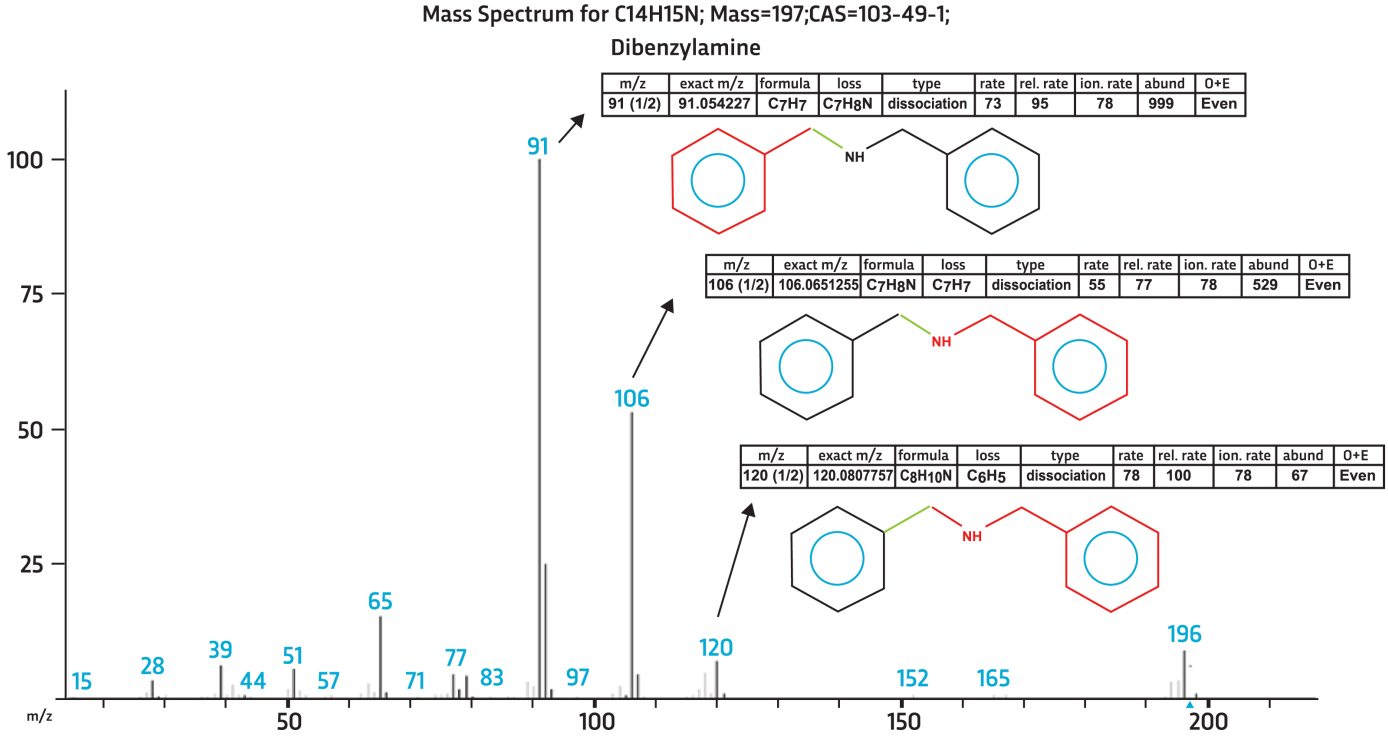


Figure 18. Explanation of fragmentation pathways observed in the EI mass spectrum of dibenzylamine using MS interpreter (version 3.4b, NIST/EPA/NIH). For instance, the base peak ion m/z 91 as well as fragment ions m/z 106 and m/z 120 can be explained by bond cleavages by dissociation at different positions of the molecule (fragment structures are displayed in red).

structural elucidation. Once the fragmentation pathways of the matching compound can be elucidated, rationalizing mutual spectral differences between the sample and reference compounds may lead to the proposal of a tentative structure. The process of rationalizing all spectral peaks for a structure candidate can be facilitated by using software packages specifically designed for *in silico* fragmentation prediction such as MS Interpreter (NIST/EPA/NIH), MS Fragmenter (ACD Labs) or Mass Frontier (*Thermo Scientific/HighChem LLC*). An example of rationalizing different mass spectral fragments using a software package can be found in Figure 18.

3.5.3 Case 3: Molecular Formula Is Identified (Accurate Mass)

In cases where the molecular formula can be determined from the mass spectrum, it is often possible to draw numerous structures which comply with the elemental composition of the molecule. Therefore, a molecular formula on its own can only correspond to a partial identification level. Databases such as PubChem, SciFinder or Chemspider could be used to search known structures. However, it should be noted that the actual structure might not be present in these databases as the number of known chemical structures is only a fraction of the total chemical space of organic compounds.[9] To reduce the number of structure candidates, mass fragments should be interpreted within the boundary of the identified molecular formula. In this case, the availability of accurate mass data increases the confidence of relating fragment ions and hence their elemental composition to the molecular structure. In addition, mass differences can be more easily linked to small functional groups. For example, an integer loss of 28 Da can either reflect the loss of ethylene (C_2H_4, 28.0313 Da) or carbon monoxide (CO, 27.9949 Da).

Other means for confirming the proposed elucidated structure include the use of available corroborating data such as supplier information on the composition of the material, identification results of other chromatographic techniques used in the characterization process or other techniques used to elucidate chemical structures (e.g. NMR). The approach of using additional evidences to support a higher level of identification is described in the following text.

PART 4

Additional Evidences Supporting Higher-Level Identifications

An identification secured by either mass spectral matching or mass spectral interpretation, is, by definition, a tentative identification as it is based on one dimension of identifying information. While it is certainly the case that tentative identifications provide the minimally acceptable input information into a toxicological safety risk assessment of extractables and leachables, greater certainty in the identity leads to greater certainty in the toxicological assessment. Thus, it is often the case that additional information about the compound of interest is pursued with the intent that such additional information would corroborate (or refute) the tentative identification. Depending on the quantity and nature of the corroborating data, tentative identities can be substantiated and therefore "elevated" to either confident or confirmed identifications.

To a certain extent, tentative identifications can be "elevated" to at least confident status using the mass spectral information itself. Thus, for example, if the same tentative identity is secured by mass spectral matching and mass spectral interpretation, then these two independent corroborating outcomes "elevate" the tentative identification secured with both processes to a confident identification.

Additionally, identifications can be also be substantiated by accumulating independent evidences or evidences from analyses which are specifically chosen to confirm a certain identification. It is evident that the more additional evidences that are gathered, the more certain the identification becomes. In this section off the text, various means of "elevating" identities are considered, and examples are provided; however, it is outside the scope of this document to provide an exhaustive list of additional evidences.

4.1 Retention Time/Index Matching

As has been previously established, mass spectrometry is the mostly used and commonly accepted means for linking an organic compound that was discovered in an extractable to leachables study to its unique identity. However, although mass spectrometry is a very powerful tool in the identification process, the technique becomes much less powerful when analyzing complex mixtures of compounds. Therefore, extracts or drug products are screened for extractables or leachables using chromatography as the "front-end" of a mass spectrometer, where the chromatographic process separates the often-complex extract or drug product mixture into individual compounds.

As a result of the chromatographic process, the compounds of interest are separated in terms of the time it takes the compounds to emerge (elute) from the chromatographic column (prior to entrance into the mass spectrometer). This elution time, referred to as the retention time, will depend on the chemical and physical nature of the compounds of interest (among other factors) and is therefore diagnostic for specific compounds. However, even with the excellent separation efficiencies (resolution) achievable by modern chromatographic methods applied to extractables/leachables screening (for example, ultra-high performance liquid chromatography, UPLC), a specific retention time is not necessarily unique to a single specific organic compound (that is, it is not uncommon that several compositionally dissimilar compounds may have comparable retention times). Thus, retention time itself is not a sufficiently

diagnostic property of a compound that it alone can be used to secure a tentative identity. Rather, retention time is corroborating information for identities secured by another means, such as mass spectral matching or interpretation.

In a certain way, use of retention time to support an identification is similar to the use of a mass spectrum to secure a tentative identity. Unlike a mass spectrum, retention time itself cannot be interpreted to produce a tentative identity. However, like a mass spectrum, the retention time can be matched to potential compound identities via a laboratory-generated database of retention times, akin to mass spectral matching. Presumably, a test compound whose mass spectrum and retention matches the mass spectrum and retention time of a reference compound in a database has been confidently established to be the reference compound.

Perhaps the single most advantageous aspect of identification corroboration via retention time matching is that the retention time is essentially a "free" piece of information. That is to say that the retention time is obtained via the same analytical activity as the mass spectrum. Securing retention time as a corroborating piece of data does not required re-analysis of the sample to secure additional information, which is required, for example, if corroborating information is obtained via a different analytical technique (for example, NMR).

For retention time to be useful as an identification tool, it is imperative that the retention time must be accurate and reproducible over time and across different instruments. However, shifts in retention time occur frequently. Routine maintenance procedures such as column trimming alter retention times. In a multi-instrument laboratory running the same method on multiple instruments, the retention times for each instrument will likely differ from each other, even when care is taken to ensure that all instruments are operated using identical conditions. These differences in retention times confound efforts to use retention time as a means of identification. This is especially true for LC-based separations.

To a certain extent, retention time differences can be managed by two methodologies, Retention Time Locking (RTL) and Relative Retention Times (RRT, alternatively referred to as Retention Index, RI). RTL is the ability to very closely match retention times on one system to those in another system by adjusting the chromatographic conditions and is more typically applicable to GC. In GC, for example, adjusting the inlet carrier gas pressure will change retention times in an even and predictable manner. Thus, retention times on a given system can be closely matched to those on another system by altering the inlet carrier gas pressure in one or both of the systems. A specific compound (usually the Internal Standard for Injection) is used for both developing the locking calibration and locking all future systems.

As noted above, absolute retention times can be irreproducible as they depend on a large variety of chromatographic factors which renders them unsuitable as a "universal" criterium for identification This shortcoming can largely be overcome by expressing retention behavior on a relative scale using retention indices (RI) or linear retention indices (LRI), which can be used as corroborating information. The advantage of the RI is that retention indices do not depend on the exact column dimensions, flows or temperature-programming. However, they do depend on the type of stationary phase (for example non-polar versus polar phases).

The retention index system was first developed by Kováts for GC-based measurements by expressing the retention time of a compound relative to the retention times of the nearest eluting n-alkanes

under isothermal conditions as Equation 1, which was adapted to Equation 2 for temperature-programmed measurements:

$$RI_x = 100\left(n + \frac{\log RT_x - \log RT_n}{\log RT_{n+1} - \log RT_n}\right) \quad \text{Equation 1}$$

$$RI_x = 100\left(n + \frac{RT_x - RT_n}{RT_{n+1} - RT_n}\right) \quad \text{Equation 2}$$

where n corresponds to the number of carbon atoms of the nearest pre-eluting n-alkane for compound x and with RT_n and RT_{n+1} correspond respectively to the retention times of the nearest n-alkanes that bracket compound x. Retention indices have been determined for reference polyaromatic hydrocarbons namely benzene (assigned index 100), naphthalene (200), phenanthrene (300), chrysene (400) and picene (500).

The NIST/EPA/NIH Mass Spectral Library for GC/MS contains a growing amount of reference RI data; for GC, the 2017 version now contains 404045 citations of RIs for 99400 compounds (72361 in the EI library). These data are collected from different contributors and are given as median value ± deviation (number

Table 6. Comparison of experimentally determined Retention Index values by Nelson Labs (based upon the recorded retention times) with the experimental and estimated Retention Index values which could be found in the NIST library.

Compound Name	Nelson Labs		NIST experim. NON-POLAR SS			NIST estim. NON-POLAR SS	
	RT (min)	RI	RI median	deviation	n	RI	CI (95%)
1-Hexanol	6.322	874	868	4	223	860	176
Cyclohexanone	7.033	900	894	4	29	891	246
Octamethyl cyclotetrasiloxane	9.567	1002	994	6	3	827	382
2-Methylbenzaldehyde	11.100	1071	1064	4	5	1095	196
4-Methylbenzaldehyde	11.417	1085	1079	1	18	1095	196
2-Acetylcyclohexanone	14.058	1217	n/a	n/a	n/a	1187	246
Caprolactam	14.817	1259	1259	11	4	1003	356
BHT	19.133	1519	1513	5	51	1668	301
Benzophenone	20.933	1642	1635	10	27	1603	246
Irgacure 184	21.783	1701	1687	n/a	1	1740	382
Tri-(2-chloroethyl) phosphate	22.700	1767	1779	3	6	n/a	n/a
Diisobutyl phthalate	24.067	1883	1870	4	32	1908	201
Palmitic acid	25.083	1963	1968	7	232	1968	220
Bisphenol A	27.750	2192	2108	0	2	2022	301
Tri-n-butyl citrate	27.817	2199	n/a	n/a	n/a	2404	382
Oleamide	29.650	2372	2386	11	2	2228	356
Antioxidant 2246	30.300	2437	2414	48	2	2788	301
Irganox 1081	33.367	2764	n/a	n/a	n/a	2939	382
Erucamide	33.583	2789	2625	0	7	2625	356
BADGE	34.733	2922	2805	n/a	1	2538	293
Bisphenol P	36.556	3131	n/a	n/a	n/a	2923	301
Irganox 1076	43.717	3615	3603	n/a	1	3823	382

of data points). It should be noted that a majority of the compounds have just one measurement, i.e. 53% in the NIST05 release, and that indices are not uniform distributed over different compounds/ compound classes.3 Experimental RI values in the NIST library are classified into three types of stationary phases:

- Semi-standard Non-Polar, i.e. poly(5% diphenyl - 95% dimethylsiloxane) columns
- Standard Non-Polar phases, i.e. poly(dimethylsiloxane) columns
- Polar phases, i.e. polyethylene glycol columns

In addition to experimental RI values, several theoretical models have been developed to estimate RIs.[11-15] Although the accuracy of the estimated RI's is generally insufficient for unambiguous identification based solely on predicted retention and matching spectrum, an estimated RI can facilitate identification as it can be used to reject of certain classes of false identifications made by GC/MS.[11]

In trying to match measured RI data to reference RI data from NIST, the following precautions need to be taken:

- Matching stationary phase, experimental versus reference data should be confirmed
- Either a standard with an n-alkane mix should be run with each sequence to setup the reference calibration which is not subject to retention time shifts or RT locking must be applied
- The certainty level of the reference RI data (deviation / confidence interval, number of entries) must be evaluated

To illustrate the effectiveness of using either RI data from an external database or an estimated RI for identification, experimentally determined RI values can be compared to reported and estimated RI values contained in an external database (e.g., the NIST/ EPA/NIH Mass Spectral Library). Such a comparison is presented below (see Table 6), using compounds of diverse chemical nature and therefore diverse retention properties. Generally, the NIST RI data agree well with experimental RI data (median $\Delta RI = 10$). However, there are notable outliers (displayed in red). The deviation with the median estimated RI (median $\Delta RI = 149$) and roughly corresponds to the half of the 95% confidence interval on those estimated RI data. Therefore, a good fit of the experimental RI-value with the obtained value from a commercial library may assist in selecting the right chemical structure from a "hit list" that has been generated via the process of mass spectral matching.

Despite the agreement noted in Table 6, identification corroboration via retention matching is most effective when the reference retention data are obtained through analysis of reference materials under standardized chromatographic and MS conditions that are identical to the conditions applied in routine laboratory operations.

Unfortunately, no universal or unified HPLC retention index system has yet been established for reversed phase, normal phase, and HILIC.5 Although differences in retention times between different instruments are higher than in LC than for GC (due to small variations between different columns, minor changes in the concentration of the organic mobile phase and other instrumental parameters such as flow rate, column temperature or pH of the mobile phase), an in-house database containing experimentally measured retention times can be leveraged to provide corroborating identification information.

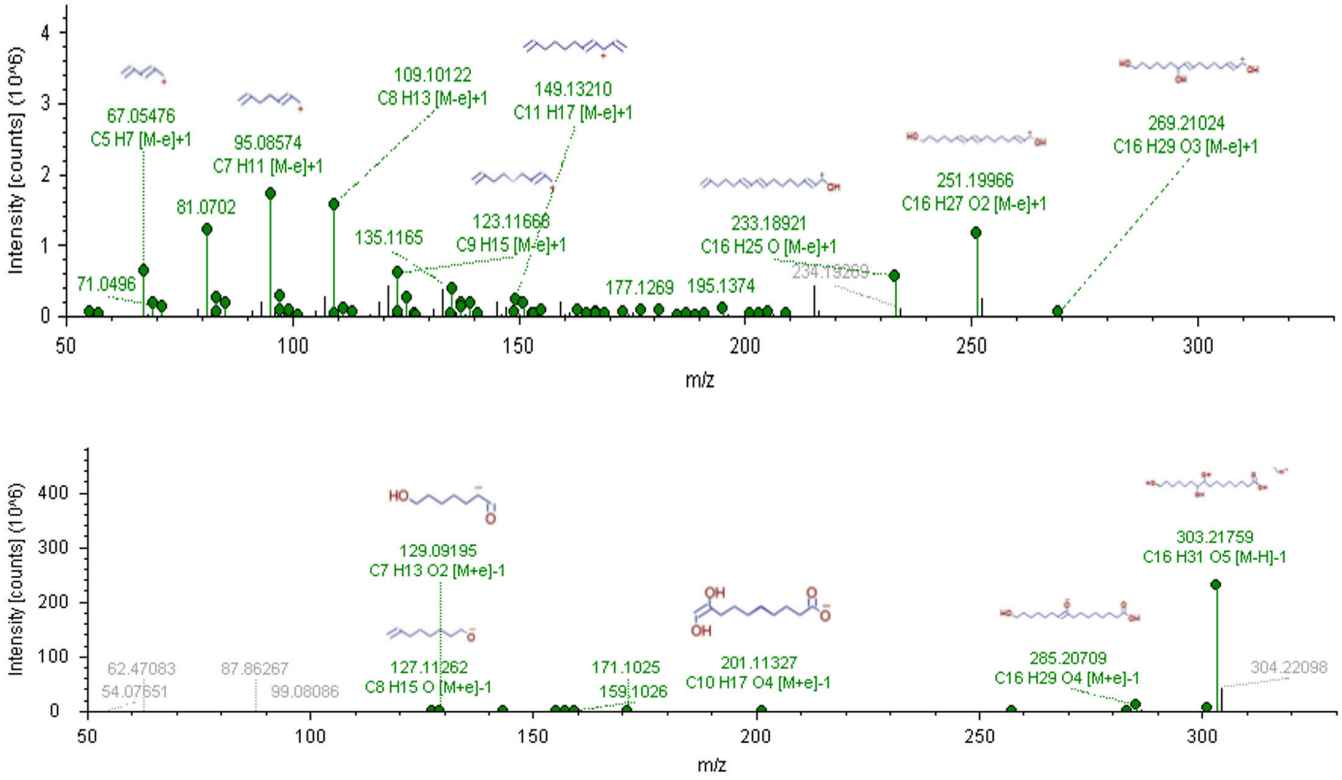


Figure 19. Annotated APCI MS/MS high resolution accurate mass fragmentation spectra (30 eV) for pseudo molecular ions top: $[M+H]^+$ at m/z 305.232 ± 0.5 m/z (positive mode) and bottom: $[M-H]^-$ at 303.217 ± 0.5 m/z (negative mode) obtained for aleuritic acid.

4.2 Tandem Mass Spectrometry

The interpretation of MS/MS (or more generally MS) spectra can either lead to the proposal of a tentative structure or further add confidence to a tentative structure that has already been proposed based on other evidence. The most common type of MS/MS analysis is the acquisition of product ion scans, which is achieved by isolating a certain precursor ion followed by fragmentation of that ion into products ions. [Depending on MS technology and instrument vendor, such MS/MS analyses can either be set up manually in a separate run or be performed along with the acquisition of screening data, for instance through selection of the top n most intense ions for isolation and fragmentation.] A good choice of precursor ion selection would be picking the molecular ion as it results in a spectrum of product ions which have an unequivocal relationship with the molecular structure. MS/MS analyses are particularly useful to obtain fragmentation data when the ionization method yields very few structurally informative fragments (e.g. APCI spectra which only contain the molecular ion). Furthermore, an MS/MS spectrum has a higher level of selectivity compared to MS^1 scan data as the in-source fragmentation in MS^1 could be obscured by other ions generated from the matrix in the course of the ionization process or by coelution with other compounds present in the sample. For example, Figure 19 represents the MS/MS annotated fragmentation spectra for the (pseudo) molecular ions for aleuritic acid, which is prone to in-source fragmentation.

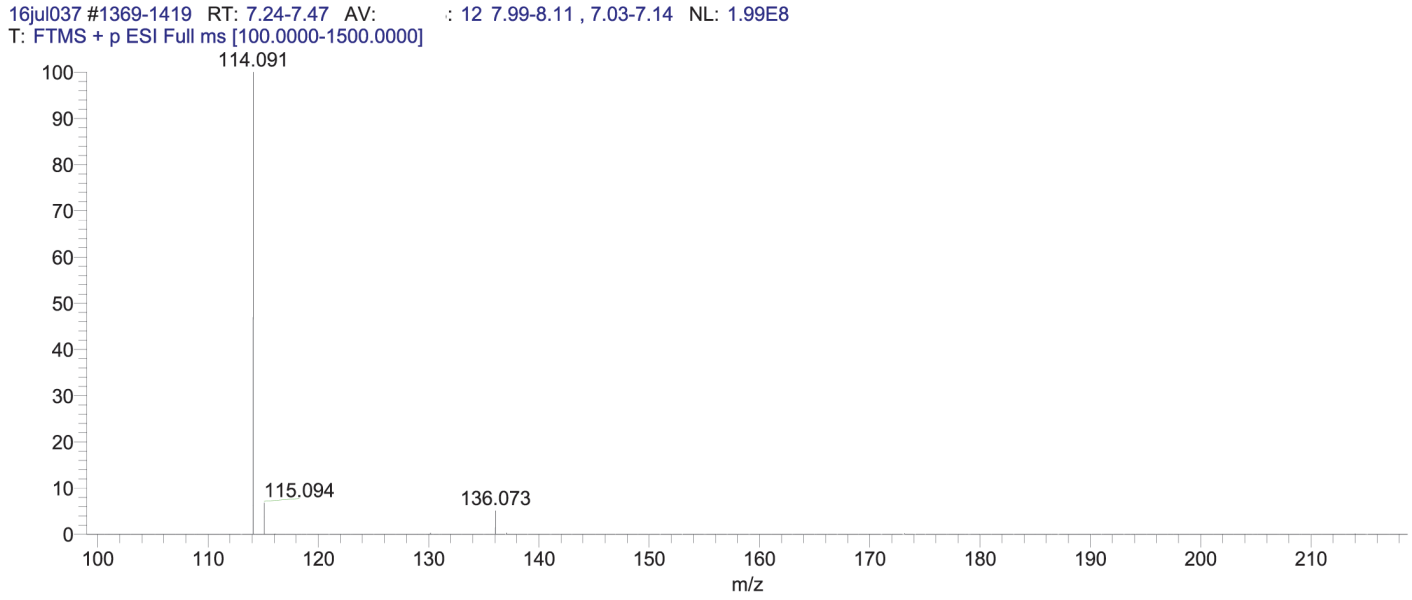


Figure 20. Corresponding mass spectrum of an extractable, detected at RT 7.34min. This mass spectrum shows the presence of a molecular ion at m/z 114.091. This assumption that this is the parent ion is confirmed by the detection of the Na-adduct at m/z 136.073

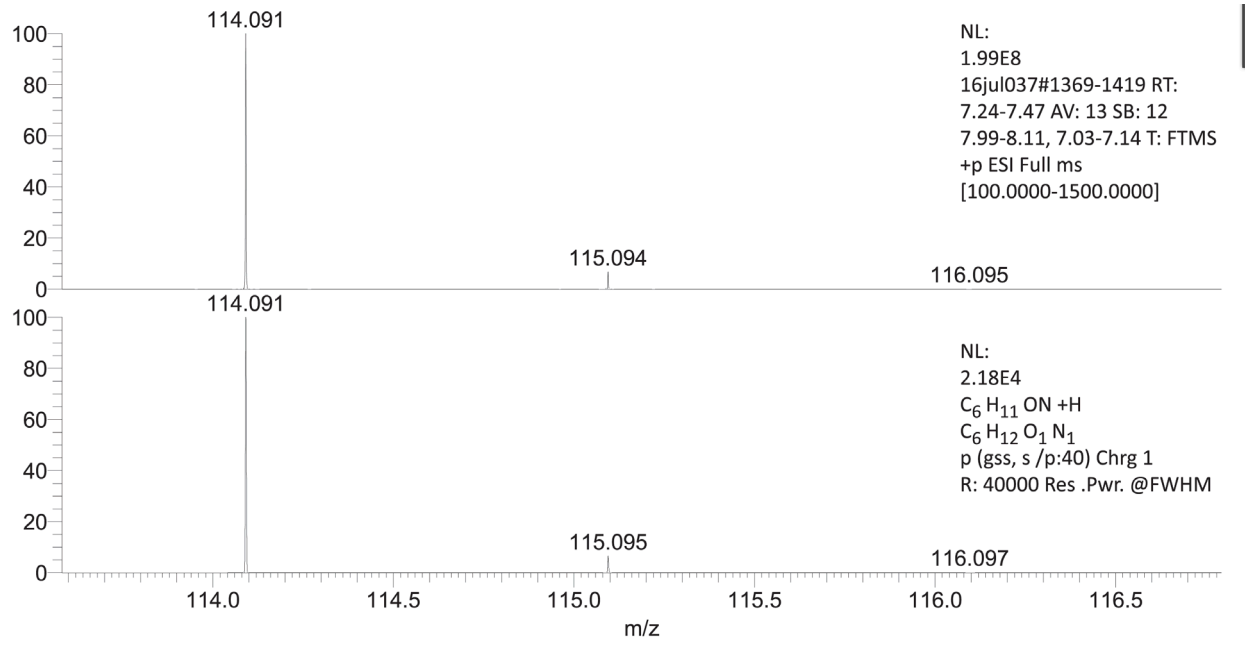


Figure 21. Verification of the isotope pattern. A simulation of how the isotope pattern could look like for the protonated C6H11ON+H (lower mass spectral isotope pattern) shows a perfect match with the isotope pattern of the detected compound (upper mass spectral isotope pattern), which confirms the suggested elemental composition.

In case of co-elution, mass spectral deconvolution is a powerful tool to resolve spectra from coeluting compounds that is effective with a vast majority of acquired spectra. However, complete resolution of complex mass chromatograms by deconvolution will not be possible in all cases.

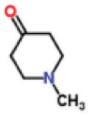
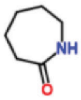
ID	Structure	Molecular Formula	Molecular Weight	# of Data Sources	# of References	# of PubMed	# of RSC
66668		$C_6H_{11}NO$	113.1576	149	236	5	66
16486 W		$C_6H_{11}NO$	113.1576	138	213	7	59
7480 W		$C_6H_{11}NO$	113.1576	137	452	279	275
7236 W		$C_6H_{11}NO$	113.1576	120	234	40	160

Figure 22. List of candidates for a C6H11NO elemental formula, generated via ChemSpider.

4.3 Additional Evidences from Orthogonal Techniques

Some compounds can be detected by multiple analytical techniques and thus it is possible that a compound could be tentatively identified by independent assessment of the evidence from each technique. When this is the case, the independent assessments (which produce the same identities) are mutually corroborative and the identification, supported by two-dimensional data is "elevated" to confident.

For example, take the relatively simple and common case where an extractable produces a response in both GC/MS and LC/MS. In this case, and without any additional testing, two tentative identities secured by both techniques independently corroborate one another, resulting in an elevated confident identity. Alternatively, a tentative identity secured by one method can be used to tentatively identify a peak that is unidentifiable by the second method.

An example of this second scenario is as follows. Screening of an extract via LC/MS (ESI+) produced a chromatogram with an extractable peak at 7.34 min and the corresponding mass spectrum for this compound is shown in Figure 20. The mass spectrum shows a (protonated) molecular ion mass ($[M+H]^+$) at m/z 114.091. The assumption that this ion establishes the nominal mass is confirmed by the detection of the Na-adduct of the molecular ion ($[M+Na]^+$) in the corresponding mass spectrum. With this information, an elemental formula of $C_6H_{11}ON$ can be calculated (using a software based elemental formula calculator) for the extractable. The suggested elemental formula is confirmed after reviewing the isotope pattern for the suggested elemental formula (see Figure 20).

While this is already very valuable information, it does not produce a tentative identity for the compound of interest until the compound's structure can be established. One mean of getting "suggestions" for the chemical structure is to consult publicly available databases, such as ChemSpider that could assist in

generating potential candidates for the compound with a confirmed elemental formula of C6H11ON. The list of candidates that is generated suggests different chemical structures that could fit with the confirmed elemental formula (see Figure 22).

At this point, the amount of information that was obtained via the LC/MS (ESI+) analysis alone does not allow a mass spectrometry expert to uniquely identify the compound. However, it is noted that analysis of the same extract by GC/MS resulted in the tentative identification of caprolactam as an extractable, which is the third compound in the candidate list generated via ChemSpider. We now have compelling evidence that the compound revealed by LC/MS is likely caprolactam. As caprolactam is a commonly encountered extractable that is commercially available as a reference standard, this inference is easily confirmed by LC/MS analysis of the reference standard.

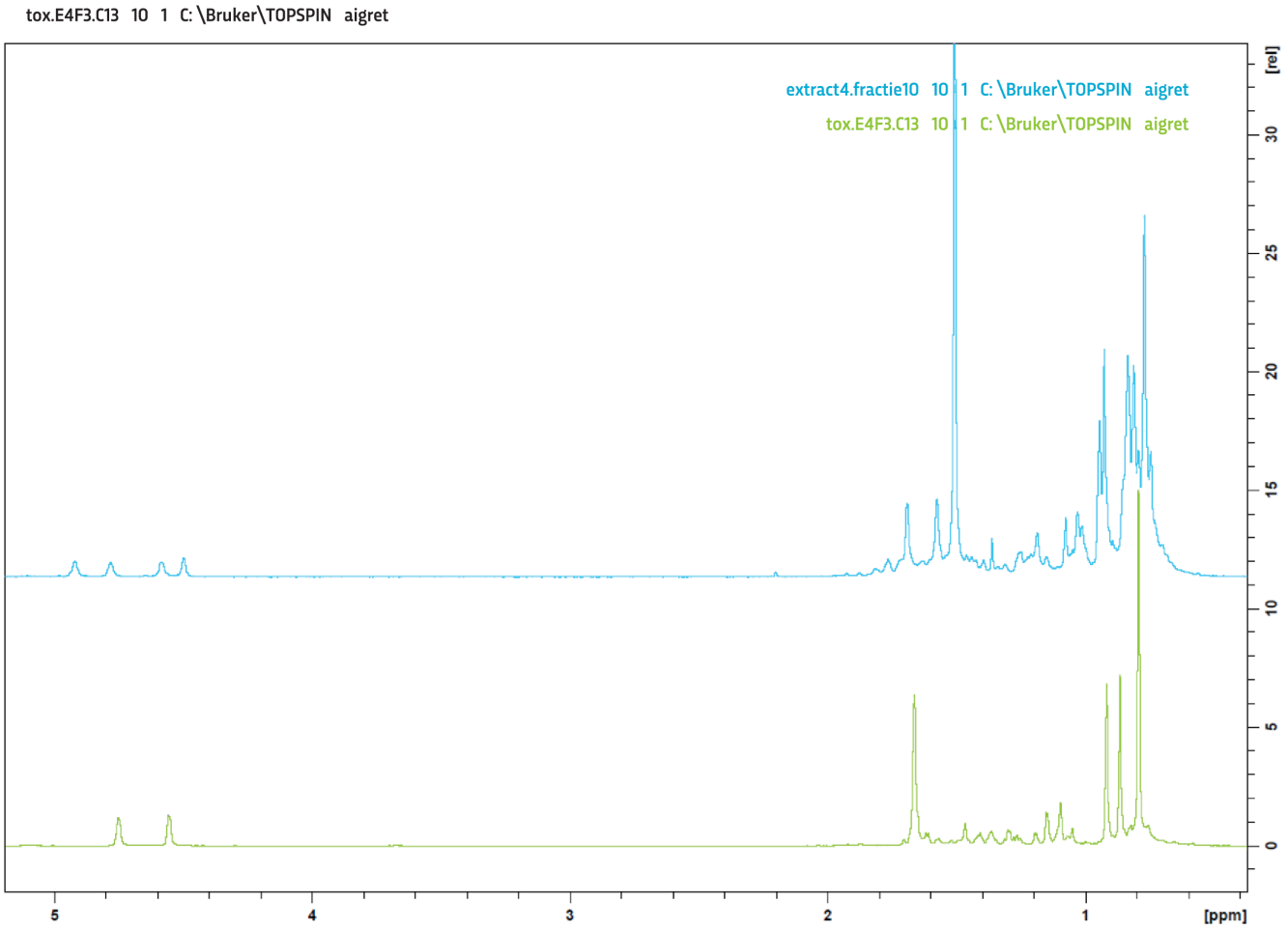


Figure 23. 1H-NMR spectra of the isolated $C_{13}H_{24}$ and the $C_{21}H_{40}$ rubber oligomers, performed by Nelson Labs in collaboration with the University of Leuven, Belgium. This NMR spectrum compares the 1H-NMR spectra of the $C_{13}H_{24}$ (top) and the $C_{21}H_{40}$ (bottom) oligomer. The $C_{13}H_{24}$ spectrum shows 2 characteristic peaks at δ = 4.6 ppm and δ = 4.8 ppm, which is typical for the 2 vinyl protons, and one peak at δ = 1.67 ppm (4H), which can be assigned to the 4 allylic protons. Furthermore, the peaks of the four methyl groups (singlets) can be identified within the aliphatic region (δ = 0.79 ppm (6H); 0.87 ppm (3H); 0.92 ppm (3H)). The interpretation of the NMR spectrum of the $C_{21}H_{40}$ oligomer is more difficult since - next to the additional peaks of multiple coupled protons of the alkyl chain - the spectrum consists of the overlaid NMR spectra of the two diastereomers. However, in analogy with the NMR spectrum of the $C_{13}H_{21}$ oligomer, the double sets of vinyl-protons, allylic protons and the methyl groups can be identified within the 1H-NMR spectrum.

Another means by which information of an orthogonal technique can assist in providing the correct identity is when compounds with the same m/z are co-eluting. This may, for instance, be the case for caprolactam and 2-methyl-1-pyrrolidinone in an LC/MS analysis. While these compounds, both with the elemental formula C6H11ON, may co-elute in the LC/MS Chromatography, they do not co-elute in GC/MS. Therefore, the identity of the detected compound in LC/MS at retention time 7.34 min with a detected m/z of 114.091 can be uniquely attributed to either caprolactam or 2-methyl-1-pyrrolidinone depending on which compound is reported in the GC/MS data.

Another manifestation of the orthogonal technique approach is the use of a non-chromatographic method, such as NMR, to independently secure an unknown's identity. This identification strategy is also described in USP <1663>[1] as follows: "Although these identification categories are based upon mass spectrometry, it is possible to use data from other analytical techniques to assist in the extractables identification. Such techniques include GC/FTIR (Fourier Transform Infrared Spectroscopy and LC/NMR (Nuclear Magnetic Resonance Spectroscopy)". In this document, we will not evaluate GC/FTIR and LC/NMR as techniques that could support a higher identification class, as the considerations for NMR that will be made below are a *fortiori* true for these techniques also.

While the power of NMR as an identification method are well-known, use of this technique in E&L laboratories is limited by certain practical realities. One such practical reality is access to NMR technology. Although access to NMR technology may be straightforward for larger pharmaceutical companies, access may be problematic for E&L labs in a contract research environment. The cost of an

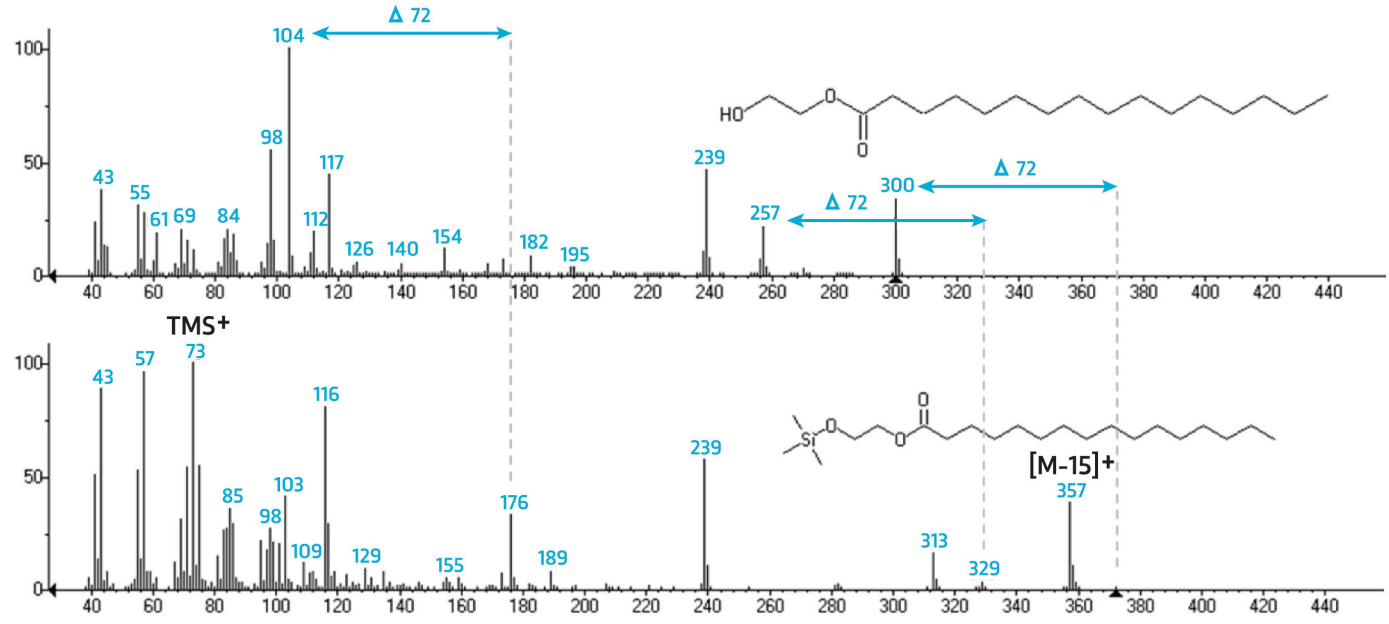


Figure 24. Comparison of the EI mass spectrum of 2-hydroxyethyl palmitate with the mass spectrum of its trimethylsilyl (TMS) derivative. A mass difference of 72 Da is observed for the molecular ion (m/z 300 versus m/z 372) and demonstrates that the molecule contains one derivatizable group (in this case a hydroxyl group). It is often observed for trimethylsilyl derivatives that the $[M-15]^+$ peak corresponding to the radical loss of a methyl group (in this case m/z 357) is more abundant than the molecular ion. Furthermore, ion m/z 73 is also diagnostic for the trimethylsilyl group.

NMR-instrument, as well as its operating cost and the level of expertise that is needed to interpret the results of an NMR spectrum prevents smaller organizations from investing readily in this option, A second practical reality is that NMR can only come to relevant conclusions if the neat "unknown compound" can be investigated. The sample amount requirements to perform an NMR experiment on this neat chemical compound – often a few milligrams of the purified "unknown compound" at least – often requires either intensive sample preparation steps, such as isolation of the compound through fraction collection.

The complexity of the NMR interpretation is illustrated in Figure 23, where the signals observed in the NMR spectrum for both the $C_{13}H_{24}$ and the $C_{21}H_{40}$ rubber oligomers are explained. As one can see, NMR is not a "magical solution" that immediately leads to a confirmed identification: the spectra need to be interpreted by an NMR-expert to come to a unique and reliable identification of the compound. In addition, no supporting libraries are available that can assist in NMR interpretation, as is the case in GC/MS, which makes the quality of an NMR interpretation highly dependent on the scientific skills of the interpreter.

Table 7. Table with a Hypothetical List of Ingredients for a Material of Construction (in this case, a polyolefin)

Compound	Synonym	Function
Polyolefin	---	Polymer
Tris-(2,4-di-tert-butylphenyl)phosphite	Irgafos 168	Antioxidant
Octadecyl 3-(3,5-di-tert-butyl-4-hydroxyphenyl)-propionate	Irganox 1076	Antioxidant
1,2,3-propanetriol-1-octadecanoate	Monostearin	Lubricant
Calcium dioctadecanoate	Ca-stearate	Acid scavenger

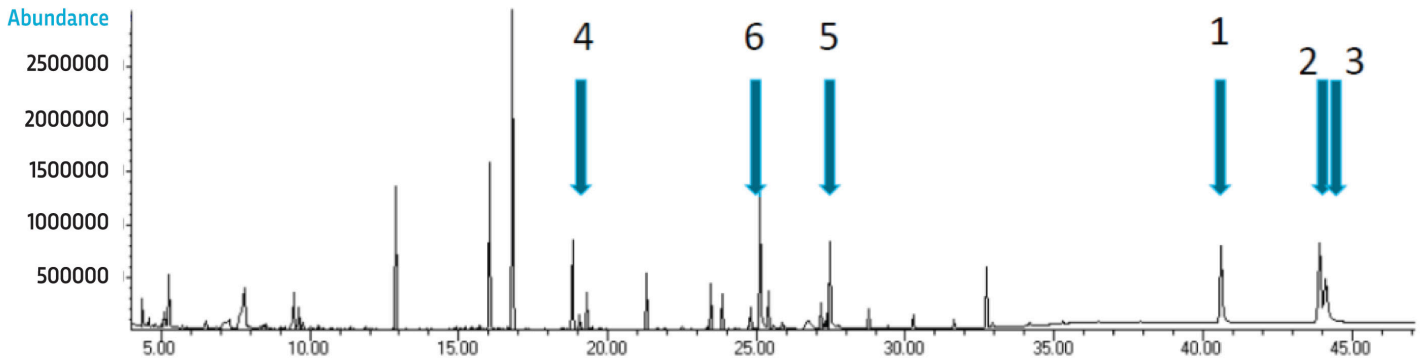


Figure 25. GC/MS Chromatogram of an organic extract of the material of Construct with a fictious composition as described in Table 7. Based upon the provided information in the ingredients list, (at least) 6 identifications can be upgraded from a tentative identification to a higher class of identification (e.g., confident) using this information.

4.4 Derivatization

Derivatization is the chemical treatment of an extract designed to convert a compound (or compounds) present in the extract to a more analytically expedient form. Derivatization is performed to increase the sensitivity, selectivity or thermal stability of a compound for a certain technique. Trimethylsilylation and methylation, for example, are common techniques in used GC/MS to increase the volatility and hence the sensitivity of polar molecules. Derivatization using halogenated acyl groups is another example and is used to increase the sensitivity for detection with an electron capture detector (ECD) or a mass spectrometer with electron capture negative chemical ionization.

Additionally, the selectivity of the derivatization reaction can also be exploited to identify the presence of certain functional groups. Trimethylsilylation, for instance, will derivatize all functional groups with active H atoms (e.g. acids, alcohols, amines) such that each active H atom is replaced by a trimethylsilyl (TMS) group. These changes will also be reflected in the mass spectrum by an increase in molecular weight of 72 Da for each TMS group. Comparison of MS chromatograms associated with non-derivatized and derivatized extract indicates whether or not the extract contains analytes whose structures include derivatizable groups and, if there is an analyte with derivatized groups, how many derivatized groups the analyte possesses. (e.g., Figure 24).

4.5 Indirect Inferences

In some cases, it is possible to support an identification with an indirect inference; that is, secondary information is used to infer whether a proposed identity is likely or not. For instance, knowledge of a test article's composition can facilitate the identification of its associated extractables, as it is likely that the extractables include the ingredients themselves or reaction products for these ingredients. Thus, the decision between two possible identities can be made on the basis of only one of the structures being related to a known test article ingredient.

As an example, consider the hypothetical list of ingredients for a polyolefin material given in Table 7. Each ingredient serves a specific purpose, either to protect the polymer from oxidation (Irganox 1076 as a primary antioxidant protecting the polymer during use, Irgafos 168 as a secondary antioxidant protecting the polymer during its manufacturing, calcium stearate as an acid scavenger) or to enhance the functionality of the polymer (monostearin as a lubricant).

This polyolefin material, with a known composition described in Table 7, is then subjected to an extraction with an organic solvent and followed by extract analysis via GC/MS. The resulting chromatogram, Figure 25, contains 6 peaks whose associated compounds can be confidently identified as follows.

Firstly, the mass spectra for the compounds 1 and 2 can be readily matched with a high match score to library spectra for Irgafos 168 and Irganox 1076. An expert review of the mass spectral matches for these two compounds leads to the conclusion that the match is sufficiently good that both compounds have been tentatively identified. However, knowing that these compounds are intentionally present in the extracted material makes it all the more likely that these tentative identities are in fact the correct identities and thus the composition information is sufficiently corroborative that the tentative identities can be "elevated" to confident identities, based on this two-dimensional corroboration.

Taking this line of reasoning further, compound 3 in Figure 25 was tentatively identified as Tris(2,4-di-tert-butylphenyl) phosphate, the well-known and

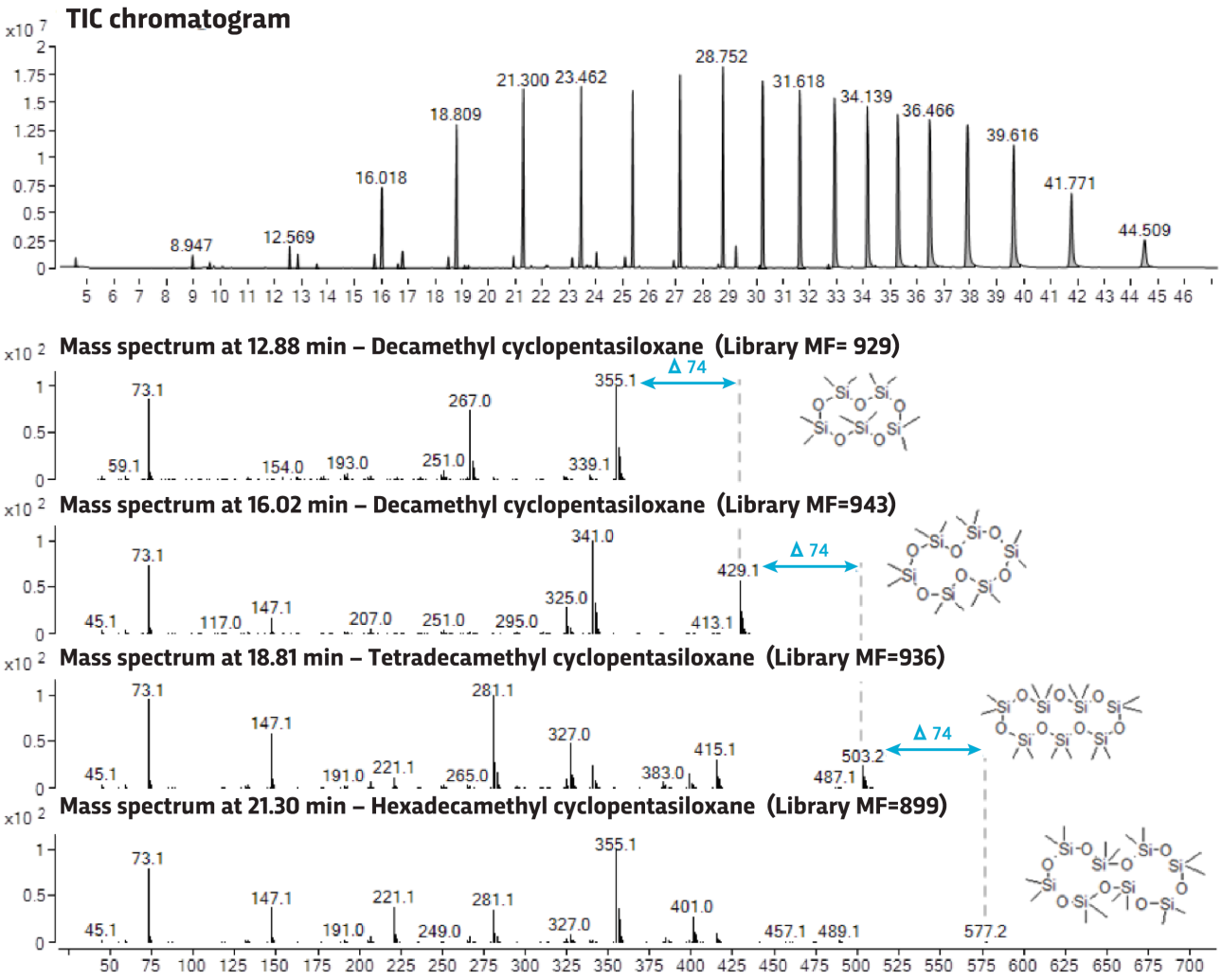


Figure 26. Identification of a homologous series of cyclic siloxane extractables. The chromatogram shows many peaks at regular retention times. The corresponding mass spectra have excellent mass spectral matches with cyclic dimethylsiloxanes of different length. One notes that the mass difference for the $[M-15]^+$ peak between each homologue (i.e. m/z 355, 429, 503 and 577) amounts to 74 Da which corresponding to one dimethylsilyloxy unit. The identification level of these homologues can thus be clearly linked to each other. That is, the certainty of identifying an initially unknown homologue increases by relating both its mass spectrum and retention time to other homologues with a confirmed (or confident) identification level. Additionally, the fact that all homologues are detected in the same test item adds confidence to the identification of each homologue.

well-characterized oxidized form of Irgafos 168, also via reviewed mass spectral matching. Given the presence of Irgafos 168 in the test material, it is very likely that the oxidized form of Irgafos 168 will also be present in the material, as it is by its sacrificial oxidation that Irgafos 168 protects the polyolefin. Thus, compound 3 is confidently identified as the oxidized form of Irgafos 168 based on corroborating information of a tentative identity based on the mass spectrum and logical inference of the presence of this compound in the test article.

A similar logic can be applied to peak 4, which again can be tentatively established to be 2,4-Di-tert-butylphenol based on an expert-verified mass spectral match. At first glance this compound is not listed as an ingredient and thus one could conclude that

the tentative identity is not corroborated by composition. However, it is well established in the chemical literature (for example, reference 16) that 2,4-Di-tert-butylphenol is a degradation product of the material ingredient Irgafos 168. In this case, the combination of compositional information and the scientific literature corroborates the tentative identification, allowing it to elevated to confident status.

Considering peak 5, it is noted that its mass spectrum shows a very good fit with mass spectrum of stearic acid. The fit is confirmed by expert review, leading to the conclusion that stearic acid is a proper tentative identification for peak 5. As was the case with peak 4, one notes that stearic acid is not listed as an intentional ingredient and at first glance the tentative identification does not appear to be corroborated by composition. However, closer examination of the ingredient list reveals that calcium stearate was added as an acid scavenger to the polyolefin. While this is not a one-on-one correlation (indeed, calcium stearate is not the same exact molecule as stearic acid), once it is understood how the "acid scavenger" mechanism works (illustrated here with acetic acid as the acid being scavenged), it becomes obvious that the acid scavenger's action results in the formation of stearic acid:

$$Ca(stearate)_2 + 2\,HCOOH \rightarrow 2\,Stearic\,Acid + Ca(HCOO)_2$$

Once again, composition corroborates a tentative identity, elevating the identity to confident status.

As a last example of how corroborating compositional information can elevate a tentative identification to confident status, consider peak 6. As was the case with the other peaks, mass spectral matching, augmented by expert review, produces a tentative identity, in this case palmitic acid. Now surely this is the best level of identification that can be obtained for this peak, as palmitic acid is clearly not a known ingredient in the tested polyolefin. But we can do better, if we think a bit. With a little digging, we can establish that calcium stearate additives are generally natural products that rarely are as pure as analytical-grade reagents. In fact, the calcium stearate additive is likely a mixture of both stearate, palmitate and even lower molecular weight fatty acid salts. Thus, the calcium stearate is a logical source of palmitic acid and once again compositional information corroborates a tentative mass spectral match identity to elevate its status to confident.

Thus, based on tentative identities secured by expert-reviewed mass spectral matching corroborated by compositional knowledge, all 6 extractables noted in Figure 25 have been confidently identified.

Even information from a partially elucidated extractables profile can either facilitate an identification or be used as collaborating information to elevate an identification. For example, consider the case where a homologous series of compounds with a certain functionality (for example, a homologous series of siloxanes) were detected and the identity of a number of those homologous compounds was confirmed via the analysis of authentic standards. An extractable from the same homologous series that was identified as a partial or a tentative identification based on the merits of its own mass spectrum could be more confidently identified on the basis of it being a member of the established homologous series of confirmed compounds.

This circumstance is illustrated in Figure 26. It is very clear that the major peaks in the chromatogram are all part of a homologous series of extractables, differing in mass 74. Via available authentic reference standards, the peaks at 12.88, 16.02 and 18.81 minutes are confirmed to be siloxanes of increasing ring size. However, the next compound in the series (peak at 21.30 min) can only be tentatively identified, via

mass spectral matching due to lack of an available reference standard. However, the fact that the compound is so clearly the "next step up" in the homologous series surely supports the proposition that the tentative identity can be elevated to at least confident status.

4.6 The Use of a Database to Capture the Identification Efforts

The practice of using corroborative data to augment and support higher level identifications, as well as the efforts to secure the identity of the compounds through mass spectral matching or mass spectral interpretation, can be quite time consuming, labor-intensive, and expensive, requiring expert scientific, process and material knowledge and advanced analytical capabilities.

It is evident that once a compound has been identified and has been assigned an elevated identification class, the supporting analytical data (such as mass spectral fragmentation or retention time) is fixed, as long as the analytical methods and instrumental settings remain unchanged. This circumstance supports the generalization that "once a compound has been identified to a certain class, it remains identified ton that class until the analytical method is changed".

Thus, there is significant value in capturing completed identifications, as it makes little sense to perform the identification exercise all over again, for each analytical event. An appropriate means of capturing identities, and documenting the identification process, is via an internal database.

Consider the example of the 2 compounds whose identities were previously elevated from tentative to confident after reviewing the list of ingredients: Irgafos 168 and Irganox 1076. If these compounds, their identities and identity class and their identifying information is captured in a database, then these compounds can be identified with their established identification class each time they are encountered in a screening study. For example, if a chromatographic peak is produced at the recorded retention time of Irganox 1010 and the peak's mass spectrum matches the recorded mass spectrum of Irganox 1010, is this not sufficient information to assign this peak a confident identity of Irganox 1010? Moreover, if the retention time and mass spectrum recorded in the database for Irganox 1010 has been confirmed by analysis of an authentic reference standard, is this not sufficient information to assign the identity a confirmed classification? Thus, one is able to provide a confirmed identity for the peak, based merely on retention time and mass spectral matching to the internal database.

The concept of "once identified always identified" is a powerful means of making identification efficient and reproducible but is only enabled if the identification information is captured in an accessible database.

4.7 Conclusion

Securing the correct identity of an extractable or leachable is essential, as it is the correct identity that enables a compound's impact assessment. If one cannot unequivocally identify a compound, the overall impact assessment of the compound will be flawed, and there is no subsequent action that can be taken in the impact assessment process to correct for a false identity.

Nevertheless, it is a practical reality that not all are extractables and leachables can be unequivocally identified, even with the best available analytical data, the most complete material and process information and the highest level of scientific appraisal. To ensure that users of an identity understand the

relative certainty that the identity is correct and to provide scientists with an aid for judging the value of the data they have collected, a hierarchy or classification of identities has been established:

- Partial: no full identity of the compound can be determined, but certain general functionalities can be ascertained
- Tentative: one-dimensional identification, only based upon one piece of information
- Confident: a two-dimensional identification, based upon at least 2 independent pieces of corroborating data
- Confirmed: a three-dimensional identification, based upon 3 or more independent and complimentary pieces of corroborating data

Clearly, the ultimate objective of the identification process is to secure a confirmed identity. When the available information is insufficient to support this level of certainty (for example, a reference standard is not available to secure the confirmation), other classes have been established to communicate the certainty in the identity, based on the amount and rigor of the supporting information.

The most likely identification class secured through the typically employed identification processes (mass spectral matching and mass spectral interpretation) is tentative. Although a tentative identification is the minimum appropriate for impact assessment, one understands that there is a possibility that the tentative identity is incorrect, leading to a flawed impact assessment. Therefore, the goal of the identification process is to secure as high an identification level as data and insight will support.

Means of "elevating" a tentative identification via corroborating information include:

- Chromatography and associated retention time considerations (e.g. "Retention Index" Matching for GC/MS)
- Tandem mass spectrometry
- Additional Evidences from orthogonal techniques
- Derivatization
- Indirect inferences

Additionally, we recognize the truth in the statement that "once a compound has been identified and assigned its highest identification class, the compound will remain identified in that identification class so long as the analytical screening methods are not materially altered". Thus, identification of the same compound in new test articles should not be a process of re-identifying the compound all over again (re-inventing the wheel) but rather a process leveraging the ability to say "I have seen and identified this compound before and thus I already know what it is". This efficient, effective and reproducible process for identification is enabled by secured identities, their identification class and their identifying information in a readily assessable and frequently used internal database.

Part 5

Final Thoughts

The use of mass spectral detectors in the screening process to account for all organic extractables and leachables that are present in a material, component or device – or compounds that may have leached out and lead to direct (for medical devices) or indirect (for drug products) patient exposure –has been widely accepted and implemented.

It is generally accepted that mass spectral matching is potentially a reliable means of identifying organic extractables and leachables via GC/MS. However, there are many circumstances where mass spectral matching will not lead to the desired outcome of an unequivocal identification. In such cases, more information about the compound's identity can be obtained via mass spectral interpretation. It should be noted that for mass spectra generated via LC/MS, mass spectral interpretation is often the only way to increase the level of identification for a detected compound, as there are no universal mass spectral libraries available that are optimized for LC/MS.

A key success factor in mass spectral interpretation is establishing the molecular ion of the detected compound. If the molecular ion is established with a High-Resolution Accurate Mass Instrument, then possible elemental compositions of the compound can be secured. In that case, additional information such as interpretation of isotopic data and interpretation of mass fragments can lead to elucidation of the compound's structure. Additionally, publicly available databases (such as PubChem, SciFinder or ChemSpider) can be consulted to look for actual structures that could potentially fit the generated mass spectrum. In addition, existing software packages - using in-silico fragmentation - can assist in rationalizing all mass spectral peaks for a structure candidate and support or reject the validity of the suggested structure. When the fragmentation supports the suggested structure (either via (manual) expert interpretation or via software supported interpretation), one can speak about a "*De Novo*" structural elucidation. In that case, the identification by structure elucidation may lead to a tentative identification. A further increase in identification level from tentative to confident or confirmed can be accomplished through the use of corroborating data or through the analysis of an authentic standard for the suggested compound.

When the elemental formula cannot be derived from the corresponding mass spectrum, then any proposal of a molecular structure would be highly speculative, and the highest level of identification that can be obtained from this information is a "partial" identification.

If "unit mass" mass spectrometers are used to collect the identifying information, a full structural identification of the compound will rarely be the outcome unless spectra of structural analogs of the compound of interest are present in a spectral library and the compound of interest is subject to similar fragmentation pathways. Generally, higher resolution, so-called accurate mass data will be required to complete the structure elucidation and thereby generate a tentative identity.

It should be emphasized that postulating a chemical structure, solely based upon mass spectral information and interpretation is not an easy task. The mass spectrometrist performing this task should be aware of the importance and consequences of postulating a defined chemical structure, as this information will be used to link the chemical compounds to its toxicological information and will be the basis for a subsequent toxicological evaluation of the compound. Cases where the wrong identity for the compound is postulated will inevitably bias the overall safety evaluation of the material, device or container/closure component or system.

Part 6

References

6.1 Cited References

1. USP <1663> Assessment of extractables associated with pharmaceutical packaging/delivery systems. USP 41/NF 36, pp. 7910-7923. Effective May 1, 2018.
2. Jenke, D.. Identification and quantitation classifications for extractables and leachables. PDA J Pharm Sci Technol. 74(2): 275-285 (2020).
3. Stein, S. Chemical substructure identification by mass spectral library searching. J Am Soc Mass Spectro. 6: 644-655 (1995).
4. Kind, T.; Fiehn, O. Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry. BMC Bioinformatics. 8: 105 (2007).
5. Watson, J.; Sparkman, O. Introduction to mass spectrometry: instrumentation, applications and strategies for data interpretation, 4th ed., Chichester: John Wiley and Sons, 2007.
6. Smith, R. Understanding mass spectra: A basic approach, 2nd ed., Hoboken, NJ: John Wiley and Sons, 2004.
7. Hites, R.; Jobs, K. Is nontargeted screening reproducible? ES&T. 52(21): 11975-11976 (2018).
8. Stein, S.; Scott, D. Optimization and testing of mass spectral library search algorithms for compound identification. J Am Soc Mass Spectro. 5: 859-566 (1994).
9. McLafferty, F.; Turecek, F. Interpretation of mass spectra, 4th ed., Sausalito: University Science Books, 1993.
10. Harrison, A.G. Chemical Ionization Mass Spectrometry, 2nd Edition, CRC Press, 1992.
11. Hoffmann, T.G.M. The next level in chemical space navigation: going far beyond enumerable compound libraries. Drug Discovery Today. 24(5): 1148-1156 (2019).
12. Milman, B. Chemical identification and its Quality Assurance, Heidelberg: Springer-Verlag, 2011.
13. Kind, T.; Fiehn, O. Advances in structure elucidation of small molecules using mass spectrometry. Bioanalytical Reviews. 2(1-4): 23-60 (2010).
14. Stein, S. An integrated method for spectrum extraction and compound identification from gas chromatography / mass spectrometry data. J Am Soc Mass Spectro. 10: 770-781 (1999).
15. Stein, S. Estimating probabilities of correct identification from results of mass spectral library searches. J Am Soc Mass Spectro. 5: 316-323 (1994).
16. Jenke, D.; Swanson, S.; Edgcomb, E.; Couch, T.; Chacko, M.; Garber, M.J.; Fang, L. Strategy for Assessing the Leachables Impact on Material Change in a Container/ Closure System. PDA J Pharm Sci Technol. 59(6): 360-379 (2005).

6.2 Additional References

1. Smith, R. Understanding mass spectra: A basic approach, 2nd ed., Hoboken, NJ: John Wiley and Sons, 2004.
2. Schymanski, E.; Jeon, J.; Gulde, R.; Fenner, K.; Ruff, M.; Singer, H.; Hollender, J. Identifying small molecules via High Resolution Mass Spectrometry: Communicating confidence. ES&T. 48(4): 2097-2098 (2014).
3. Sica, V.; Krivos, K.; Kiehl, D.; Pulliam, C.; Henry, I.; Baker, T. The role of mass spectrometry and related techniques in the analysis of extractable and leachable chemicals. Mass Spectro Rev. 9999: 1-15 (2019).
4. Stein, S.; Scott, D. Optimization and testing of mass spectral library search algorithms for compound identification. J Am Soc Mass Spectro. 5: 859-566 (1994).
5. Stein, S.; Wallace, W. NIST Standard Reference Database 1A - User's Guide NIST/EPA/NIH Mass Spectral Library (NIST17) and NIS Mass Spectral Search Program (version 2.3), Gaithersburg : U.S. Department of Commerce - National Institute of Standards and Technology, 2017.

6. Moorthy, A.; Wallace, W.; Kearsley, A.; Tchekhovskoi, D.; Stein, S. Combining fragment-ion and neutral-loss matching during mass spectral library searching: A new general purpose algorithm applicable to illicit drug identification. Anal Chem. 89(24): 13261-16268 (2017).

7. Sparkman, D. Evaluating electron ionization mass spectral library search results. J Am Soc Mass Spectro. 7: 313-318 (1996).

8. Samanipour, S.; Martin, J.; Lamoree, M.; Reid, M.; Thomas, K. Letter to the Editor: Optimism for nontarget analysis in environmental chemistry. ES&T. 53(10): 5529-5530 (2019).

9. Hites, R.; Jobst, K. Response to Letter to the Editor: Optimism for Nontarget analysis in environmental chemistry. ES&T. 53(10): 5531-5533 (2019).

10. Hites, R.; Jobs, K. Is nontargeted screening reproducible?". ES&T, 52(21): 11975-11976 (2018).

11. McLafferty, F.; Zhang, M.; Stauffer, D.; Loh, S. Comparison of algorithms and databases for matching unknown mass spectra. J Am Soc Mass Spectro. 9: 92-95 (1998).

12. U.S. Environmental Protection Agency, "USEPA Contract laboratory program National functional guidelines for organic data review (EPA540/R-99/008)," 1999.

13. Babushok, V.; Linstrom, P.; Reed, J.; Zenkevich, I.; Brown, R.; Mallar, W.; Stein, S. Development of a database of gas chromatographic retention properties of organic compounds. J Chomatogr A. 1157: 414-421 (2007).

14. Stein, S.; Babushok, R.; Brown, R.; Linstrom, P. Estimation of Kóvats retention indices using group contributions. J Chem Infor Modelling. 47(3): 975-980 (2007).

15. Wen, Y.; Talebi, M.; Amos, R.; Szucs, R.; Dolan, J.; Pohl, C.; Haddad, P. Retention prediction in reversed phase high performance liquid chromatography using quantitative structure-retention relationships applied to the Hydrophobic Subtraction Mode. J Chromatogr A. 1541(16): 1-11 (2018).

16. Aalizadh, R.; Nika, M.; Thomaidis, N. Development and application of retention time prediction models in the suspect and non-target screening of emerging contaminants. J Haz Mat. 363: 277-285 (2019).

17. Kaliszan, R. QSSR: Quantitative Structure-(Chromatographic) Retention Relationships. Chem Rev. 107(7): 3212-3246 (2007).

18. Hébérger, K. Quantitative structure - (chromatographic) retention relationships: QSRR. J Chromatogr A. 1158(1-2): 273-305 (2007).