match no. 24 pp. 181-189 1989

GROUP THEORETICAL DETERMINATION OF MOLECULAR FORMULAS FOR GIVEN HIGH-RESOLUTION MASS

Tetsuo MORIKAWA
Department of Chemistry, Joetsu University of Education
Joetsu, Niigata Prefecture 943, Japan

(Received: September 1988)

(Abstract) It is shown that decimal fractions observed for molecules /fragments in high-resolution mass spectrometry have a cyclic group structure. Expansion of the decimal fraction into a finite continued fraction gives an approximate proper fraction whose denominator decides the order of the cyclic group. When a prime number is selected as the order, the group becomes a Galois field (a finite field). Group-theoretical consideration for chemical compounds leads to a linear congruence (an equation in several unknowns) that determines the numbers of isotopes/fragments in the corresponding molecule. The congruent equation is solvable with the multiplication (and/or division) of integers.

1. Introduction

Chemists can observe the molecular weight of an unknown organic compound by high-resolution mass spectroscopy. The numerical value is based on the standard isotope carbon-twelve (^{12}C = 12 mass units, exact). Hence they encounter the problem: Determine the molecular formula using only one numerical value. Here the term "molecular formula" means a set of isotopes/atoms and of the number of each. No good algorithm for solving this problem is known; any method in use is fundamentally by trial and error. Molecular weights for various compounds have been calculated; each table is a massive volume for organic compounds containing only a few isotopes such as ^{12}C, ^{1}H, ^{16}O, and ^{14}N. One of the famous tables was made by Beynon[1]. Lederberg[2] somewhat improved the situation; however, his method also uses a large list of numerical values.

The purpose of the present paper is to solve the problem described above by algebraic theory. We note the fact that an observed molecular weight is made up of two parts, an integer and a decimal fraction. In the following it will be shown that the part of decimals in molecules has a group theoretical structure; the integer part (or the total mass number) is out of the focus in this paper[3]. A cyclic additive group[4] of order (or period) m, Z_m = { 1, 2, ..., m-1, 0}, plays an important role; Z_m is isomorphic to the quotient group Z/mZ ; that is, $Z_m \cong Z/mZ$. Here Z = {..., -2, -1, 0, 1, 2, ...} is an additive group composed of rational integers, and mZ = {..., -2m, -m, 0, m, 2m, ...} is a subgroup of Z. Z_m becomes a Galois field if m is a prime number.

2. Group Structure of Decimal Fractions for Isotopes/Fragments

We consider a series of molecules. Particular attention in this chapter is paid to only one kind of isotopes (or molecular fragments) in the molecules. Let d be a decimal fraction for the relative atomic weight of an isotope in question; for example, d = 0.007825037 in mass units[5] for ^{1}H(= 1.007825037). A set of numbers, { x·d : x=1, 2, 3, ... }, for the series can then be constructed; x represents the number of isotopes. Obviously this set is an additive semigroup generated by d. We restrict every number in the set by assuming that when x·d becomes a numerical value containing an integer (positive), then the integer part is removed from that x·d ; this is denoted by {x·d}. The assumption is considered proper under the condition that each high-resolution mass can be separated into two numbers, a positive integer and a decimal fraction. Then the set consists of at most finite members because {x·d} ultimately becomes zero under the assumption; the set is abbreviated as <d>.

The following is a simple algorithm for determining the number of members of <d>. First, d is converted into a proper (rational) fraction; second, the proper fraction is transformed into an irreducible fraction r/m (m > r > 0), i.e., both numerator and denominator are divided by the greatest common divisor[6]; then, m is the answer. The reason is that m is the smallest number by which the multiplication of the proper fraction for d becomes an integer. In order to get the proper fraction r/m approximating a decimal fraction d we use a method of continued fraction[7] ; in this process the numerator is relatively prime to the denominator.

$$d = \cfrac{1}{t_1 + \cfrac{1}{t_2 + \cfrac{1}{t_3 + \ddots + \cfrac{1}{t_n}}}}$$

where t_n is called the n th term.

Calculated results for atomic weights[5] of isotopes and molecular fragments are shown below using the n th term approximation. Data in parentheses represent the error, the absolute value |d - r/m| in 10^{-6} mass units.

^{1}H = 1.007825037: <1/127> (49), <1/128> (13), <4/511> (2.8), <5/639> (0.3), ...

^{13}C = 13.003354839 : <1/298> (0.9), ...

^{14}N = 14.003074008 : <1/325> (2.9), <3/976> (0.2), ...

^{16}O = 15.99491464 : (Note8), <195/196> (17), <196/197> (9), <391/393> (3.7), <587/590> (0.6), ...

$^{12}C^{1}H_2$ = 14.015650074: <1/63> (223), <1/64> (25), <9/575> (2), <10/639> (0.6), ...

$^{14}N^{1}H$ = 15.010899045: <1/91> (90), <1/92> (29), <4/367> (0.1), ...

It is clear that a mapping between Z_m and <r/m> is isomorphic, which is defined, for example, by $f(k) = [k \cdot r]_m/m$ onto <r/m>, $Z_m \ni k$, where $[k \cdot r]_m$ is the remainder of the division of $k \cdot r$ by m; $<r/m> \cong Z_m$. In other words, the set <r/m> is a cyclic group of order m.

3. Linear Congruence in One Unknown

It is thus obtainable an equation that determines the number of isotopes/fragments. Let d_W be the decimal fraction of observed mass for an unknown compound, and let x be the unknown number of isotopes/fragments. We have

$$d_W = \{x \cdot d\} = \{x \cdot r/m\} = [r \cdot x]_m/m$$

which means a linear congruence

$$r \cdot x \equiv h \pmod{m}$$

where $h = m \cdot d_W$ is chosen as an integer. This congruence is always solvable for x because r and m are relatively prime[9].

The following is an easy way to treat actual problems. We previously calculate $r \cdot x \equiv 1 \pmod{m}$, in which none of the parameters is related to d_W; the solution is denoted by a ; namely, $r \cdot a \equiv 1 \pmod{m}$. Then the multiplication of these two congruences, $r \cdot x \equiv r \cdot a \cdot h \pmod{m}$, leads to

$$x \equiv a \cdot h \pmod{m}$$

If $a \cdot h > m$, then $[a \cdot h]_m$ is calculated. Values of a, for example, are: 128 for <4/511>, 128 for <5/639>, 651 for <3/976>, 195 for <195/196>, 196 for <196/197>, 196 for <391/393>, 393 for <587/590>, 64 for <9/575>, 64 for <10/639>, and 92 for <4/367>.

Example 1. $d_W = 0.2348$ is given, and a hydrocarbon containing only ^{12}C and ^{1}H is assumed. To this case <4/511> and a = 128 are applied. Then, $511 \cdot 0.2348 = 199.98 \ldots$, h = 120; therefore, $x \equiv 128 \cdot 120 \equiv 30 \pmod{511}$, the number of hydrogens. Alternatively, the same solution results using <1/128>, namely, $128 \cdot 0.2348 = 30.05 \ldots$, h = 30; therefore, $x \equiv 30 \pmod{128}$. The number of carbons can now be easily determined from the integer part of the molecular mass.

Example 2. The molecular weight 248.2140 for a compound containing ^{12}C, ^{1}H, and only one ^{16}O, is assumed. Using <1/128> we have 128(0.2140 + (1 - 0.99491)) = 28.04..., and h = 28. The number of carbons is (248 - 28 - 16)/12 = 17; therefore, the molecular formula is expressed as $C_{17}H_{28}O$.

It is well-known that Z_p is one of the Galois fields if p is a prime number. In Galois fields the operations of addition, subtraction, multiplication, and division, i.e., the four arithmetic operations, are possible, just as in the set of rational numbers[10]. We have tried to find approximate fractions r/p for several isotopes/fragments, where a fraction r/p approximating a given d satisfies $|d - r/p| < |d - r'/p'|$ for any other fraction r'/p' ; p and p' are prime numbers with 3 places. The results are as follows. The absolute value |d - r/p| in 10^{-6} mass units is given in parentheses.

^{1}H : <3/383> (7.9), $3^{-1} = 128$.
^{13}C : <2/599> (16), $2^{-1} = 300$.
^{14}N : <3/977> (3.4), $3^{-1} = 326$.
^{16}O : <978/983> (1.1), $978^{-1} = 393$.
$^{12}C^{1}H_2$: <6/383> (16), $6^{-1} = 64$.
$^{14}N^{1}H$: <4/367> (0.1), $4^{-1} = 92$.

4. Simultaneous Equations for Isotopes/Fragments in Molecules

As Example 2 shows, the linear congruence with one unknown can be applied to molecules containing several kinds of isotopes/fragments. This chapter deals with another way to solve the problem simultaneously.

Let us consider the case in which there are two kinds of isotopes /fragments in molecules. Then the decimal fraction d_w for an observed molecular weight must be equal to one of the terms in the set $\{ x_1d_1 + x_2d_2 : x_1, x_2 = 1, 2, 3, \ldots \}$. This set under the same assumption as in chapter 2 is designated by $<r_1/m_1, r_2/m_2>$, where r_j/m_j is an approximate proper fraction for d_j. If each of the proper fractions in the previous chapters is used, then the period of $<r_1/m_1, r_2/m_2>$ is too long for calculation; that is because the order of this set is decided by the greatest common divisor of m_1 and m_2. A new decimal fraction appropriate for the set must be derived from d_1 and d_2.

Applying to d_1 and d_2 the same algorithm as for the greatest common divisor of two integers, and neglecting the remainder of decimals in the course of iterations, we can obtain an approximate decimal d for d_1 and d_2, where d_j is divisible by d approximately; namely, d_j/d is expressed as an integer b_j. An approximate proper fraction for that d can now be calculated by the continued fraction method; it is written as r/m. Thus the irreducible form r_j/m_j is obtained from the proper fraction b_jr/m.

We get

$$\begin{aligned} dw &= \{x_1r_1/m_1 + x_2r_2/m_2\} \\ &= \{b_1r\, x_1/m + b_2r\, x_2/m\} \\ &= [b_1r\, x_1 + b_2r\, x_2]_m/m \end{aligned}$$

which gives a congruence in two unknowns

$$r(b_1x_1 + b_2x_2) \equiv h \pmod m$$

where $h = m \cdot d_w$ is selected as an integer.

Similar treatment is possible for three kinds of isotopes/fragments.

The following are examples calculated for molecules containing several isotopes.

	$<r/m>$	$<r_1/m_1, r_2/m_2, \ldots>$	Congruent Equations
^1H and ^{13}C	<1/897>	<7/897,1/299>	$7x_1 + 3x_2 \equiv h \pmod{897}$
^1H and ^{14}N	<1/716>	<3/358,1/179>	$6x_1 + 4x_2 \equiv h \pmod{716}$
^1H and ^{16}O	<1/985>	<8/985,196/197>	$8x_1 + 980x_2 \equiv h \pmod{985}$
^1H, ^{14}N and ^{16}O	<1/716>	<3/358,1/358,178/179>	$6x_1 + 2x_2 + 712x_3 \equiv h \pmod{716}$

Note: $897 = 3 \cdot 13 \cdot 23$, $716 = 2^2 \cdot 179$, $985 = 5 \cdot 197$.

Example3. The molecular weight 249.2754 for a hydrocarbon (^{13}C enriched) is assumed. Then $897 \cdot 0.2754 = 247.03\ldots$, $h = 247$; $7x_1 + 3x_2 \equiv 247 \pmod{897}$. Putting $x_1 = 3k + 1$, we obtain $3x_2 \equiv 3(80 - 7k) \pmod{897}$, which gives $x_2 \equiv 80 - 7k \pmod{299}$. The chemical restrictions for hydrocarbons are $2x_2 + 2 \geq x_1$, and even x_1; therefore, the molecular formula is written as $^{13}C_{17}{}^1H_{28}$.

Notes and References

1. J. H. Beynon, "Mass Spectrometry and its Applications to Organic Chemistry", Elsevier, Amsterdam,1960.
2. J. Lederberg, "Computation of Molecular Formulas for Mass Spectrometry", Holden-Day, San Francisco, 1964.
3. A method for determining molecular formulas using only the integer part has been discussed, for example, in J. Chem. Educ., 63, 1053 (1986).

4. Cyclic additive groups. Refer to G. Birkhoff and T. C. Bartee, "Modern Applied Algebra", McGraw-Hill, New York 1970, chap. 7.

5. Relative atomic weights are listed in "Element by Element Review of Their Atomic Weights", Pure & Appl. Chem. 56, 695(1984), IUPAC.

6. The greatest common divisor of two positive integers can be calculated, for example., by Euclid's algorithm. H. Davenport, "The Higher Arithmetic", Cambridge Univ., Cambridge, 1962, chap. 1.

7. Continued fractions. Chap. 4. in Ref. 6

8. The first approximation is equal to 1.

9. Congruent equations can be solved, for example, by Euclid's algorithm; chap. 2 in Ref. 6.

10. Galois fields. Chap.12 in Ref 4.