match no. 23 pp. 35-46 1988

ALGEBRAIC GENERATION OF RATIONAL FORMULAS FOR ACYCLIC CHEMICAL COMPOUNDS

Tetsuo Morikawa

Department of Chemistry, Joetsu University of Education,

Joetsu, Niigata Prefecture 943, Japan

(Received: March 1988)

(Abstract) A method of molecule-generating function is developed for constructing and enumerating all the rational formulas for given acyclic molecules such as alkanes, alkenes, alkynes, alkyl radicals, and alkyl alcohols. A generating function for the alkanes is of the form $H^2(1 - O)^{-1}(1 - NH)^{-1}(1 - CH^2)^{-1}$, where the exponents of the parameters C, N, O, and H, respectively, represent the numbers of chemical functional groups $>C<$, $>CH-$, $-CH_2-$, and $-CH_3$; every term in the power series expansion of the generating function can be interpreted as not only a molecular formula that is realizable as at least one hydrogen-suppressed graph (carbon-tree) but also a rational formula that is made up of the functional groups. If each parameter in the above function is replaced by t, then the coefficient of t^n in the resulting function is just the total number of rational formulas for the alkanes C_nH_{2n+2}.

1. Introduction

In this paper the term "molecular formula" means only a collection of chemical symbols that constitutes at least one molecular structure (graph), and if a given set of chemical functional groups forms a molecular formula, then the set is called "a rational formula[1]". The concept of rational formula plays an important role in modern chemical analysis such as nuclear magnetic resonance and infrared spectrometries[2] ; the resulting spectra help us find partial structures in the molecule corresponding to an unknown chemical substance, each partial structure often being expressed as a chemical functional group. It should be noted that there are different kinds of rational formulas with the same molecular formula, and that in many cases there are more than two constitutional isomers with the same rational formula. For example, the molecular formula C_3H_8O gives the rational formulas $(C_3H_7)OH$ and $(C_3H_8)O$; the former alcohol has two constitutional isomers, and the latter ether has only one constitutional isomer.

This paper will present a method which generates and enumerates rational formulas for given acyclic molecules. The method involves two major steps: First, a symbolic function, which may be called a molecule-generating function[3], is established for a class of rational formulas in question, and then the function is expanded into a power series whose coefficients answer the problem.

2. Molecule-Generating Function

How to construct a function that generates rational formulas for the alkanes, typical saturated molecules in organic chemistry, will be described in some detail. Assume that the general formula C_nH_{2n+2} (except the case n = 1) is composed of functional groups $>C<$, $>CH-$, $-CH_2-$, and $-CH_3$ (with valencies 4, 3, 2, and 1) ; these functional groups are denoted by C, N, O, and H, respectively.

The number of carbon atoms in the alkane is equal to the summation over all the numbers of the functional groups :

$$n_C + n_N + n_O + n_H = n$$

where each of the subscript n's is used to designate the number of symbols. Similarly the number of hydrogen atoms in the alkane is represented by:

$$n_N + 2n_O + 3n_H = 2n + 2$$

Subtracting twice the former equation from the latter equation, we have the relation

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

Note that neither n nor n_O makes any contribution to this relation, and that $n_H \geqq 2$; these facts will be used in the following.

If $n_H = 2$, then the relation gives $n_C = 0$ and $n_N = 0$; this can be denoted by H^2. Similarly, if $n_H = 3$, then $n_C = 0$ and $n_N = 1$; thus using the power expression we have H^3N. The addition of them leads to:

$$
\begin{aligned}
& H^2 \\
& + H^3N \\
& + H^4(N^2 + C) \\
& + H^5(N^3 + NC) \\
& + H^6(N^4 + N^2C + C^2) \\
& + H^7(N^5 + N^3C + NC^2) \\
& + H^8(N^6 + N^4C + N^2C^2 + C^3) \\
& + H^9(N^7 + N^5C + N^3C^2 + NC^3) \\
& + \ldots
\end{aligned}
$$

$$= \frac{H^2}{1 - NH} + \frac{CH^4}{1 - NH} + \frac{C^2H^6}{1 - NH} + \frac{C^3H^8}{1 - NH} + \ldots.$$

$$= H^2(1 - NH)^{-1}(1 - CH^2)^{-1} = H^2M(C, H, N)$$

Summing up over all the multiplications of $M(C, H, N)$ by 1, O^1, O^2, O^3, ..., we obtain the function $H^2M(C, H, N, O)$ required for the alkanes, where

$$M(C, H, N, O) = (1 - O)^{-1}(1 - NH)^{-1}(1 - CH^2)^{-1}$$

Notice that each term in $H^2M(C, H, N, O)$ can be considered to be a molecular formula that is realizable as hydrogen-suppressed graphs (carbon-trees[4]).

The general formula C_nH_{2n+1} can be rewritten as $(C^nH^{2n+2})H^{-1}$ in the power expression, so that $H^1M(C, H, N, O)$ can be used for calculation; in a similar manner, $H^3M(C, H, N, O)$ for C_nH_{2n+3}.

3. Rational-Formula Generation for the Alkanes

We can now describe how to generate rational formulas for the alkanes. The parameters C, H, N in M(C, H, N) are replaced by Ct, Nt, and Ht because each parameter includes one carbon atom in the alkane. Here t is a variable.

$$M(Ct, Ht, Nt) = 1 + NHt^2 + CH^2t^3 + N^2H^2t^4 + \dots$$

$$= \sum_{n=0}^{\infty} M_n(C, H, N)t^n$$

In order to obtain the general form $M_n(C, H, N)$ we have to count the number of rational formulas; hence each parameter is set as 1 ; then

$$M(t, t, t) = (1 - t^2)^{-1}(1 - t^3)^{-1} = 1 + t^2 + t^3 + \dots$$

$$= \sum_{n=0}^{\infty} M_n(1, 1, 1)t^n$$

The numerical coefficient $M_n(1, 1, 1)$ can easily be determined as follows[5] :

$$M_n(1, 1, 1) = \begin{cases} k & r < s \\ k + 1 & r \geqq s \end{cases}$$

where $n = 2(3k + r) + s$, $0 \leqq r < 3$, $0 \leqq s < 2$ $(n \geqq 0)$.

An algorithm to solve this problem is very simple : Divide n by 2, and get the quotient and the remainder s; divide the above quotient by 3, and get the quotient k and the remainder r; compare r with s, and select k or k + 1.

The integer n in $M_n(C,H,N)$ is equal to twice the number of NH plus treble the number of CH^2, so that $M_n(C, H, N)$ for $n \geqq 2$ is expressed using the parameters k, r, and s as

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

$$+ \dots + (NH)^{3+r-s}(CH^2)^{s+2(k-1)}$$

$$= C^sH^{3k+r+s}N^{3k+r-s} + C^{s+2}H^{3(k-1)+r+s+4}N^{3(k-1)+r-s}$$

$$+ \dots + C^{s+2(k-1)}H^{3+r+s+4(k-1)}N^{3+r-s} ,$$

and if $r \geqq s$, then

$$(NH)^{r-s}(CH^2)^{s+2k} \quad \text{or} \quad C^{s+2k}H^{r+s+4k}N^{r-s}$$

is added. Note: $M_1(C,H,N) = 0$, and $M_0(C,H,N) = 1$.

We can thus obtain the coefficient of t^n in the power series expansion of $M(Ct, Ht, Nt, Ot)$,

$$M_n(C,\ H,\ N,\ O) = O^n M_0(C,\ H,\ N) + O^{n-1}M_1(C,\ H,\ N)$$

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

$$= O^n + O^{n-2}M_2(C,\ H,\ N) + O^{n-3}M_3(C,\ H,\ N)$$

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

The recursion formula is given by

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

It is clear that every term in $H^2 M_{n-2}(C,\ H,\ N,\ O)$ corresponds to a rational formula for the alkane C_nH_{2n+2} ($n \geqq 2$), and that the total number of rational formulas is given by $M_{n-2}(1,1,1,1)$, that is, by the coefficient of t^{n-2} in

$$M(t,\ t,\ t,\ t) = (1 - t)^{-1}(1 - t^2)^{-1}(1 - t^3)^{-1}$$

$$= 1 + t + 2t^2 + \ldots$$

$$= \sum_{n=0}^{\infty} M_n(1,1,1,1)t^n$$

The recursion formula is then

$$M_n(1,\ 1,\ 1,\ 1) - M_{n-1}(1,\ 1,\ 1,\ 1) = M_n(1,\ 1,\ 1)$$

which gives $M_n(1, 1, 1, 1) = \sum_{i=0}^{n} M_i(1,1,1)$.

Example. $C_{10}H_{22}$ (decanes). $10 - 2 = 8 = 2(3\cdot 1 + 1) + 0$, $k = 1$, $r = 1$, $s = 0$. We have $(CH_3)_6(CH)_4$, $C_2(CH_3)_7CH$, $C(CH_3)_6(CH)_2(CH_2)$, $(CH_3)_5(CH)_3(CH_2)_2$, $C_2(CH_3)_6(CH_2)_2$, $C(CH_3)_5(CH)(CH_2)_3$, $(CH_3)_4(CH)_2(CH_2)_4$, $C(CH_3)_4(CH_2)_5$, $(CH_3)_3CH(CH_2)_6$, and $(CH_3)_2(CH_2)_8$. The total number is $M_8(1,1,1,1) = 10$. Note: The number of constitutional isomers $= 75$ (Ref. 4).

Example. $C_{50}H_{102}$ (pentacontanes).

$$M_{48}(1,1,1,1) - M_{47}(1,1,1,1) + M_{47}(1,1,1,1) - M_{46}(1,1,1,1)$$

$$+ \ldots + M_6(1,1,1,1) - M_5(1,1,1,1) = (8 + 1) + \sum_{k=1}^{7}(6k + 5) =$$

212, $M_5(1,1,1,1) = 5$. Then $M_{48}(1,1,1,1) = 212 + 5 = 217$.

4. Enumeration for Alkens, Alkynes, and Related Compounds

(a) Alkynes. The general formula C_nH_{2n-2} (alkynes) can be divided into two separate parts $C_{n-2}H_{2(n-2)+2}(C\equiv C)$ and $C_{n-2}H_{2(n-2)+1}(C\equiv CH)$. Evidently $H^2M_{(n-2)-2}(C, H, N, O)$ can be applied to the former, and the latter can be calculated by use of $H^1M_{(n-2)-1}(C,H,N,O)$. The total number of rational formulas then equals $M_{n-4}(1,1,1,1) + M_{n-3}(1,1,1,1)$.

Example. $C_{12}H_{22}$ (dodecynes). $H^2M_8(C, H, N) = H^6N^4 + C^2H^7N$;

TABLE 1. Calculated values for $M_n(C,H,N)$.

n	$M_n(C, H, N)$
0	1
1	0
2	NH
3	CH^2
4	N^2H^2
5	CH^3N
6	$C^2H^4 + N^3H^3$
7	CH^4N^2
8	$C^2H^5N + N^4H^4$
9	$C^3H^6 + CH^5N^3$
10	$C^2H^6N^2 + N^5H^5$
11	$C^3H^7N + CH^6N^4$
12	$C^4H^8 + C^2H^7N^3 + N^6H^6$
13	$C^3H^8N^2 + CH^7N^5$
14	$C^4H^9N + C^2H^8N^4 + N^7H^7$
15	$C^5H^{10} + C^3H^9N^3 + CH^8N^6$
16	$C^4H^{10}N^2 + C^2H^9N^5 + N^8H^8$
17	$C^5H^{11}N + C^3H^{10}N^4 + CH^9N^7$
18	$C^6H^{12} + C^4H^{11}N^3 + C^2H^{10}N^6 + N^9H^9$
19	$C^5H^{12}N^2 + C^3H^{11}N^5 + CH^{10}N^8$
20	$C^6H^{13}N + C^4H^{12}N^4 + C^2H^{11}N^7 + N^{10}H^{10}$
21	$C^7H^{14} + C^5H^{13}N^3 + C^3H^{12}N^6 + CH^{11}N^9$
22	$C^6H^{14}N^2 + C^4H^{13}N^5 + C^2H^{12}N^8 + N^{11}H^{11}$
23	$C^7H^{15}N + C^5H^{14}N^4 + C^3H^{13}N^7 + CH^{12}N^{10}$
24	$C^8H^{16} + C^6H^{15}N^3 + C^4H^{14}N^6 + C^2H^{13}N^9 + N^{12}H^{12}$
25	$C^7H^{16}N^2 + C^5H^{15}N^5 + C^3H^{14}N^8 + CH^{13}N^{11}$
26	$C^8H^{17}N + C^6H^{16}N^4 + C^4H^{15}N^7 + C^2H^{14}N^{10} + N^{13}H^{13}$
27	$C^9H^{18} + C^7H^{17}N^3 + C^5H^{16}N^6 + C^3H^{15}N^9 + CH^{14}N^{12}$
28	$C^8H^{18}N^2 + C^6H^{17}N^5 + C^4H^{16}N^8 + C^2H^{15}N^{11} + N^{14}H^{14}$
29	$C^9H^{19}N + C^7H^{18}N^4 + C^5H^{17}N^7 + C^3H^{16}N^{10} + CH^{15}N^{13}$
30	$C^{10}H^{20} + C^8H^{19}N^3 + C^6H^{18}N^6 + C^4H^{17}N^9 + C^2H^{16}N^{12} + N^{15}H^{15}$
31	$C^9H^{20}N^2 + C^7H^{19}N^5 + C^5H^{18}N^8 + C^3H^{17}N^{11} + CH^{16}N^{14}$
32	$C^{10}H^{21}N + C^8H^{20}N^4 + C^6H^{19}N^7 + C^4H^{18}N^{10} + C^2H^{17}N^{13} + N^{16}H^{16}$
33	$C^{11}H^{22} + C^9H^{21}N^3 + C^7H^{20}N^6 + C^5H^{19}N^9 + C^3H^{18}N^{12} + CH^{17}N^{15}$
34	$C^{10}H^{22}N^2 + C^8H^{21}N^5 + C^6H^{20}N^8 + C^4H^{19}N^{11} + C^2H^{18}N^{14} + N^{17}H^{17}$

TABLE 2. Calculated $M_n(1,1,1,1)$ in the form $n/M_n(1,1,1,1)$.

	r= 0	0	1	1	2	2
k	s= 0	1	0	1	0	1
0	0/1	1/1	2/2	3/3	4/4	5/5
1	6/7	7/8	8/10	9/12	10/14	11/16
2	12/19	13/21	14/24	15/27	16/30	17/33
3	18/37	19/40	20/44	21/48	22/52	23/56
4	24/61	25/65	26/70	27/75	28/80	29/85
5	30/91	31/96	32/102	33/108	34/114	35/120
6	36/127	37/133	38/140	39/147	40/154	41/161
7	42/169	43/176	44/184	45/192	46/200	47/208
8	48/217	49/225	50/234	51/243	52/252	53/261
9	54/271	55/280	56/290	57/300	58/310	59/320
10	60/331	61/341	62/352	63/363	64/374	65/385
11	66/397	67/408	68/420	69/432	70/444	71/456
12	72/469	73/481	74/494	75/507	76/520	77/533
13	78/547	79/560	80/574	81/588	82/602	83/616
14	84/631	85/645	86/660	87/675	88/690	89/705
15	90/721	91/736	92/752	93/768	94/784	95/800
16	96/817	97/833	98/850	99/867	100/884	101/901

therefore, $(CH_3)_6(CH)_4(C\equiv C)$ and $C_2(CH_3)_7(CH)(C\equiv C)$ are possible. $H^1M_9(C,H,N) = CH^6N^3 + C^3H^7$; then, $C(CH_3)_6(CH)_3(C\equiv CH)$ and $C_3(CH_3)_7(C\equiv CH)$, and so on. $M_{8-4}(1, 1, 1, 1) + M_{8-3}(1, 1, 1, 1) = 4 + 5 = 9$.

(b) Alkenes, Alkyl Alcohols, and Alkyl Radicals.

The alkyl alcohol $C_nH_{2n+1}OH$ has three types; the primary alcohol $(C_{n-1}H_{2(n-1)+1})(CH_2OH)$, the secondary alcohol $(C_{n-1}H_{2(n-1)+2})(CHOH)$, and the tertiary alcohol $C_{n-1}H_{2(n-1)+3}(COH)$. Thus $H^1M_{(n-1)-1}(C,H,N,O)$, $H^2M_{(n-1)-2}(C,H,N,O)$, and $H^3M_{(n-1)-3}(C,H,N,O)$ are used. A similar treatment is possible for the alkyl radical C_nH_{2n+1} which gives $(C_{n-1}H_{2(n-1)+1})(C^{\cdot}H_2)$, $(C_{n-1}H_{2(n-1)+2})(C^{\cdot}H)$, and $(C_{n-1}H_{2(n-1)+3})(C^{\cdot})$; also to the alkenes C_nH_{2n} the present method can easily be applied.

Note. Professor Balaban suggested me, using Polya's theorem, that the number of valence-isomeric classes for adamantane $C_{10}H_{16}$ is 18; refer to "Conclusions" in Rev. Roum. Chim., 31(8), 795 (1986). Our equations for calculation must be generalized as follows. Hydrocarbons in the general form C_nH_m are structurally possible only if $m = 2(n + 1 - u)$ and $u \geqq 0$; then,

$$n_H = 2n_C + n_N + 2(1 - u)$$

Thus $H^2M(C,H,N)$ in the text is replaced by $H^{2(1-u)}M(C,H,N)$, where $$M(C, H, N) = \sum_{\substack{x \geq 0 \\ 2x+y \geq 2(u-1)}}^{\infty} \sum_{y \geq 0}^{\infty} (CH^2)^x (NH)^y$$

For adamantane we take terms ($n \leq 10 - 2(1 - 3) = 14$ and $n_H \geq 2(3 - 1) = 4$) from Table 1; the number is just 18.

References

1. Oxford English Dictionary, 3. b. spec. in Chem. and Med. "1985 Daubeny, C. G. B. Atomic The. ix.(ed. 2) 297. By rational, in contradiction to empirical, formulae, we mean expressions of the manner in which the respective atoms are combined or grouped together, and not merely of the number of atoms of each of the ingredients present."

2. R. M. Silverstein, G. C. Bassler and T. C. Morrill, Spectrometric Identification of Organic Compounds, Fourth edition (John Wiley & Sons, New York, 1981).

3. Generating functions. Refer to Graph Theory 1736 - 1936, N. L. Biggs, E. K. Lloyd and R. J. Wilson, (Clarendon, Oxford 1977), pp. 40 - 44.

4. D. H. Rouvray and A. T. Balaban, in Application of Graph Theory, ed. R. J. Wilson and L. W. Beineke, (Academic Press, London, 1979), Chap. 7.

5. T. Morikawa, MATCH 22, 67(1987).