Isomer Enumeration of Alkenes and Aliphatic Cyclopropane Derivatives

Chin-yah Yeh

Department of Anesthesiology, The University of Utah School of Medicine, Salt Lake City, Utah 84132

Received January 12, 1996[⊗]

Two ways of evaluating alkenes taking *cis−trans* isomerism into account are pursued, one using the Redfield-Pólya theorem, the other doing without. The geometric isomers of aliphatic cyclopropane derivatives are also enumerated and tabulated.

The Cayley[1] formula for counting rooted trees has been used to enumerate isomers of homologous series of saturated aliphatic compounds with either a functional group or a distinct part that can be marked as a root of the tree. Radicals, monoalcohols, glycols, esters,[2] and chiral alkanes[3] have been enumerated. The first part of this paper is to report a recount of alkenes[2,4,5] with geometric isomerism due to a double bond being taken into consideration. Two methods of enumeration are used here; one is derived directly by taking the structural dissymmetry around double bond, the other incorporates the Redfield-Pólya theorem[6,7] into the Cayley formula. Both lead to equivalent results. In the second part of this note, the latter method is applied to count saturated aliphatic cyclopropane derivatives, again, with geometric isomerism taken into account.

Let us introduce the notations through the example of the alkane series.[8,9] The right hand side of the recurrence relation for counting rooted trees

$$\sum_n b_n x^n = x \prod_n (1 - x^n)^{-b_n} \tag{1}$$

is expanded and terms are collected, subject to an additional constraint $\Sigma_n i_n \leq 3$, which reflects the fact that the branching of the alkane series at each node is limited to three times. This leads to a function

$$f_3(x) = x \sum_{N=1}^{\infty} \left[\sum_{\substack{\Sigma_n i_n \leq 3, \\ \Sigma_n n i_n = N-1}} \prod_{n=1}^{N-1} \binom{b_n + i_n - 1}{i_n} \right] x^{N-1} \tag{2}$$

The second summation in eq 2 is meant to collect all the products that meet the two constraints listed below the summation sign. Explicitly this function is

$$f_3(x) = x + x^2 + 2x^3 + 4x^4 + 8x^5 + 17x^6 + 39x^7 +$$
$$89x^8 + 211x^9 + 507x^{10} + 1238X^{11} + 3057x^{12} +$$
$$7639x^{13} + 19241x^{14} + 48865x^{15} + 124906x^{16} +$$
$$321198x^{17} + 830219x^{18} + 2156010x^{19} + 5622109x^{20} +$$
$$14715813x^{21} + ... \tag{3}$$

Read[2] combines this function with Redfield-Pólya theorem and has obtained a counting series for alkanes [eq 4.1 in ref 2]; his notation $A(x)$ corresponds to $1 + f_3(x)$ in this paper.

On the other hand, carrying out the expansion of eq 1 again, by using the expansion coefficients found in $f_3(x)$ as exponents, but with the constraint changed to $\Sigma_n i_n \leq 4$, results in another function $g(x) = \Sigma g_n x^n$, which is equivalent to Read's[2] function $P(x)$, without the recourse to Redfield-Pólya theorem. Each coefficient g_n is the count of singly carbon-labeled alkanes. The rest of the problem is to delabel the rooted trees, which has been counted by $g(x)$, in order to yield the constitutional isomer count of alkanes.

The same procedure can be used to simplify the counting of alkenes or allenes. This time the constraint $\Sigma_n i_n \leq 2$ instead of $\Sigma_n i_n \leq 4$ is taken to obtain $g(x)$. Explicitly, $g(x) = x + x^2 + 2x^3 + 3x^4 + 7x^5 + 14x^6 + 32x^7 + 72x^8 + 171x^9 + 405x^{10} + ...$ Now each coefficient g_n becomes the count of carbenes of the constitution $C_n H_{2n}$; note that it is also the count of alkynes $C_{n+1} H_{2n}$. Alkenes are made of two carbenes. Therefore, the counting series of alkenes is obtained by multiplying $g(x)$ by itself, except that those made of two different moieties are doubly counted. Consequently, taking the fact that $g(x^2)$ counts the alkenes with two identical moieties, the function

$$F(x) = \frac{1}{2}[g^2(x) + g(x^2)] \tag{4}$$

yields the net counts of the alkene series. The results, which will not be repeated here, comply with those in literature.[2,4,5] By the same token, allenes are made of two carbenes bridged by a carbon atom; their counting series is the function $xF(x)$. However, squaring $g(x)$ implies a free-rotating, ball-joint type linking of the carbenes. Therefore, neither *cis−trans* isomerism in alkenes nor chirality in allenes is accounted. Chirality causes little change in physical properties except chiro-optical ones, hence can practically be ignored. But *cis−trans* isomerism exerts a much stronger force in structural discrimination and is worth the endeavor to resolve in enumeration.

Prior to the launching into the formulation for counting alkenes and aliphatic cyclopropane derivatives, a caveat is warranted. The function $f_3(x)$ as given in eq 2, on which the present calculation is based, accounts for only the constitutional isomers formed from the skeletons of saturated aliphatic hydrocarbons. It is the geometric isomerism of the core, being it either the olefinic group or the cyclopropanyl group, that is concerned in this note. Intuitively, one anticipates to find four isomers of butene, including a *cis−*

trans pair, in lieu of the previously reported number,[2,4,5] three, which ignores geometric isomerism.

There are two schemes to incorporate *cis−trans* isomerism. One is imposing dissymmetry on the function $F(x)$. The term dissymmetry dictates that both carbenes have to be asymmetrical in order to produce *cis−trans* isomers. For this purpose, $g(x)$, as obtained by resetting the constraint to $\Sigma_n i_n \leq 2$ in eq 2, is separated into two parts, $g_s(x) + g_a(x)$, each counting symmetrical and asymmetrical carbenes respectively. $g_s(x)$ is found by counting twin trees (namely, those with two identical branches at the root), which is incidentally $x + x f_3(x^2)$. Equation 4 is modified to tend the dissymmetry caused by *cis−trans* isomerism as follows.

$$F_a(x) = \frac{1}{2}[g^2(x) + g(x^2) + g_a^2(x) + g_a(x^2)] \quad (5)$$

As $g_a(x) = g(x) - x[1 + f_3(x^2)]$ is used, eq 5 turns into

$$F_a(x) = g^2(x) + g(x^2) + \frac{x^2}{2}\{[1 + f_3(x^2)]^2 -$$
$$[1 + f_3(x^4)]\} - xg(x)[1 + f_3(x^2)] \quad (6)$$

The difference $F_a(x) - F(x)$ marks the numbers of *cis−trans* pairs.

The other scheme is to use Redfield-Pólya theorem. With the core symmetry of olefins being D_{2h}, there are four distinct symmetry elements that permute the four surrounding alkyl groups. Hence, the counting series of alkenes becomes

$$F_b(x) = \frac{1}{4}\{[1 + f_3(x)]^4 + 3[1 + f_3(x^2)]^2\} \quad (7)$$

By not allowing the rotation around double bond, *cis−trans* isomerism is reinstated. Equations 6 and 7 differ in that eq 6 counts the two carbon atoms at the ethylenic group as part of carbenes, whereas eq 7 does not count the core. Hence one needs an extra factor x^2 to balance the equality; namely, $F_a(x) = x^2 F_b(x)$, which gives

$$g(x) = \frac{x}{2}\{[1 + f_3(x)]^2 + [1 + f_3(x^2)]\} \quad (8)$$

Equation 8 demonstrates that the enumeration series of carbenes and alkynes are also expressible in terms of the function $f_3(x)$.

If one deliberately assumes a free twisting of the double bond and ignores the geometric isomerism, the symmetry is expanded to $C_2 \times D_{2h}$, of which the set of distinct components for permuting the side groups on alkenes forms the wreath product $C_2[C_2]$, a group of eight elements. The cycle index for this group,[2,7] $Z(C_2[C_2])$, counts the constitutional isomers of alkenes and turns out to be a form equivalent to eq 4, where $g(x)$ is now a functional, $g(x) = x[A^2(x) + A(x^2)]/2$ and $A(x) = 1 + f_3(x)$. Obviously, the resultant $g(x)$ agrees with eq 8; it does also with the one obtained by resetting the constraint to $\Sigma_n i_n \leq 2$ in eq 2.

For the next simple series in paraffins, cyclopropane derivatives, the core has a D_{3h} structure from which constitutional, geometric, and optical isomers can be formed. Each of these three isomerisms is covered by a different symmetry, which leads to a different counting function.[7,10]
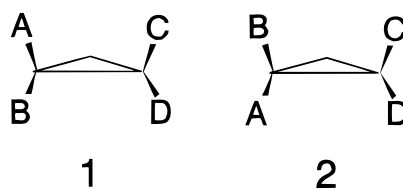

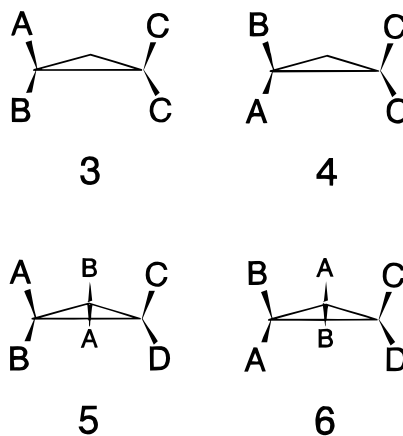**Figure 1.** Diastereoisomers of cyclopropane derivatives.

Figure 2. Enantiomeric pairs of cyclopropane derivatives.

By the reason stated previously, our interest is in geometric isomerism; for this, one uses the D_{3h} symmetry proper. The counting series is

$$F_c(x) =$$
$$\frac{1}{12}[f_3^6(x) + 3f_3^2(x)f_3^2(x^2) + 4f_3^3(x^2) + 2f_3^2(x^3) + 2f_3(x^6)] \quad (9)$$

This function distinguishes diastereoisomers like **1** and **2**, shown in Figure 1, but not enantiomeric pairs like **3** and **4**, or **5** and **6**, shown in Figure 2, since planar symmetries, but no permutations due to the free linking at each of the three nodes of the cyclopropanyl core, are assumed in the formulation. Similar to the discussion in the previous paragraph on alkenes, one could loosen the distinction between geometric isomers and count only constitutional isomers formed by the cyclopropane core. The symmetry is the wreath product $D_3[C_2]$. The cycle index has been listed[7,10] and expressed in terms of f_i, which needs to be replaced by $f_3(x^i)$ as defined in this paper for enumeration. However, the result is the same as[11] monocyclic carbon skeletons of ring size 3 and will not be repeated here.

Equations 7 and 9 are used to enumerate alkenes and cyclopropane derivatives, respectively. The code written in Turbo Pascal is concise and runs efficiently. The double summation occurred in eq 2 is simplified by carrying through the partition of cardinal numbers.[8] Variables are compiled in extended double precision with 64 bit mantissa. For such a precision, 19−20 significant digits can be accommodated; this means that the truncation or rounding error begins to show up at $N \sim 45$ for the enumeration of alkenes and cyclopropane derivatives. Table 1 lists the result. Now, butenes are properly counted to include geometric isomerism: four isomers in contrast to the previously reported number,[2,4,5] three.

It has been shown that the equivalent of $1 + f_3(x)$, for example, $A(x)$ in ref 2 or $r(x)$ in ref 7, is useful in enumerating many hydrocarbons and their derivatives. Here it is demonstrated that the use of the function $f_3(x)$ can be

Table 1. Numbers of Geometric Isomers of Alkenes and Aliphatic Cyclopropane Derivatives

N	alkenes $C_{N+2}H_{2N+4}$	cyclopropane derivatives $C_{N+3}H_{2N+6}$
1	1	1
2	4	4
3	6	8
4	17	22
5	36	51
6	92	136
7	215	335
8	542	871
9	1327	2217
10	3354	5749
11	8429	14837
12	21494	38636
13	54889	100622
14	141317	263381
15	365014	690709
16	947897	1817544
17	2470047	4793449
18	6462205	12675741
19	16959708	33592349
20	44651511	89223734
21	117881286	237455566
22	312031372	633176939
23	827908711	1691377956
24	2201600927	4525792533
25	5866623538	12129365576
26	15663033725	32556355947
27	41893093576	87508275471
28	112237942135	235529797422
29	301176236399	634737263112
30	809365327210	1712636129045
31	2178075539348	4626281275138
32	5869096262229	12510365756211
33	15834612290414	33865360093125
34	42771287499006	91762986828843
35	115658841829654	248877328616873
36	313084872544919	675600998747977
37	848357647297063	1835543905996132
38	2300955216224250	4991055157076187
39	6246400103897940	13581809637822792
40	16971703333467254	36986746709502343

extended to enumerate carbenes, alkynes, allenes, butatriene derivatives,[12] and their homologous series and to incorporate geometric isomerism in the enumeration of alkenes and saturated aliphatic cyclopropane derivatives. It suffices to claim that $f_3(x)$ is the basic unit of counting the constitutional structures of saturated aliphatic hydrocarbon skeletons. $f_3(x)$ can be conveniently obtained by constraining the Cayley expansion in the fashion of eq 2, as demonstrated in this note and ref 8. Although $f_3(x)$ cannot be explicitly expressed, it can readily be retrieved through a simple algorithm.[8]

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) Cayley, A. On the analytical forms called trees. *Am. J. Math.* **1881**, *4*, 266−268.
(2) Read, R. C. The enumeration of acyclic chemical compounds, in *Chemical Applications of Graph Theory*, Balaban, A. T., Ed.; Academic Press: **1976**; pp 25−62.
(3) Robinson, R. W.; Harary, F.; Balaban, A. T. The numbers of chiral and achiral alkanes and monosubstituted alkanes. *Tetrahedron* **1976**, *32*, 355−361.
(4) Henze, H; Blair, C. M. The number of structurally isomeric hydrocarbons of the ethylene series. *J. Am. Chem. Soc.* **1933**, *55*, 680−686.
(5) Lederberg, J.; Sutherland, G. L.; Buchanan, B. G.; Feigenbaum, E. A.; Robertson, A. V.; Duffield, A. M.; Djerassi, C. Applications of artificial intelligence for chemical inference. I. The number of possible chemical compounds. Acyclic structures containing C, H, O, and N. *J. Am. Chem. Soc.* **1969**, *91*, 2973−2981.
(6) Redfield, J. H. The theory of group-reduced distributions. *Am. J. Math.* **1927**, *49*, 433−455.
(7) Pólya, G. Combinatorial enumeration of groups, graphs, and chemical compounds. *Acta Math.* **1938**, *68*, 145−254.
(8) Yeh, C.-Y. Isomer enumeration of alkanes, labeled alkanes, and monosubstituted alkanes. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 912−913.
(9) Erratum in ref 8: In line 7, read $= \Sigma b_n x_n$ as $= \Sigma b_n x^n$.
(10) Taylor, W. J. Application of Polya's theorem to optical, geometrical, and structural isomerism. *J. Chem. Phys.* **1943**, *11*, 532.
(11) Parks, C. A.; Hendrickson, J. B. Enumeration of monocyclic and bicyclic carbon skeletons. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 334−339.
(12) Naturally an extension of allenes. The author is indebted to Professor S. J. Cyvin for this point.