## Diophantine Approach to Isotopic Abundance Calculations

## Chang S. Hsu

Analytical Division, Exxon Research and Engineering Co., Clinton Township, Route 22 East, Annandale, New Jersey 08801

Polynomial approaches have been used in the literature for calculating the isotopic abundance of a given formula containing polyisotopic elements. These approaches limit the number of elements in the formula being calculated. A more general and flexible computer program based on a Diophantine algorithm has been developed. The equation used for calculating the fractional abundance is converted into a logarithmic form to avoid the use of a Gamma function, which limits the number of occurrences of an element one may calculate by a computer program. To reduce computation time and error for heavy isotopic clusters, a method similar to that described by Yergey is used. In the program, the formula weight is defined and calculated. The mass assignment of the isotopic cluster is corrected for the mass defect of the elements so that the calculated output can be directly compared with low-resolution mass spectrometric measurements. This program is written in FORTRAN V for use on a Data General Eclipse C/350 computer. It is capable of handling all of the elements, from hydrogen to uranium, with stable, naturally occurring isotopes. The program is useful for identification or confirmation of organometallics and coordination compounds, as well as for middle-mass and polymeric materials whose mass spectra contain complex isotopic patterns.

Mass spectrometry was originally used by physicists as a convenient and effective way to detect isotopes. It was later developed to analyze organic materials and volatile organometallics. When these materials contain elements with multiple isotopes, their mass spectra can be very complex.

With the improvements in mass spectrometric instrumentation, ions in the "middle-mass" range (1000-10000 amu) can be detected (1). In these middle-mass ions, the number of occurrences of each element is usually large. Low abundance isotopes, which do not play a significant role in low-mass ions, can make significant contributions to the mass spectral pattern as their number of occurrences increases. For example, <sup>13</sup>C contributes only 1.1% relative to <sup>12</sup>C in nature. In the styrene molecular ion, C<sub>8</sub>H<sub>8</sub>, the most abundant isotopic peak is at m/z 104 and contains only  $^{12}\mathrm{C}$  atoms. The abundance of the C<sub>8</sub>H<sub>8</sub> ion containing one <sup>13</sup>C atom is 8.8% relative to the most abundant isotopic peak. However, for polystyrene with a degree of polymerization of 100 (C<sub>804</sub>H<sub>810</sub>), the isotopic peak containing one <sup>13</sup>C atom is about nine times as abundant as the isotopic peak containing only  ${}^{12}\mathrm{C}$  atoms. The most abundant ionic species in the molecular ion cluster has a formula  ${}^{12}\text{C}_{795}{}^{13}\text{C}_{9}{}^{1}\text{H}_{810}$  (1, 2). Therefore, complex isotopic distributions are also observed in middle-mass and heavy ions

Precise masses of different isotopic combinations are all different. For those isotopic combinations with the same nominal mass, the mass spread is often less than 50 ppm (4). Middle-mass and heavy ions, especially, often require ultrahigh resolution (>100 000) to resolve the peaks corresponding to individual isotopic combinations (5). Identification of the isotopic composition by high-resolution mass spectrometry

is, therefore, impractical. Isotopic clusters are usually identified by low-resolution mass spectrometric measurements. An isotopic peak is actually the overall unresolved isotopic combinations appearing at a particular nominal mass.

To identify complex isotopic clusters obtained from low-resolution mass spectrometric measurements, the theoretical isotopic distributions of the molecules or fragments are needed for comparisons. Calculations of isotopic abundance are usually long and tedious, almost always requiring a computer program. Several computer programs for these isotopic abundance calculations are available in the literature (4, 6–12), but most of them have only narrow capability. The number of occurrences of each element or the number of elements is usually limited. Furthermore, the mass defects of the elements are not considered in these calculations. In middle-mass ion clusters, the measured mass of an isotopic peak can be significantly different from the sum of the nominal masses of the isotopes. This discrepancy results when the mass defects of the elements are not considered.

Because of the wide range in mass in many isotopic clusters, it would be convenient to use formula weight, a single nominal mass, to describe the clusters. In this paper, the formula weight is defined as the sum of the products of the nominal mass of the most abundant isotope times the number of occurrences of that element in the molecule or fragment (13) corrected by the mass defects. This formula weight is not necessarily the same as the nominal mass of either the most abundant or the lightest isotopic peak in the cluster.

A computer program has been developed to handle all of the elements, from hydrogen to uranium, that have stable, naturally occurring isotopes. The program calculates the isotopic distribution of a given formula, defines the formula weight, and corrects the nominal mass by considering mass defects. There are no limits on mass number, number of elements, or the number of occurrences of each element. The program is written in FORTRAN V for use on a Data General Eclipse C/350 computer.

# DIOPHANTINE APPROACH TO ISOTOPIC-ABUNDANCE CALCULATIONS

The only input needed by the computer progam is the chemical formula of a molecule or a fragment. Other information needed for the calculations (the mass of the lightest isotope, the number of isotopes, the natural abundances of all the isotopes, and the mass defects for all the elements) is stored in the DATA blocks of the program. The number of isotopes is defined as one plus the difference in masses of the heaviest and the lightest isotopes of the element. This includes some isotopes that do not actually occur in nature, with their fractional abundances set to zero. The values of the natural abundances of the isotopes are taken from ref 14.

Mathematically, isotopic abundance calculations involve probablity theory and combinatorial analysis. Methods of calculation can be found in several references (15–17). All of them use polynomial approaches to find isotopic distributions among elements in a given formula, but they limit the number of elements in the formula being calculated. For each additional element, an additional factor  $(x_1 + x_2 + ...)$ , where  $x_1, x_2, ...$  represent the isotopes of the element x, is needed

Chart I

$m_{_1}$	$m_{_2}$	$m_3$
2	0	0
1	1	0
1	0	1
0	2	0
0	1	1
0	0	2

for polynomial expansion and collection of terms (12). In the early 1970s, a different approach was developed to provide simplicity and flexibility to the calculations (18); in this approach, a Diophantine equation was used to calculate all of the possible isotopic combinations of a given element. The isotopic distribution of each element in a given formula was then convoluted to give the fractional abundance of the isotope cluster. However, no details of this approach were published.

Diophantine equations can be defined generally as a set of algebraic equations with integer coefficients for which integer solutions are desired (19). The number of solutions to a specific form of Diophantine equation varies depending on the values of parameters (other than the variables) in the equation. A Diophantine restraint equation can be expressed by

$$\sum_{i=1}^{k} m_i = n \tag{1}$$

where n is the number of atoms (occurrences) of each element in a formula, k is the number of isotopes of that element, and  $m_i$  is the number of occurrences of the ith isotope of the element.

In the computer program we describe here, the number of isotopes is determined on a continuous mass basis. For example, chlorine has two naturally occurring isotopes,  $^{35}$ Cl and  $^{37}$ Cl. In the program, the number of isotopes for chlorine is defined as 3 (k = 37 - 35 + 1).

All of the possible isotopic combinations for a chlorine molecule are represented by the Diophantine output shown in Chart I, where  $m_1$  represents  $^{35}$ Cl,  $m_2$  represents  $^{36}$ Cl, and  $m_3$  represents  $^{37}$ Cl. A total of six combinations is found.

Searching for these isotopic combinations by computer is not straightforward. Figure 1 is a flow chart of the Diophantine subroutine, in which L counts the number of isotopic combinations of an element being calculated, and K1 and K2 are the control parameters. N and NI are the number of occurrences (atoms) and the number of isotopes, respectively, of the element. M(I) is the number of occurrences of the Ith isotope, where I=1,2,...,NI. In Figure 1, the M(I) values are printed after completion of each search. The isotopic calculation program does not print the M(I) values, but proceeds with the abundance calculations for each combination.

For each combination of the isotopes for a given element, the fractional abundance can be calculated by the following equation:

$$F_{i} = \frac{n!}{m_{1}!m_{2}! \dots m_{n}!} f_{1}^{m_{1}} f_{2}^{m_{2}} \dots f_{n}^{m_{n}}$$

$$= \frac{n!}{\prod_{j=1}^{k} m_{j}!} \prod_{j=1}^{k} (f_{j})^{m_{j}}$$
(2)

where  $F_i$  is the fractional abundance of the *i*th combination,  $f_j$  is the fractional abundance of the *j*th isotope, and n, k, and  $m_j$  are defined as in eq 1.

The factorials can be evaluated by using the Gamma function library program, which is stored in the computer. The largest factorial which the Data General Eclipse C/350 computer can evaluate is  $\Gamma(58) = 57! = 4.05 \times 10^{76}$ . Above

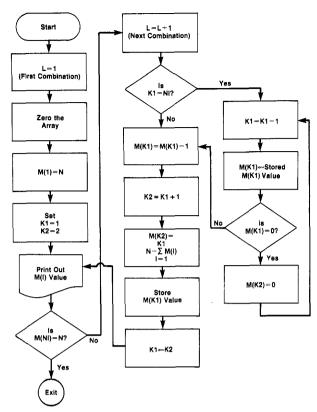


Figure 1. Flow chart of the Diophantine subroutine.

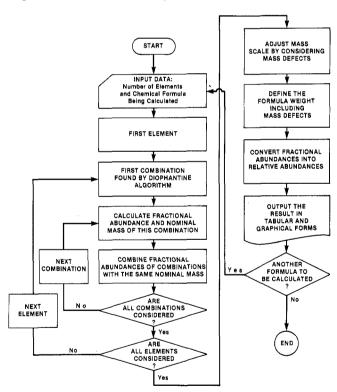


Figure 2. Algorithm of the isotopic abundance convolution program.

this, the computer will "overflow". This limits the maximum number of occurrences for an element to 57. To overcome this limitation, eq 1 is converted into a more generally useful logarithmic form

$$\log F_i = \sum_{\sigma=1}^n \log \sigma - \sum_{j=1}^k \left[ \sum_{i=1}^{m_j} \log i \right] + \sum_{j=1}^k [m_j \log f_j]$$
 (3)

The nominal mass of this combination is also calculated. Calculations of the fractional abundances and nominal masses are repeated for all of the other combinations. The fractional

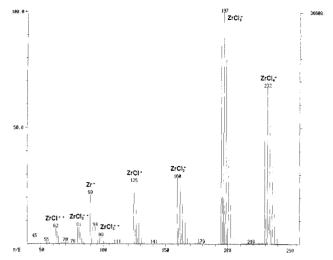


Figure 3. 70-eV electron-impact ionization mass spectrum of zirconium chloride subjected to direct-insertion probe analysis.

abundances of the same nominal masses are summed. After all of the possible combinations are considered, an isotopic distribution for that element results. The procedure is repeated for the other elements and, finally, the fractional abundances of all the nominal masses in the cluster of the given formula are obtained. This is a convoluted mass spectrum of the given formula.

The mass assignments and the formula weight are then corrected by considering the mass defects of all of the elements in the calculations. Since the mass defects of the isotopes for a given element do not differ significantly (14), only the mass defect of the most abundant isotope of each element is considered.

Finally, fractional abundances are converted into relative abundances. The results of the isotopic abundance convolution are presented in both tabular and graphical (bar plot) forms, and the formula weight is included.

Figure 2 shows the algorithm of the isotopic abundance convolution program run interactively on a Data General Eclipse C/350 computer. The chemical formula was input via a Data General Dasher D200 terminal in a repetitive A2,I3 format. The I3 format is adequate to calculate any isotopic distribution for the clusters measurable by current mass spectrometers. Copies of the listing are available from the author.

## CALCULATION OF FRACTIONAL ABUNDANCES FOR LARGE ISOTOPIC CLUSTERS

For calculating the fractional abundances of a large isotopic cluster, the straightforward logarithmic approach becomes inefficient, because both computation time and error increase due to numerous repetitive calculations.

Let us assume that a given formula contains 500 atoms of an element with two isotopes. The first three isotopic combinations obtained by eq 1 are

$$m_1 \ ({\rm isotope} \ 1) \qquad m_2 \ ({\rm isotope} \ 2) \\ 500 \qquad 0 \\ 499 \qquad 1 \\ 498 \qquad 2$$

With the direct logarithmic approach, there are more than 2000 computer manipulations (call subroutine for logarithm, additions, subtractions, and multiplications) for EACH combination. Thus, more than 6000 calculations are required for these three combinations.

The number of calculations can be reduced substantially if we use the ratio of two consecutive combinations, a method

```
THE CONVOLUTED MASS SPECTRUM FOR ZR 1CL 3
MZZ
           198
199
200
   .04686
.21577
.01499
   .11495
 201
202
 203
        12.91
 204
   .00000
   .00621
        2.15
SUM = 1.00000
       345.80
*** THE FORMULA WEIGHT IS
THE CONVOLUTED MASS SPECTRUM FOR ZR 1CL
 M/Z
    F. A.
        R.A.
           SUM = 1.00000 365.83
```

Figure 4. Computer outputs of the calculated isotopic patterns for ZrCl<sub>3</sub><sup>+</sup> (left) and ZrCl<sub>4</sub><sup>+</sup> (right).

\*\*\* THE FORMULA WEIGHT IS

THE CONVOLUTED MASS SPECTRUM FOR C 24 H 20GE 1

\*\*\* THE FORMULA WEIGHT IS 382 \*\*\*

Figure 5. Calculated mass spectrum for the molecular ion of tetraphenylgermane.

THE CONVOLUTED MASS SPECTRUM FOR C 14 H 20 O 3MO 1GE 1

```
F.A.
398
     .02753
               18.28
               .00444
399
400
     .04620
.10589
.07552
               30.68
70.32
50.15
401
402
403
404
406
406
407
408
409
410
     .13979
.10328
.15059
              92.83
69.59
100.00
     .07654
.11814
.03097
               52.22
78.45
20.57
                       .04945
.00773
.00719
               32.84
413
414
415
     .00100
                  .73
     .00012
                  .08
     .00000
     .99970
              663.86
```

\*\*\* THE FORMULA WEIGHT IS 408 \*\*\*

**Figure 6.** Calculated mass spectrum for the molecular ion of  $\pi$ -C<sub>5</sub>H<sub>5</sub>(CO)<sub>3</sub>MoGe(C<sub>2</sub>H<sub>5</sub>)<sub>3</sub>.

similar to that suggested by Yergey (20). For the ith combination, the fractional abundance is

$$F_i = \frac{n!}{\prod_{i=1}^k m_j!} \prod_{j=1}^k (f_j)^{m_j} = n! \prod_{j=1}^k \frac{(f_j)^{m_j}}{(m_j)!}$$
(4)

Similarly, the fractional abundance of the (i + 1)th combination is

$$F_{i+1} = n! \prod_{j=1}^{k} \frac{(f_j')^{m_j'}}{(m_j')!}$$
 (5)

If we divide eq 5 by eq 4, we obtain

$$\frac{F_{i+1}}{F_i} = \prod_{j=1}^k \frac{(m_j)! \ (f_j')^{m_{j'}}}{(f_i)^{m_j}(m_{j'})!} = \prod_{j=1}^k \rho_j \tag{6}$$

where  $\rho_j = [(m_j)!(f_j')^{m_j}]/[(f_j)^{m_j}(m_j')!]$  and  $f_j = f_j'$  for each isotope. The value of  $\rho_j$ , where j = 1, 2, ..., k, can be obtained by one of the following three equations, depending on the values of  $m_j$  and  $m_j'$ :

$$\rho_j = \frac{(m_j)(m_j - 1) \dots (m_j' + 1)}{(f_j)^{(m_j - m_j')}} \quad \text{if } m_j > m_j' \quad (7)$$

$$\rho_j = 1 \qquad \text{if } m_j = m_j' \tag{8}$$

$$\rho_j = \frac{(f_j)^{(m_j - m_j)}}{(m_j / (m_j - 1) \dots (m_j + 1)} \quad \text{if } m_j < m_j$$
 (9)

These equations can be converted to logarithmic forms as shown in eq 10-12, respectively.

$$\log \rho_j = \sum_{\sigma=m_j'+1}^{m_j} \log \sigma - (m_j - m_j') \log f_j \qquad (10)$$

$$\log \rho_i = 0 \tag{11}$$

$$\log \rho_j = (m_j' - m_j) \log f_j - \sum_{\sigma = m_j + 1}^{m_j} \log \sigma \qquad (12)$$

For any isotopic combination with  $m_j = n$  for  $j = \lambda$  and  $m_j = 0$  for  $j \neq \lambda$ , where  $\lambda = 1, 2, ..., k$ , the fractional abundance can be obtained by combining eq 4 with eq 1

$$F_i = f_{\lambda}^n$$

or

$$\log F_i = n \log f_{\lambda} \tag{13}$$

With this modified approach, the total number of calculations for the first three combinations of n = 500 and k = 2 is reduced from more than 6000 to fewer than 25, thereby reducing both computation time and error.

## APPLICATIONS

One of the uses of this computer program is to recognize the isotopic patterns of clusters containing polyisotopic elements. Usually, a complete formula of the molecule or fragment is not needed. For organic compounds, the most

```
THE CONVOLUTED MASS SPECTRUM FOR C 804 H 810
а
 MZZ
    F.A.
        R.A.
10464
    000004
          .08 1
10465
    .00036
         3.50 1***
10467
10468
    .00507
    .01156
10469
    .02108
            10470
10471
10472
    .03198
    .04153
        87.42
            10473
10474
10475
   .04751
.04303
.03540
            | 本学文法查查者表考生表考生的方式。
            90.59
        10476
10477
10478
            .02665
   .01850
        **********
10479
10480
   .00715
         8.46
12481
   .00212
            | ****
10482
10483
         1.05
.47
.20
   .00050
10484
    .00022
   .00009
10486
   .00004
          .08 1
   .00001
10487
10489
   .00000
   .35867 754.97
*** THE FORMULA WEIGHT IS 10464 ***
THE CONVOLUTED MASS SPECTRUM FOR C 804 H 810
b|
M/Z
    F.A.
10464
   .00011
10465
   .00101
10466
10467
        10468
   .03224
10469
    05876
   .08915
   .11579
.13143
        10473
        10473
   .13245
        100.00
10474
   .11998
10475
10476
   .07431
        10477
   .05159
10478
10479
   .01994
           ! *******
! ***
10480
   .01121
         8.46
         2.23 144
10432
   .00295
         1.05
10483
    .00139
         .47
10485
   .00026
10486
   .00011
10487
   .00002
         .01
10489
   .00001
SHM =
   .99997 754.97
```

Figure 7. Calculated isotopic pattern for a polystyrene oligomer ion,  $C_{804}H_{810}$  (degree of polymerization = 100): (a) using eq 3; (b) using eq 10-13.

characteristic isotopic patterns are observed in the mass spectra of chlorine- and/or bromine-containing compounds. These patterns can be used to identify the distribution of chlorine/bromine in organic molecules because they are not significantly altered by the presence of other elements, such as carbon, hydrogen, oxygen, and nitrogen. Beynon has calculated the chlorine/bromine isotopic patterns for as many as eight halogen atoms (17). Similary, for most organometallics the isotopic patterns can be recognized by determining the patterns produced by the polyisotopic transition metals that are present (4).

\*\*\* THE FORMULA WEIGHT IS 10464 \*\*\*

A typical comparison of experimental and calculated results is shown in Figures 3 and 4. Figure 3 is a 70-eV electron-impact-ionization mass spectrum of zirconium chloride acquired on a Finnigan 4000 mass spectrometer. The most significant ions in the spectrum are the clusters corresponding to  ${\rm ZrCl_3}^+$  and  ${\rm ZrCl_4}^+$ . Figure 4 shows the calculated isotopic

patterns of these two clusters in both tabular and graphic forms. There is excellent agreement between the calculated and experimental isotopic patterns for these zirconium chloride species.

Formula weight, as defined in the introduction, is the sum of the products of the precise mass of the most abundant isotope times its occurrence in a given formula. It can be the same as the nominal mass of the lightest, most abundant, or any isotopic peak. As in most cases, the formula weights of the zirconium chloride ion clusters are the same as the nominal masses of the lightest isotopic peaks (195 amu for  $ZrCl_3^+$  and 230 amu for  $ZrCl_4^+$ ). For tetraphenylgermane (12),  $(C_6H_5)_4Ge$ , the formula weight is the same as the nominal mass of the most abundant isotopic peak (382 amu), as shown in Figure 5. In other cases, formula weight is not the same as the nominal mass of either the lightest or the most abundant isotopic peak, as shown in Figure 6 for  $\pi$ - $C_5H_5(CO)_3MoGe$ -

 $(C_2H_5)_3$  (408 amu) (4). When the formula weight is different from the nominal mass of the lightest isotopic peak, the isotopic peak corresponding to the formula weight is not monoisotopic (containing only one isotope contributed by each element).

To demonstrate the capabilities of this program in calculating "middle-mass" and heavy ion clusters, a theoretical isotopic pattern was calculated for a polystyrene oligomer with a degree of polymerization of 100,  $C_{804}H_{810}$ . Figure 7 shows the results of this calculation. If only the nominal masses of carbon and hydrogen are used, the formula weight is 10 458 amu. When the mass defects of carbon and hydrogen are considered, the formula weight is 10464 amu. The most abundant isotopic peak in this oligomer ion cluster is found at 10473 amu, as pointed out by Matsuo et al. (2).

Figure 7 demonstrates the problems of using any polyisotopic material as a high-mass standard, for instance, polystyrene as suggested by Matsuo et al. Isotopic patterns in the mass spectra of these materials are extremely complex and may be distorted by mass spectrometric measurements. In the C<sub>804</sub>H<sub>810</sub> ion cluster, the intensities of several isotopic peaks are so similar that the computer may select the wrong isotopic peak for mass calibration. The problem is similar in the mass spectra of several oligomer ion clusters of hydrocarbons or perfluorohydrocarbons with carbon numbers greater than 90, and it can become even more complicated if protonated or hydride-abstracted molecular ions are present.

The results in Figure 7a were obtained by using eq 3 for the abundance calculations. The fractional abundance at each nominal mass is the probability of finding all of the ionic species with the specific nominal mass in an isotopic cluster. The sum of all of the fractional abundances in an isotopic cluster should be unity. By use of the straightforward logarithmic method of eq 3, the total fractional abundance of the polystyrene cluster is 0.359 because of the loss of significance from numerous calculations. With the modified logarithmic method of eq 10-13, the total fractional abundance is 1.000, as shown in Figure 7b. However, the same relative abundances are obtained by either method. The computation time is reduced from 4.5 min for the straightforward logarithmic method to less than 25 s for the modified method.

## CONCLUSION

This paper introduces a Diophantine approach for calculating distributions of stable, naturally occurring isotopes for a given chemical formula by computer. The most obvious uses of the computer program described are for the mass spectral identification and confirmation of organometallics, coordination compounds, and middle-mass and polymeric materials. This program can be extended to the deconvolution of overlapped isotopic clusters, the analysis of isotopic labeling, and the calculations of isotopic distributions in various types of mass spectrometry/mass spectrometry (MS/MS) spectra (21).

### ACKNOWLEDGMENT

The author wishes to thank Robert W. Kiser of the University of Kentucky for his encouragement and advice in writing the initial program. Discussion with T. R. Ashe and R. J. Pancirov of Exxon Research and Engineering Co. during manuscript preparation are also appreciated.

### LITERATURE CITED

- (1) Fenselau, C. Anal. Chem. 1982, 54, 105A-116A
- Matsuo, T.; Matsuda, H.; Katahuse, I. Anal. Chem. 1979, 51, 1329-1331.
- Yergey, J.; Heller, D.; Hansen, G.; Cotter, R. J.; Fenselau, C. Anal. Chem. 1983, 55, 353-356.
- (4) Carrick, A.; Glockling, F. J. Chem. Soc. A 1967, 40–42.
   (5) Fenselau, C.; Yergey, J.; Heller, D. Int. J. Mass Spectrom. Ion. Phys. 1983, 53, 5–20.
- Lee, J. D. Talanta 1973, 20, 1029-1038.
- VanderVelde, G. Quant. Chem. Progr. Exch. 1974, 11, 262
- Dombek, B. D.; Lowther, J.; Carberry, E. J. Chem. Educ. 1971, 48,
- Dombek, B. D.; Lowther, J.; Carberry, E. Quant. Chem. Progr. Exch. 1976, 11, 294,
- Mattson, B.; Carberry, E. J. Chem. Educ. 1973, 50, 511.
- Mattson, B.; Carberry, E. Quant. Chem. Progr. Exch. 1976, 11, 298. Brownawell, M. L.; San Filippo, J., Jr. J. Chem. Educ. 1982, 59,
- 663-665.
- (13) Robinson, R. J.; Warner, C. G.; Gohlke, R. S. J. Chem. Educ. 1970, 47, 467–468.
- Weast, R. C. "CRC Handbook of Chemistry and Physics", 63rd ed.;
- Chemical Rubber Publishing Co.: Boca Ratan, FL, 1982-1983.

  (15) Kiser, R. W. "Introduction to Mass spectrometry and Its Applications":
- Prentice-Hall: Englewood Cliffs, NJ, 1965; pp 229–231.

  Margrave, J. L.; Polansky, R. B. J. Chem. Educ. 1962, 39, 335–337.

  Beynon, J. H. "Mass Spectrometry and Its Application to Organic Chemistry"; Elsevier: Amsterdam, 1960; pp 71–83.
- (18) Hsu, C. S. Ph.D. Dissertation, University of Kentucy, 1974.
  (19) "Van Nostrand's Scientific Encyclopedia", 5th ed.; Van Nostrand-Reinhold: New York, 1976; p 805. Yergey, J. A. *Int. J. Mass Spectrom. Ion Phys.* 1983, *52*, 337–349. Singleton, K. E.; Cooks, R. G.; Wood, K. V. *Anal. Chem.* 1983, *55*,
- 762-764.

RECEIVED for review November 9, 1983. Accepted March 5,