



EPA Public Access

Author manuscript

J Breath Res. Author manuscript; available in PMC 2019 November 15.

About author manuscripts

Submit a manuscript

Published in final edited form as:

J Breath Res. ; 13(1): 012001. doi:10.1088/1752-7163/aae8c3.

Beyond monoisotopic accurate mass spectrometry: ancillary techniques for identifying unknown features in non-targeted discovery analysis

Joachim D Pleil¹, M Ariel Geer Wallace, and James McCord

Exposure Methods and Measurements Division, National Exposure Research Laboratory, US Environmental Protection Agency, NC27711, United States of America

Abstract

High-resolution mass spectrometry (HR-MS) is an important tool for performing non-targeted analysis for investigating complex organic mixtures in human or environmental media. This perspective demonstrates HR-MS compound identification strategies using atom counting, isotope ratios, and fragmentation pattern analysis based on ‘exact’ or ‘accurate’ mass, which allows analytical distinction among mass fragments with the same integer mass, but with different atomic constituents of the original molecules. Herein, HR-MS technology is shown to narrow down the identity of unknown compounds for specific examples, and ultimately inform future analyses when these compounds reoccur. Although HR-MS is important for all biological media, this is particularly critical for new methods and instrumentation invoking exhaled breath condensate, particles, and aerosols. In contrast to standard breath gas-phase analyses where 1 mass unit (Da) resolution is generally sufficient, the condensed phase breath media are particularly vulnerable to errors in compound identification because the larger organic non-volatile molecules can form identical integer mass fragments from different atomic constituents which then require high-resolution mass analyses to tell them apart.

Keywords

accurate mass; monoisotopic mass; compound identification; chemical features; high resolution MS; isotope ratios; mass fragments

Premise

High-resolution mass spectrometry (HR-MS) has introduced an additional dimension for identifying unknown organic compounds in complex mixtures. Termed ‘non-targeted analysis’, this approach is defined as agnostic with respect to analytes; the samples are processed with no preconception as to content and the results are considered to be unidentified chemical ‘features’. These features are then post-processed to assign chemical formulae based on the HR-MS results (Schymanski *et al* 2015, Sobus *et al* 2018).

Original content from this work may be used under the terms of the Creative Commons Attribution 3.0 licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

¹ Author to whom any correspondence should be addressed. pleil.joachim@epa.gov.

This perspective is intended to explain the value and general methods for HR-MS in investigating complex organic mixtures as in human blood, breath and urine, and demonstrate the value of some compound identification strategies with specific examples. The primary HR-MS parameter, referred to as ‘exact’, or ‘accurate’ mass, allows analytical distinction among mass fragments with the same integer mass, but with different atomic constituents of the original molecules. The most basic discrimination occurs when separating ionic fragments with resolution at the fourth or fifth decimal place. This is a vast improvement over standard single-digit mass units (integer mass) instrumentation. The underlying concepts and physics principles of exact mass analysis have been described in detail in a predecessor article (Pleil and Isaacs 2016). Such ‘monoisotopic’ analysis is only the first, and most basic step. Once features are located and assigned to a chromatographic retention time or retention index, they can be further investigated by their molecular fragmentation patterns, the exact masses of the fragments, and their relative isotopic abundances (McLafferty *et al* 1999). This perspective describes how to exploit new technology for identifying unknown organic constituents within complex matrices, develop reasonable confidence in their identity, and ultimately inform future analyses when these compounds reoccur. Although HR-MS is crucial for all biological media, this is particularly important for new methods and instrumentation invoking exhaled breath condensate, particles, and aerosols as a diagnostic biological medium complementing blood and urine analysis (Ladva *et al* 2000, Zamuruyev *et al* 2016, Ghio *et al* 2017, Sauvain *et al* 2017, Winters *et al* 2017, Wallace and Pleil 2018a, 2018b). In contrast to standard breath gas-phase analyses where 1 Da resolution is generally sufficient, the condensed phase media are particularly vulnerable to errors in identification as the larger organic molecules contained therein have more possibilities for forming integer mass fragments.

Various forms of HR-MS are now being employed in new breath applications. A cursory search of recent articles finds that real-time and gas chromatographic (GC) instruments are employing time-of-flight (ToF) as a replacement for linear quadrupole detectors to improve discrimination of ionic fragments and that liquid chromatography HR-MS applications using ToF, MS-MS and orbitrap instruments are becoming more prevalent (Herbig *et al* 2009, Sukul *et al* 2015, Nizio *et al* 2016, Peralbo-Molina *et al* 2016, Andra *et al* 2017, Li *et al* 2017, Bregy *et al* 2018, Singh *et al* 2018).

Overview

As discussed in recent conceptual articles, chemical toxicity testing and human disease diagnosis have evolved beyond simple targeted analysis of chemicals of exposure, their chemical biomarkers, and certain endogenous response metabolites (Krewski *et al* 2010, Ala-Korpela *et al* 2012, Teeguarden *et al* 2016, Vineis *et al* 2017). Basically, samples are analyzed for as many compounds as feasible within a particular laboratory’s capability, and subsequently subjected to statistical analysis to identify ‘features’ characteristic of the behavior under investigation. The simplest version of this process is designating ‘case-control’ pairs, where one sample is treated or exposed in some fashion. At this point, the results are compared and unknowns are differentiated between cases and controls by identity, relative concentration, or correlation and are subsequently further explored. The methodology has been implemented for a wide range of sample types including human

blood, breath, and urine, as well as for in vitro systems investigating chemical changes in cell-lines, tissue biopsies, and bacterial/ fungal cultures (Vorst *et al* 2005, Aura *et al* 2008, Croley *et al* 2012, Kerian *et al* 2015).

Until recently, discrimination among sample groups has relied on targeted compounds, or those that were readily identified by existing methodologies. Data post-processing for such targeted experiments is relatively straightforward, and has for the most part been streamlined with computational tools. The advent of HR-MS, along with extraordinary advances in sensitivity, has resulted in an explosion of available data, and a concomitant burden on researchers in deciphering their meaning. Currently, the newest non-targeted (discovery) analyses require detailed supervision from subject matter experts to provide defensible results. A non-targeted experiment refers to one aiming to observe as many chemicals as possible from a complex sample mixture; in practice complete detection of all chemicals in a mixture is not possible with a single technique. Non-targeted approaches allow for sophisticated data mining and multivariate analysis to tease apart individual compounds associated with sample groups, but appropriate choice of techniques and experimental design is non-trivial. The state-of-the-art technology for interpreting HR-MS data is still under development; instrument manufacturers each have proprietary software/firmware and numerous open-source data analysis packages exist, which further complicates the task of creating a single data analysis workflow.

As HR-MS technologies are becoming more commonplace in the analytical laboratory, we have developed some general guidance for the analyst community as to how to implement compound identification techniques beyond rudimentary library searches of exact match candidates. Specifically, we present a series of examples where uncertainty is reduced for assigning chemical structure and formula by implementing exact mass fragments and isotope ratios. We further suggest how software products could assist in automating aspects of decision-making for identifying unknown chemicals.

High-resolution mass spectrometry: what is 'high-res'?

When comparing mass spectrometry instrumentation, there are several critical parameters of merit. The most central is the mass resolving power of the platform defined as the ability to successfully distinguish two closely separated masses. The International Union of Pure and Applied Chemistry (IUPAC) definition of MS resolution (R) is $R = M/\Delta M$ where M is the measured mass, and ΔM is the separation required to distinguish two peaks at a certain height from the baseline, similar to the definition of peak resolution in chromatography (IUPAC 1997). An instrument might be discussed in terms of the minimum separation required to resolve two equal height mass peaks (e.g. mass resolution of 0.001 Da at 250 Da). IUPAC further defines a single peak measurement methodology, also called mass resolving power, as $M/\Delta M$ where ΔM is the peak width at a specified height (figure 1). Because resolving power is an instrument performance parameter, it is the most commonly quoted value, and is traditionally measured at the full width at half maximum height (FWHM) of the MS peak (figure 1). Mass resolution can be measured at any degree of peak separation but most frequently at 10% of the maximum peak height, equivalent to the full peak width at 5% peak height for an isolated gaussian peak. It is worth noting that in

common usage the terms resolution and resolving power are used interchangeably but should always specify the means of determination and the target mass (e.g. $R_{FWHM} @ m/z 200 = 100\,000$) because resolution/resolving power values are dependent on the mass measured and the height of peak measurement. For the remainder of this manuscript R will refer to FWHM resolving power at the discussed mass.

An example calculation for theoretical peak pairs is shown in figure 1. A peak with a mass of 400.0000 and an observed peak width at half height of 0.5 Da has a single peak resolving power calculated as:

$$R_{FWHM} = \frac{M}{\Delta M} = \frac{400}{0.5} = 800.$$

The resolution of this peak from another peak at 401.0000 at 10% of the chromatographic peak width is 1 Da and has an R value calculated as:

$$R_{10\%} = \frac{m_1}{m_2 - m_1} = \frac{400}{401 - 400} = 400.$$

At a significantly higher resolving power the peak widths decreased substantially (figure 1, right) and the effective mass resolution likewise decreased (0.02 Da for the figure shown).

Given the incremental progress of mass spectrometry instrumentation over the years, there is a constantly moving goalpost for describing when an instrument or spectrum is ‘high-resolution.’ This is further complicated by instrument manufacturers changing target masses for quoted resolution, and intermixed usage of FWHM and 10% valley definitions. Nevertheless, there are broad thresholds of resolution at which increasing amount of chemical information can be gained. (Marshall *et al* 2002, Marshall and Hendrickson 2008) At $R \sim 1000$ nominal masses are distinguishable, (i.e., isotopic peaks of a single spectrum) and at $R \sim 100\,000$ isotopologues differing only in the presence of nominally identical isotopes begin to separate (e.g., $^{13}\text{C}_{12}\text{C}_4$ $^{14}\text{NH}_5$ and $^{12}\text{C}_5$ $^{15}\text{NH}_5$, see figure 2).

For the purposes of this commentary, any measurements capable of resolving the isotopic fingerprint of molecules can apply the compound identification strategies discussed. We also note that the mass described by significant digits beyond integer mass may be referred to as ‘exact mass’ or as ‘accurate mass’. In general, ‘exact mass’ is the fragment mass calculated from a known chemical formula, where as the term ‘accurate mass’ refers to a measurement with high precision; however, they are used interchangeably for the purposes of identifying compounds.

High-resolution mass spectrometry: a brief history

Exact (or accurate) mass spectrometers are not new; they trace their origins to ‘one of a kind’ cyclotron and magnetic sector instruments at major research centers that were used to separate inorganic radio-isotopes and organic molecules (e.g., Beynon 1956, Beynon 1959, Henning *et al* 1981, De Laeter and Kurz 2006). A timeline of the history of exact mass

measurement is available from the archives of the American Society for Mass Spectrometry; the Society attributes the initial achievements of high resolving power mass spectrometry to E O Lawrence and M S Livingston in 1932, and lists some of the major technical advances for Time of Flight (ToF-MS) in 1956, double-magnetic sector geometries from 1957 to 1960, and Fourier transform ion cyclotron resonance from 1965 to 1968 (Grayson 2008).

The history of commercial HR-MS products began in the 1940's. The Chemical Electrostatics Corporation entered the MS commercial market in 1943 with the 21–101 Mass Spectrometer based on magnetic sector technology; it was used to assess petroleum hydrocarbons and had an estimated mass resolution of 0.05 Da at 250 Da ($R \sim 4000$) (Carlson *et al* 1960). The Omegatron, based on vacuum tube technology, was developed by University of Minnesota in 1949 as an MS instrument designed primarily as a residual gas analyzer separating unit Da gases for vacuum applications (Zdanuk *et al* 1960); a patent filed in 1957 indicates that it was improved to achieve resolution to ~ 0.01 Da @ 250 Da ($R \sim 25\,000$) (McNarry and Hobson 1957). Subsequently other instruments entered the commercial HR-MS market including the Bendix ToF-MS (Wiley 1956), the Finnigan MAT Sector in 1978 (Huebschmann 2011), and the Bruker FTICR in 1983, based on research by Mel Comisarow and Alan Marshall (Comisarow and Marshall 1974). These legacy commercial instruments had resolving powers up to 10 000 (~ 1 ppm).

Contemporary HR-MS systems almost always incorporate a high-performance gas or liquid chromatography platform for analyte separation coupled to a TOF or FT detector. Some exceptions are the use of ionization sources (e.g., MALDI, SELDI, APCI, PTR, etc.) without a separation step that are directly coupled to the MS (Byrdwell 2001, Petricoin and Liotta 2003, Jordan *et al* 2009, Gaugg *et al* 2016, Ruhaak *et al* 2016). High-performance TOF instruments routinely generate resolutions of 10 000+, and research platforms exceeding 40 000, with very fast cycle times for fragmentations scans. While cutting edge research on FTICR continues, producing multi-million resolving power instruments (Hendrickson *et al* 2015), an electrostatic trapping FT instrument, the Orbitrap (Makarov 2000), has become the most prevalent FT-MS platform. Orbitrap instruments likewise can perform rapid MS/MS fragmentation with scalable MS and MS/MS resolutions from 10 000 to 100 000+.

The prevalence of these high resolving power instruments, coupled with the power of accurate mass analysis and achievable MS/MS fragmentation duty cycles allows for analysts to achieve an unprecedented degree of information about unknown chemical species.

Interpreting high-resolution features: isotope ratios

The power of HR-MS measurement is the ability to resolve combinations of elements that differ in weights much less than 1.67×10^{-27} kg, or 1 Da. This enables very fine measurements of the exact mass, from which much information can be extracted (Pleil and Isaacs 2016), but also enables elemental composition analysis based on elemental isotopes.

Each signal in a mass spectrometer is the measurement of a single type of ion composed of a linear combination of elemental isotopes. The mass of a molecule composed entirely of the most abundant stable isotope of each constituent element is known as the mono-isotopic

mass (A) and is but one of many masses that can be measured for a molecule. Because isotopes of common elements such as carbon, sulfur, and many halogen species occur at appreciable levels in nature, molecules incorporating one or more higher mass stable isotopes are encountered whenever a molecular sample is measured. The variation in the number of neutrons contained in elemental isotopes creates atomic combinations with masses that differ nominally by a single Da, and these peaks are identified based on the nominal mass shift relative to the monoisotopic mass (i.e. $A - 1$, $A + 1$, $A + 2$ etc.). The theoretical spectrum for a given molecule is thus the combination of all the peaks for all the isotopic combinations of atoms making up that molecule. Although rudimentary assignments of an atom can be made using isotope ratios using single Da resolution, it requires high-resolution to confirm which atoms are actually responsible for the isotopic ratio. For example, figure 3 shows the resolved structure for the two different $A + 2$ possibilities within the per-fluorinated compound $C_{12}H_{12}F_{17}NO_5S$, wherein the sulfur ^{34}S and or two carbon ^{13}C 's could each contribute to the $A + 2$ isotopic peak. Note that with sufficient resolution, one can distinguish the contribution from ^{34}S versus that from $^{13}C_2$, which could not be done with single Da resolution:

$$\begin{aligned} \text{Exact Mass } (^{34}S - ^{32}S) &= 33.967\ 867\ 004 \\ &- 31.972\ 071\ 174 = 1.995\ 795\ 83 \end{aligned}$$

and from $^{13}C_2$:

$$\begin{aligned} \text{Exact Mass } (^{13}C_2 - ^{12}C_2) &= 2 \times (13.003\ 354\ 84 \\ &- 12.000\ 000\ 00) = 2.006\ 709\ 68, \end{aligned}$$

as shown in the inset centered a 607 Da.

The exact mass and relative abundance of atomic isotopes in nature is well characterized (table 1, Berglund and Wieser 2011). This means that calculating a theoretical mass spectrum is a straightforward, but computationally taxing problem of combinatorics, which has many implemented solutions (Valkenborg *et al* 2012). For the purposes of molecular formula generation, it is therefore possible to compare empirical chemical spectra against theoretical distributions for molecular formula generation. This is necessary for the assignment of molecules of middling complexity, because even an excellent <1 ppm mass accuracy is often insufficient to uniquely resolve the majority of chemicals currently known (Kind and Fiehn 2006). Note that the average molecular weight that appears in the standard Periodic Table of the Elements and is familiar to most chemists or biologists, is, in-fact, the weighted average of these many isotopic combinations.

The number of formula matching and scoring algorithms is extensive and constantly evolving, thanks in no small part to entries in the yearly Critical Assessment of Small Molecule Identification (CASMI) contest (<http://casmi-contest.org/>), which specifically invites the development of new formula generation software and techniques. Nevertheless, some of the basics for molecular assignment based on elemental composition are amenable

to manual inspection of isotopic patterns and can be useful in the assignment of molecular formulae. It is worth noting that many of these strategies can be applied even to low resolution full-scan data, but this is difficult in complex samples where isotope peaks are convoluted with other compounds.

Example 1. Carbon Counting

For a typical organic molecule, the major impact on the A+1 abundance is the amount of carbon contained in the molecule. This is because carbon is generally the most prevalent individual atom in an organic structure, and the relative abundance of Carbon-13 (1.07%) is substantially higher than that from ^2H (0.1%), ^{15}N (0.3%), and ^{17}O (0.03%). Under this enormous simplifying assumption, the abundance of the A+1 peak is simply the probability of having only one ^{13}C in the molecule. For example, in a molecule with 8 carbon atoms, this can be approximated by the multi-nomial expansion (Valkenborg *et al* 2012) as having a likelihood of 8.3% and allows rough estimation of the carbon content of a molecule. Figure 4 shows the A+1 ion for four different configurations of organic molecules with 8 carbons. Although the relative prevalence of the A+1 varies slightly due to the other atoms, it maintains a value around 8%, indicating the presence of 8 carbon atoms in each chemical regardless of the other atoms. We note that the exact mass difference of 1.003 35 Da helps confirm that the isotope is ^{13}C , rather than from ^2H 's, ^{15}N 's, or ^{17}O 's.

Example 2. Sulfur Counting

As previously mentioned, the use of isotopic abundances can be of substantial value in narrowing the possible molecular formulae for a given mass. For example, a compound with an accurate mass of 307.083 906 has five possible formula matches within 5 ppm, and two within 1 ppm instrument resolution, when searched against the US Environmental Protection Agency's (EPA) CompTox Chemistry Dashboard (<https://comptox.epa.gov/dashboard>), a publicly available database containing over 760 000 chemicals (Williams *et al* 2017). The A + 1 peaks for these molecules have theoretical values which closely coincide, as is expected given the similar number of carbon atoms in these molecules. The A + 2 peak however, shows significant variation depending on the inclusion of sulfur or chlorine. Sulfur is common in biomolecules and has an A + 2 isotope with abundance ~5%, meaning that for a small molecule the A + 2 peak is similar or higher in abundance to the A+1 peak. Given an empirical spectrum with a relative A+1 and A + 2 abundance of 14% and 6% we could safely conclude this molecule is likely $\text{C}_{10}\text{H}_{17}\text{N}_3\text{O}_6\text{S}$, Glutathione as shown by the red curve in figure 5.

Example 3. Halogen Counting

Several halogens offer very distinct isotopic patterns that are apparent to the naked eye. Both Bromine and Chlorine have major A + 2 isotopes with large abundances at ~25% for ^{37}Cl and ~50% for ^{81}Br . Consequently, an A + 2 peak of substantial magnitude can be representative of a compound containing one or more Cl or Br species. Inclusion of multiple halogen atoms yields complex splitting patterns in the MS spectrum, with spacing every two Daltons (figures 6, 7).

Interpreting high-resolution features: fragmentation patterns

When organic compounds are fragmented, whether in-source by electron ionization (EI) or chemical ionization, or through intentional MS/MS, it is possible to form charged fragments, each with their characteristic isotopic features. Even at low resolution, the patterns of these fragments serve as an additional dimension for identifying compounds, especially when the molecular ion is missing or uncertain. High resolution MS can help distinguish between multiple possibilities for the atomic composition of a particular fragment just as easily as a molecular species. For example, a mass fragment at 85 Da could be either C_6H_{13} or C_5H_9O at 1 Da resolution. However, at high resolution the quandary is resolved, as the accurate masses 85.101 725 and 85.065 339, respectively, are easily separated.

Much like the isotopic mass patterns discussed in the previous section, accurate mass of fragments serves as identifiers of molecular substructures. For larger molecules with many fragments, this allows a plausible, but complicated process to reassemble the original chemical structure. Instead, patterns are most frequently compared against library spectra to generate a list of potential structures and/or substructures from the detected fragments. A standard library search (at unit resolution) generally gives a long list of potential chemical formulae, but high-resolution fragment spectra avoids many assignment issues regardless of the fragmentation method or structural assignment approach.

Example 1. Precursor discrimination in MS/MS

MS analysis with a triple-quadrupole mass spectrometer (QqQ) requires the ionization and selection of a precursor molecule with a single quadrupole, which is passed into a second collision cell and fragmented, before fragments are isolated by a third and final quadrupole section. The process of selecting both precursor and fragment at a known collision energy is intended to be very specific, even with low resolution quadrupoles. Nevertheless, false positives can occur between compounds of similar precursor mass with non-specific fragment transitions.

Consider the following three chemicals, taurodeoxycholic acid (TDCA), perfluorooctanesulfonic acid (PFOS), and 8-(acetyloxy)-1,3,6-pyrenesulfonic acid (1,3,6-PSA) which might coelute in food and fish samples contaminated with fluorinated compounds. (table 2).

Each of the three compounds exhibit a similar precursor mass within a 1 Da isolation window and share a common mass fragment transition at $499 > 80$ due to the production of an SO_3^- ion fragment. The QqQ transition is thus non-specific at low resolution and requires further comparison to reference standards or the inclusion of secondary transitions to ensure the identity of the species and the purity of the transition observed. Using a higher resolution instrument, the transitions are obviously distinct. PFOS and TDCA transitions of $498.9 > 80.0$ and $498.2 > 80.0$ are resolvable at $R \sim 1000$, while PFOS and 1,3,6-PSA require a slightly higher resolution of $R \sim 30\ 000$ to distinguish the respective transitions of $498.9375 > 79.9568$ and $498.9463 > 79.9568$. Note that a typical HR-MS instrument still uses a low resolution quadrupole for precursor isolation, so coeluting species would not have isolated

precursors, but accurate assignment of distinct precursor masses in a precursor scan allows for recognition of false assignments.

Example 2. Substructure Fragments

Distinguishing sub-structural fragments of an unknown compound allows for unequivocal identification of the structure. Consider a compound with the monoisotopic mass 270.0892, assigned empirical formula of $C_{16}H_{14}O_4$, and an EI fragment spectrum as shown in figure 8.

A search of the US EPA CompTox Chemistry Dashboard yields 228 chemicals with this empirical formula, but significant variation in structure that can be investigated by fragmentation. Even at low resolution, the observable 193 fragment corresponds to the loss of a phenyl ring (loss of 77), allowing significant narrowing of the search space. The 139 fragment is more difficult to assign given only a low-resolution spectrum, as it could correspond to the loss of either C_9H_7O (131.0497) or $C_{10}H_{11}$ (131.0861). With high resolution mass measurement, the exact mass differential of 131.0492 is consistent with the C_9H_7O loss. The fragmentation is thus consistent with a chalcone backbone with dihydroxy and methoxy functionalization on a single phenyl ring, such as Cardamonin shown in figure 8. Further details on structural elucidation of complex structures based on MS/MS are beyond the scope of this manuscript, but comparison with reference spectra allows for validation even in the absence of sufficient expertise for *de novo* elucidation.

Interpreting high-resolution features: automated data reduction and high-resolution search algorithms

So far, the identifications schemes have been implemented manually, that is, each analytical feature was investigated individually using the expertise of the researcher with assistance from mass look-up tables and database searches. With the advent of higher sensitivity instrumentation, we are now faced with thousands of features per sample for hundreds or more samples from any given study. As such, the major HR-MS manufacturers, such as Agilent Technologies (Santa Clara, CA, USA), ThermoFisher Scientific (Waltham, MA, USA), SCIEX (Framingham, MA, USA), LECO (Saint Joseph, MI, USA), Shimadzu (Nakagyo-ku, Kyoto, Japan), and Waters (Milford, MA, USA), have all developed their own proprietary software specific for their platforms to automate identifications of unknowns. In addition, academic researchers have been working on more generic algorithms that can receive input from different instrumental data streams. Regardless of the exact implementation, the overarching goal is to employ the subject matter expertise described above in an automated fashion to expedite compound identification.

Briefly, some of the major software packages for feature extraction, compound identification and annotation, as well as statistical analysis and visualization include the following:

- MassHunter Profinder (Agilent Technologies)
- Mass Profiler Professional (Agilent Technologies)
- Unknowns Analysis (Agilent Technologies)
- BioConfirm Software (Agilent Technologies)

- Mass Frontier Spectral Interpretation Software (ThermoFisher Scientific)
- Compound Discoverer Software (small molecule identification) (ThermoFisher Scientific)
- TargetQuan3 Software (identify persistent organic pollutants) (ThermoFisher Scientific)
- ToxFinder (ThermoFisher Scientific)
- PeakView Software (SCIEX)
- MarkerView Software (SCIEX)
- XCMSPlus Software (SCIEX)
- ChromaTOF (LECO)
- LabSolutions Insight (Shimadzu)
- ChromaLynx (Waters)
- MarkerLynx (Waters).

Each vendor has their own preferred software packages for data reduction and analysis, and no two packages provide all of the same features and capabilities. Software selections depend upon the instrumentation being used for analysis as well as the needs of the researcher. Additionally, computer-based spectral fragmentation software methods, including CFMID, MAGMA, and MaxQuant are available to assist with data analysis (Cox and Mann 2008, Allen *et al* 2014). MaxQuant consists of a series of algorithms that can be used for peak detection and quantification as well as identification of fragmentation spectra (Cox and Mann 2008). CFM-ID is a web server that can be used for peak annotation, spectra prediction, and metabolite identification (Allen *et al* 2014). XCMS software is also now available as an online tool that can be used to analyze non-targeted LC/MS metabolomics data through feature detection, retention time correction, and alignment as well as providing a platform for statistical analyses and data visualization tools (Tautenhahn *et al* 2012).

General mass spectral libraries such as the National Institute of Standards and Technology library can be used to search non-targeted mass spectral data. Additional databases, such as the U.S. EPA CompTox Chemistry Dashboard can be implemented into the data processing workflow to help characterize identified compounds of interest. Compounds can be searched by chemical name, CASRN, DSSTox ID, MSready formulae, exact formulae, monoisotopic mass, or InChIKey. A list of tentative compound identifications based on the results of MS data post-processing can be retrieved from the Dashboard. Additional compound information, including presence in lists, number of data sources, National Health and Nutrition Examination Survey predicted exposure, and number of PubMed articles can also be downloaded from this site, as well as specific properties of compounds of interest. Online databases, including the U.S. EPA CompTox Chemistry Dashboard, US EPA DSSTox, and ChemSpider can be used to rank order unknown compounds based on monoisotopic masses and chemical formulae as well as retrieve important chemical identification information, properties, and structural images (Rager *et al* 2016, McEachran *et al* 2017, Sobus *et al* 2017). Metabolomics databases, such as the human metabolome database (<http://hmdb.ca/>),

METLIN (https://metlin.scripps.edu/landing_page.php?pgcontent=mainPage), and the Golm Metabolome Database (<http://gmd.mpimp-golm.mpg.de/>) can be used to search high resolution tandem MS/MS, GC/MS, and LC/MS spectra for known and unknown human metabolites (Rathahao-Paris *et al* 2016).

Summary and recommendations

Accurate (or exact) mass from HR-MS brings an additional dimension to the identification of complex organic molecules. The use of monoisotopic exact mass alone is a great improvement over standard single Da unit resolution, and often narrows compound identity from hundreds of possibilities to a dozen or so. Herein, we discussed the value of going beyond this single monoisotopic focus and incorporated additional knowledge to narrow the possibilities even further.

There are two basic techniques that help confirm the suspected identities of unknown features based on their monoisotopic mass library search:

- count major constituent atoms such as carbon, sulfur, or halogens based on their isotopic abundances and exact mass differences between isotopes;
- identify the various molecular fragments from ionization by their exact mass and gain additional information to reconstruct the original molecule based on its building blocks.

These methods can be implemented on a molecule-by-molecule basis by the researcher and are generally a first approach for initial evaluation of analytical features thought to be of importance. For example, in a case-control evaluation, a handful of features may be differentiated between the groups, and a hands-on identification effort is called for.

However, this is not practical for processing hundreds (possibly thousands) of features from many samples. As described briefly, efficient data reduction requires specialized software. There are many entries in the mass spectrometry software arena that are constantly being updated and improved. The primary focus currently is building reference data used to confirm the calculations based on empirical spectra.

As HR-MS instruments improve in sensitivity and mass accuracy, the combination of exact mass, knowledge of relative isotope abundances and the increasing size of confirmatory databases will continue to improve identification of unknowns in complex biological samples. The primary importance of this process is to document as many subsets of biological exposome data as possible to develop a comprehensive understanding of human systems biology. As we better understand the connections between health-related stressors and the human response through investigations of subtle biochemical perturbations, we will be able to develop interventions to preserve public health.

Acknowledgments

The authors are grateful for expert advice from Mark Strynar, Seth Newton, and Jon Sobus of US EPA. This article was reviewed in accordance with the policies of the National Exposure Research Laboratory, US Environmental Protection Agency, and approved for publication. Mention of trade names or commercial products does not

constitute endorsement or recommendation for use. Theoretical spectra presented throughout the manuscript were prepared using the *enviPat* in R (DOI: [10.1021/acs.analchem.5b00941](https://doi.org/10.1021/acs.analchem.5b00941)) at the presented resolving power.

References

- Ala-Korpela M, Kangas AJ and Soininen P 2012 Quantitative high-throughput metabolomics: a new era in epidemiology and genetics *Genome Medicine* 4 36 [PubMed: 22546737]
- Allen F, Pon A, Wilson M, Greiner R and Wishart D 2014 CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra *Nucleic Acids Research* 42 W94–9 [PubMed: 24895432]
- Andra SS, Austin C, Patel D, Dolios G, Awawda M and Arora M 2017 Trends in the application of high-resolution mass spectrometry for human biomonitoring: an analytical primer to studying the environmental chemical space of the human exposome *Environment International* 100 32–61 [PubMed: 28062070]
- Aura AM, Mattila I, Seppänen-Laakso T, Miettinen J, Oksman-Caldentey KM and Oreši M 2008 Microbial metabolism of catechin stereoisomers by human faecal microbiota: comparison of targeted analysis and a non-targeted metabolomics method *Phytochemistry Letters* 1 18–22
- Berglund M and Wieser M 2011 Isotopic compositions of the elements (IUPAC technical report) *Pure Appl. Chem* 83 397–410
- Beynon JH 1956 The use of the mass spectrometer for the identification of organic compounds *Microchimica Acta*. 44 437–53
- Beynon JH 1959 High resolution mass spectrometry of organic materials *Advances in Mass Spectrometry* 328–54
- Bregy L, Nussbaumer-Ochsner Y, Sinues PM, García-Gómez D, Suter Y, Gaisl T, Stebler N, Gaugg MT, Kohler M and Zenobi R 2018 Real-time mass spectrometric identification of metabolites characteristic of chronic obstructive pulmonary disease in exhaled breath *Clinical Mass Spectrometry* 7 29–35
- Byrdwell WC 2001 Atmospheric pressure chemical ionization mass spectrometry for analysis of lipids *Lipids* 36 327–46 [PubMed: 11383683]
- Carlson EG, Paulissen GT, Hunt RH and O’Neal MJ 1960 High resolution mass spectrometry. Interpretation of spectra of petroleum fractions *Anal. Chem* 32 1489–94
- Comisarow MB and Marshall AG 1974 Fourier transform ion cyclotron resonance spectroscopy *Chem. Phys. Lett* 25 282–3
- Cox J and Mann M 2008 MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification *Nat. Biotechnol* 6 1367
- Croley TR, White KD, Callahan JH and Musser SM 2012 The chromatographic role in high resolution mass spectrometry for non-targeted analysis *J. Am. Soc. Mass Spectrom* 23 1569–78 [PubMed: 22711514]
- De Laeter J and Kurz MD 2006 Alfred Nier and the sector field mass spectrometer *Journal of Mass Spectrometry* 41 847–54 [PubMed: 16810642]
- Gaugg MT, Gomez DG, Barrios-Collado C, Vidal-de-Miguel G, Kohler M, Zenobi R and Sinues PM 2016 Expanding metabolite coverage of real-time breath analysis by coupling a universal secondary electrospray ionization source and high resolution mass spectrometry—a pilot study on tobacco smokers *J. Breath Res* 10 016010 [PubMed: 26866308]
- Ghio A, Madden MC and Esther CR 2017 Transition and post-transition metals in exhaled breath condensate *J. Breath Res* 12 027112
- Grayson MA 2008 In search of accurate mass: the unending quest for higher resolving power, Media. American Society of Mass Spectrometry archives: St. Louis, MO ASMS (<https://asms.org/docs/history-posters/in-search-of-accurate-massdenver-2008.pdf?sfvrsn=2>)
- Hendrickson CL, Quinn JP, Kaiser NK, Smith DF, Blakney GT, Chen T, Marshall AG, Weisbrod CR and Beu SC 2015 21 tesla fourier transform ion cyclotron resonance mass spectrometer: a national resource for ultrahigh resolution mass analysis *J. Am. Soc. Mass Spectrom* 26 1626–32 [PubMed: 26091892]

- Henning W, Kutschera W, Paul M, Smither RK, Stephenson EJ and Yntema JL 1981 Accelerator mass spectrometry and radioisotope detection at the Argonne FN tandem facility Nucl. Instrum. Methods 184 247–68
- Herbig J, Müller M, Schallhart S, Titzmann T, Graus M and Hansel A 2009 On-line breath analysis with PTR-TOF J. Breath Res 3 027004 [PubMed: 21383459]
- Huebschmann H-J 2011 A Brief History of Thermo Fisher (High Resolution) Mass Spectrometry in Bremen. Media (Austin, TX: ThermoFisher Scientific) pp 1–40 (http://apps.thermoscientific.com/media/SID/IOMS/PDF/niagara2011/1_Huebschmann_History_of_MS_in_Bremen.pdf)
- IUPAC 1997 Compendium of Chemical Terminology 2nd edn (the 'Gold Book'), Compiled by ed McNaught AD and Wilkinson A (Oxford: Blackwell Scientific Publications)
- Jordan A, Haidacher S, Hanel G, Hartungen E, Märk L, Seehauser H, Schotchkowsky R, Sulzer P and Märk TD 2009 A high resolution and high sensitivity proton-transfer-reaction time-of-flight mass spectrometer (PTR-TOF-MS) International Journal of Mass Spectrometry 286 122–8
- Kerian KS, Jarmusch AK, Pirro V, Koch MO, Masterson TA, Cheng L and Cooks RG 2015 Differentiation of prostate cancer from normal tissue in radical prostatectomy specimens by desorption electrospray ionization and touch spray ionization mass spectrometry Analyst 140 1090–8 [PubMed: 25521825]
- Kind T and Fiehn O 2006 Metabolomic database annotations via query of elemental compositions: mass accuracy is insufficient even at less than 1 ppm BMC Bioinform. 7 234
- Krewski D et al. 2010 Toxicity testing in the 21st century: a vision and a strategy Journal of Toxicology and Environmental Health, Part B 13 51–138 [PubMed: 20574894]
- Ladva CN. et al. 2000; Metabolomic profiles of plasma, exhaled breath condensate, and saliva are correlated with potential for air toxics detection. J. Breath Res. 12:016008.
- Li X, Huang L, Zhu H and Zhou Z 2017 Direct human breath analysis by secondary nano-electrospray ionization ultrahigh-resolution mass spectrometry: Importance of high mass resolution and mass accuracy Rapid Commun. Mass Spectrom 31 301–8 [PubMed: 27859758]
- Makarov A 2000 Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis Anal. Chem 72 1156–62 [PubMed: 10740853]
- Marshall AG, Hendrickson CL and Shi SD-H 2002 Peer reviewed: scaling ms plateaus with high-resolution FT-ICRMS Anal. Chem 74 A–25259
- Marshall AG and Hendrickson CL 2008 High-resolution mass spectrometers Annu. Rev. Anal. Chem 1 579–99
- McEachran AD, Sobus JR and Williams AJ 2017 Identifying known unknowns using the US EPA's comptox chemistry dashboard Analytical and Bioanalytical Chemistry 409 1729–35 [PubMed: 27987027]
- McLafferty FW, Stauffer DA, Loh SY and Wesdemiotis C 1999 Unknown identification using reference mass spectra. Quality evaluation of databases J. Am. Soc. Mass Spectrom 10 1229–40 [PubMed: 10584326]
- McNarry LR and Hobson JP 1957 Omegatron with orbit increment detection Patent National Research Council, Ottawa, Ontario, Canada Filed 5 7, 1957, Ser. No. 657,616 4 Claims. (CI. 250–419).US Patent,
- Nizio KD, Perrault KA, Troobnikoff AN, Ueland M, Shoma S, Iredell JR, Middleton PG and Forbes SL 2016 In vitro volatile organic compound profiling using GC × GC-TOFMS to differentiate bacteria associated with lung infections: a proof-of-concept study J. Breath Res 10 026008 [PubMed: 27120170]
- Peralbo-Molina A, Calderón-Santiago M, Priego-Capote F, Jurado-Gámez B and de Castro ML 2016 Metabolomics analysis of exhaled breath condensate for discrimination between lung cancer patients and risk factor individuals J. Breath Res 10 016011 [PubMed: 26866403]
- Petricoin EF and Liotta LA 2003 Mass spectrometry-based diagnostics: the upcoming revolution in disease detection Clinical Chemistry 49 533–4 [PubMed: 12651801]
- Pleil JD and Isaacs KK 2016 High-resolution mass spectrometry: basic principles for using exact mass and mass defect for discovery analysis of organic molecules in blood, breath, urine and environmental media J. Breath Res 10 012001 [PubMed: 26954610]

- Rager JE et al. 2016 Linking high resolution mass spectrometry data with exposure and toxicity forecasts to advance high-throughput environmental monitoring *Environment International* 88 269–80 [PubMed: 26812473]
- Rathahao-Paris E, Alves S, Junot C and Tabet JC 2016 High resolution mass spectrometry for structural identification of metabolites in metabolomics *Metabolomics* 12 10
- Ruhaak LR, van der Burgt YEM and Cobbaert CM 2016 Prospective applications of ultrahigh resolution proteomics in clinical mass spectrometry *Expert Review of Proteomics* 13 1063–71 [PubMed: 27798968]
- Sauvain JJ, Suarez G, Edmé JL, Bezerra OM, Silveira KG, Amaral LS, Carneiro AP, Chérot-Kornobis N, Sobaszek A and Hulo S 2017 Method validation of nanoparticle tracking analysis to measure pulmonary nanoparticle content: the size distribution in exhaled breath condensate depends on occupational exposure *J. Breath Res* 11 016010 [PubMed: 28054515]
- Schymanski EL et al. 2015 Non-target screening with high-resolution mass spectrometry: critical review using a collaborative trial on water analysis *Anal Bioanal Chem* 407 6237–55 [PubMed: 25976391]
- Singh KD, del Miguel GV, Gaugg MT, Ibañez AJ, Zenobi R, Kohler M, Frey U and Sinues PM 2018 Translating secondary electrospray ionization–high-resolution mass spectrometry to the clinical environment *J. Breath Res* 12 027113 [PubMed: 29411710]
- Sobus JR et al. 2018 Integrating tools for non-targeted analysis research and chemical safety evaluations at the US EPA *Journal of Exposure Science & Environmental Epidemiology* 28 411–26 [PubMed: 29288256]
- Sukul P, Trefz P, Kamysek S, Schubert JK and Miekisch W 2015 Instant effects of changing body positions on compositions of exhaled breath *J. Breath Res* 9 047105 [PubMed: 26582820]
- Tautenhahn R, Patti GJ, Rinehart D and Siuzdak G 2012 XCMS Online: a web-based platform to process untargeted metabolomic data *Anal. Chem* 84 5035–9 [PubMed: 22533540]
- Teeguarden JG et al. 2016 Completing the link between exposure science and toxicology for improved environmental health decision making: The aggregate exposure pathway framework *Environ. Sci. Technol* 50 4579–86 [PubMed: 26759916]
- Valkenburg D, Mertens I, Lemièrre F, Witters E and Burzykowski T 2012 The isotopic distribution conundrum *Mass Spectrometry Reviews* 31 96–109 [PubMed: 21590704]
- Vineis P et al. 2017 The exposome in practice: design of the EXPOsOMICS project *International Journal of Hygiene and Environmental Health* 220 142–51
- Vorst O, De Vos CH, Lommen A, Staps RV, Visser RG, Bino RJ and Hall RD 2005 A non-directed approach to the differential analysis of multiple LC–MS-derived metabolic profiles *Metabolomics* 1 169–80
- Wallace MA and Pleil JD 2018a Evolution of clinical and environmental health applications of exhaled breath research: review of methods and instrumentation for gas-phase, condensate, and aerosols *Analytica Chimica Acta* 1024 18–38 [PubMed: 29776545]
- Wallace MA and Pleil JD 2018b Dataset of breath research manuscripts curated using PubMed search strings from 1995–2016 *Data in Brief* 18 1711–24 [PubMed: 29904671]
- Wiley WC 1956 Bendix time-of-flight mass spectrometer *Science* 124 817–20 [PubMed: 17774254]
- Williams AJ. et al. 2017; *Journal of Cheminformatics*. 9:61. [PubMed: 29185060]
- Winters BR, Pleil JD, Angrish MM, Stiegel MA, Risby TH and Madden MC 2017 Standardization of the collection of exhaled breath condensate and exhaled breath aerosol using a feedback regulated sampling device *J. Breath Res* 11 047107 [PubMed: 28894051]
- Zamuruyev KO. et al. 2016; Human breath metabolomics using an optimized non-invasive exhaled breath condensate sampler. *J. Breath Res*. 11:016001. [PubMed: 28004639]
- Zdanuk EJ, Bierig R, Rubin LG and Wolsky SP 1960 An omegatron spectrometer, its characteristics and application *Vacuum* 10 382–9

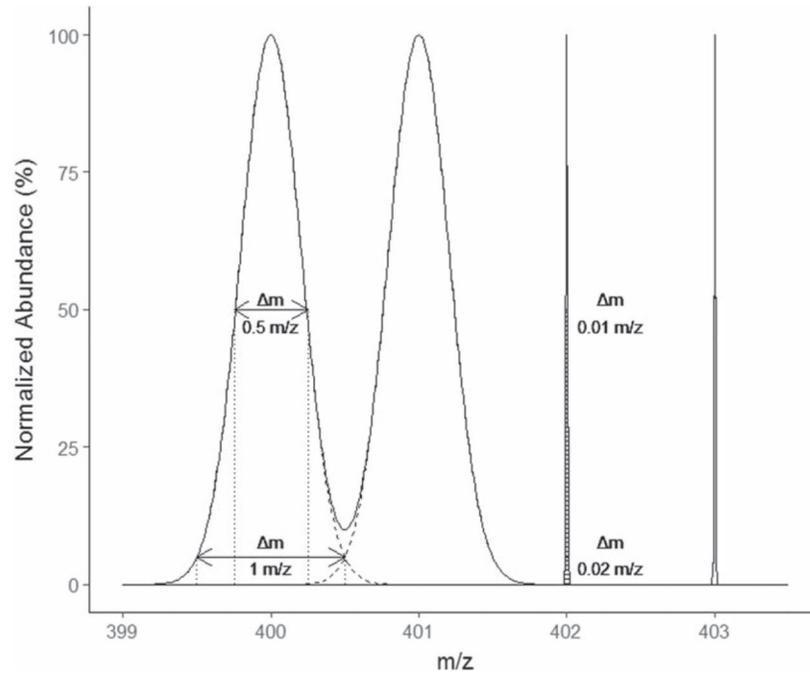


Figure 1. Mass resolution definitions. Simulated FWHM and 10% valley Δm measurements for a pair of isotopes at $R_{FWHM} \sim 1000$ (left) and $R_{FWHM} \sim 40\,000$ (right).

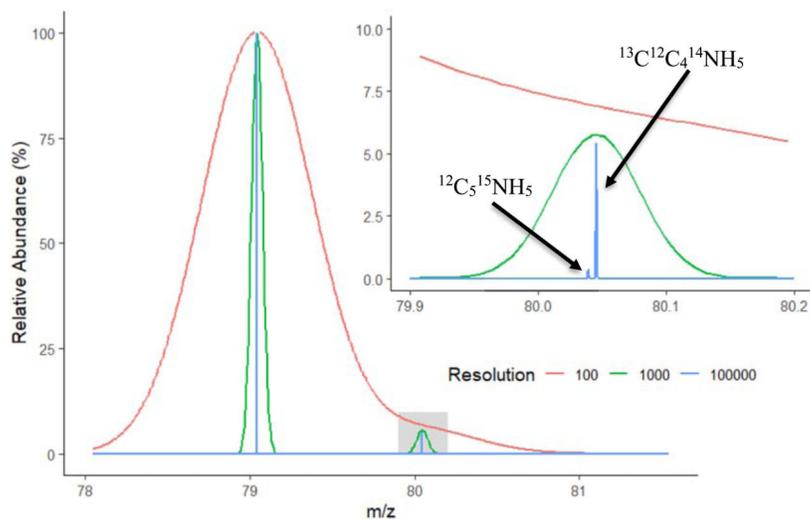


Figure 2. Theoretical MS spectrum of Pyridine (C_5H_5N) at increasing resolving power. At sufficiently high resolving power the mass peaks corresponding to $^{13}C^{12}C_4^{14}NH_5$ and $^{12}C_5^{15}NH_5$ can be distinguished.

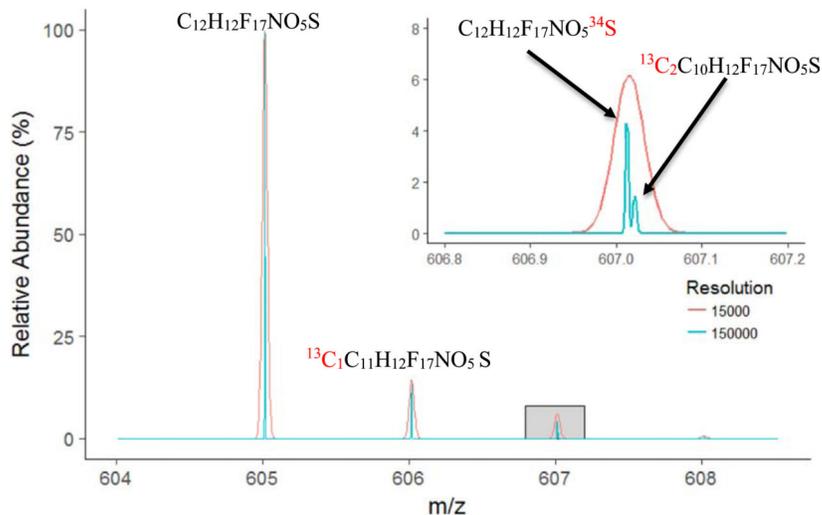


Figure 3. Theoretical High (R = 15 000) and Ultra-High (R = 150 000) resolution spectra of molecular formula $C_{12}H_{12}F_{17}NO_5S$. The additional information from exact masses of different isotopic constituents further confirms the monoisotopic identification of the original compound by showing the shift from the ^{34}S versus the two ^{13}C 's isotopes in the 607 Da centered peak (inset). Note: The peak labeled $^{13}C_2C_{10}H_{12}F_{17}NO_5S$ is not pure and contains further contributions that cannot be resolved under the conditions indicated.

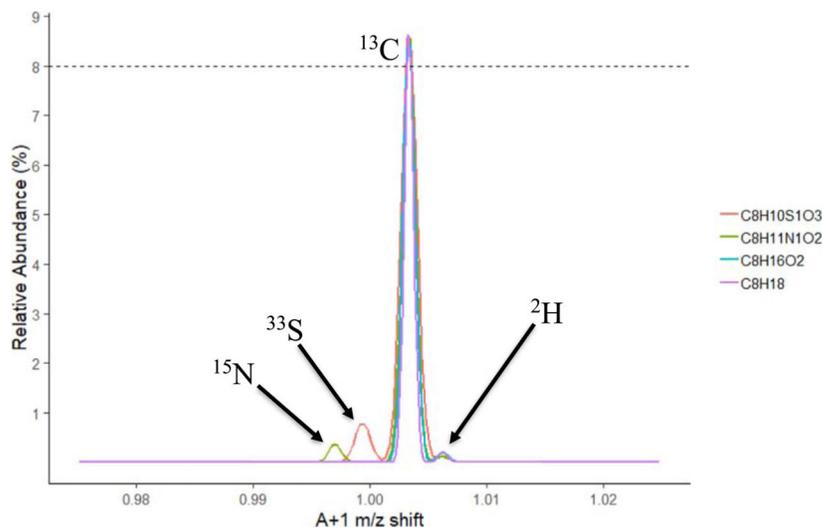


Figure 4.

Theoretical shift of the A+1 peaks for four organic molecules compared to the calculated value for a molecule containing only carbon and hydrogen. In each case, the spacing between the monoisotopic mass (relative abundance 100%) and the ^{13}C A+1 peak shown is 1.003 35 Da, and the relative abundance at ~8% indicates eight carbons. Fine isotope structure from ^2H , 1.006 28 Da, from ^{15}N , 0.997 03 Da, and from ^{33}S , 0.999 387 Da, can be observed with sufficient resolving power.

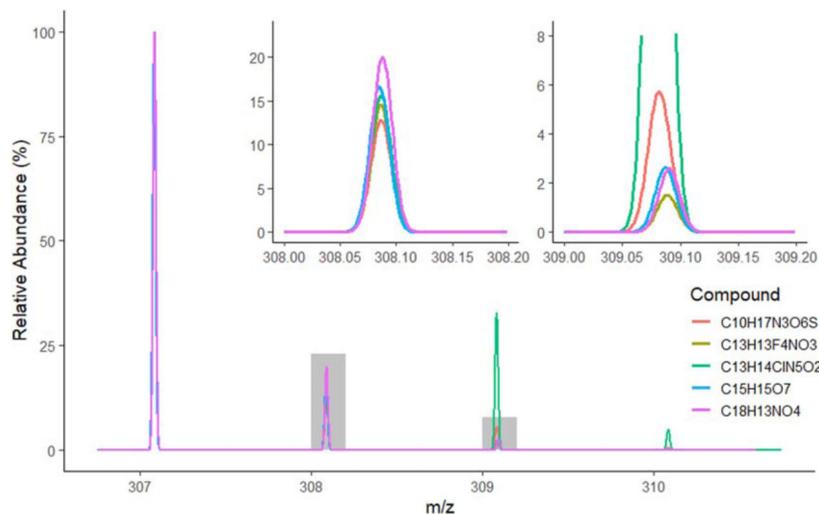


Figure 5.

Theoretical mass spectral patterns ($R \sim 15\,000$) for five molecular formulae with exact mass 307.083 906 (within 5ppm instrument resolution). Insets show the A+1 and A + 2 isotope peaks. In the second extracted panel, the relative abundance of the A + 2 peaks for the five candidate compounds range from 2% to 30%. The 30% value indicates chlorine, and the entries below 3% eliminate sulfur (at 4.25%). Given that the true empirical spectrum has ~6% A + 2, and that sulfur and oxygen have A + 2 abundances of 4.25% and 0.25%, respectively, the relative A + 2 abundance for (O6 + S) should match 5.75%. Furthermore, the spacing between sulfur A and A + 2 is 1.995 795 83 whereas the shift for oxygen is 2.004 244 99, so the slight shift of the red trace confirms the sulfur atom, and the confirmed candidate is C₁₀H₁₇N₃O₆S, Glutathione.

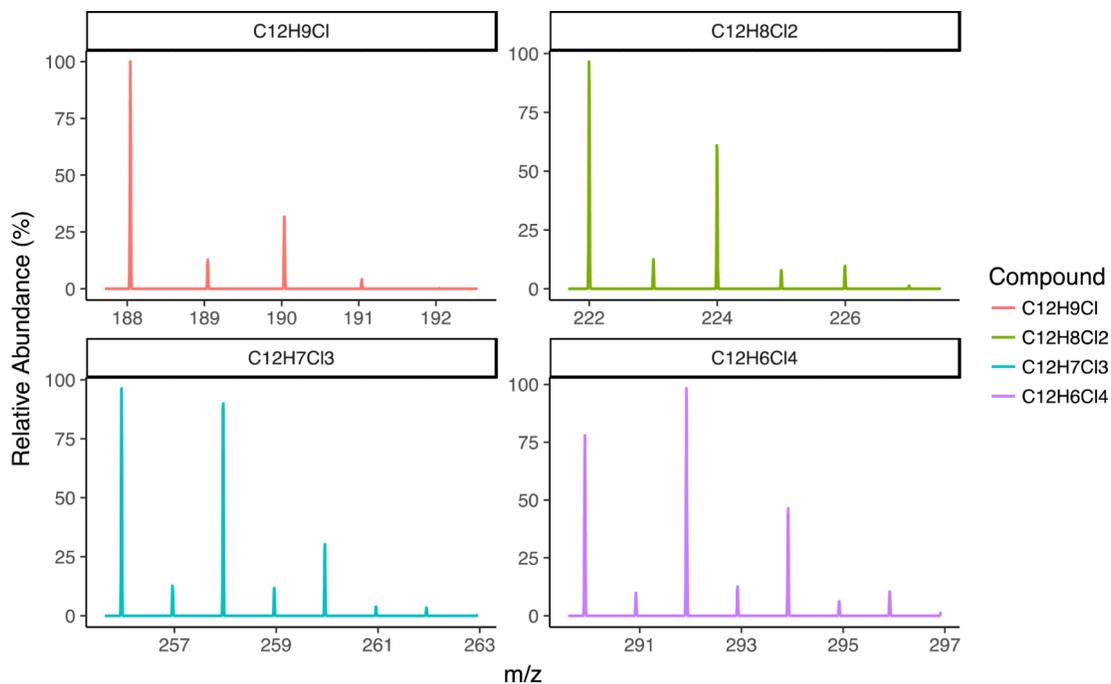


Figure 6.

Using isotopic abundance patterns to count the number of chlorine atoms in a C₁₂ compound: at each individual mass fragment, exact mass can be further investigated to confirm the chemical formula. Chlorine has an abundance of A + 2 isotopes of ~25%, so a spacing at 1.997 375 78 Da with relative abundance of ~25% indicates 1 chlorine, 63% indicates 2 chlorines, 100% indicates 3 chlorines, etc. The spacing of A+1 at 1.003 354 84 Da, and the relative abundance at ~12% indicates 12 carbons.

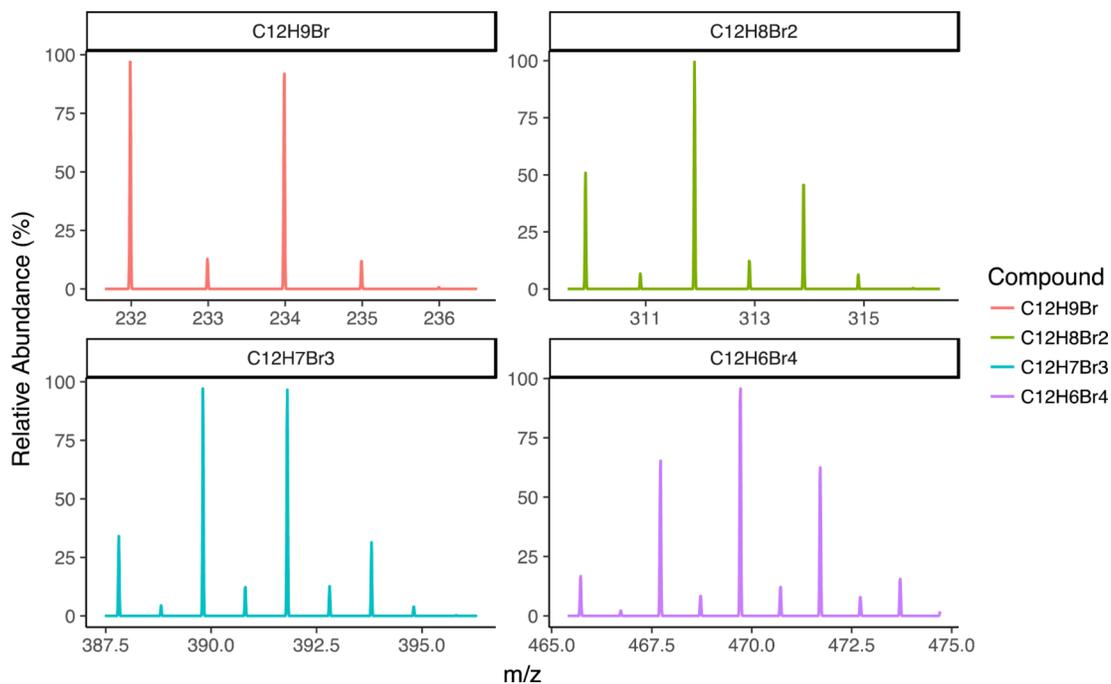


Figure 7.

Similarly to figure 6, isotopic abundance patterns are used to count the number of bromine atoms in a C_{12} compound; at each monoisotopic mass fragment, exact mass can be further investigated to confirm the chemical formula. Bromine has an abundance of $A + 2$ isotopes of $\sim 50\%$ with a spacing of 1.997 9521 Da; as such abundances of $A/(A + 2) = 1$ indicates 1 bromine, and further ratios of $A/(A + 2) = 2$ indicates 2 bromines, $A/(A + 2) = 3$ indicates 3 bromines, etc.

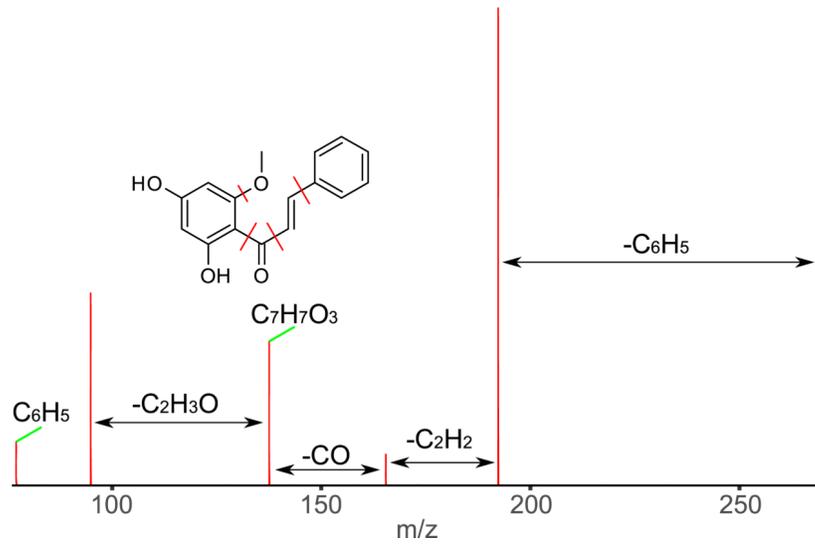


Figure 8. Theoretical EI spectrum of Cardamonin, with chemical formula $C_{16}H_{14}O_4$ and monoisotopic mass 270.0892 indicating major fragments and losses.

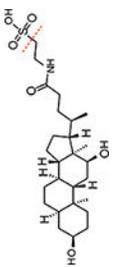
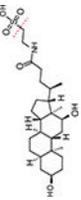
Table 1.

Relative abundance values for stable isotopes of common organic elements. Derived from Isotopic compositions of the elements 2009 (Berglund and Weiser 2011).

Element	Isotope	Exact Mass (Da)	Composition Fraction
Hydrogen			
¹ H	A	1.007 825 032	0.998 85
² H	A+1	2.014 101 778	0.001 15
Carbon			
¹² C	A	12.000 000 00	0.9893
¹³ C	A+1	13.003 354 84	0.0107
Nitrogen			
¹⁴ N	A	14.003 074 00	0.996 36
¹⁵ N	A+1	15.000 108 90	0.003 64
Oxygen			
¹⁶ O	A	15.994 914 62	0.997 57
¹⁷ O	A+1	16.999 131 76	0.000 38
¹⁸ O	A+2	17.999 159 61	0.002 05
Fluorine			
¹⁹ F	A	18.998 403 162	1
Phosphorus			
³¹ P	A	30.973 761 998	1
Sulfur			
³² S	A	31.972 071 174	0.9499
³³ S	A+1	32.971 458 910	0.0075
³⁴ S	A+2	33.967 867 004	0.0425
Chlorine			
³⁵ Cl	A	34.968 526 82	0.7576
³⁷ Cl	A+2	36.965 902 60	0.2424
Bromine			
⁷⁹ Br	A	78.918 3376	0.5069
⁸¹ Br	A+2	80.916 2897	0.4931

Table 2.

Three organic molecules demonstrating low resolution MS/MS transitions of 499 > 80 from the loss of sulfonate.

Name	Formula	Exact Mass	Structure
TDCA	C ₂₆ H ₄₅ N ₂ O ₆ S	499.29675	
PFOS	C ₈ H ₁₇ O ₃ S	499.93749	
8-(acetyloxy)-1,3,6-PSA	C ₁₈ H ₁₂ O ₁₁ S ₃	499.95417	