Nelson Labs®
A Sotera Health company

A DISCUSSION OF:

Good Identification Practices For Organic Extractables & Leachables Via Mass Spectrometry

PART IV OF IV: Additional Evidences Supporting Higher Level Identifications

Authors:
Dr. Dennis Jenke
Dr. Piet Christiaens
Dr. Philippe Verlinde
Dr. Jan Baeten
Dr. Ward D'Autry
Dr. Jean-Marie Beusen
James O. Mullis

Sales **Europe +32 16 40 04 84** | InfoEurope@**Nelsonlabs**.com
Sales **US +1 801-290-7502** | Sales@**NelsonLabs**.com

www.**NelsonLabs**.com

ADDITIONAL EVIDENCES SUPPORTING HIGHER LEVEL IDENTIFICATIONS

OPENING THOUGHTS

Identification of extractables and leachables is a critical aspect for a substances' toxicological safety risk assessment, as identity establishes the substance's inherent toxicity. Nelson Labs has generated a series of white papers focusing on identification and, more specifically, on the process by which mass spectral data and other supporting evidence is used to secure, judge, and justify complete and correct identities for organic extractables or leachables surfaced by chromatographic screening analyses. Part I of this series introduced the concept of identification and established its critical role in safety assessment. It also described the various means of securing identities, discussed the concept of identification classes, and proposed an identification classification. The importance of confidence in identification was emphasized and the identification process was delineated via an identification decision tree. In Part II, the process of securing a compound's identity via mass spectral matching to mass spectral libraries was considered; specifically addressing the strengths, points of attention, and potential pitfalls of this strategy. In Part III, the identification strategy called *Mass Spectral Interpretation* was considered; where *Mass Spectral Interpretation* is the process of securing a compound's identity solely by expert interpretation of the information present in the compound's mass spectrum.

An identification secured by either mass spectral matching (Part II) or mass spectral interpretation (Part III), is by definition a TENTATIVE identification as it is based on a single dimension of identifying information. While TENTATIVE identifications provide the minimally acceptable input to a toxicological safety risk assessment of extractables and leachables, greater certainty in the identity leads to greater certainty in the toxicological assessment. Thus, additional information about the compound of interest is often pursued to 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 *(Part II)* and via a well-documented mass spectral interpretation of the mass fragments *(Part III)*, then these two independent corroborating outcomes may "elevate" the TENTATIVE identification secured with both processes to a CONFIDENT identification.

Additionally, identifications can be substantiated by accumulating independent evidences

"Thus, for example, if the same TENTATIVE identity is secured by mass spectral matching (Part II) and via a well-documented mass spectral interpretation of the mass fragments (Part III), then these two independent corroborating outcomes may "elevate" the TENTATIVE identification secured with both processes to a CONFIDENT identification."

or evidences from analyses which are specifically chosen to confirm a certain identification. Logically, the more additional evidences that are gathered, the more certain the identification becomes. In this current Part IV, various means of "augmenting" identities are considered, and examples are provided. However, providing an exhaustive list of additional evidences is outside the scope of this document.

1. CHROMATOGRAPHY AND ASSOCIATED RETENTION TIME CONSIDERATIONS (E.G. "RETENTION INDEX" MATCHING FOR GC/MS)

As established previously, mass spectrometry is the most commonly used and accepted means for linking a discovered extractable or leachable to its unique identity. However, although mass spectrometry is a very powerful tool for identification, the technique becomes much less powerful when analyzing complex mixtures of compounds. Therefore, extracts or drug products are screened for extractables and 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 individually eluting compounds.

As a result of the chromatographic process, the compounds of interest are separated in terms of the time it takes them to elute from the chromatographic column *(prior to entrance into the mass spectrometer)*. This elution time, referred to as the retention time, will depend both on the chemical and physical nature of the compounds of interest and their interaction with the selected stationary phase, the dimensions and temperature *(program)* of the chromatographic column, and the mobile phase selected for the chromatographic method. The elution time for a compound in a chromatographic method may therefore be diagnostic. 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)*, specific retention times are 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 sufficiently diagnostic that it can be used alone 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 way, the 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

"This elution time, referred to as the retention time, will depend both on the chemical and physical nature of the compounds of interest and their interaction with the selected stationary phase, the dimensions and temperature (program) of the chromatographic column, and the mobile phase selected for the chromatographic method. The elution time for a compound in a chromatographic method may therefore be diagnostic."

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 that has a mass spectrum and retention that matches the mass spectrum and retention time of a reference compound in a database is recognized as the reference compound and is therefore considered as a CONFIRMED identity [1].

Perhaps the greatest advantage of identification corroboration via retention time matching is that the retention time is essentially a "free" piece of information. That is, the retention time is obtained via the same analytical activity as the mass spectrum. Unlike other corroborating information, obtained via a different analytical technique (for example, NMR), securing retention time as a corroborating piece of data does not require re-analysis of the sample to obtain this additional information.

For the retention time to be useful as an identification tool, it must be accurate and reproducible over time and across different instruments running the same chromatographic methods. However, shifts in retention time occur frequently as a result of routine maintenance procedures such as column trimming. In a multi-instrument laboratory where multiple instruments are running the same chromatographic method, 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.

"Presumably, a test compound that has a mass spectrum and retention that matches the mass spectrum and retention time of a reference compound in a database is recognized as the reference compound and is therefore considered as a CONFIRMED identity."

To a certain extent, two methodologies can manage retention time differences , 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 obtained on another system by adjusting the chromatographic conditions; this is typically more 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 systems. A specific *compound (usually the Internal Standard for Injection)* is used for both developing the locking calibration and locking all future systems.

"Unlike other corroborating information, obtained via a different analytical technique (for example, NMR), securing retention time as a corroborating piece of data does not require re-analysis of the sample to obtain this additional information."

As noted above, absolute retention times can be irreproducible as they depend on a variety of chromatographic factors which renders them unsuitable as a "universal" criterion 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 using the retention indices as supporting identification information is that the RI-values do not depend on the exact column dimensions, flows, or temperature programming. However, they do depend on the type of stationary phase used *(for example non-polar versus polar phases).*

> ***"The advantage of using the retention indices as supporting identification information is that the RI-values do not depend on the exact column dimensions, flows, or temperature programming."***

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. This is shown in Equation 1, which was adapted from Equation 2 for temperature-programmed measurements

Equation 1

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

Equation 2

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

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*. Lee 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 retention index data that were either determined experimentally or were estimated using theoretical models. The 2020 version of this MS library contains 447,289 citations of experimental RIs for 114,629 compounds. These experimental RI data are collected from different contributors and are given as median value ± deviation *(number of data points).* It should be noted that a majority of the compounds have just one measurement and that indices are not uniformly distributed over different compounds or compound classes [2]. Experimental RI values in the NIST library are classified into three types of stationary phases:

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

In addition to experimental RI values, several theoretical models have been developed to estimate RIs [3, 4, 5, 6, 7]. 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 is suitable for the rejection of certain false identifications made by GC/MS [3].

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

- Matching stationary phase should be ascertained.
- Either a standard with an *n*-alkane mix should be run with each sequence to set up 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.

"The NIST/EPA/NIH Mass Spectral Library for GC/MS contains a growing amount of retention index data that were either determined experimentally or were estimated using theoretical models. The 2020 version of this MS library contains 447,289 citations of experimental RIs for 114,629 compounds. These experimental RI data are collected from different contributors and are given as median value ± deviation (number of data points)."

Table 1 illustrates the effectiveness of using RI data contained in publicly available commercial MS libraries *(e.g., the NIST/EPA/NIH Mass Spectral Library)* as potential corroborating information in support of a mass spectral identification. In this table, the experimental RI values, derived from the Nelson Labs GC/MS screening methods, are compared to *(1)* the experimental RI data and *(2)* the "estimated" RI data, both reported in the NIST/EPA/NIH Mass Spectral Library. In the selection of compounds represented in Table 1, care was taken to include compounds of diverse chemical nature and associated variation in retention properties. Except for 2 outliers *(displayed in red),* the experimental RI data in NIST agree very well with experimental RI data which were derived from the Nelson Labs GC/MS "locked" screening method *(median ΔRI =*

10). Therefore, a good fit of the experimental RI value of a query compound – detected in an extraction study – with the obtained experimental RI value from a commercial library may assist in selecting the right chemical structure in a "hitlist" that is generated in the process of mass spectral matching.

However, as it can be observed in Table 1, the correlation between the calculated or "estimated" RI data from the NIST library and the experimental RI data which were derived from the Nelson Labs GC/MS screening method, is substantially lower. Therefore, it is concluded that the accuracy of the "estimated" RI values in a commercial library is too low to support a higher confidence in the initially secured identification of a compound. However, the "estimated" NIST RI values may be used to invalidate a suggested identification *(e.g. obtained via mass spectral matching)* if the recorded RI value of the compound is deviates substantially from the "estimated" RI value from NIST.

Compound Name	Nelson Labs		NIST experimental NON-POLAR SS			NIST estimated 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	2793	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

Table 1. *Comparison of experimentally determined Retention Index values by Nelson Labs (based upon the recorded retention times) with the experimental and "estimated" Retention Index (RI) values which could be found in the NIST library (SS: Stationary Phase; CI: Confidence Interval)*

Despite the agreement noted in Table 1, identification corroboration via retention matching is most effective when the reference retention data are not derived from an external *source (such as the NIST MS Library),* but when these retention data are obtained through analysis of authentic reference standards using the same chromatographic screening method used in routine laboratory operations for the analysis of the extracts or drug products.

Unfortunately, no universal or unified HPLC retention index system has been established for reversed phase, normal phase, and HILIC [8]. Although differences in retention times across instruments are higher 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.

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. The molecular ion is good choice of precursor ion as it results in a spectrum of product ions which are unequivocally related to the molecular structure. MS/MS analyses are particularly useful in obtaining 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 1 represents the MS/MS annotated fragmentation spectra for the *(pseudo)* molecular ions for aleuritic acid, which is prone to in-source fragmentation.

In case of co-elution, mass spectral deconvolution is a powerful tool to resolve spectra from coeluting compounds and is effective with a vast majority of acquired spectra *(Part III of this series on mass spectral interpretation)*. However, complete resolution of complex mass chromatograms by deconvolution will not be possible in all cases.

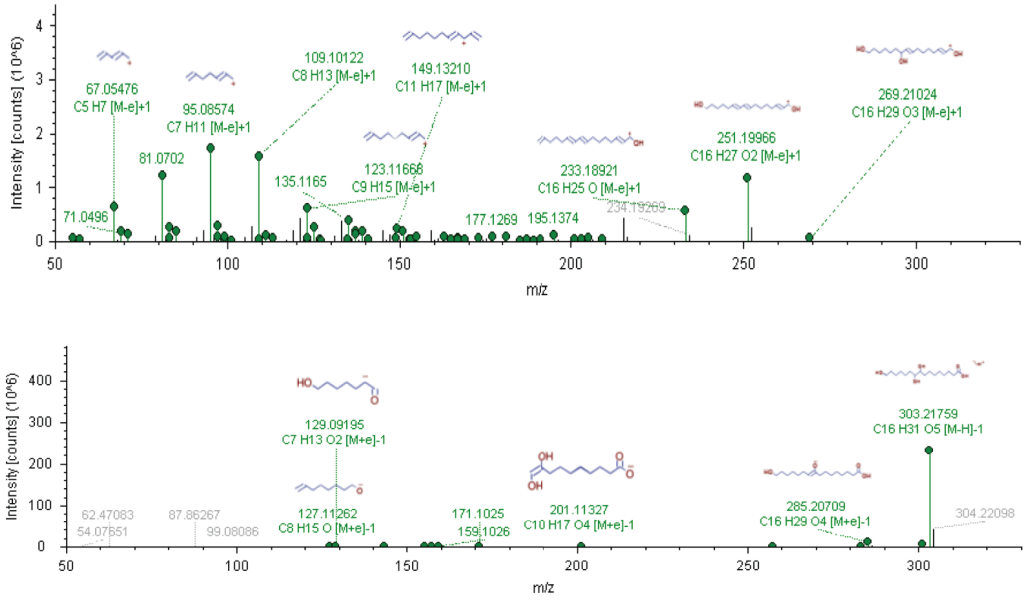

Figure 1. *Annotated APCI MS/MS high resolution accurate mass fragmentation spectra (30 eV) for pseudo molecular ions top: $[M+H]^+$ at m/z 305.232 ± 0.5 m/z (positive mode) and bottom: $[M-H]^-$ at 303.217 ± 0.5 m/z (negative mode) obtained for aleuritic acid.*

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.

Following is an example of this second scenario: Screening of an extract via LC/MS *(ESI+)* produced a chromatogram with an extractable's peak at 7.34 min whose corresponding mass spectrum is shown in Figure 2. The mass spectrum shows a (protonated) molecular ion mass *($[M+H]^+$)* at m/z 114.091. The hypothesis 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 3)*.

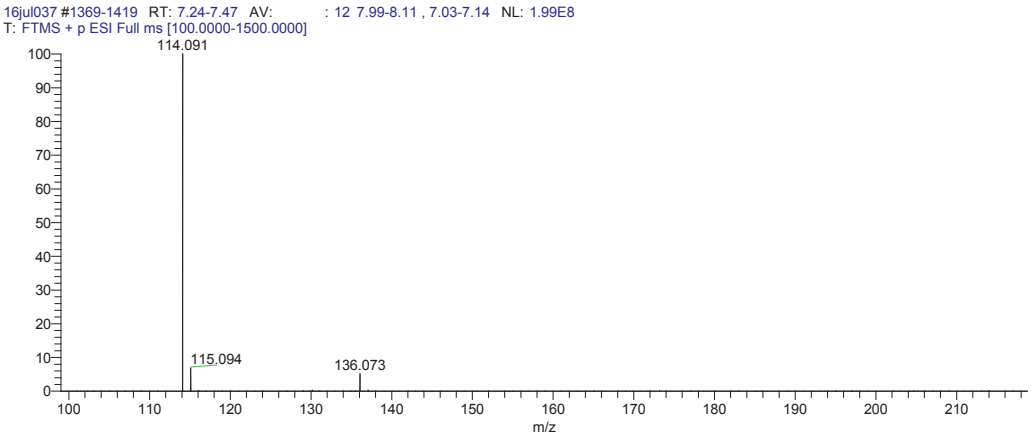

Figure 2. *Corresponding mass spectrum of the compound, detected at RT 7.34min. This mass spectrum shows the presence of a molecular ion at m/z 114,091. This assumption that this is the parent ion is confirmed by the detection of the Na-adduct at m/z 136,073.*

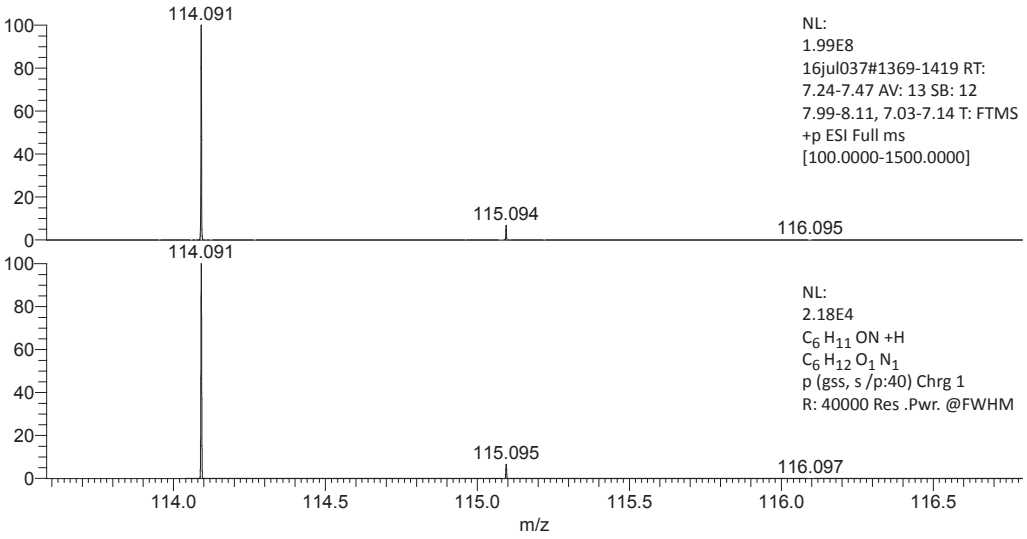

Figure 3. *Verification of the isotope pattern. A simulation of how the isotope pattern could look for the protonated $C_6H_{11}ON+H$ (lower mass spectral isotope pattern). This shows a perfect match with the isotope pattern of the detected compound (upper mass spectral isotope pattern), which confirms the suggested elemental composition.*

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 means of obtaining "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 $C_6H_{11}ON$. The list of candidates that is generated suggests different chemical structures that could fit with the established elemental formula *(see Figure 4)*.

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		$C_6H_{11}NO$	113.1576	138	213	7	59
7480		$C_6H_{11}NO$	113.1576	137	452	279	275
7236		$C_6H_{11}NO$	113.1576	120	234	40	160

Figure 4. *List of candidates for a $C_6H_{11}NO$ elemental formula, generated via ChemSpider*

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 option of the candidate list generated via ChemSpider. This is 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.

Another circumstance where 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 $C_6H_{11}ON$) may co-elute in the LC/MS chromatogram, 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 data, such as NMR, to independently secure an unknown's identity. This identification strategy is described in the USP <1663>, *"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 some of the considerations for NMR, made below, are *a fortiori* true for these high-end techniques also.

While the power of NMR as an identification method is well known, the use of this technique in E&L laboratories is limited by certain practical realities such as access to NMR technology. Although access to NMR technology may be straightforward for larger pharmaceutical companies, it may be problematic for E&L labs in a contract research environment. The cost of an 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 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 requirements to perform an NMR experiment on this neat chemical compound – often a few milligrams of the purified "unknown compound" at least – may require intensive sample preparation steps, such as isolation of the compound through fraction collection.

The complexity of the NMR interpretation is illustrated in Figure 5, 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. It becomes obvious that NMR is not a "magical solution" that immediately leads to a confirmed identification; rather, 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.

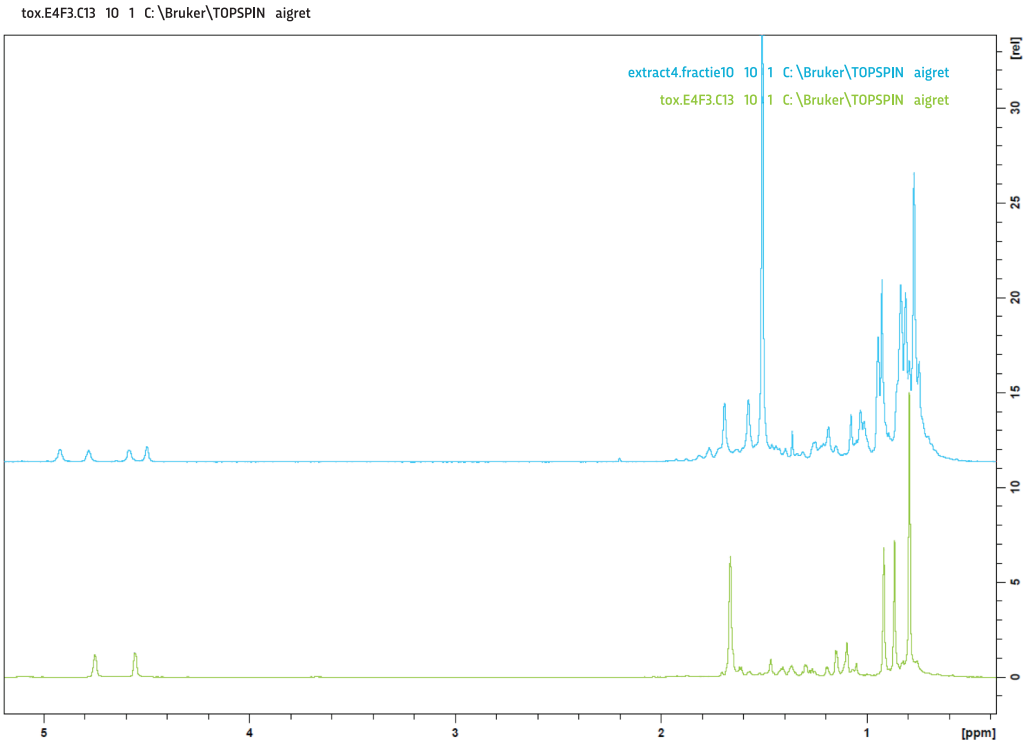

Figure 5. *^{1}H-NMR spectra of the isolated $C_{13}H_{24}$ and the $C_{21}H_{40}$ rubber oligomers, performed by Nelson Labs in collaboration with the University of Leuven, Belgium. This NMR spectrum compares the ^{1}H-NMR spectra of the $C_{13}H_{24}$ (top) and the $C_{21}H_{40}$ (bottom) oligomer. 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 ^{1}H-NMR spectrum.*

4. DERIVATIZATION

Derivatization is the chemical treatment of an extract designed to convert an extracted compound *(or compounds)* 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 derivatization techniques used in 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 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 chromatograms and MS spectra from both the non-derivatized and derivatized extract indicates whether the extract contains analytes whose structure includes derivatizable groups; and if there is an analyte with derivatized groups, how many derivatized groups the analyte possesses. *(e.g. Figure 6).*

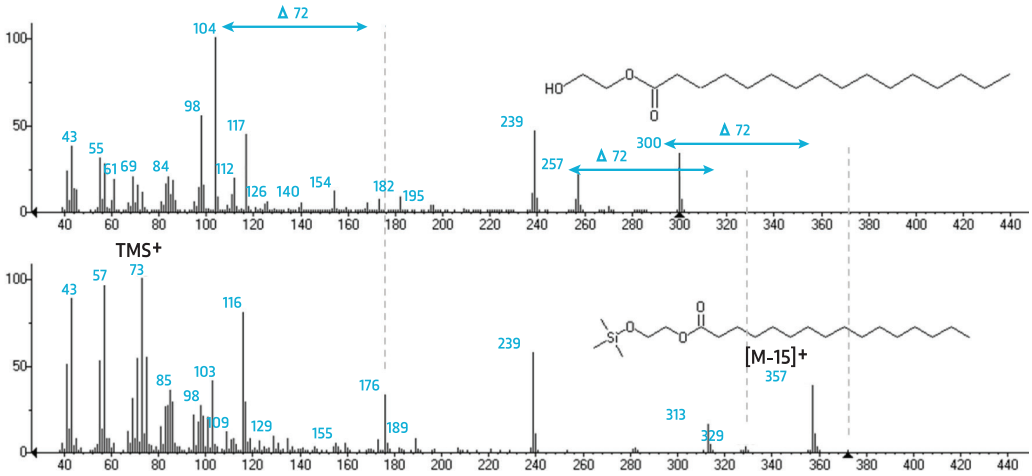

Figure 6. *Comparison of the EI mass spectrum of 2-hydroxyethyl palmitate with the mass spectrum of its trimethylsilyl (TMS) derivative. A mass difference of 72 Da is observed for the molecular ion (m/z 300 versus m/z 372) and demonstrates that the molecule contains one derivatizable group (in this case 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.*

5. INDIRECT INFERENCES

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

To illustrate, a hypothetical list of ingredients for a polyolefin material is given in Table 2. Each ingredient serves a specific purpose, to either protect the polymer (in this case *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)*.

Compound Name	CI (95%)	RI
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

Table 2. *Table with a hypothetical list of ingredients for a material of Construction (in this case, a polyolefin)*

This polyolefin material, with a composition described in Table 2, is then subjected to an extraction with an organic solvent followed by extract analysis via GC/MS. The resulting chromatogram, Figure 7, contains 6 peaks with associated compounds that can be confidently identified as follows.

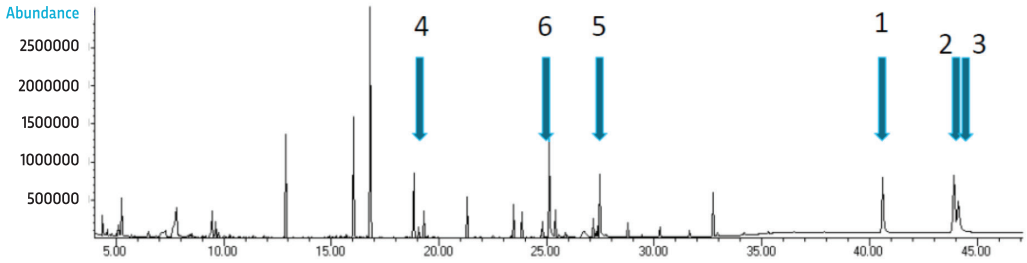

Figure 7. *GC/MS Chromatogram of an organic extract of the material of Construct with a fictious composition as described in Table 2. Based upon the provided information in the ingredients list, (at least) 6 identifications can be upgraded from a TENTATIVE identification to a higher class of identification (i.e. CONFIDENT) using this information.*

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 leads to the conclusion that the match is sufficient for both compounds to be 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 was tentatively identified, via reviewed mass spectral matching, as Tris (2,4-di-tert-butylphenyl) phosphate, the well-known and well-characterized oxidized form of Irgafos 168. 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 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 [9])* 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 be elevated to CONFIDENT status.

Considering peak 5, note that its mass spectrum shows a very good fit with the mass spectrum of stearic acid. The fit is confirmed by expert review, leading to the conclusion that stearic acid is a proper TENTATIVE identification, obtained through mass spectral matching. As was the case with peak 4, 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 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 action of this acid scavenger results in the formation of stearic acid:

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

Again, composition corroborates a TENTATIVE identity, elevating the identity to CONFIDENT status.

Lastly, 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 this is surely 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 with a little digging, it can be established that calcium stearate additives are generally natural products that are rarely as pure as analytical-grade reagents *(for example)*. 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 7 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 8. 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.

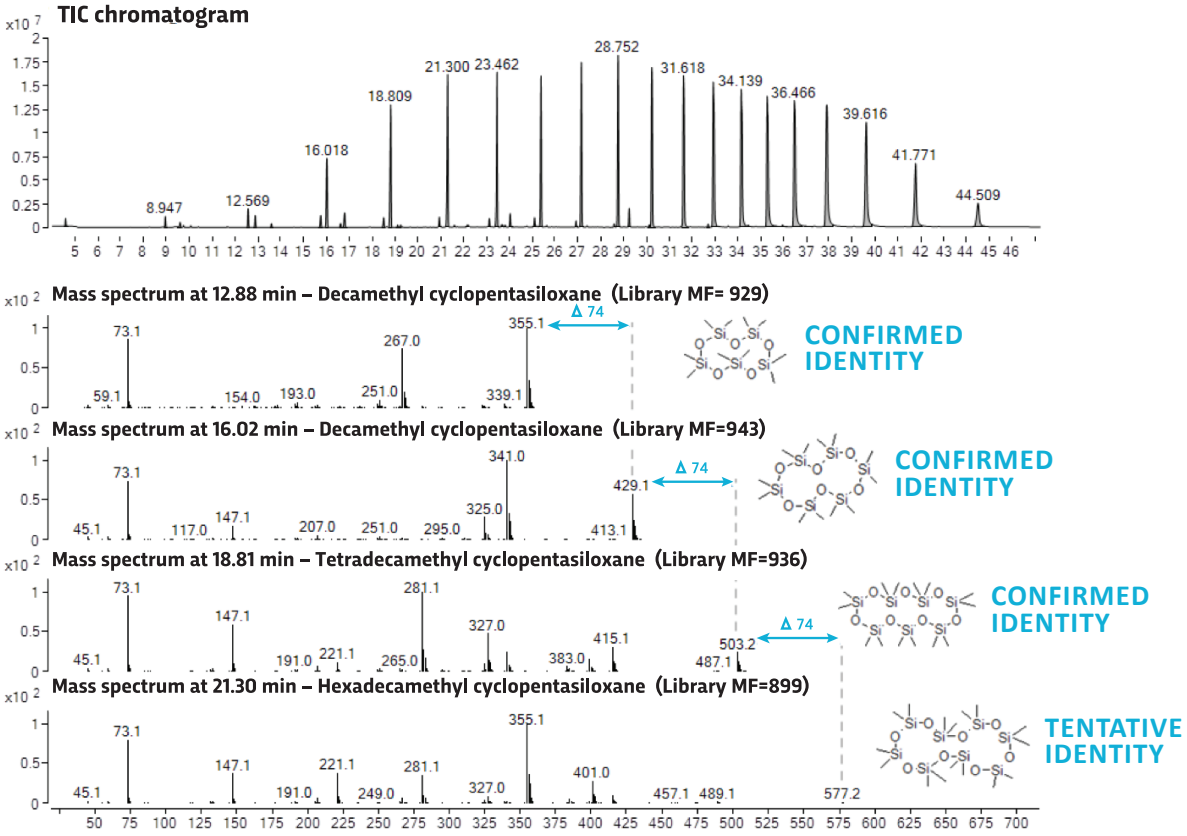

Figure 8. *Identification of a homologous series of cyclic siloxanes in a silicone sheet extract. The chromatogram (top) shows many peaks which are separated by regular time intervals. The corresponding mass spectra (only four spectra displayed) have excellent mass spectral matches with cyclic dimethylsiloxanes of different length. It can be clearly observed in the above spectra 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. 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 all homologues are detected in the same test item adds confidence to the identification.*

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 *(Part II)* or mass spectral interpretation *(Part III)* can be quite time consuming, labor intensive, and expensive requiring expert scientific process and material knowledge as well as 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)* and corroborative data 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 in that class until the analytical method is changed".

> ***"An appropriate means of capturing identities, and documenting the identification process, is via the development of an internal database."***

Thus, there is significant value in collating completed identifications, as it makes little sense to perform the identification exercise over again for each analytical event even for the evident extractables. An appropriate means of capturing identities, and documenting the identification process, is via the development of 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 1076 and the peak's mass spectrum matches the recorded mass spectrum of Irganox 1076, this should be sufficient information to assign this peak a CONFIDENT identity of Irganox 1076. Moreover, if the retention time and mass spectrum recorded in the database for Irganox 1076 has been confirmed by analysis of an authentic reference standard, this should be enough information to assign the identity a CONFIRMED classification. Thus, you can provide a CONFIRMED identity for the peak, based merely on

> ***"Moreover, if the retention time and mass spectrum recorded in the database for Irganox 1076 has been confirmed by analysis of an authentic reference standard, this should be enough information to assign the identity a CONFIRMED classification. Thus, you can provide a CONFIRMED identity for the peak, based merely on retention time and mass spectral matching to the internal database, without having to run the authentic standard each time the peak is encountered."***

retention time and mass spectral matching to the internal database, without having to run the authentic standard each time the peak is encountered.

The concept of "once identified always identified" is a powerful means of making identification efficient and reproducible but is only possible if the identification information is captured in an accessible database. [10]

7. CONCLUSION

Securing the correct identity of an extractable or leachable is essential, as the correct identity 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 what may be 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 collected data, the following 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 complementary 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 typical identification processes *(mass spectral matching and mass spectral interpretation)* is TENTATIVE. Although a TENTATIVE identification is the minimum appropriate identification level for impact assessment, you should understand 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.

In this Part IV of Nelson Lab's series on Good Identification Practices, means of "elevating" a TENTATIVE identification via corroborating information were discussed, including:

- Chromatography and associated retention time considerations *(e.g. "Retention Index" Matching for GC/MS),*
- Tandem mass spectrometry,
- Additional evidences from orthogonal techniques,
- Derivatization, and
- 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 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 collating identities, their identification class, and their identifying information in a readily assessable and frequently used internal database.

This Part IV completes our series on this topic. It is our hope that the individual Parts have provided E&L practitioners sound and practical knowledge, practices, and insights that can be leveraged to reproducibly and unequivocally produce the highest confidence identities for the greatest number of extractables or leachables likely to be encountered in drug products, medical devices and their packaging, manufacturing, and as appropriate, their delivery systems.

BIBLIOGRAPHY

[1] USP 40-NF 35, "<1663> Assessment of extractables associated with pharmaceutical packaging/delivery systems".

[2] S. Stein, "Chemical substructure identification by mass spectral library searching," *Journal of the American Society for Mass Spectrometry,* vol. 6, pp. 644-655, 1995.

[3] T. G. M. Hoffmann, "The next level in chemical space navigation: going far beyond enumerable compound libraries," *Drug Discovery Today,* vol. 24, no. 5, pp. 1148-1156, 2019.

[4] B. Milman, Chemical identification and its Quality Assurance, Heidelberg: Springer-Verlag, 2011.

[5] T. Kind and O. Fiehn, "Advances in structure elucidation of small molecules using mass spectrometry," *Bioanalytical Reviews,* vol. 2, no. 1-4, pp. 23-60, 2010.

[6] S. Stein, "An integrated method for spectrum extraction and compound identification from gas chromatography / mass spectromtery data," *Journal of the American Society for Mass Spectrometry,* vol. 10, pp. 770-781, 1999.

[7] S. Stein, "Estimating probabilities of correct identification from results of mass spectral library searches," *Journal of the American Society for Mass Spectrometry,* vol. 5, pp. 316-323, 1994.

[8] J. Watson and O. Sparkman, Introduction to mass spectrometry: instrumentation, applications and strategies for data interpretation, 4th ed., Chichester: John Wiley and Sons, 2007.

[9] D. Jenke, S. Swanson, E. Edgcomb, T. Couch, M. Chacko, M. J. Garber, L. Fang, "Strategy for Assessing the Leachables Impact on Material Change in a Container/CLosure System", *PDA Journal of Pharmaceutical Science and Technology*, vol. 59, no. 6, pp. 360-379, 2005.

[10] P. Christiaens, J.-M. Beusen, P. Verlinde, J. Baeten, D. Jenke, "Identifying and Mitigating Errors in Screening for Organic Extractables adn Leachables: Part 2 - Errors in Inexact Identification and Inaccurate Quantification", *PDA Journal of Pharmaceutical Science and Technology*, vol. 74, pp. 108-133, 2020.