



MALDI-8030

Benchtop MALD-TOF
Mass Spectrometer

Affordable, Class-Leading Workhorse

Shimadzu's dual-polarity, benchtop linear **MALDI-8030 mass spectrometer** delivers outstanding performance in a user-friendly, compact design, making it an ideal choice for today's increasingly demanding laboratories.

- Dual mode (positive/negative ion) MALDI-TOF
- 200 Hz solid-state laser, 355 nm
- Fast sample introduction
- Integrated barcode reader
- Low maintenance and quiet operation

Tailored MALDI Solutions software features data security for regulatory compliance facilitated with password-controlled user login, encrypted data and user traceability.

Learn more about Shimadzu's MALDI-8030.

Visit us online at www.ssi.shimadzu.com

Shimadzu Scientific Instruments Inc., 7102 Riverwood Dr., Columbia, MD 21046, USA



Rapid Commun. Mass Spectrom. 2016, 30, 2561–2567
(wileyonlinelibrary.com) DOI: 10.1002/rcm.7748

A computer program to simplify analysis of mass scan data of organometallic compounds from high-resolution mass spectrometers

Michael C. Doran^{1*} and Kelly L. LeBlanc²

¹Water Quality Centre, Trent University, Peterborough, ON, Canada, K9L 0G2

²Environmental and Life Sciences Graduate Program, Trent University, Peterborough, ON, Canada, K9L 0G2

RATIONALE: Software accompanying high-resolution mass spectrometers, particularly that used for the analysis of organometallic compounds, has lagged the technology of the instruments themselves. We have developed a computer program that partially fills this gap.

METHODS: Given the user's expectation for the number of atoms of a target element likely to be in an ion, the program calculates isotopologue mass differences for combinations of that element's isotopes and their expected intensity ratios relative to the most abundant isotopologue. These values are compared with mass differences and intensity ratios found in the experimental mass scan data and these metrics feed into a four-factor scoring model which ranks the ions as to the likelihood of each containing the specified number of the target atoms. The program was tested using experimental data obtained for selenomethionine.

RESULTS: Across a broad range of sample concentrations, the program consistently ranked selenomethionine at or near the top of the list of ions that passed the screening and ranking process. Mass scan data files in excess of 24,000 records were analyzed in less than one second.

CONCLUSIONS: The program is quick and efficient at scanning voluminous experimental data files for the presence of ions containing the expected number of atoms of a target element. Best results were obtained the scarcer the target element and the more isotopes it comprised. Copyright © 2016 John Wiley & Sons, Ltd.

Modern Fourier transform mass spectrometry (FTMS) instruments, including ion cyclotron resonance and Orbitrap instruments, allow mass measurement accuracy that has formerly been unavailable with older unit resolution quadrupole instruments. Mass measurements accurate to a few parts per million (ppm) allow an intelligent guess to be made of the molecular formula of unknown molecules, especially at lower masses, but these instruments produce a torrent of data and are not always equipped with software to process the data adequately. Particularly problematic is the analysis of low concentrations of organometallic analytes within complex matrices such as many types of environmental samples. This need has been partially satisfied with the development of a program that winnows mass scan data obtained from these instruments to search for patterns among masses of isotopologues of an organometallic compound with the simple goal of determining if ions possessing a particular isotopic makeup are present in a sample.

Our survey of analytical mass spectrometry-related software found many easily and often freely available programs but most were skewed toward proteomics analysis. A few, such as mMass,^[1] multiMS-toolbox^[2] and Universal

Mass Calculator,^[3] claim to perform calculations similar to those presented here, such as isotopologue mass, mass differences and abundances, but none is capable of searching output data sets for isotopic mass differences specific to organometallic compounds.

EXPERIMENTAL

Reagents

Selenomethionine (C₅H₁₁NO₂Se, >99% as D,L-selenomethionine) was obtained from Sigma Aldrich (Oakville, Ontario, Canada). Samples were diluted from an aqueous 1000 ppm stock standard to 1, 10 and 50 ppb and 1, 25 and 100 ppm, in a matrix of 50% methanol (HPLC grade, Sigma Aldrich) and 0.5 mM NaOH (from 50% w/w stock, certified, Thermo Fisher Scientific, Mississauga, Ontario, Canada) in Milli-Q water (Millipore, Etobicoke, Ontario, Canada).

Instrumentation

Samples were analyzed using a Bruker solarix XR FTICR (Fourier transform ion cyclotron resonance) mass spectrometer (Solarix), equipped with a 7 T magnet, following direct infusion at 3 μL/min into an electrospray ionization (ESI) source with a capillary voltage of 6000 V and a nebulizer gas pressure of 1 bar. For all samples, the accumulation time

* Correspondence to: Michael C. Doran, Water Quality Centre, Trent University, Peterborough, ON, Canada K9L 0G2.
E-mail: michael.doran@trentu.ca

was optimized such that approximately 1×10^9 ions remained in the ICR cell at all times. The free induction decay was 1.0486 s for the scan range m/z 53.76–1000 and external mass calibration resulted in mass errors of less than 1 ppm throughout this range. Twenty-five scans were acquired for each sample.

Description of the program

The program, Winnow, written in C^[4] and compiled using the Open Watcom 32-bit C compiler,^[5] runs as a standalone executable in the Windows command shell, taking user-specified input parameters from the keyboard or from a user-written batch file. All testing of the program has been done in the Windows 7 operating system. The program has successfully analyzed data for up to five atoms of the target element from both the Thermo Scientific Q Exactive (Orbitrap) and the Solarix instruments. The former, a hybrid quadrupole-orbitrap instrument, is capable of producing mass scan data with masses accurate in the range of four decimal places and the latter, an FTICR instrument, in the range of five decimal places. Both instruments' software is capable of producing mass scan data as a text list, sorted in ascending order by mass. Although other data are available, the Winnow program uses only the mass and intensity components of the data. In principle, any element that has more than one isotope could be analyzed but to date the program has been used primarily for the analysis of certain organometallic compounds containing metals such as Se, Ru, Mo and U.

In the command shell of Windows, the user inputs the program name (Winnow), the name of the input data file and the analysis mode (1 for positive ESI; -1 for negative). The ESI mode is used to calculate the starting mass for the search and neutral monoisotopic masses (of interest mostly when organic molecules are being analyzed) but is not required for the type of calculations performed for the search. Also input on the command line are the atomic symbol of the element of interest, the number of atoms of that particular element that are expected to be present in the target molecule and the error on mass accuracy in ppm (obtained from the external calibration data of the instrument, e.g. for the Solarix a typical number might be 0.7 ppm). The final input parameter is the user-defined output file name. The output file is written in comma separated values (.CSV) format for easy entry into Microsoft Excel. The number of decimal places in the data is deduced by the program from the input data; for the Orbitrap it is four and for the Solarix it is five.

Calculation of isotopologue masses

In general, for some N combination of the p isotopes of an element and for some set of integer proportions of those isotopes, the column vector of the number of each isotope in the isotopologue, $S = (n_1, n_2, \dots, n_p)$, can be written, where $\sum_{k=1}^p n_k = N$; $n_k \geq 0$. Writing the corresponding column vector of isotopic masses as $M = (m_1, m_2, \dots, m_p)$, the total mass of the isotopologue equals $S^T M$.

Calculation of isotopologue abundances

To illustrate, the expected abundance of the isotopologue $^{80}\text{Se}^a \text{Se}^b$ is the joint probability of obtaining a atoms of ^{80}Se on a successive random draws from a natural population of

selenium atoms, followed by b atoms of ^{78}Se on b successive random draws from a natural population of selenium atoms. Letting A represent the natural abundance of an isotope, the probability P of forming one permutation of this isotopologue is therefore:

$$P = A(^{80}\text{Se})^a \cdot A(^{78}\text{Se})^b \quad (1)$$

The number of ways isotopes may be combined in an isotopologue depends on the number of atoms of a given isotope that are present in the isotopologue (n) and on the number of isotopes of that element from which an isotope can be chosen (x). In general, the number of possible combinations, C , of the different isotopes of an element within a molecule, without regard to their uniqueness, is:

$$C = x^n \quad (2)$$

In the case of selenium, there are six naturally occurring isotopes, and the number of possible combinations of the six isotopes equals 6^n , some combinations being redundant.

For n atoms chosen from among x isotopes, the total number of choices, excluding redundancies, is:

$$\binom{x+n-1}{n} \quad (3)$$

In the case of selenium, with its six isotopes, Eqn. (3) shows that molecules containing two atoms of that element can form 21 selenium isotopologues.

Pascal's triangle can be used to account for the redundant combinations of isotopes when only two isotopes occur naturally (e.g. carbon). A more complicated formula must be used for elements such as selenium that have more than two isotopes in order to account for the redundancies. In the case of an element such as selenium with its six isotopes, the form of the Pascal coefficient, P , is:

$$P = \binom{n}{m} \binom{m}{l} \binom{l}{k} \binom{k}{j} \binom{j}{i} \quad (4)$$

For example, for $^{82}\text{Se}_a \text{Se}_b^{80}\text{Se}_c \text{Se}_d^{78}\text{Se}_e \text{Se}_f^{76}\text{Se}_g \text{Se}_h^{74}\text{Se}_i$, where $a + b + c + d + e + f = n$, if $b = 1$, $c = 1$ and $a, d, e, f = 0$, then the abundance of the $^{80}\text{Se}^a \text{Se}^b$ isotopologue is ($n = 2$, $m = n - a$, $l = m - b$, $k = l - c$, $j = k - d$, $i = l - e$):

$$\begin{aligned} & \binom{2}{0} \binom{2}{1} \binom{1}{1} \binom{0}{0} \binom{0}{0} \cdot 0.4961 \cdot 0.2377 \\ &= 2 \cdot 0.117923 \\ &= 0.235846 \end{aligned}$$

In the case of two atoms of selenium, a similar calculation is done for each of the remaining 20 isotopologues.

Given the element and the number of atoms of that element that are expected to be in the target molecule, all of the possible isotopologues involving isotopes of that element are assembled and their respective masses and natural abundances are calculated, as described above.^[6] Typical results of these calculations are demonstrated in Table 1 for Se_2 . For a single atom of the element of interest, the program searches for simple *isotope* mass differences, but we will use

Table 1. Di-selenium isotopologue masses and abundances

Isotopologue (lightest to heaviest)	Neutral mass (Da)	Abundance (most abundant in bold)	Mass difference vs. the most abundant (Da)
$^{74}\text{Se}_2$	147.84495	0.00008	-11.98809
$^{76}\text{Se}^{74}\text{Se}$	149.84169	0.00167	-9.99135
$^{77}\text{Se}^{74}\text{Se}$	150.84239	0.00136	-8.99065
$^{76}\text{Se}_2$	151.83843	0.00878	-7.99461
$^{78}\text{Se}^{74}\text{Se}$	151.83979	0.00423	-7.99325
$^{77}\text{Se}^{76}\text{Se}$	152.83913	0.01430	-6.99391
$^{78}\text{Se}^{76}\text{Se}$	153.83652	0.04455	-5.99652
$^{80}\text{Se}^{74}\text{Se}$	153.83900	0.00883	-5.99404
$^{77}\text{Se}_2$	153.83983	0.00582	-5.99321
$^{78}\text{Se}^{77}\text{Se}$	154.83722	0.03627	-4.99582
$^{78}\text{Se}_2$	155.83462	0.05650	-3.99842
$^{80}\text{Se}^{76}\text{Se}$	155.83574	0.09297	-3.99730
$^{82}\text{Se}^{74}\text{Se}$	155.83918	0.00155	-3.99386
$^{80}\text{Se}^{77}\text{Se}$	156.83644	0.07570	-2.99660
$^{80}\text{Se}^{78}\text{Se}$	157.83383	0.23585	-1.99921
$^{82}\text{Se}^{76}\text{Se}$	157.83591	0.01636	-1.99713
$^{82}\text{Se}^{77}\text{Se}$	158.83662	0.01332	-0.99642
$^{80}\text{Se}_2$	159.83304	0.24612	0.00000
$^{82}\text{Se}^{78}\text{Se}$	159.83401	0.04150	0.00097
$^{82}\text{Se}^{80}\text{Se}$	161.83322	0.08662	2.00018
$^{82}\text{Se}_2$	163.83340	0.00762	4.00036

the term *isotopologue* to generalize, whether one or more atoms are involved. In a typical application, the most abundant isotopologue is determined and denoted the 'key isotopologue' – many of the subsequent calculations will be done relative to this key isotopologue. The mass difference of each isotopologue relative to the key isotopologue is calculated and these values drive the search process. The reason for choosing the most abundant isotopologue as key is that, at low intensity, background signals or noise could be mistaken for isotopologue signals; measurement of an isotopic mass offset to a spurious signal would most likely only reveal another spurious signal. We postulate that the most abundant isotopologue should produce the most intense signal and provide the most reliable isotopologue mass offset data. The abundance of the key isotopologue is used as the denominator of the ratio with each of the other isotopologues' abundance, resulting in an intensity ratio vector. The expected intensity ratios in the intensity ratio vector will later be compared with those calculated analogously for qualifying masses found by the search algorithm and will be used in the calculation of the intensity ratios scoring parameter. An important assumption throughout is that all of the observed ions are singly charged.

It is probably easiest to illustrate the operation of the program with an example of a dataset from the Solarix instrument. In this example, the search was for occurrences of ions in a negative ESI mass scan that contained two selenium atoms. With no foreknowledge of what the total mass of those ions might be, a logical place to begin sifting through the data is at the mass of the most abundant di-selenium isotopologue comprising only two atoms of ^{80}Se with total mass of 159.83304 Da (rounded to five decimal places). In negative ESI mode, the mass of an electron is added to this giving a mass of 159.83359 Da (in positive mode, it would be subtracted).¹⁷ The program 'presumes' that the first eligible mass in the data list (the first value ≥ 159.83359 Da)

contains this isotopic combination. The program then tests that presumption by searching for the other expected masses of isotopologues for singly negatively charged ions containing two selenium atoms, which would occur at the known mass differences from the first eligible mass. Whether or not other elements are present in the mass being analyzed is unimportant as we are only calculating mass differences and looking for those that match the pattern expected for molecules containing two atoms of selenium. The program starts by assuming the target mass is the current mass plus the first mass difference on the list (which is sorted from most negative to most positive). In the example, the current mass is assumed to be accurate to ± 0.7 ppm, which sets the upper and lower mass range for itself. The mass difference is added to each of these limits and the resulting masses each are also adjusted for ± 0.7 ppm resolution to give the final search range. A sample calculation of a search range for the next highest mass isotopologue relative to the key isotopologue is shown in Table 2.

After the mass range has been searched, the process is repeated for each of the remaining mass differences for each of the isotopologues. Whenever a candidate is found within the target mass search range it is counted and its intensity recorded. Then the next mass is chosen from the input data and the process continues sequentially for each mass in the mass scan. If more than one mass is found in the target mass search range for any given isotopologue, all are counted and reported in the output file but only the mass that is closest to the target mass, along with its intensity, is kept and used for calculations of goodness of fit to the expected mass and intensity. The closest mass of each isotopologue found in the mass search range is compared to the mass expected for that isotopologue and all such mass differences are aggregated into a single root-mean-square (RMS) mass error statistic. The ratio of the intensity of each isotopologue mass found to that of the key isotopologue is calculated (the key isotopologue's

Table 2. Sample calculation of the mass search range for di-selenium compounds (negative ESI mode; all masses in Da; resolution = ± 0.7 ppm)

Search mass ^a	Search mass \times (1 \pm 0.7 ppm)	Search mass + 0.00097 ^b	Search mass + 0.00097 \times (1 \pm 0.7 ppm)	((Search mass + 0.00097) \times (1 \pm 0.7 ppm)) \times (1 \pm 0.7 ppm)	Final search range
				159.84411	
		159.84389	159.84400		159.84367 to 159.84411
			159.84378		
	159.84303			159.84367	
159.84292	159.84281				

^aThis is the lightest mass in the Solarix data set that equals or exceeds the mass of ⁸⁰Se₂ plus the mass of an electron; 159.83359 Da in total.
^b0.00097 is the mass difference between ⁸⁰Se₂ and the next heaviest isotopologue, ⁸²Se⁷⁸Se, in Da.

intensity, the denominator, is assumed always to exist and be non-zero, given the assumption that the starting search mass is that of the key isotopologue) and compared with the expected ratio; all of the resulting differences are aggregated into a RMS intensity error statistic.

Scoring

Four parameters make up the scoring objective function: the simple percentage of possible isotopologues found, the abundance-weighted percentage of isotopologues found, the percentile rank of RMS mass errors for the isotopologues found (smallest receiving the highest score) and the percentile rank of RMS intensity ratio errors found (smallest receiving the highest score). The first two scoring parameters appear to be redundant but are only partially so. Candidates that have numerically the most mass-related isotopologues in the data should be favored over those that have fewer; candidates that have the most abundant and therefore the most easily observed mass-related isotopologues in the data should also be favored over those that are less abundant and more likely to be due to spurious signals. Candidates that have mass-related isotopologues in the data at masses closest to the expected masses are favored over those with larger mass errors. Finally, candidates that have mass-related isotopologues in the data whose intensity relative to the key

isotopologue is closest to the pattern expected in the natural population of expected isotopologues are favored over those whose intensity ratios are not as close.

All four of the sub-scores are compounded and the fourth root taken to give an aggregate score from 0 to 100. This scoring method is more rigorous than a simple average because, for example, a zero score on any of the four scoring parameters will result in an aggregate score of zero. The higher the score, the more attractive the observed mass is as a candidate containing the specified number of atoms of the selected element. All masses in the experimental data set that show evidence of at least one mass-related isotopologue in the data are ranked and sorted from best to worst on aggregate score before being printed to the output file. Table (3) shows that, in one particular case, the input data file comprised 24,802 records, which produced an output file of 3600 scored and sorted records. Many of the lowest-ranked output records are probably spurious and, in practice, likely only the top 10 to 30 will be of interest to the user. We chose to show the complete, ranked data set and leave it to the user's experience and discretion to decide which records are actually useful.

For the Solarix instrument, the output file shows the raw mass and intensity data along with the resolution in the original data file (see Supporting Information). The program calculates the mass adjusted for ESI by either adding or removing the mass of a proton to the ion depending on which

Table 3. Selenium-targeted search using selenomethionine standards of varying concentration: results for monoisotopic selenomethionine

Selenomethionine sample concentration	[C ₅ H ₁₀ NO ₂ ⁸⁰ Se] ⁺ found at mass (Da)	Mass error (ppm)	Rank among all ions in mass scan	Total number of masses in data file
1 ppb	195.98855	1.7	1	4,764
10 ppb	195.98847	1.3	1	9,222
50 ppb	195.98861	2.0	1	6,211
1 ppm	195.98849	1.4	1	24,802
25 ppm	195.98816	-0.3	4	1,878
100 ppm	195.98862	2.0	2	9,204

Table 4. Selenium-targeted search using selenomethionine standards of varying concentration: results for the ^{13}C isotopologue

Selenomethionine sample concentration	$^{13}\text{C}^{12}\text{C}_4\text{H}_{10}\text{NO}_2^{80}\text{Se}^-$ found at mass (Da)	Mass error (ppm)	$^{12}\text{C}_4^{13}\text{CH}_{10}\text{NO}_2^{80}\text{Se}^-$ intensity as % ^a of $^{12}\text{C}_5\text{H}_{10}\text{NO}_2^{80}\text{Se}^-$ intensity	Rank among all ions in mass scan
1 ppb	-	-	-	-
10 ppb	196.99184	1.3	4.86%	2
50 ppb	196.99200	2.1	5.13%	2
1 ppm	196.99186	1.4	5.55%	2
25 ppm	196.99156	-0.1	5.96%	1
100 ppm	196.99210	2.6	5.71%	3

^aThis is expected to be 5.56% for a molecule containing 5 carbon atoms.

ESI mode was used, and then calculates the monoisotopic neutral mass based on the proton-adjusted mass of the isotopologue comprising only the most abundant isotope of the target element. These last two columns will normally show the same values; they are only intended for use with samples that may contain hydrogen and should otherwise be ignored. The number of masses found for each expected isotopologue mass difference is shown along with the percentage of isotopologues found and the abundance-weighted percentage of total possible mass represented by those isotopologues. Next, for each isotopologue mass difference that produced an isotopically related mass in the data, that mass is shown and can be easily checked in the original dataset. Following this, the ppm mass error is shown for each mass, then the RMS of the ppm mass differences and the corresponding percentile rank. (In the case of a tie in any RMS measure, the Winnow program assigns an equal percentile rank to each of those observations.) Next, the ratio of intensity versus that of the key isotopologue is shown for each mass found, then the difference from the expected ratio for each, then the RMS of the differences of these intensity ratios from those expected and finally the percentile rank of the RMS scores. The final column is the aggregate score, described previously.

RESULTS AND DISCUSSION

The Winnow program was tested by running a series of mass scans on the Solarix instrument using solutions of varying concentrations of selenomethionine. Resulting mass scan files (i.e. the input files for the Winnow program) were the sum of 25 scans and comprised from 1878 to 24,802 rows of data and all program runs executed in less than one second.

The output of experimental data was searched for ions containing a single selenium atom and Table 3 shows that the Winnow program found the selenium standard in position 1 in four of the six standard samples and in positions 2 and 4 in the remaining samples. For organic samples, when the sample intensity is sufficient, it should be possible to see the ^{13}C isotopologue of the target molecule in a position near the base peak for the target compound. Table 4 shows that the ^{13}C isotopologue for selenomethionine was found five times out of six: in position 1 once, in position 2 three times and in position 3 once. When performing a mass scan it cannot be relied upon that fragmentation will occur in the source, but

under the conditions of the tests, the methylselenide ion was produced and in three of the five instances when it was detected, it was ranked in position 3; in the remaining two instances it was ranked in position 4 (Table 5).

Table 6 shows a comparison of the experimental isotopologue mass and abundance data with the corresponding calculated (i.e. theoretical) data, which was the result of the successful search for selenomethionine in the 1 ppm standard solution.

It is expected that the ability of the program to detect the presence of target ions efficiently should depend directly on the mass resolution of the instrument. Mass resolution is most commonly defined as the centroid mass divided by full peak width at half maximum peak height (FWHM). It is easy to show that in terms of mass and its standard deviation (assuming mass measurements are normally distributed) the FWHM boundaries occur at the population's centroid mass $\pm \sigma \sqrt{-2 \ln(0.5)}$ (where σ is the standard deviation of the population of masses), or ± 1.18 standard deviations (i.e. 76% of the masses measured will lie between the two mass boundaries). A centroid mass reported at a resolution as thus defined is an estimate that carries with it an implied estimate of the likelihood that a particular experimentally measured mass is equal to a specific exact mass. As peak width widens, taking in increasing numbers of discrete, nearby, but possibly unrelated masses, the probability that the centroid of the peak equals a specific target mass diminishes. Thus, the search for a particular mass and its related isotopologues, at a particular threshold of mass precision, becomes increasingly inefficient as resolution declines.

Table 5. Under the conditions of the mass scan, fragmentation occurred in the source. The $[\text{CH}_3^{80}\text{Se}]^-$ fragment was observed in subsequent product ion scans

Selenomethionine sample concentration	$[\text{CH}_3^{80}\text{Se}]^-$ fragment found at mass (Da)	Mass error (ppm)	Rank among all ions in mass scan
1 ppb	-	-	-
10 ppb	94.94059	0.4	3
50 ppb	94.94062	0.7	4
1 ppm	94.94060	0.5	3
25 ppm	94.94052	-0.3	3
100 ppm	94.94051	-0.4	4

Table 6. Masses and abundances found for selenomethionine for the 1 ppm standard solution

Selenomethionine isotopologue	Expected mass (Da)	Mass found (Da)	Mass resolution ^a	Expected abundance	Abundance found
C ₅ H ₁₀ NO ₂ ⁷⁴ Se ⁻	189.99418	189.99446	296,604	0.0089	0.0071
C ₅ H ₁₀ NO ₂ ⁷⁶ Se ⁻	191.99091	191.99119	291,238	0.0937	0.0801
C ₅ H ₁₀ NO ₂ ⁷⁷ Se ⁻	192.99162	192.99189	289,876	0.0763	0.0682
C ₅ H ₁₀ NO ₂ ⁷⁸ Se ⁻	193.98901	193.98928	290,157	0.2377	0.2264
C ₅ H ₁₀ NO ₂ ⁸⁰ Se ⁻	195.98822	195.98849	286,022	0.4961	0.5196
C ₅ H ₁₀ NO ₂ ⁸² Se ⁻	197.98840	197.98868	283,535	0.0873	0.0986
Total				1.0000	1.0000

^aThe experimental key isotopomer mass is shown in bold. Resolution is measured as mass divided by full peak width at half maximum peak height.

The mass measurements presented here were done on the Solarix instrument and were recorded by the instrument to five decimal places. For our measurements at various concentrations, the average mass resolution for monoisotopic selenomethionine was 288,069. The program does not use the instrument's mass resolution directly in its search algorithm, relying instead on the error on mass accuracy found during instrument calibration. Although mass resolution and error on mass measurement are two different things, low error on mass measurement cannot be obtained without high mass resolution. The uncertainty in the accuracy of the average mass of a distribution of masses should be approximately proportional to the uncertainty that a mass reported to a certain number of decimal places equals a theoretical target mass calculated at high precision.

A simulation was done to test the efficiency of the program at placing target ions that are present in the sample into the highest rankings, given a particular instrument mass resolution capability. As a proxy for mass measurements done at varying levels of resolution, the selenomethionine mass data were successively rounded to fewer decimal places, aggregating the intensities for resulting equal masses into a single mass and intensity pair at each step. Table 7 shows that there was little difference between the rankings at four and five decimal places of reported mass precision. At fewer than four decimal places of reported mass precision the ability of the program to rank target ions efficiently fell quickly. It is important to note that the target ion used here comprised only one atom of the target element; for higher numbers of target atoms, especially those with a large number of isotopes, the

program's ability to discriminate would likely be improved with higher mass resolution and with more decimal places of precision for the reported masses.

As a further test of the program, Winnow was used to search the experimental mass scan data of the 1 ppm standard solution for selenomethionine using a single nitrogen atom, a pair of oxygen atoms and five carbon atoms, consistent with the structure of selenomethionine (Fig. 1). The results are summarized in Table 8. Generally, it is expected that the program will be more successful the scarcer the element targeted by the search because of fewer interferences occurring in the mass scan data. It is also expected that best results will be obtained when the target element comprises more isotopes versus fewer, providing more sources of information, and when its isotopic distribution is most uniformly distributed, thus providing fewer weak signals. These generalizations are supported by our results: when nitrogen was targeted, Winnow ranked the ⁸²Se-selenomethionine isotopologue in position 1; a search for di-oxygen isotopologue ions ranked ⁷⁸Se-selenomethionine in

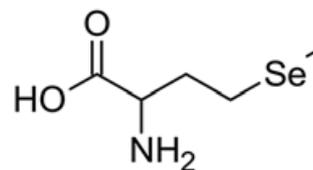

Figure 1. Structure of selenomethionine.

Table 7. Rank of ions found in a Se-targeted search using selenomethionine standards of varying concentration: results for monoisotopic selenomethionine at varying decimal places of simulated reporting precision (assumed to be proportional to instrument resolution)

ppb SeMet \ Target Mass	196	196.0	195.99	195.988	195.9882	195.98822
1 ppb	467	809	2635	46	1	1
10 ppb	399	173	1	1	1	1
50 ppb	355	144	5	1	1	1
1 ppm	496	195	84	15	1	1
25 ppm	121	32	38	4	1	4
100 ppm	199	107	51	16	1	2
Average	340	243	469	14	1	2

Table 8. Searches based on alternative elements

Elemental search pattern	Top three selenomethionine isotopologue ions found	Rank
N	$[\text{C}_5\text{H}_{10}^{14}\text{NO}_2^{82}\text{Se}]^-$	1
	$[\text{C}_5\text{H}_{10}^{14}\text{NO}_2^{80}\text{Se}]^-$	2
	$[\text{C}_5\text{H}_{10}^{14}\text{NO}_2^{78}\text{Se}]^-$	4
O ₂	$[\text{C}_5\text{H}_{10}\text{N}^{16}\text{O}_2^{78}\text{Se}]^-$	2
	$[\text{C}_5\text{H}_{10}\text{N}^{16}\text{O}_2^{82}\text{Se}]^-$	35
	$[\text{C}_5\text{H}_{10}\text{N}^{16}\text{O}_2^{80}\text{Se}]^-$	36
C ₅	$[\text{C}_5\text{H}_{10}\text{NO}_2^{82}\text{Se}]^-$	8
	$[\text{C}_5\text{H}_{10}\text{NO}_2^{76}\text{Se}]^-$	43
	$[\text{C}_5\text{H}_{10}\text{NO}_2^{78}\text{Se}]^-$	45

position 2 and a search for 5-carbon isotopologue ions ranked ⁸²Se-selenomethionine in position 8. To further confirm the importance of elemental scarcity, the number of isotopes and the relative abundance of the isotopes of the target element, more tests should be run using other elements and different types of samples.

CONCLUSIONS

This program is not intended to be a stand alone analytical tool, but to complement the (somewhat limited) capabilities of the software provided with modern, high-resolution FTMS instruments. As noted above, the user is generally only interested in the first 10 to 30 masses listed in the output file – this number can be reduced even further based on the user's knowledge of the compounds being sought within the dataset, poor aggregate scores or low numbers found of expected isotopologues. The mass(es) generated by this program, as well as the knowledge that a peak of interest contains a specific element, can be used in conjunction with the instrument software's molecular formula generator, if available, to predict its identity. While this identity must still be confirmed based

on the mass spectrum, this program saves the user countless hours of meticulously and laboriously searching through data to visually identify isotopic signals of analytes hidden among background noise.

The program will be made available for download on the Water Quality Centre website.^[8]

Acknowledgements

The authors thank Dr. Naomi Stock for many helpful suggestions and Dr. Breda Novotnik for the use of data collected in Trent University's Water Quality Centre Lab 2.

REFERENCES

- [1] mMass – open source mass spectrometry tool. www.mmass.org.
- [2] S. Hrdlickova Kuckova, G. Rambouskova, R. Hynek, P. Cejnar, D. Oltrogge, R. Fuchs. Evaluation of mass spectrometric data using principal component analysis for determination of the effects of organic lakes on protein binder identification. *J. Mass Spectrom.* **2015**, *50*, 1270.
- [3] M. Letzel. Universal Mass Calculator (UMC) for Windows (Version 3.7.1.6). Westfälische Wilhelms-Universität Münster. www.uni-muenster.de/Chemie.oc/ms/en/links.html.
- [4] P. Godfrey. *A Note on the Computation of the Convergent Lanczos Complex Gamma Approximation*. Intersil Corp., Palm Bay, FL, **2001**.
- [5] Open Watcom Integrated Development Environment (version 1.9). <http://www.openwatcom.org/>.
- [6] Table of isotopic masses and natural abundances. www.chem.ualberta.ca/~massspec/atomic_mass_abund.pdf.
- [7] I. Ferrer, E. M. Thurman. Letter to the editor. *Rapid Commun. Mass Spectrom.* **2007**, *21*, 2538.
- [8] www.trentu.ca/wqc.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's website.