match no. 25 pp. 165-176 1990

ENUMERATION OF RATIONAL FORMULAS (REPRESENTATIVES FOR VALENCE ISOMERIC CLASSES) FOR MOLECULAR COMPOUNDS

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

(Received: March 1990)

(Abstract) A method is developed that helps us enumerate all the rational formulas for given chemical compounds containing carbon, hydrogen, oxygen, and chlorine atoms; the rational formula can be interpreted as a representative for a class of valence isomers which have been introduced by A. T. Balaban. This method makes use of generating functions such as $L(u;C,H,N)M(C,H,N,O)$ for hydrocarbons, where $L(u;C,H,N) = (1 - CH^2)(C^{u-1}N^2 - N^{2u})/(C - N^2) + C^{u-1}$, and $M(C,H,N,O) = 1/\{(1 - CH^2)(1 - NH)(1 - O)\}$; the powers of the parameters C, H, N, and O indicate the numbers of chemical functional groups >C<, $-CH_3$, >CH-, and $-CH_2-$, respectively; u is the cyclomatic number in graph theory.

1. Introduction

Balaban[1-4] called "valence isomers" a subclass of constitutional (structural) isomers which have a common set of vertex degrees (valencies) in hydrogen-depleted (molecular) graphs. One class of valence isomers for a given molecular formula can then be constructed by means of partition of groups having degree 1 (terminal groups)

separately from groups having degrees 2, 3, or 4 ; in the chemical notation, the class is represented by a type of molecular formula in which the atoms are grouped together by round brackets. Such a representative can thus be regarded as being a rational formula[5]. For example, two compounds, 2-methyl-3-pentene and trimethylcyclopropane, are valence isomers for the molecular formula C_6H_{12}; they share the same rational formula $(CH)_3(CH_3)_3$ and the same set {3, 3, 3, 1, 1, 1} of valencies.

The present note will describe a generating function method applicable to the enumeration of rational formulas for given molecular formulas containing both several kinds of and any number of atoms. This method is more general than that for acyclic hydrocarbons reported in Reference 5.

2. Rational-Formula Generating Functions for Hydrocarbons

Our starting point of the discussion is to assume that a given hydrocarbon C_nH_m is made up of functional groups >C<, >CH-, $-CH_2-$, and $-CH_3$; they have valency 4, 3, 2, and 1, and are symbolized by C, N, O, and H, respectively. Not all given sets of valencies are realizable as multigraphs (or molecular constitutional structures); mathematical restriction should be imposed on the set[6]. The necessary and sufficient condition has been established by Senior[7]: (i) The sum of valencies is an even number. (ii) The sum of valencies is greater than or equal to twice the maximum valency. (iii) The sum of valencies is greater than or equal to twice the number of valencies minus 1. We need to satisfy Senior's condition for a given set of valencies.

Senior's restriction (iii) for C_nH_m states that the number $4n + m - 2(n + m - 1) = 2n - m + 2$ is non-negative; we can put it into $2u \geq 0$ because of restriction (i); u is just the cyclomatic number[8] in graph theory. Note that the inequality $u \leq n$ follows from $m \geq 2$.

It is possible to derive two equations

$n_C + n_N + n_O + n_H = n$ from the number of carbon atoms, and

$n_N + 2n_O + 3n_H = m$ from the number of hydrogen atoms

where n_C, n_N, n_O, and n_H stand for the numbers of the functional groups. Subtraction of the latter equation from twice the former one leads to

$$2n_C + n_N - n_H = 2n - m$$

which is free of n_O. Therefore, we obtain the Diophantine equation

$$n_H + 2u = 2n_C + n_N + 2$$

in which the variables are all non-negative. Each solution of the Diophantine equation, if it fulfils Senior's restriction (ii), can be regarded clearly as being a rational formula that is realizable as at least one molecular structure. We try to express the solution using a power term of the parameters C, N, O, and H, and then to add all the power terms together; the final expression is a form of generating function.

In a particular case $u = 0$ the Diophantine equation is read as $n_H = 2n_C + n_N + 2$, for which the generating function has been reported in Reference 5; that is, $H^2M(C,H,N)$, where $M(C,H,N) = 1/\{(1 - CH^2)(1 - NH)\}$. For $u = 1$ we have the Diophantine equation $n_H = 2n_C + n_N$ for which the

generating function is simpler than that for the case $u = 0$; namely, $M(C,H,N)$.

Let us consider a general case for given $u > 1$; then the Diophantine equation becomes $n_H + 2(u - 1) = 2n_C + n_N$. If n_C is equal to 0, then a sequence of power terms

$$N^{2(u-1)},\ N^{2(u-1)}NH,\ N^{2(u-1)}(NH)^2,\ N^{2(u-1)}(NH)^3,\ \ldots \text{ (ad infinitum)}$$

is determined in a way that n_H runs from zero to infinity; the sum of them leads to a series $N^{2(u-1)}/(1 - NH)$. In a similar manner, if n_C equals 1, then a series of power terms for the running of n_H and n_N is calculated:

$$N^{2(u-2)}C + N^{2(u-2)}CNH + N^{2(u-2)}C(NH)^2 + \ldots = N^{2(u-2)}C/(1 - NH)$$

If $n_C = k < u - 1$ in general, we can obtain a series $N^{2(u-k-1)}C^k/(1 - NH)$. It is readily seen that if $n_C = u - 1$, the sum of power terms for $n_H = 0, 1, 2, 3, \ldots$ takes the form $C^{u-1}/(1 - NH)$. If $n_C > u - 1$, then the series

$$C^uH^2/(1 - NH),\ C^uH^2CH^2/(1 - NH),\ C^uH^2(CH^2)^2/(1 - NH),\ \ldots$$

are obtainable for the running of n_C; the summation of them is given by $C^uH^2/\{(1 - CH^2)(1 - NH)\}$.

Summing all the above series up from $n_C = 0$ to infinity we have

$$\{N^{2(u-1)} + N^{2(u-2)}C + N^{2(u-3)}C^2 + \ldots + N^2C^{u-2} + C^{u-1} + C^uH^2/(1-CH^2)\}/(1- NH)$$

$= \{(C^{u-1}N^2 - N^{2u})/(C - N^2) + C^{u-1}/(1 - CH^2)\}/(1 - NH)$

$= \{(1 - CH^2)(C^{u-1}N^2 - N^{2u})/(C - N^2) + C^{u-1}\}/\{(1 - CH^2)(1 - NH)\}$

This function is denoted by L(u;C,H,N)M(C,H,N), where

$$L(u;C,H,N) = (1 - CH^2)(C^{u-1}N^2 - N^{2u})/(C - N^2) + C^{u-1}$$

Multiplying L(u;C,H,N)M(C,H,N) by 1, O^1, O^2, O^3, ..., and adding them together, we get the required function L(u;C,H,N)M(C,H,N,O) for C_nH_m, where M(C,H,N,O) = M(C,H,N)/(1 - O).

Table 1. List of L(u;C,H,N) for u = 0 to 10.

u	L(u;C,H,N)
0	H^2
1	1
2	$(1 - CH^2)N^2 + C$
3	$(1 - CH^2)(N^4 + N^2C) + C^2$
4	$(1- CH^2)(N^6 + N^4C + N^2C^2) + C^3$
5	$(1- CH^2)(N^8 + N^6C + N^4C^2 + N^2C^3) + C^4$
6	$(1- CH^2)(N^{10} + N^8C + N^6C^2 + N^4C^3 + N^2C^4) + C^5$
7	$(1- CH^2)(N^{12} + N^{10}C + N^8C^2 + N^6C^3 + N^4C^4 + N^2C^5) + C^6$
8	$(1- CH^2)(N^{14} + N^{12}C + N^{10}C^2 + N^8C^3 + N^6C^4 + N^4C^5 + N^2C^6) + C^7$
9	$(1- CH^2)(N^{16} + N^{14}C + N^{12}C^2 + N^{10}C^3 + N^8C^4 + N^6C^5 + N^4C^6 + N^2C^7) + C^8$
10	$(1-CH^2)(N^{18} + N^{16}C + N^{14}C^2 + N^{12}C^3 + N^{10}C^4 + N^8C^5 + N^6C^6 + N^4C^7 + N^2C^8) + C^9$

3. Methods of Enumeration for Hydrocarbons

All the rational formulas for a hydrocarbon C_nH_m $(2u = 2n - m + 2 \geq 0)$ can be produced by the coefficient of t^n in the expansion of

L(u;Ct,Ht,Nt)M(Ct,Ht,Nt,Ot) because only one carbon atom is involved in each functional group. The function L(u;Ct,Ht,Nt) is the numerator in the totally generating function, so that it affects the starting power of the expansion; M(Ct,Ht,Nt,Ot) specifies an increase in the powers. The expansion of M(Ct,Ht,Nt,Ot) has been written[5] as

$$M(Ct,Ht,Nt,Ot) = \sum_{n=0}^{\infty} M_n(C,H,N,O)t^n,$$

$$M_n(C,H,N,O) = M_n(C,H,N) + O^1M_{n-1}(C,H,N) + O^2M_{n-2}(C,H,N) + \ldots + O^{n-2}M_2(C,H,N) + O^n,$$

$$M_n(C,H,N) = (NH)^{r-s}(CH^2)^s\{ (NH)^{3k} + (NH)^{3(k-1)}(CH^2)^2 + \ldots + (NH)^3(CH^2)^{2(k-1)}\},$$

and if $r \geq s$, then $(NH)^{r-s}(CH^2)^s(CH^2)^{2k}$ is added to $M_n(C,H,N)$; $n = 2(3k + r) + s$, $0 \leq r < 3$, and $0 \leq s < 2$.

In order to get the number of rational formulas for the hydrocarbon C_nH_m with u we replace each parameter in the totally generating function by t; namely, the coefficient of t^n in the expansion of L(u;t,t,t)M(t,t,t,t) indicates the number of rational formulas. Using $M_n(1,1,1) = k$ $(r < s)$, $M_n(1,1,1) = k + 1$ $(r \geq s)$, and summing them up from k = 0 to k-1, we have

$$(0\cdot6 + 5) + (1\cdot6 + 5) + (2\cdot6 + 5) + \ldots + ((k-1)6 + 5) = 3k^2 + 2k$$

The coefficient $M_n(1,1,1,1)$ of t^n in the expansion of M(t,t,t,t) is thus calculated as in Table 2.

Table 2. List of $M_n(1,1,1,1)$ for $n = 2(3k + r) + s$, $0 \le r < 3$, $0 \le s < 2$.

	$s = 0$	$s = 1$
$r = 0$	$3k(k + 1) + 1$	$(3k + 1)(k + 1)$
$r = 1$	$(3k + 2)(k + 1)$	$(3k + 3)(k + 1)$
$r = 2$	$(3k + 4)(k + 1)$	$(3k + 5)(k + 1)$

Hence by use of the function

$$L(u;t,t,t) = t^{u-1} + t^{u} + t^{u+1} + t^{u+2} - t^{2u-1} - t^{2u} - t^{2u+1}$$

the general solution that gives the number of rational formulas for C_nH_m with u is expressed by

$$M_{n-u+1}(1,1,1,1) + M_{n-u}(1,1,1,1) + M_{n-u-1}(1,1,1,1) + M_{n-u-2}(1,1,1,1)$$
$$- M_{n-2u+1}(1,1,1,1) - M_{n-2u}(1,1,1,1) - M_{n-2u-1}(1,1,1,1)$$

where $M_q(1,1,1,1)$ are ignored if $q < 0$ has happened in practice. It should be noted that Senior's restriction (ii) comes into effect when small molecules are treated; for example, refer to Example (b) in the next Chapter.

4. Calculation Examples for Hydrocarbons

We now make use of the generating function method in order to determine all rational formulas for given hydrocarbons.

(a) The molecular formula $C_{2p}H_{2p}$, $u = p + 1$, is given. In the particular case of C_4H_4 ($p=2$) the generating function is written as

$$C^2M_2(C,H,N,O) + CN^2M_1(C,H,N,O) + N^4M_0(C,H,N,O) = C^2(NH + O^2) + CN^2O + N^4;$$

in chemical words, these power terms can be interpreted as $C_2(CH)(CH_3)$, $C_2(CH_2)_2$, $C(CH)_2(CH_2)$, and $(CH)_4$, respectively; the number of rational formulas is $M_{4-2}(1,1,1,1) + M_{4-3}(1,1,1,1) + M_{4-4}(1,1,1,1) = 2 + 1 + 1 = 4$.

In the case of C_6H_6 ($p = 3$), the generating function becomes

$$C^3M_3(C,H,N,O) + C^2N^2M_2(C,H,N,O) + CN^4M_1(C,H,N,O) + N^6M_0(C,H,N,O)$$
$$= C^3(CH^2 + NHO + O^3) + C^2N^2(NH + O^2) + CN^4O + N^6,$$

and $3 + 2 + 1 + 1 = 7$. The above results for C_4H_4 and C_6H_6 agree with those reported by A. T. Balaban[3].

The effective terms in $L(p+1;C,H,N)$ with $p > 3$ take the form

$$N^{2p} + N^{2(p-1)}C + N^{2(p-2)}C^2$$
$$+ (1 - CH^2)\{N^{2(p-3)}C^3 + N^{2(p-4)}C^4 + \dots + N^2C^{p-1}\} + C^p$$

in which the total power of each term is at most the number $2p$ of carbon atoms; then $t^p + t^{p+1} + t^{p+2} + t^{p+3}$. Therefore, the general solution for $p > 3$ is expressed by

$$C^pM_p(C,H,N,O) + N^2C^{p-1}M_{p-1}(C,H,N,O) + \dots + N^{2(p-1)}CM_1(C,H,N,O)$$
$$+ N^{2p}M_0(C,H,N,O) - CH^2\{N^{2(p-3)}C^3M_0(C,H,N,O) + N^{2(p-4)}C^4M_1(C,H,N,O) + \dots$$
$$+ N^2C^{p-1}M_{p-4}(C,H,N,O)\},$$

and the number of rational formulas is equal to

$$Mp(1,1,1,1) + Mp\text{-}1(1,1,1,1) + Mp\text{-}2(1,1,1,1) + Mp\text{-}3(1,1,1,1)$$

For example, C_8H_8 (p = 4):

$C^4M_4(C,H,N,O) + N^2C^3M_3(C,H,N,O) + N^4C^2M_2(C,H,N,O) + N^6CM_1(C,H,N,O)$
$+ N^8M_0(C,H,N,O) - CH^2N^2C^3M_0(C,H,N,O)$
$= C^4(N^2H^2 + CH^2O + NHO^2 + O^4) + N^2C^3(CH^2 + NHO + O^3) + N^4C^2(NH + O^2)$
$+ N^6CO + N^8 - CH^2N^2C^3$
$= C^4N^2H^2 + C^5OH^2 + C^4NO^2H + C^4O^4 + C^3N^3OH + C^3N^2O^3 + C^2N^5H + C^2N^4O^2$
$+ CN^6O + N^8$,

and the number 4 + 3 + 2 + 1 = 10.

Another example $C_{10}H_{10}$ (p = 5):

$C^5M_5(C,H,N,O) + N^2C^4M_4(C,H,N,O) + N^4C^3M_3(C,H,N,O) + N^6C^2M_2(C,H,N,O)$
$+ N^8CM_1(C,H,N,O) + N^{10}M_0(C,H,N,O) - CH^2N^4C^3M_0(C,H,N,O)$
$- CH^2N^2C^4M_1(C,H,N,O)$
$= C^5(CH^3N + N^2H^2O + CH^2O^2 + NHO^3 + O^5) + N^2C^4(N^2H^2 + CH^2O + NHO^2 + O^4)$
$+ N^4C^3(CH^2 + NHO + O^3) + N^6C^2(NH + O^2) + N^8CO + N^{10} - CH^2N^4C^3$
$- CH^2N^2C^4O$
$= C^6NH^3 + C^5N^2OH^2 + C^6O^2H^2 + C^5NO^3H + C^5O^5 + C^4N^3O^2H + C^4N^2O^4$
$+ C^4N^4H^2 + C^3N^5OH + C^3N^4O^3 + C^2N^7H + C^2N^6O^2 + N^8CO + N^{10}$,

and 5 + 4 + 3 + 2 = 14.

(b) We consider next the case where the molecular formula is written as C_nH_{2n}, u = 1. Every rational formula is able to be calculated by means of $M_n(C,H,N,O)$ because $L(1;C,H,N) = 1$. There is only one solution O^2 for C_2H_4 although $M_2(C,H,N,O) = NH + O^2$, because Senior's restriction (ii) is required. Senior's restriction (ii) for C_3H_6 also excludes the first power term in $M_3(C,H,N,O) = CH^2 + NHO + O^3$.

The general expression for $n > 3$ is exactly $M_n(C,H,N,O)$; the number of rational formulas is given by $M_n(1,1,1,1)$; for example, $M_{50}(1,1,1,1) = 234$ for $C_{50}H_{100}$. Note that the above result is different from the general answer $H^4M_{n-2-4}(C,H,N,O) + H^3M_{n-2-3}(C,H,N,O) + H^2M_{n-2-2}(C,H,N,O) + H^1M_{n-2-1}(C,H,N,O)$ for alkenes C_nH_{2n} in Reference 5; these functions are derived from four compounds $C_{n-2}H_{2n}(C{=}C)$, $C_{n-2}H_{2n-1}(HC{=}C)$, $C_{n-2}H_{2n-2}(H_2CC)$, and $C_{n-2}H_{2n-3}(H_3CC)$.

(c) We are given the molecular formula C_nH_{2n-2}, $u = 2$. It follows from $L(2;C,H,N) = (1 - CH^2)N^2 + C$ that if $2 \le n \le 4$, the generating function is $C^1M_{n-1}(C,H,N,O) + N^2M_{n-2}(C,H,N,O)$, and that if $n > 4$, the generating function becomes

$$C^1M_{n-1}(C,H,N,O) + N^2M_{n-2}(C,H,N,O) - CH^2N^2M_{n-5}(C,H,N,O),$$

and the number of rational formulas is $M_{n-1}(1,1,1,1) + M_{n-2}(1,1,1,1) - M_{n-5}(1,1,1,1)$. For example, C_5H_8:

$C^1M_4(C,H,N,O) + N^2M_3(C,H,N,O) - CH^2N^2M_0(C,H,N,O)$

$= C(N^2H^2 + CH^2O + NHO^2 + O^4) + N^2(CH^2 + NHO + O^3) - CH^2N^2$

$= CN^2H^2 + C^2OH^2 + CNO^2H + CO^4 + N^3OH + N^2O^3$;

these power terms correspond to the results derived by A. T. Balaban[3]. Another example, $C_{50}H_{98}$. The number of rational formulas equals $M_{49}(1,1,1,1) + M_{48}(1,1,1,1) - M_{45}(1,1,1,1) = 225 + 217 - 192 = 250$.

(d) The molecular formula $C_{10}H_{16}$ (involving adamantanes), $u = 3$.

$L(3;t,t,t) = t^2 + t^3 + t^4 - t^6 - t^7$.

$M_{10-2}(1,1,1,1) + M_{10-3}(1,1,1,1) + M_{10-4}(1,1,1,1) - M_{10-6}(1,1,1,1) - M_{10-7}(1,1,1,1)$

$= 10 + 8 + 7 - 4 - 3$

= 18, the number of valence isomeric classes; refer to Conclusion in Reference 4.

The molecular formula $C_{14}H_{20}$ (involving diamantanes), u = 5.

$L(5;t,t,t) = t^4 + t^5 + t^6 + t^7 - t^9 - t^{10} - t^{11}$.

$M_{14-4}(1,1,1,1) + M_{14-5}(1,1,1,1) + M_{14-6}(1,1,1,1) + M_{14-7}(1,1,1,1)$

$- M_{14-9}(1,1,1,1) - M_{14-10}(1,1,1,1) - M_{14-11}(1,1,1,1)$

$= 14 + 12 + 10 + 8 - 5 - 4 - 3$

$= 32$.

The compound $C_{22}H_{28}$ (such as tetramantanes), u = 9.

$M_{22-9+1}(1,1,1,1) + M_{22-9}(1,1,1,1) + M_{22-9-1}(1,1,1,1)$

$+ M_{22-9-2}(1,1,1,1) - M_{22-18+1}(1,1,1,1) - M_{22-18}(1,1,1,1)$

$- M_{22-18-1}(1,1,1,1)$

$= 24 + 21 + 19 + 16 - 5 - 4 - 3$

$= 68$.

5. Enumeration for Compounds Containing Oxygens and Chlorines

The method described in the previous chapters can easily be applied to the enumeration of rational formulas for compounds containing heteroatoms. We first treat the simplest case where the molecular formula C_nH_mO containing only one oxygen atom is given; it becomes separated into two distinct cases; the one is $(C_nH_m)O$, ethers, and the other is $C_nH_{m-1}(OH)$, alcohols. The answer for the former case is of course equivalent to the one for hydrocarbons. For the latter the Diophantine equation becomes $n_H + 2u = 2n_C + n_N + 1$ where $4n + m - 1 + 1 - 2(n + m-1 + 1 -1) = 2n - m + 2 = 2u$ because the valency of OH is 1. The integer 1 in this Diophantine equation slides powers of the parameters C, H, N; in other words, the integer affects only the numerator of the rational-formula generating function. We have

$H/\{(1 - CH^2)(1 - NH)\}$ for $u = 0$, $\{N + HC/(1 - CH^2)\}/(1 - NH)$ for $u = 1$, $\{N^3 + NC + HC^2/(1 - CH^2)\}/(1 - NH)$ for $u = 2$, $\{N^5 + N^3C + NC^2 + HC^3/(1 - CH^2)\}/(1 - NH)$ for $u = 3$, ... , $\{N(N^{2(u-1)} + N^{2(u-2)}C + ... + N^2C^{u-2} + C^{u-1}) + HC^u/(1 - CH^2)\}/(1 - NH)$ for given u.

A similar treatment is possible for the molecular formula C_nH_mCl.

The molecular formula $C_nH_{m-2}(OH)_2$ is given. Then the Diophantine equation has the form $n_H + 2u = 2n_C + n_N$. Therefore, the totally generating function is written as $L(u+1;C,H,N)M(C,H,N,O)$ for given u.

References

1. A. T. Balaban, Rev. Roum. Chim. 11, 1097 (1966).
2. A. T. Balaban, Rev. Roum. Chim. 15, 463 (1970).
3. A. T. Balaban, Rev. Roum. Chim. 18, 635 (1973).
4. A. T. Balaban, Rev. Roum. Chim. 31, 795 (1986).
5. T. Morikawa, MATCH 23, 35 (1988). The line "1985 Daubeny . . . " of page 46, in Reference 1, should be read as "1850 Daubeny . . .".
6. F. Harary, "Graph Theory", Addison-Wesley Pub., Reading (1972), Chap. 6, p. 57.
7. J. K. Senior, Amer. J. Math. 73, 663 (1951).
8. N. L. Biggs, E. K. Lloyd, and R. J. Wilson, "Graph Theory 1736-1936", Clarendon Press, Oxford (1977), Chap. 8, p. 132. See also A. T. Balaban, in "Chemical Applications of Graph Theory" edited by A. T. Balaban, Academic Press, London (1976), Chap. 5, p. 102.