

---

---

# Isotopic Compositions and Accurate Masses of Single Isotopic Peaks

Alan L. Rockwood

ARUP Institute for Clinical and Experimental Pathology, Salt Lake City, Utah, USA

Jordan R. Van Orman and David V. Dearden

Department of Chemistry and Biochemistry, Brigham Young University, Provo, Utah, USA

---

This paper presents and proves an algorithm to calculate the isotopic composition of individual (nominal) isotopic peaks. From this information one can calculate the accurate masses of isotopic peaks. This opens the way to use accurate mass measurements to determine chemical compositions of compounds using non-mono-isotopic peaks. The algorithm is computationally efficient and rigorously correct in the absence of roundoff error. Highly effective error correction strategies are described to detect and correct computational errors arising in practical calculations. Results from theoretical calculations of isotopic masses for a krypton inclusion complex agreed well with experimental measurements. (J Am Soc Mass Spectrom 2004, 15, 12–21) © 2004 American Society for Mass Spectrometry

---

High accuracy mass measurements by mass spectrometry are often used to determine the chemical composition of compounds [1]. Standard practice is to select the mono-isotopic peak because it is the only peak for which the accurate mass can be conveniently calculated from the chemical formula. However, for isotopically complex molecules the mono-isotopic mass peak may be too weak to be useful. As a partial solution to this problem Senko et al. have proposed the use of “averagine” modeling as a way of estimating the unobservable mono-isotopic masses [2]. While extremely useful, this technique is designed to handle cases for which averagine is a good model, i.e., proteins of typical composition, and is not designed as a general-purpose tool.

An alternative approach would be to calculate the accurate masses of nominal isotopic peaks, and then compare the calculated values to measured values. [In this paper, when we refer to the accurate mass of an isotopic peak we are referring to the accurate mass of a nominal isotopic peak, for example the accurate mass of  $^{29}\text{CO}$ , including all isotopic contributions, rather than the accurate masses of individual contributions to the isobar, such as  $^{13}\text{C}^{16}\text{O}$  and  $^{12}\text{C}^{17}\text{O}$ .] One approach to calculating accurate isotopic masses is to first determine the isotopic composition of an isotopic peak, and then calculate the accurate mass of the isotopic peak based on the isotopic composition and the accurate elemental isotopic masses. In principle this is a solved problem.

The polynomial method can be used to generate a peak list. Each element of the peak list is associated with a specific isotopic composition and probability. The accurate mass of a nominal isotopic peak would be calculated by sorting through the list to organize the list into isobars and then computing the mass of a nominal isotopic peak as a probability-weighted sum of all the terms within an isobaric group.

Although the polynomial method works well for small molecules, the recent trend in mass spectrometry has been toward larger and larger systems. For these, the polynomial method becomes increasingly complex and impractical. The basic problem is a combinatorial explosion of terms when the method is applied to isotopically complex molecules. This leads to computational inefficiencies, both in terms of speed and memory usage. The only known method of dealing with this problem is to omit the calculation of some terms, a process known as “pruning”. An unavoidable byproduct of pruning is a loss of accuracy [3–7]. The loss of accuracy affects the peak probabilities and average mass more strongly than the individual isotopic peak positions, but pruning does affect the isotopic peak positions as well. The loss of accuracy might be acceptable in some cases, but this has not been validated in the literature for isotopically complex systems. Furthermore, even with pruning there is a loss of computational efficiency as isotopic complexity increases, and speed may be crucial in some applications.

The present paper introduces a method for calculating isotopic compositions of isotopic peaks that avoids the need for polynomial calculations. The method is both accurate and computationally efficient. The algorithm itself is exact, and errors that creep into practical

---

Published online November 19, 2003

Address reprint requests to Dr. A. L. Rockwood, ARUP Institute for Clinical and Experimental Pathology, 500 Chipeta Way, Salt Lake City, UT 84108, USA. E-mail: rockwoal@aruplab.com

**Table 1.** Masses and abundances used for calculations in this paper

Nuclide	Mass (isotope.dat)	Mass (XMASS)	Abundance
<sup>1</sup> H	1.0078246	1.007825032	0.99985
<sup>2</sup> H	2.0141021	2.014101778	0.00015
<sup>12</sup> C	12.0000000	12.000000000	0.988930
<sup>13</sup> C	13.0033554	13.003354838	0.011070
<sup>14</sup> N	14.0030732	14.003074005	0.996337
<sup>15</sup> N	15.0001088	15.000108898	0.003663
<sup>16</sup> O	15.9949141	15.994914622	0.997590
<sup>17</sup> O	16.9991322	16.999131501	0.000374
<sup>18</sup> O	17.9991616	17.999160419	0.002036
<sup>35</sup> Cl	34.9688531	34.968852707	0.755290
<sup>37</sup> Cl	36.9659034	36.9659026	0.244710
<sup>39</sup> K	38.963707	38.963706861	0.932581
<sup>40</sup> K	39.963999	39.963998672	0.000117
<sup>41</sup> K	40.961825	40.961825972	0.067302
<sup>79</sup> Br	78.918336	78.918337647	0.5069
<sup>81</sup> Br	80.916289	80.91629106	0.4931
<sup>78</sup> Kr	77.914	77.920386271	0.0035
<sup>80</sup> Kr	79.916380	79.91637804	0.0225
<sup>82</sup> Kr	81.913482	81.913484601	0.116
<sup>83</sup> Kr	82.914135	82.914135952	0.115
<sup>84</sup> Kr	83.911507	83.911506627	0.570
<sup>85</sup> Kr	85.910616	85.910610313	0.173

calculations, such as roundoff error, are easily handled by very effective error detection and correction strategies.

## Theoretical Development

### The Algorithm

Unless otherwise noted an isotopic distribution will be understood to represent a discrete probability distribution, implying that the sum over the distribution equals one. For presentation purposes peaks will be labeled by integer mass (nucleon number) unless otherwise noted. For example, the nucleon number of the mono-isotopic peak of CO<sub>2</sub> is 44, so its integer mass is 44.

All elemental isotopic abundances and masses used in this paper are given in Table 1. Unless otherwise indicated all masses used in calculations in this paper are taken from the column labeled “Mass (isotope.dat)”, which is excerpted from the isotope table normally provided with the computer program “Mercury6”. Mercury6 is based on work presented elsewhere [4].

The algorithm is based on the idea that one can determine the isotopic composition of a peak by randomly removing an atom from the molecule. Each isotope of each element has a probability associated with it, and this probability, together with the chemical formula of the compound, provides the information needed to calculate the isotopic composition. One approach could be a Monte Carlo sampling procedure. However, that approach is subject to statistical uncertainty. The following algorithm performs a direct calculation of the probabilities (Steps 1–5), and then folds

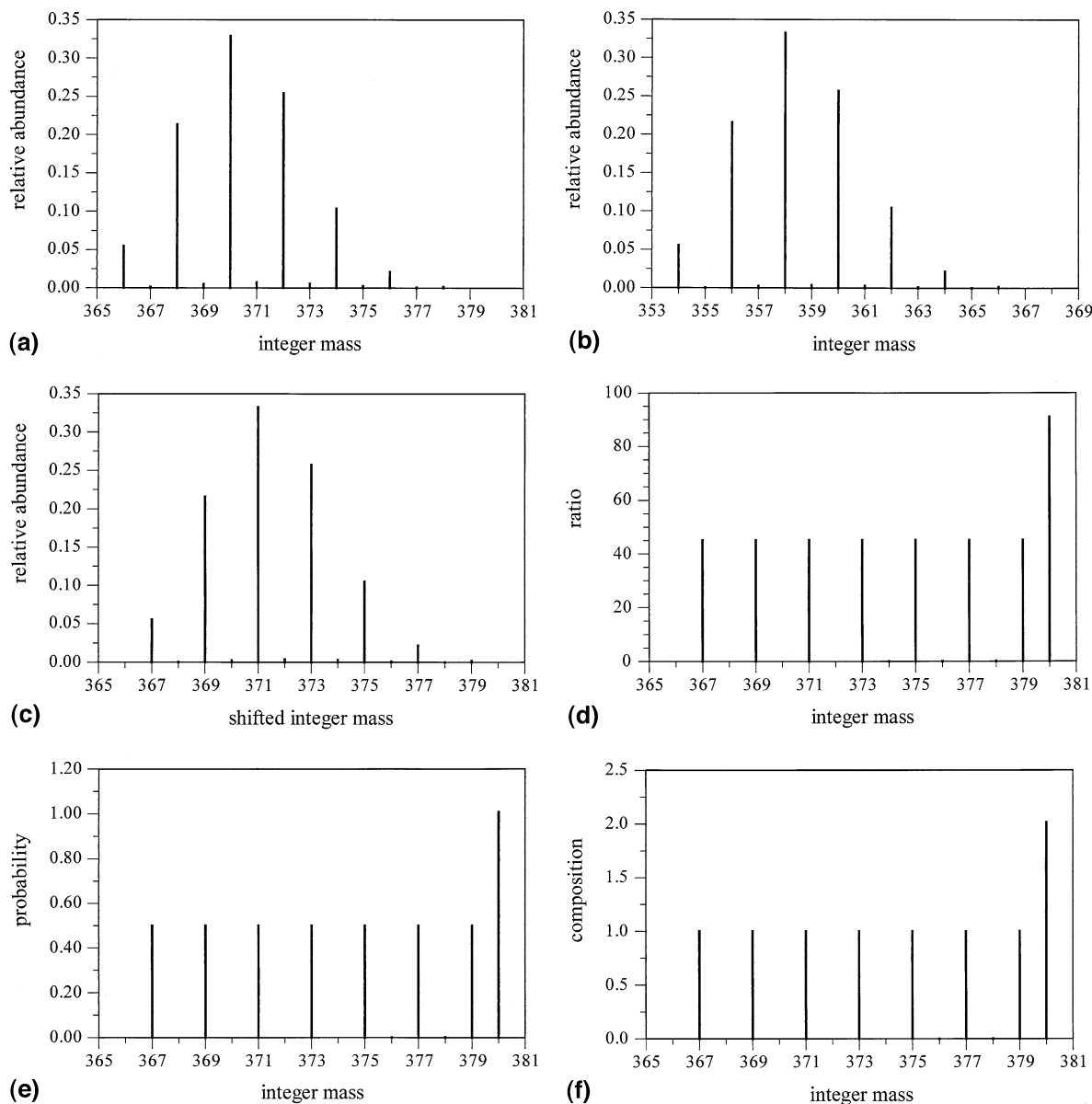
the result with the chemical formula to yield the isotopic composition of the isotopic peaks (Step 6).

The compound under consideration will be called the “parent compound”. Tribromo-trichloro-ethane (C<sub>2</sub>Br<sub>3</sub>Cl<sub>3</sub>) will serve as an example of a parent compound. In the algorithm described below one divides molecules of the parent compound into two unequal pieces, one (the “atom”) consisting of a single atom of a particular element, and the other (the “product compound”) consisting of the remaining atoms. For example, if we are computing the contribution of carbon isotopes to C<sub>2</sub>Br<sub>3</sub>Cl<sub>3</sub> then the atom and product compound are C and CBr<sub>3</sub>Cl<sub>3</sub> respectively.

The algorithm proceeds as follows. Step 1: Compute the isotopic distribution of the parent compound. Figure 1a illustrates Step 1 for C<sub>2</sub>Br<sub>3</sub>Cl<sub>3</sub>. Step 2: Remove the atom and compute the isotopic distribution of the product compound. In our example C is removed, so the product compound is CBr<sub>3</sub>Cl<sub>3</sub>. Figure 1b illustrates this step. Step 3: Add the mass of the isotope of interest to the mass scale of the isotopic distribution calculated in Step 2. For the <sup>13</sup>C isotope Step 3 yields a result illustrated in Figure 1c. (Figures 1c–f look like mass spectra, but they are not. Instead they represent intermediate and final results of calculations.) Step 4: Divide the intensities of the peaks in Figure 1c by the intensities of the corresponding peaks in Figure 1a. The result is demonstrated in Figure 1d. Step 5: Multiply this result by the elemental isotopic abundance for the isotope of interest for the atom removed in Step 2. In our example, we multiply by 0.01107, which is the elemental isotopic abundance of <sup>13</sup>C. Results are shown in Figure 1e. Step 6: Multiply the result by the number of atoms of the element under consideration in the formula for the parent compound. In our example we multiply by 2, which is the number of carbon atoms in the parent compound. Results are shown in Figure 1f. The numbers represented in this figure are the <sup>13</sup>C compositions of the isotopic peaks of the original compound.

Interesting features of the <sup>13</sup>C composition are as follows. The highest-mass isotopic peak has a <sup>13</sup>C composition of 2 because the highest-mass peak is made of all-heavy isotopes, so both carbon atoms must be <sup>13</sup>C. The odd-mass peaks all have a <sup>13</sup>C composition of 1. This is because the hypothetical Br<sub>3</sub>Cl<sub>3</sub> molecule has a strictly even-mass isotopic distribution. Therefore, odd mass peaks for C<sub>2</sub>Br<sub>3</sub>Cl<sub>3</sub> must be the result of the addition of carbon, and the only odd masses possible from two carbon atoms are via one <sup>13</sup>C. The even-mass peaks have a very small contribution from <sup>13</sup>C. This is because <sup>13</sup>C forms a small percentage of the total carbon, and the probability of finding two <sup>13</sup>C atoms in a molecule is very low.

The <sup>12</sup>C composition would be calculated the same way, except that the shift in mass scale in Step 2 would be 12, and the probability used in Step 5 would be 0.98893. The composition for the isotopes of the other elements would be calculated by an analogous procedure.



**Figure 1.** Illustration of steps in algorithm to calculate isotopic composition of single isotopic peaks. (a) Step 1, isotopic distribution of parent compound,  $C_2Br_3Cl_3$ . (b) Step 2, isotopic distribution of product compound formed by removing one carbon atom,  $CBr_3Cl_3$ . (c) Step 3, distribution in Step 2, shifted by 13 mass units, the isotopic mass of  $^{13}C$ . (d) Step 4, peaks in Figure 1c divided by peaks in Figure 1a. (e) Step 5, peaks in Step 4 multiplied by elemental isotopic abundance of  $^{13}C$ . (f) Step 6, peaks in Step 5 multiplied by 2, the number of carbon atoms in  $C_2Br_3Cl_3$ . This represents the isotopic composition of  $^{13}C$  in each of the isotopic peaks of  $C_2Br_3Cl_3$ .

The resulting isotopic composition from applying the algorithm can then be used to calculate accurate isotopic masses. The procedure is to first select a nominal isotopic peak of the molecule. Then one calculates the elemental isotopic composition for that peak. One then multiplies these isotopic compositions by accurate elemental isotopic masses. Finally, one adds these results to compute the accurate mass. Succinctly stated, the accurate mass of a molecular isotopic peak is the dot product of an isotopic composition vector and an accurate mass vector.

### *Proof*

A somewhat difficult proof was previously outlined using Fourier transforms [8]. Here we present a simpler proof based on the idea that taking a molecule apart into an atom and a remaining portion is the inverse process of combining these same two components to form the whole molecule. Chemically, these two inverse processes are dissociation and recombination, respectively. Mathematically, these processes correspond to deconvolution and convolution, respectively. Because

**Table 2.** Combinations of isotopic peaks of  $\text{CBr}_3\text{Cl}_3$  and C to form  $^{370}(\text{C}_2\text{Br}_3\text{Cl}_3)$ 

	$\text{C}_2\text{Br}_3\text{Cl}_3$ (parent compound)		$\text{CBr}_3\text{Cl}_3$ (product compound)		C (atom)	
	Integer mass	Probability	Integer mass	Probability	Integer mass	Probability
Combination 1	370	2.675386e-5	357	2.416798e-3	13	1.107e-2
Combination 2	370	3.287701e-1	358	3.324504e-1	12	9.8893e-1

deconvolution is the inverse of convolution we can investigate the deconvolution process, the process of interest here, by examining certain intermediate terms of a convolution calculation.

The proof is via derivation of the algorithm. Assume for illustration purposes that we are interested in the  $^{13}\text{C}$  isotopic composition of  $\text{C}_2\text{Br}_3\text{Cl}_3$ . The proof is broken into two main parts. The first seeks to answer the following question: If we randomly remove an atom of a selected element (e.g., C) from a selected isotopic peak of the parent compound (e.g., the isotopic peak of  $\text{C}_2\text{Br}_3\text{Cl}_3$  having an integer mass of 370), what is the probability that it will be a specific isotope of the element (e.g.,  $^{13}\text{C}$ )? This relates to the deconvolution process mentioned above.

In the second part of the proof one uses this probability, together with the number of atoms of the element of interest in the chemical formula (e.g., 2, if one is interested in the carbon isotopes of  $\text{C}_2\text{Br}_3\text{Cl}_3$ ), to calculate the isotopic composition of the selected isotopic peak of the parent compound (e.g., the isotopic peak of  $\text{C}_2\text{Br}_3\text{Cl}_3$  having an integer mass of 370).

We rely on recent results on isotopic patterns of product ions from the dissociation of single isotopic peaks [9]. To the extent possible symbols in the present paper are harmonized with those of reference [9]. However, the naming conventions differ slightly. Generally speaking, the word “ion” is removed from the naming conventions of the present paper because charge is not relevant to the calculations. A vector denoted by  $h$  in both papers refers to the isotopic distribution of the parent species which we refer to as “parent compound” in the present paper and “parent ion” in the previous paper. Vector  $h$  represents a discrete probability distribution, so its elements are non-negative and add to one. The index into the vector represents integer mass, and the value of the corresponding vector element represents the relative abundance of the isotopic peak expressed as a probability. The vector  $h$  for  $\text{C}_2\text{Br}_3\text{Cl}_3$  is shown in Figure 1a.

Vector  $g$  represents the isotopic distribution of the “product compound”. In the previous paper this was called the “product ion”. In the present paper,  $g$  serves only as a computational tool and does not imply the physical existence of the product compound. The vector  $g$  for  $\text{CBr}_3\text{Cl}_3$  (which is  $\text{C}_2\text{Br}_3\text{Cl}_3$  minus a carbon atom) is shown in Figure 1b.

Vector  $f$  represents the isotopic distribution of the “atom”. The analogous vector in the previous paper was the isotopic distribution of the “complementary product”. We use the term “atom” in the present paper

to emphasize the concept that during the derivation we conceptually remove an atom from the parent compound, whereas in the previous paper the complementary product was allowed to be polyatomic. Throughout the presentation we use a simplified notation in order to avoid a proliferation of subscripts by omitting any specifier of the chemical element under consideration.

We define a matrix,  $C$ , whose elements are defined by the following relationship:

$$C(i, j) = g(i)f(j) \quad (1)$$

In reference 9 this matrix was called by several names, the most relevant to the present discussion being “probability matrix”.

Each combination of  $i$  and  $j$  in eq 1 represents a specific way to generate molecules of the parent compound. Table 2 summarizes all combinations that can generate  $^{370}(\text{C}_2\text{Br}_3\text{Cl}_3)$  by combining isotopic peaks of C and  $\text{CBr}_3\text{Cl}_3$ . Choosing  $j = 13$  corresponds to a selection of  $^{13}\text{C}$ , where the upper-left superscript indicates integer mass, and choosing  $i = 357$  corresponds to selection of  $^{357}(\text{CBr}_3\text{Cl}_3)$ , where again the upper left superscript indicates integer mass. The combined choice of  $(i, j) = (357, 13)$  relates to the  $^{370}(\text{C}_2\text{Br}_3\text{Cl}_3)$  parent isotopic peak. There is also a second way of generating  $^{370}(\text{C}_2\text{Br}_3\text{Cl}_3)$ , which is to combine  $^{12}\text{C}$  with  $^{358}(\text{CBr}_3\text{Cl}_3)$ . The  $(i, j) = (358, 12)$  combination has a matrix element (i.e., probability) of 2.675386e-5, which is the result of multiplying 2.416798e-3 with 1.107e-2 from the table, and the  $(i, j) = (358, 12)$  combination has a matrix element of 3.287701e-1, which is the result of multiplying 3.324501e-1 with 9.8893e-1 from the table. Because carbon contains only two isotopes there are only two non-zero contributions to  $^{370}(\text{C}_2\text{Br}_3\text{Cl}_3)$ . If carbon were a three-isotope element then there could be as many as three ways to contribute to any individual isotopic peak of the parent ion.

Removal of an atom from the parent compound is the reverse of the process of combining the atom with the product compound to form the parent compound. Consequently, if we dissociate the full isotopic profile of the parent compound, producing the atom and the product compound, each microscopic dissociation pathway (specified by a combination of  $i$  and  $j$ ) contributes a relative amount equal to  $C(i, j)$ . For example, from the previous paragraph, the matrix element  $C(358, 12) = 0.3287701$  gives the probability of the pathway for the removal of  $^{12}\text{C}$  from  $^{370}(\text{C}_2\text{Br}_3\text{Cl}_3)$  to form  $^{358}(\text{CBr}_3\text{Cl}_3) + ^{12}\text{C}$ , and the matrix element  $C(357, 13) = 2.68\text{e-}5$  gives

the probability of the pathway for the removal of  $^{13}\text{C}$  from  $^{370}(\text{C}_2\text{Br}_3\text{Cl}_3)$  to form  $^{357}(\text{CBr}_3\text{Cl}_3) + ^{13}\text{C}$ .

These probabilities are normalized to the full isotopic profile of the parent compound. We must renormalize them to account for the fact that we are sampling from only part of the parent compound distribution. In our example we would divide by 0.3287969, which is the fraction of the parent compound distribution concentrated in the mass 370 peak. The value of 0.3287969 is the sum of 0.3287701 and 2.68e-5, the matrix elements for the two ways of making  $^{370}(\text{C}_2\text{Br}_3\text{Cl}_3)$  discussed above. The general formula for the process just described is

$$\begin{aligned} R(k, j) &= C(k - j, j) / \sum_n C(k - n, n) \\ &= C(k - j, j) / h(k) \end{aligned} \quad (2)$$

where  $k$  is the molecular weight of the selected isotopic peak of the parent compound and  $j$  is the atomic weight of the isotope of the atom under consideration.  $R(k, j)$  is the probability that the atom will be of isotopic mass  $j$  if it is removed at random from a parent ion of isotopic mass  $k$ . In this expression we have taken advantage of the fact that

$$h(k) = \sum_n C(k - n, n) \quad (3)$$

where  $n$  is a dummy variable used to generate the sum.

To this point we have shown that  $R(k, j)$  is the probability that an atom (e.g., C) removed at random from an isotopic peak of a parent ion will be of isotopic mass  $j$  (e.g.,  $^{13}\text{C}$ ). Eq 2 therefore represents completion of part one of the derivation.

For clarification, let us now take a moment and relate the quantities discussed so far to specific steps in the algorithm.  $C(k-j, j)$  of eq 3 is equivalent to the results of Steps 2 and 5 of the algorithm. This can be seen by looking at the factors that define the matrix elements of  $C$  in eq 1. The first factor,  $g$ , is identified with the isotopic abundance of the product compound in Step 2 of the algorithm. The second factor,  $f$ , is identified with the isotopic abundance of an isotope of an atom in Step 5 of the algorithm. The quantity  $h(k)$  is equivalent to the results of Step 1 of the algorithm.  $R(k, j)$  is equivalent to the results of Step 4 of the algorithm. The mass shift described in Step 3 of the algorithm is implicit in eq 3. Of the steps in the algorithm, only Step 6 is missing to this point.

Now that we know  $R(k, j)$  is the probability that an atom (e.g., C) removed at random from an isotopic peak of a parent ion will be of isotopic mass  $j$  (e.g.,  $^{13}\text{C}$ ), we have only to convert this probability into an actual number, such as the number of  $^{13}\text{C}$  atoms in an isotopic peak. To do this we multiply the probability,  $R(k, j)$ , by  $N$ , the number of atoms in the chemical formula of the

parent compound for the element being considered. This yields the final result,  $Q(k, j)$ .

$$Q(k, j) = N \cdot R(k, j) \quad (4)$$

Eq 4 is equivalent to Step 6 of the algorithm, and the proof is therefore complete.

## Computational Methods

The algorithm described above can be broken into two phases. In the first phase (Steps 1 and 2 above) the isotopic distributions are calculated for the parent compound and product compound. Methods that produce centroided results (stick spectra) are ideal for these calculations. [4, 6]. In this paper we use a computer program, "Mercury6". Unless otherwise noted, elemental isotopic abundances and accurate masses used in the calculations were taken from the isotope.dat file that Mercury6 uses as input (Table 1). Mercury6 generates an output file containing mass/intensity pairs. The isotopic distribution of the atom is also required, but this does not require a separate calculation since it can be found in tables. The resulting distribution was normalized to unit probability and the remaining calculations done using Psi-Plot [10], which has extensive capabilities for vector operations. A few miscellaneous calculations were done with hand calculators (HP 12C and Casio fx-991N).

## Experimental Methods

All experiments were performed using a Bruker Daltonics (Billerica, MA) model APEX 47e Fourier transform mass spectrometer (FTMS), equipped with a 4.7 tesla superconducting magnet and an ion source external to the high-field region of the magnet, and the "infinity" trapping cell [11] design. Ions were generated using microelectrospray of a solution of decamethylcurcubit[5]uril (hereafter referred to by the abbreviation mc5; molecular formula:  $\text{C}_{40}\text{H}_{50}\text{O}_{10}\text{N}_{20}$ ; 0.1 mg mL $^{-1}$ , synthesized as has been described [12]) and potassium acetate (2 mM) in ethanol, into which Kr gas had been bubbled. The microelectrospray source has been described [13]. The ions were electrostatically injected into the FTMS trap. Following injection, the potentials on both trapping plates were adjusted to a value of +1.0 V. The trapped ions were allowed to react with approximately  $5 \times 10^{-8}$  mbar of neutral 18-crown-6 (Sigma, St. Louis, MO) for 1.0 s prior to ion excitation via a chirped RF pulse and acquisition of the time-domain transient signal. The instrument was controlled using a MIDAS data system [14] (software version 3.18). The acquired spectra consisted of 128 k of time-domain data points. The time-domain spectra were processed and mass calibrations were performed using Bruker Daltonics' XMASS package (version 5.0.10); processing consisted of a single zero-fill, fol-

**Table 3.** Calculated isotopic compositions of isotopic peaks of  $C_2Br_3Cl_3$ 

Integer mass	Composition					
	$^{12}C$	$^{13}C$	$^{79}Br$	$^{81}Br$	$^{35}Cl$	$^{37}Cl$
366	2.00000006	0.00000000	3.00000001	0.00000000	3.00000009	0.00000000
367	0.99999889	1.00000140	3.00000345	0.00000000	3.00000250	0.00000000
368	1.99993557	0.00006442	2.24987157	0.75012839	2.75016059	0.24983940
369	1.00000047	1.00000022	2.24984717	0.75015255	2.75015201	0.24984736
370	1.99983724	0.00016274	1.57871795	1.42128202	2.42136339	0.57863660
371	1.00000036	1.00000036	1.57866364	1.42133669	2.42133635	0.57866332
372	1.99967600	0.00032398	1.01445041	1.98554958	1.98571159	1.01428843
373	1.00000005	1.00000054	1.01435908	1.98564174	1.98564125	1.01435907
374	1.99938525	0.00061477	0.57879726	2.42120275	1.42151013	1.57848993
375	0.99999940	1.00000082	0.57866343	2.42133751	1.42133598	1.57866399
376	1.99877636	0.00122362	0.25004868	2.74995132	0.75056306	2.24943684
377	1.00000171	1.00000026	0.24984834	2.750115146	0.75014748	2.24984517
378	1.99691047	0.00308857	0.00038608	2.99961642	0.00115655	2.99884241
379	0.99991862	0.99997577	0.00000000	2.9994022	0.00000000	2.99990906
380	0.00000000	2.01474162	0.00000000	3.02422326	0.00000000	3.01734260

lowed by Gaussian multiplication (with parameters  $LB = -10.0$  and  $GB = 0.5$ ) and magnitude-mode Fourier transformation. The mass scale was calibrated internally using the “cal2” method supplied with XMASS; internal calibration peaks were the mono-isotopic ions of  $mc5(K^+)_2$ ,  $N_2@mc5(K^+)_2$ , and  $mc5\cdot K^+$ . Reference masses for the calibration were from the XMASS periodic table [15] and are listed in Table 1 of this paper. Peak positions for all spectra were determined using the XMASS apex method.

## Discussion

### Results of Sample Calculation

Table 3 presents the calculated isotopic compositions of the isotopic peaks of  $C_2Br_3Cl_3$ . Peaks are labeled according to their integer masses.

### Errors and Uncertainties

Although the algorithm presented above is exact, it is subject to roundoff error. Psi-plot is not a significant source of computational error because it performs its calculations to >14 significant figures. Mercury6 generates the main source of error when it writes its output file because it scales the base peak to 100% and rounds its output file to 6 decimal places. This is particularly significant for the weaker isotopic peaks. In addition to computational error, there are additional sources of uncertainty that may affect the comparison to experimental results.

*Error Detection.* Fortunately,  $C_2Br_3Cl_3$  is simple enough to verify using hand calculations by the polynomial method. Some entries in Table 3 contain significant errors, although the algorithm itself is exact.

One way to gauge errors in the calculations is to look for internal inconsistencies in the calculated isotopic compositions. In general the Table 3 entries for isotopic

composition are non-integer numbers. For example, the 370-nucleon peak has an isotopic composition of  $^{12}C_{1.99983724}$   $^{13}C_{0.00016274}$   $^{79}Br_{1.57871795}$   $^{81}Br_{1.42128202}$   $^{35}Cl_{2.42136339}$   $^{37}Cl_{0.57863660}$ . For each element these non-integer numbers must add to yield an integer number reflective of the chemical formula for the compound. This applies to each isotopic peak. For most isotopic peaks the errors were well under 1 ppm. This error test can be applied without recourse to any alternative methods of calculating isotopic compositions, such as the polynomial method.

A second test for errors is to calculate an integer molecular weight (nucleon number) from the isotopic composition, and compare it to the correct integer molecular weight. This error test can also be applied without recourse to any alternative method of calculating isotopic compositions. Errors in integer molecular weights were well within 1 ppm for most isotopic peaks in our example. The calculated integer molecular weight is found taking the dot product of two vectors, the first containing the isotopic composition, and the second being the corresponding integer elemental atomic isotopic masses. For example the composition vector for the 370-nucleon peak of  $C_2Br_3Cl_3$  is [1.99983724, 0.00016274, 1.57871795, 1.42128202, 2.42136339, 0.57863660], which was taken from Row 5 of Table 3, and the integer mass vector is [12, 13, 79, 81, 35, 37]. To the accuracy of the numbers listed in Table 3, the calculated integer mass of the 370-nucleon peak is 369.99999702 for an error of  $-8.1$  ppb. Two peaks in particular had significant errors, the mass 379 peak ( $-25$  ppm) and the mass 380 peak (0.74%). These are both very low abundance peaks, and the errors result from roundoff error in the output file of Mercury6.

A third error check is to compare isotopic compositions calculated by the present method with those calculated using the polynomial method. A calculation of this type for  $^{370}(C_2Br_3Cl_3)$  was performed by hand. A comparison is given in Table 4.

**Table 4.** Fractional errors in isotopic composition for  $^{370}(\text{C}_2\text{Br}_3\text{Cl}_3)$ 

Isotopic species	Expected composition <sup>a</sup>	Fractional error
$^{12}\text{C}$	1.99983726	-1.2E-08
$^{13}\text{C}$	0.00016274	8.6E-07
$^{79}\text{Br}$	1.57871794	5.9E-09
$^{81}\text{Br}$	1.42128206	-3.0E-08
$^{35}\text{Cl}$	2.42136343	-1.7E-08
$^{37}\text{Cl}$	0.57863657	4.9E-08

<sup>a</sup>Expected composition calculated by hand by the polynomial method.

A fourth error check is to compare accurate masses computed using the polynomial method to accurate masses computed using the present method. This method uses the information generated by the third error checking method, and therefore adds no additional information. Nevertheless, it is instructive because the errors determined by this method were almost identical in value to those determined by error test number 2. This therefore hints at a method for correcting molecular mass errors. As an example of an accurate mass calculation, the elemental isotopic composition of the 370 peak of  $\text{C}_2\text{Br}_3\text{Cl}_3$  is given in Row 5 of Table 3. Accurate masses of the elemental isotopes are given in column 2 of Table 1. Taking the dot product of the isotopic composition matrix with the exact mass matrix gives a value of 369.6569487 for the accurate mass 370 peak of  $\text{C}_2\text{Br}_3\text{Cl}_3$ . The error in this calculation was virtually identical to the error in the integer mass calculation presented above, -8.1 ppb. Table 5, column 2 presents results for selected isotopic peaks of  $\text{C}_2\text{Br}_3\text{Cl}_3$ . (The error listed in Table 5 for the 370-nucleon peak, -9.3 ppb, disagrees with the -8.1 ppb of the sample calculation just presented because the calculations in the Table were performed with more significant figures.)

**Error correction.** Three simple strategies can help control the effects of computational error. The first is to normalize the isotopic composition to give the correct integer values in the chemical formula. This provides correction against errors in the calculation of isotopic composition. For example, the computed  $^{12}\text{C}$  and  $^{13}\text{C}$  composition of  $^{379}(\text{C}_2\text{Br}_3\text{Cl}_3)$  added to give a value of  $\text{C}_{1.99989439}$  rather than the correct value of  $\text{C}_2$ . One

would simply rescale the isotopic composition of all the carbon isotopes so they add to the correct value of 2.

The second strategy is to rescale the accurate mass of an individual isotopic peak as follows:

$$m_{\text{new\_accurate}} = m_{\text{calc\_accurate}} \cdot m_{\text{true\_integer}} / m_{\text{calc\_integer}} \quad (5)$$

Where  $m_{\text{new\_accurate}}$  is an improved mass estimation,  $m_{\text{calc\_accurate}}$  is the "accurate" molecular mass one would calculate using the accurate elemental isotopic masses and the calculated isotopic composition of the peak,  $m_{\text{true\_integer}}$  is the true integer molecular weight of the isotopic peak, and  $m_{\text{calc\_integer}}$  is the molecular weight one would calculate from the integer elemental atomic masses and the calculated isotopic composition of the peak. This strategy enables better mass calculations without adjusting the computed molecular formulas, though it does nothing to correct the calculated isotopic composition.

The third strategy is a combination approach that sequentially applies the first strategy and then the second strategy. Applying any of the three strategies generally yields improved mass accuracy over an uncorrected calculation. Table 5 provides a sample calculation for five selected isotopic peaks of  $\text{C}_2\text{Br}_3\text{Cl}_3$ . Masses determined by the polynomial method are given for comparison. In general, the third strategy provided the most accurate results where the worst-case difference was 0.3 ppb. This was followed by the second strategy (-50 ppb worst case), and the first strategy (76 ppb worst case). Each of the three strategies provided sub-ppm accuracy, even in the case of  $^{380}(\text{C}_2\text{Br}_3\text{Cl}_3)$ , where the uncorrected mass calculation was 0.74% too high.

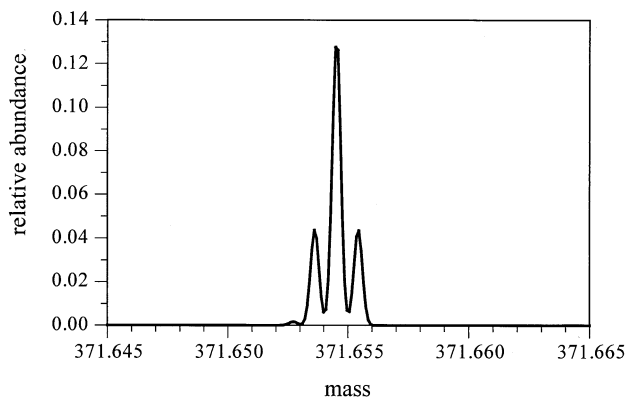
**Isotopic Variability.** The natural variability of isotopic abundances (such as the  $^{13}\text{C}/^{12}\text{C}$  abundance ratios) affects the average mass of an ion. For example, measurements of average molecular weights of proteins are limited to an accuracy of ~8 ppm due to the variability of the  $^{13}\text{C}/^{12}\text{C}$  abundance ratio [16]. One must also ask whether the variability of isotopic abundance ratios will also affect the accurate masses of individual isotopic peaks. Consider for example the 100-mer of polyethylene glycol, the protonated ion of which has a chemical

**Table 5.** Fractional errors of corrected masses of selected isotopic peaks of  $\text{C}_2\text{Br}_3\text{Cl}_3$ 

Integer mass	Expected accurate mass <sup>a</sup>	Fractional error (uncorrected) <sup>b</sup>	Fractional error (strategy 1)	Fractional error (strategy 2)	Fractional error (strategy 3)
366	365.66156730	1.3E-08	0	1.8E-12	0
367	366.66492270	9.9E-07	3.4E-09	-5.9E-11	1.5E-11
370	369.6595168	-9.3E-09	2.3E-11	-8.1E-14	-8.1E-14
379	378.64993260	-2.5E-05	7.6E-08	-2.0E-09	3.2E-10
380	379.65328800	7.4E-03	0	-5.0E-08	0

<sup>a</sup>Expected accurate masses calculated by polynomial method.

<sup>b</sup>Uncorrected masses are those calculated using the results from Table 3. Corrected masses are calculated by three strategies explained in paper.

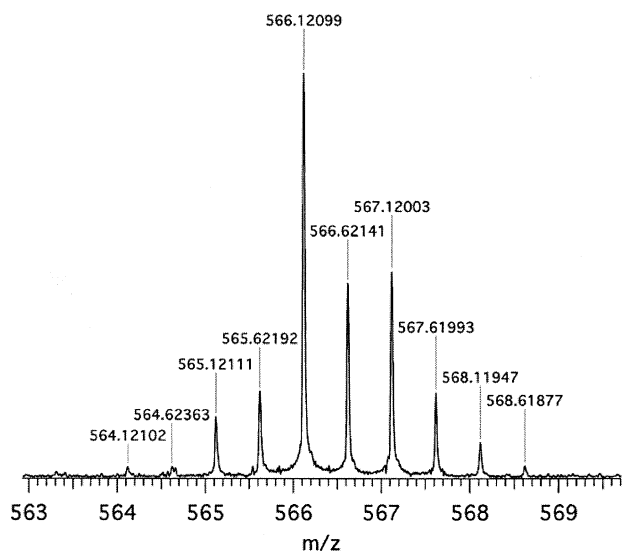


**Figure 2.** Ultrahigh resolution calculation of isotopic fine structure of isotopic peak of mass 371.655 of  $C_2Br_3Cl_3$ . Calculations performed using algorithm in reference [17] and normalized to unit area.

formula of  $C_{200}H_{403}O_{101}^+$ . Using the methods presented in this paper, and applying error correction strategy number 2, we calculated the accurate masses of all isotopic peaks of abundance greater than 0.01% of the base peak. The calculation was done twice, once using a  $^{13}C$  isotopic abundance of 1.1070% and once using an altered abundance of 1.2177%, a relative increase of 10%. As expected, the altered  $^{13}C$  isotopic abundances introduced mass shifts for the individual isotopic peaks. However, the calculated mass shifts were very small, being less than 0.15 ppm in the worst case.

*Statistical uncertainty.* In comparing theoretical calculations to experimental measurements, invariably experimental errors must be dealt with. Most are beyond the scope of this paper, but one relates directly to the mathematical properties of the isotopic distributions and can therefore be discussed here. In general a nominal isotopic peak is not composed of a single contribution, but rather on the millidalton scale most isotopic peaks consist of a cluster of closely spaced peaks [17–20]. A calculated example is shown in Figure 2. In most cases this fine structure is not resolved, though some ultrahigh resolution experiments have been reported in which this fine structure was resolved [20]. If the isotopic distribution for a single nominal isotopic peak is normalized, it can be treated as a probability distribution. In an experimental measurement of a nominal isotopic peak ions are selected at random from the distribution. Hence, the measured mass (the mean of the random selection from the distribution) will be uncertain on the millidalton scale. The statistical uncertainty of the measured mass,  $\sigma_{meas}$  depends on two factors: (1) The number of ions measured in the nominal isotopic peak,  $N_{ion}$ , and (2)  $\sigma_{ion}$ , the standard deviation of the distribution for the nominal isotopic peak:

$$\sigma_{meas} = \sigma_{ion} / \sqrt{N_{ion}} \quad (6)$$



**Figure 3.** Fourier transform mass spectrum of the isotopic cluster of  $Kr@mc5(K^+)_2$ .

This relationship sets an absolute statistical limit on the measurable mass accuracy. The standard deviation of the isotopic peak shown in Figure 2 is 0.63 millidalton, or 1.7 ppm. Therefore, achieving a standard deviation of 0.1 ppm in mass accuracy for measurements of this isotopic peak would require sampling approximately 300 ions.

*Elemental isotopic mass uncertainties.* Uncertainties in the masses of the elemental isotopes affect the masses of isotopic peaks computed from elemental compositions, though they do not affect the calculation of the elemental compositions themselves. These errors are easily estimated by simple formulas for propagation of errors, which will not be given here.

### Experimental Results

Agreement between the reference  $m/z$  values and the results of the calibration is about 0.1 ppm over a range of about 500  $m/z$  units. The calibrated spectrum was used to obtain accurate mass measurements for the various peaks of the isotope cluster corresponding to the  $Kr@mc5(K^+)_2$  inclusion complex. The region of the spectrum comprising the complex is given in Figure 3. A list of the measured accurate  $m/z$  values is included in Table 6. Also included are the theoretical accurate  $m/z$  values, calculated using reference masses and abundances from the XMASS periodic table, excerpted in Table 1. Included in the calculation are the charge (+2) and the electron mass. Error correction strategy number 3, described above, was used. However, corrections were small (0.2 ppm or less) for each of the peaks listed.

Theory and experiment agree at the low ppm level. Errors ranged from +1.4 to –3.9 ppm, taking the theoretical values as the reference point. The difference between the theoretical and experimental results



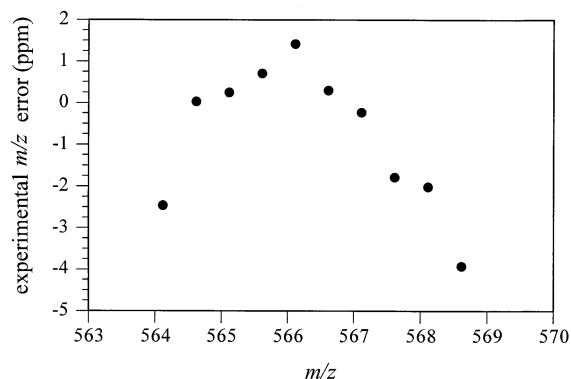
**Table 6.** Accurate  $m/z$  values for peaks in the  $\text{Kr@mc5(K}^+)_2$  spectrum. Peaks are doubly charged

Peak	Measured $m/z$	Theoretical $m/z$	$m/z$ error (ppm)
1	564.12102	564.12242	-2.47
2	564.62363	564.62362	+0.02
3	565.12111	565.12097	+0.24
4	565.62192	565.62152	+0.70
5	566.12099	566.12019	+1.41
6	566.62141	566.62125	+0.29
7	567.12003	567.12017	-0.24
8	567.61993	567.62095	-1.79
9	568.11947	568.12062	-2.02
10	568.61877	568.62100	-3.93

showed systematic trends, as seen in Figure 4. The origin of these trends is unclear. Two possibilities include space charge induced frequency shifts in the ICR cell (possibly including “local” frequency perturbations [21]) and background noise, both of which could affect the peak positions, especially for small peaks.

### Strategies for Comparison Between Experiment and Theory

Computer programs for accurate mass comparisons generally perform a semi-exhaustive search of all possible compositions to match experimental results to possible chemical compositions. The search is generally restricted to a certain mass error band and to a subset of all possible isobars, based on composition constraints. A similar semi-exhaustive search strategy could be used for isotopic masses. For high molecular weight molecules, which in general contain complex isotopic structure, the use of multiple isotopic peaks would be highly advantageous because the multitude of isotopic peaks provides redundant information. However, because of this redundancy, the comparison becomes more complicated. Instead of a single parameter to compare (agreement between experimental and theoretical mono-isotopic masses), there are multiple parameters to compare (agreement of experimental and theoretical

**Figure 4.** Systematic  $m/z$  error in experimental masses of  $\text{Kr@mc5(K}^+)_2$ .

masses for several isotopic peaks), and consequently several different forms of comparison are possible. Most involve calculating the mass difference between each experimental peak and the theoretical peaks, and then applying some kind of averaging method to the data to provide a global comparison.

Rather than discuss all possibilities we will mention one. From statistical considerations the best gauge of error is probably an RMS average with the terms weighted according to the square root of intensity. In the present example this method produces a 1.4 ppm average error.

One might ask what is the level of accuracy required for accurate mass measurements for large molecules. Clearly this will be dependent on the context. Traditional practice for small to medium molecular weight molecules has been to specify low ppm mass accuracy in order to provide an unambiguous composition. For large molecules the requirement is not as clear-cut. For example, protein compositions are restricted to combinations of amino acids and certain modifications. This greatly limits the composition space that must be searched and reduces the likelihood of ambiguity in deducing composition. Furthermore, even a non-definitive result may be very useful. For example, if one were interested in a protein generated by genetic engineering methods the composition may already be “known” with a fair degree of confidence. Mass spectrometry would then be used as an additional confirmatory tool. Suppose the accurate mass measurement were to eliminate just 95% of possible isobaric proteins. The product identity would then be confirmed to a very high confidence level (much greater than 95%), even if some isobaric proteins were not excluded by the mass measurement alone. This is analogous to the problem of diagnosing illness based on a laboratory test in which prior probabilities interact with test performance to influence the “predictive value positive” of the test [22].

### Computational Efficiency

In a semi-exhaustive search, which requires repeated calculations over a large number of trial compositions, computational efficiency is extremely important. For example, consider two algorithms, one taking 0.5 s for a single composition and a second taking 0.05 s. Both algorithms would be considered “fast” when run on a single instance, but if the semi-exhaustive search included 200 different compositions then the practical difference between these two “fast” algorithms (100 s versus 10 s) becomes dramatic.

We estimate upper limits for run times achievable for a fully integrated computer program from the following. To completely characterize the isotopic composition of all the isotopic peaks of a compound one must calculate the isotopic distribution of the parent compound and the isotopic distributions for several product compounds, one for each chemical element in the formula. The remaining operations require relatively

little computational effort and can be ignored in the estimate. Consequently, one can estimate the computational time to fully implement the algorithm in a computer program by simply summing the time required to calculate the isotopic distributions for the parent compound and product compounds. We have tested the computational speed for calculating the isotopic distribution of the hypothetical compound  $C_{50,000}H_{50,000}N_{50,000}O_{50,000}$  using the mercury6 program running on an 864 MHz Pentium-based PC. This hypothetical compound has a molecular weight above 2,000,000. This calculation required less than 0.055 s. Therefore, taking an extremely conservative estimate of 0.1 s of computational time per distribution, and assuming as many as nine chemical elements in a compound, we estimate a worst-case computational time of <1 s. Even this estimate may be 2 or more orders of magnitude too conservative, based on results from rescaling the chemical formula of the test compound to  $C_{5,000,000}H_{5,000,000}N_{5,000,000}O_{5,000,000}$ . However, full verification of timing results must await the development of a fully integrated computer program implementing the algorithm.

Based on its computational efficiency, the present method should lend itself well to applications requiring a semi-exhaustive search. The method is also advantageous for applications that do not require repetitive calculations because it gives accurate results at high speed, even when applied blindly. By contrast, polynomial methods using pruning require some art in their application because they trade speed for accuracy.

## Conclusions

The simple algorithm presented here enables the calculation of isotopic compositions of single isotopic peaks. From isotopic compositions one may then calculate accurate masses of single isotopic peaks, making accurate mass measurements useful, even for compounds where the mono-isotopic peak is not the dominant peak in the mass spectrum. The algorithm itself is exact. Errors introduced during practical calculations, such as roundoff errors, are controllable by highly effective error detection and correction strategies. The algorithm produces results that agree well with calculations based on the polynomial method. However, the present algorithm is computationally more efficient than the polynomial method and may therefore be applied to systems of virtually any isotopic complexity. The isotopic masses of a krypton inclusion complex were measured experimentally. Low ppm agreement was observed between theory and experiment.

## References

- McLafferty, F. W.; Turecek, F. Interpretation of Mass Spectra; University Science Books: Mill Valley, CA, 1993; 4th ed.; 107–109.
- Senko, M. W.; Beu, S. C.; McLafferty, F. W. Determination of Monoisotopic Masses and Ion Populations for Large Biomolecules from Resolved Isotopic Distributions. *J. Am. Soc. Mass Spectrom.* **1995**, *6*, 229–233.
- Rockwood, A. L. Relationship of Fourier Transforms to Isotope Distribution Calculations. *Rapid Commun. Mass Spectrom.* **1995**, *9*, 103–105.
- Rockwood, A. L.; Van Orden, S. L. Ultrahigh-Speed Calculation of Isotope Distributions. *Anal. Chem.* **1996**, *68*, 2027–2030.
- Rockwood, A. L.; Van Orden, S. L.; Smith, R. D. Rapid Calculation of Isotope Distributions. *Anal. Chem.* **1995**, *67*, 2699–2704.
- Kubinyi, H. Calculation of Isotope Distributions in Mass Spectrometry. A Trivial Solution for a Non-Trivial Problem. *Anal. Chim. Acta* **1991**, *247*, 107–119.
- Yergey, J. A. A General Approach to Calculating Isotopic Distributions for Mass Spectrometry. *Int. J. Mass Spectrom. Ion Phys.* **1983**, *52*, 337–349.
- Rockwood, A. L. Isotopic Composition of Isotopic Peaks. *Proceedings of the 45th ASMS Conference*; Palm Springs, CA, June, 1997; p 1221.
- Rockwood, A. L.; Kushnir, M. M.; Nelson, G. J. Dissociation of Individual Isotopic Peaks: Predicting Isotopic Distributions of Product Ions in MS<sup>n</sup>. *J. Am. Soc. Mass Spectrom.* **2003**, *14*, 311–322.
- Psi-Plot computer program, version 6.5 for Windows, Poly Software International Inc.: Pearl River, NY.
- Caravatti, P.; Allemann, M. The “Infinity Cell.” A New Trapped-ion Cell with Radiofrequency Covered Trapping Electrodes for Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Org. Mass Spectrom.* **1991**, *26*, 514–518.
- Flinn, A.; Hough, G. C.; Stoddart, J. F.; Williams, D. J. Decamethylcucurbit[5]uril. *Angew. Chem. Int. Ed. Engl.* **1992**, *31*, 1475–1477.
- Liang, Y.; Bradshaw, J. S.; Dearden, D. V. The Thermodynamic Basis for Enantiodiscrimination: Gas Phase Measurement of the Enthalpy and Entropy of Chiral Amine Recognition by Dimethyldiketopyridino-18-crown-6. *J. Phys. Chem. A* **2002**, *106*, 9665–9671.
- Senko, M. W.; Canterbury, J. D.; Guan, S.; Marshall, A. G. A High-Performance Modular Data System for Fourier Transform Ion Cyclotron Resonance Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 1839–1844.
- Audi, G.; Wapstra, A. H. The 1995 Update to the Atomic Mass Evaluation. *Nucl. Phys. A* **1995**, *595*, 409–480.
- Beavis, R. C. Chemical Mass of Carbon in Proteins. *Anal. Chem.* **1993**, *65*, 496–497.
- Rockwood, A. L.; Van Orden, S. L.; Smith, R. D. Ultrahigh Resolution Isotope Distribution Calculations. *Rapid Commun. Mass Spectrom.* **1996**, *10*, 54–59.
- Werlen, R. C. Effect of Resolution on the Shape of Mass Spectra of Proteins: Some Theoretical Considerations. *Rapid Commun. Mass Spectrom.* **1994**, *8*, 976.
- Senko, M. W. IsoPro Computer Program, version 2.0.
- Shi, S. D.-H.; Hendrickson, C. L.; Marshall, A. G. Counting Individual Sulfur Atoms in a Protein by Ultrahigh Resolution FTICR Mass Spectrometry; Experimental Resolution of Isotopic Fine Structure in Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 11532–11537.
- Masselon, C.; Tolmachev, A. V.; Anderson, G. A.; Harkewicz, R.; Smith, R. D. Mass Measurement Errors Caused by “local” Frequency Perturbations in FTICR Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2002**, *13*, 99–106.
- Shultz, E. K. Selection and Interpretation of Laboratory Procedures. *Tietz Textbook of Clinical Chemistry*; In Burtis, C. A.; Ashwood, E. A., Eds.; W. B. Saunders Company: Philadelphia, PA, 1999; 3rd ed.; 314–316.