Calculation of Mass and Abundance for Polyisotopic Ion by Low Resolution Mass Spectrometry

YU, Chong-Tian(余旻天) **GUO, Yin-Long***(郭寅龙)
Shanghai Institute of Organic Chemistry, Chinese Academy Sciences, Shanghai 200032, China
ZHANG, Zun-Jian(张尊建) **XIANG, Bing-Ren**(相秉仁) **AN, Deng-Kui**(安登魁)
Department of Pharmaceutical Analysis, China Pharmaceutical University, Nanjing, Jiangsu 210009, China

The new algorithm calculated mass and abundance for polyisotopic ion by low-resolution mass spectrometry was developed based on probability theory. The results of mass and abundance data of polyisotopic ion calculated by the algorithm were coincided with the experimental values. By comparison with the Kubinyi's method, our algorithm is simpler and easier to master for operator.

Keywords Mass spectrometry, algorithm, isotope distribution, weighting method

Introduction

Every specific chemical formula is corresponding to its characteristic isotope patterns of peaks, and these spectra are easy for us to obtain by low-resolution mass spectrometer. However, for some reasons, sometimes we need to know the more exact isotopic pattern. We can resort to the computer by contriving an appropriate algorithm because the rapid development of computer technique makes the mass calculation procedure of isotopic distribution be substituted by computer.

In recent years, various programs have been described for calculating the isotopic distributions derived from the formula of the molecules. The formula often used is the expansion of the following product of polynomial expression:[1]

$$(a_1 + a_2 + a_3 + \cdots)^m (b_1 + b_2 + b_3 + \cdots)^n (c_1 + c_2 + c_3 + \cdots)^l \cdots$$

where $a_1, a_2, a_3, \cdots, b_1, b_2, b_3, \cdots, c_1, c_2, c_3, \cdots$ represent the individual isotopes of the elements in the molecule, and the exponents m, n, l, *etc.*, are the number of atoms of each element present in the molecule. However, it will be spent a long time in calculating and large number of permutation will be generated when we deal with multinominal molecule or macromolecule by the above algorithm. Therefore, we have to limit the numbers of atoms sometimes. For example, Carrick and Glocking[2] designed a program calculating distributions of organometallic compounds, which retains the information about the individual peaks at each mass unit, but only suits to calculate the molecules which contain no exceeding three elements and five atoms for each element.

Another algorithm has ever been mentioned by Kubinyi,[3] instead of developing such polynomials, isotopic pattern of the first atom is split by that of the second atom, leading to the masses and their intensity ratios of two atoms. Then the third atom split these values again, and the new mass and intensity are exported. The procedure repeated atom by atom until all the polyiosotopic atoms have considered. However, without any further modifications of the algorithm, it is a laborious and very time-consuming procedure. Moreover, simply summing the terms by the nominal mass can lead to a little deviation.

Thus, we improve this algorithm by amalgamating the terms by the weighing method in which the mass of

* E-mail: ylguo@pub.sioc.ca.cn
Received October 9, 2000; revised and accepted January 8, 2001.
Project supported by the Natural Science Foundation of Shanghai (No. 00ZA14077), China.

every term is the vector, and the intensity% is the weight. This method not only offers us more accurate value but also avoids the trouble calculations and the broader isotope distribution.

Algorithm

The improved method applied to calculate the isotope patterns is based on probabilistic theory. First, we get the equation of new isotope distribution:

$$\begin{cases} P'_n(k) = P_o(j) \times P(i) \\ m'_n(k) = m_o(j) + m(i) \end{cases} \quad (i = 1, 2, \ldots, i_n)\ (j = 1, 2, \ldots, j_n) \tag{1}$$

in which "$m'_n(k)$" and "$P'_n(k)$" represent the new mass and abundance of the kth peak when adding a new atom, and $m_o(j)$, $P_o(j)$ is the old mass value and abundance of the kth peak as no adding new atom. "i_n", "j_n" represent separately number of isotope of specific element, the peak number which we want to get. Furthermore, the terms of "$P(i)$" and "$m(i)$" express the probability and mass for ith kind of isotope of the new atom that we want to add. Then the calculation begins from the first atom; next it is split by the pattern of the second atom, atom by atom, until all the polyisotopic atoms have been considered. Because in low resolution mass spectrometer, the peak with the same nominal mass can not be split, then the combination is generated by the under equation basing on weighing method:

$$\begin{cases} P_n(k) = P'_n(k_1) + P'_n(k_2) + \cdots + P'_n(k_{n-1}) + P'_n(k_n) \\ m_n(k) = (P'_n(k_1) \times m'_n(k_1) + P'_n(k_2) \times m'_n(k_2) + \cdots + P'_n(k_{n-1}) \times m'_n(k_{n-1}) \\ \qquad + P'_n(k_n) \times m'_n(k_n))/P_n(k) \end{cases} \tag{2}$$

in this procedure, all peaks (k_1, k_2, $\cdots$, k_n) with the same nominal mass are summed to one peak which generates new mass ($m_n(k)$) and new abundance ($P_n(k)$). For example, if the unlimited resolution is permitted by mass spectrometer, chemical formula "CH" may generate two peaks ^{13}CH and ^{12}CD which nominal mass is m/z 14. Usually, we only can be told the isotope pattern by unit mass in low-resolution mass spectrometer. Thus these two terms can be summed to one term by nominal mass 14. It is noticed that the value of mass that was applied to calculate is the real mass of the isotope of atom and not the nominal mass. Next step we can prove the similar equation below which deal with multi-peak with the same nominal mass:

$$\begin{cases} P_n(k) = P_n(k_{n-1}) + P'_n(k_n) \\ m_n(k) = (P_n(k_{n-1}) \times m_n(k_{n-1}) + P'_n(k_n) \times m_n(k_n))/P_n(k) \end{cases} \tag{3}$$

the Eq. (3) can be derived from Eq. (2). First, when n is equal to 1 or 2, it is obvious Eq. (3) is true. Similarly, we can extend it to n or $n+1$. It shows that this kind of algorithm is universally applicable in all situations, thus this algorithm is tenable. Next step we can realize this algorithm by stepwise procedure in computer.

In the computer program, the end result is the isotope distribution pattern of molecule. In the case of most calculation, the number of peaks was set to 10, which is supported by most chemical formula. However, if the polyisotopic chemical formula was calculated, the number of peaks should be set in accordance with the integral mass range of isotope pattern which will be mentioned in follow discussion. Maybe, it was unfortunate that the isotope distribution pattern of molecule given is not in high-resolution. In fact, the calculation value is enough for us to analyze in the low-resolution mass spectrometer. Because the peak with the same nominal mass can not be separated completely even if in high-resolution mass spectrometer (*e.g.* magnetic sector MS).

Experimental

Mass spectrometer

The experiment was performed on a Hewlett-Packard 5989A Quadruple mass spectrometer with a manual controllable heated sampler (scientific instrument services incorporation). The system was operated in positive ion mode. Typical source parameters were Emission current = 300 mA. Electronic impact voltage = 70 V. Mass data were acquired in scan mode scanning from 40 to 600 amu.

Sample preparation

About 0.1 mg sample (no further preparation) was laid on the bottom of the little glass tube. Then put the tube into the concave in the head of sampler. Direct insert the sampler into the ionization source, then quickly raise its temperature to 150℃.

Result and discussion

For validating the program, we examine it by some organometallic compounds, for example, ferrocene. This simple example helps to illustrate the prominent of this algorithm because of its especial polyisotopic elements. Comparing with the experimental values of mass spectrometry (see Table 1.), it is interesting to find the calculated results are coincided well with those of the experimental.

Table 1 Relative intensity and mass of organometallic compounds by experiment or calculation

Chemical formula	Exp. value of mass[a]	Calcd. value of mass	Exp. value of relative abundance[a]	Calcd. value of relative abundance
$C_{15}H_{12}BrFeN$	339	338.9500	7.48	5.97
	340	339.9582	No signal	1.04
	341	340.9504	100.00	100.00
	342	341.9533	20.12	19.65
	343	342.9484	92.58	94.22
	344	343.9512	17.75	18.39
	345	344.9529	No signal	2.03
	346	345.9545	No signal	0.16
$C_{15}H_{13}BrFeN_2$	354	353.9659	8.81	5.97
	355	354.9690	No signal	1.06
	356	355.9613	97.30	100.00
	357	356.9641	21.45	20.03
	358	357.9593	100.00	94.29
	359	358.9620	20.05	18.75
	360	359.9637	2.88	2.10
	361	360.9651	No signal	0.16
$Sn_2(C_4H_9)_2H_3$	345	344.9711	1.80	1.48
	346	345.9723	1.82	0.94
	347	346.9704	5.39	3.53
	348	347.9716	6.73	2.26
	349	348.9690	17.02	16.34
	350	349.9698	13.66	14.54
	351	350.9684	42.62	43.86
	352	351.9695	40.69	37.62
	353	352.9684	95.95	91.54
	354	353.9697	57.08	56.87
	355	354.9685	100.00	100.00
	356	355.9701	45.91	42.95
	357	356.9690	79.47	81.74
	358	357.9719	18.68	16.50
	359	358.9706	30.48	32.71

Contituned

Chemical formula	Exp. value of mass[a]	Calcd. value of mass	Exp. value of relative abundance[a]	Calcd. value of relative abundance
$Sn_2(C_4H_9)_2H_3$	360	359.9732	10.41	8.38
	361	360.9715	20.95	22.49
	362	361.9749	2.06	2.06
	363	362.9728	2.90	3.06
	364	363.9762	No signal	0.28
	365	364.9747	1.80	1.88

[a] The experiment proceed in HP5989A MS. The value was the mean of 10 times survey, and relative standard error was 0.05.

The results achieved with this program are more accurate than other algorithm used before. Usually, the data accuracy can only be kept to 1 decimal digit and effect by the pruning threshold factor,[3] however, our program offered mass data 3 decimal digits and circumscribed by the more direct factor which was easy to master for operator (the comparison result see Table 2). Because the carbon isotope ratio we applied is different to the Kubinyi applied, the isotope abundance result generated a little difference between these two algorithms.

Table 2 Comparison of results obtained with our algorithm and Kubinyi's for bovine insulin, $C_{254}H_{377}N_{65}O_{75}S_6$

Run A[a]		Run B[b]	
Mass	Isotope abundance	Mass	Isotope abundance
5729.613	14.50	5729.6	14.66
5730.617	46.79	5730.6	47.10
5731.620	82.33	5731.6	81.58
5732.625	100	5732.6	100.00
5733.621	96.89	5733.6	96.62
5734.625	78.34	5734.6	77.91
5735.625	54.76	5735.6	54.32
5736.628	33.88	5736.6	33.53
5737.629	18.88	5737.6	18.64
5738.632	9.60	5738.6	9.45
5739.631	4.49	5739.6	4.42
5740.634	1.95	740.6	1.92
5741.634	0.79	5741.6	0.78
5742.637	0.30	5742.6	0.30
5743.635	0.11	5743.6	0.11
5744.639	0.04	5744.6	0.04
5745.638	0.01	5745.6	0.01

[a] Calculated by our program, peak number = 17, $^{13}C/^{12}C = 1.112$. [b] Calculated by Kubinyi's algorithm, pruning threshold factor = 0.000001, $^{13}C/^{12}C = 1.08$.

Another advance of our algorithm is its simple operation. To decrease the calculation time without leading to less accurate result, the pruning threshold factors of the Kubinyi's method must be optimized. Before optimization, the knowledge about the program and composition of formula should be considered. However, our algorithm decreases the effect of artificial factor to the calculation, and converts this complex arithmetic problem to be a simple thing.

Program and limitation

The program that calculates the molecular ion isotopic distributions consists of a main program and two

subroutines. A molecular formula is input within main program ISOTOP, then will be got its numbers and abundance of atoms of each element present in the molecule in subroutine FIND. The major calculation is performed in subroutine CALCULATE, then the isotope pattern was induced from it and stored into the file RESULT.TXT. The numbers of peak would be asked input before calculation, and it was followed by the norm that it ought to be set to just twice by the width of isotope range of all elements in the chemical formula ($p_n = 2 \times \sum$width of isotope range of every element). Or else, it will result in big error.

There seems to be no practical limit to the number of polyisotopic atoms in the molecule. However, because of the limitation of computer we can not give the abundance value less than 1/exp(103), which often occur from large molecule whose mass is greater than 100,000. But the practical range in analyzing the mass of chemical does not exceed 1000 by the low-resolution mass spectrometer, even the mass range of FTMS does not exceed 100,000.

It is interesting that the theory isotope distribution is not in full coinciding with the value of experiment. That maybe arises from the experimental error. However, except for this reason, another major reason is that there are other quasi-molecular ion, for example $(M-H)^+$ and $(M+H)^+$ *etc*., whose isotope distribution also exists in the mass spectra. Thus, the experimental isotope pattern observed is the summation of isotope distribution which generated from all molecular ion. But fortunately, we may calculate the individual patterns of quasi-molecular ion respectively, then we combine them by weighting method. The third reason must also be noted. Some elements, carbon being the most prominent element, vary in their isotope distributions depending on the source of the material. For synthetic compounds (if the source is mainly atmospheric carbon dioxide and/or limestone), the $^{12}C/^{13}C$ ratio may be as high as 100/1.12, whereas organic matter coming from plants or animals may have a $^{12}C/^{13}C$ ratio of about 100/1.08.[3] However, the calculations in the program are done with a ratio of 100/1.1112. The influence of such small changes in the isotope abundance on the final result is relatively large for large compounds.

References

1 Yergey, J. A. *Int. J. Mass Spectrom. Ion Phys.* **1983**, *52*, 337.
2 Cärrick, A.; Glocking, F. *J. Chem. Soc. A* **1967**, 40.
3 Kubinyi, H. *Anal. Chim. Acta* **1991**, *247*, 107.

(E200010206 SONG, J.P.; DONG, L.J.)