Isomer Enumeration of Alkanes, Labeled Alkanes, and Monosubstituted Alkanes

Chin-yah Yeh

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

Received June 14, 1995[⊗]

A simple algorithm for counting constitutional isomers of alkanes, single C-atom isotopically labeled alkanes, and monosubstituted alkanes is reported.

Enumeration of alkanes and alkyl derivatives has long been of interest.[1−6] Despite the claim several times in the literature that there is no simple formula for this problem, the present author is to report a simple solution. Although only halfway analytical, the solution suits computer programming.

Cayley's[1] expansion series for counting rooted trees, $f(x) = \sum b_n x_n$, where b_n is found by the recursion formula

$$f(x) = x\prod_{n=1}^{\infty}(1 - x^n)^{-b_n} \tag{1}$$

is readjusted as follows. First, eq 1 is expanded and terms are collected, subject to a newly imposed constraint $\sum_n i_n \leq m$. This constraint is used twice throughout the algorithm, each time with a different m, which is designated as the maximal vertex degree of the root in rooted trees. The expansion leads to a different function

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

The second summation means to collect all the products that meet the two constraints listed below the summation sign. Note that the first $m + 1$ terms of $f_m(x)$ and $f(x)$ are equal. Second, b_n's are found by setting $m = 3$ and reiterating. This gives a new expansion $f_3(x) = \sum b_n x^n$. The root has three connections or less, whereas the rest of the vertices can have four. Third, with the found b_n's, carry out the expansion of eq 2 again, but with the constraint changed to $m = 4$. This time the constraint dictates that both the root and other vertices can have four connections. The collection results in a function $g(x) = \sum g_n x^n$. $f_3(x)$ is the count of constitutional isomers for monosubstituted alkanes or alkyl radicals, whereas $g(x)$ is the count for singly labeled (isotopically, for instance) alkanes. Note that $g(x) \neq f_4(x)$.

Testing the constraint $\sum_n i_n \leq m$ for carrying out the summation in eq 2 takes $(m + 1)^{N-1}$ operations for a given N. To override the testing, one could use the fact that diagrams for partitioning cardinal numbers, which can be better visualized in terms of Young diagrams (no symmetrization procedures intended, however), give one-to-one correspondence to all the combinations meeting the constraint. In other words, to find the coefficient of x^N in $f_3(x)$, one partitions the number $N-1$ into four descending numbers $k_1k_2k_3k_4$ representing the term $x^{k_1}x^{k_2}x^{k_3}x^{k_4}$, then seeks the product of binomial coefficients for each of these terms, as expressed by b_n and i_n in eq 2, and sums over all these terms. Since there are at most four nonzero i_n's in each term represented by such a Young diagram, their subscripts are disregarded, and these four are readdressed as $i1$ to $i4$ (see Table 1), not necessarily all being nonzero; their values are conveniently found with the aid of the Young diagram as follows. If k_1–k_4 are distinctive, their i_n values are 1. On the other hand, if a partitioning results in two or more equal numbers among k_1–k_4, they are gathered as a group, the i_n value of the first in the group is the size of the group, values of the rest being zero. For instance, if $k_1 = k_2 \neq k_3$, then $i1 = 2$ and $i2 = 0$. Because only partial order exists among Young diagrams, counting them is still cumbersome for computers. Fortunately one only has to count down to four rows deep (three rows for finding b_n's), a problem now tangible to computer.

Table 1. Turbo Pascal Code for the Results in Table 2

```
($N+)
  Uses dos, crt;
  Const size=101;
  Var N, j, k, k1, k2, k3, k4, k1p, k2p, k3p, k4p, i1, i2, i3, i4: integer;
      b, g, f32, capf: array[1..size] of extended;
      f3: array[0..size, 0..4] of extended;
  Label 1, 9;
Begin clrscr;
  For N:=0 to size Do For j:=0 to 4 Do f3[N,j]:=1;
  For N:=1 to size Do begin
    k1:=N-1; k2:=0; k3:=0; k4:=0;
    k1p:=k1; k2p:=k2; k3p:=k3; k4p:=k4;
    g[N]:=0; f3[N,1]:=0;
1:  If (k1<k1p) then begin k1p:=k1; k2p:=k2; k3p:=k3; k4p:=k4; End;
    i1:=1+Ord(k1=k2)+Ord(k1=k3)+Ord(k1=k4);
    i2:=Ord(k1>k2)*Ord(k2>0)*(1+Ord(k2=k3)+Ord(k2=k4));
    i3:=Ord(k2>k3)*Ord(k3>0)*(1+Ord(k3=k4));
    i4:=Ord(k3>k4)*Ord(k4>0);
    If (k4=0) then f3[N,1]:= f3[N,1] + f3[k1,i1]*f3[k2,i2]*f3[k3,i3];
    b[N]:= f3[N,1];
    g[N]:= g[N]+f3[k1,i1]*f3[k2,i2]*f3[k3,i3]*f3[k4,i4];
    If (N<=size div 2) then f3[N,2]:=b[N]*(b[N]+1)/2;
    If (N<=size div 3) then f3[N,3]:=b[N]*(b[N]+1)*(b[N]+2)/6;
    If (N<=size div 4) then f3[N,4]:=b[N]*(b[N]+1)/2*(b[N]+2)/3*(b[N]+3)/4;
    If (k1<=k4+1) then Goto 9;
    If (k2<=k4+1) then begin
      Dec(k1);
      If (2*k1>N-1) then k2:=N-1-k1 Else k2:=k1;
      If (3*k1>N-1) then k3:=N-1-k1-k2 Else k3:=k1;
      If (4*k1>N-1) then k4:=N-1-k1-k2-k3 Else k4:=k1;
      Goto 1; End;
    If (k3<=k4+1) then begin
      k:=k2+k3+k4;
      Dec(k2);
      If (2*k2>k) then k3:=k-k2 Else k3:=k2;
      If (3*k2>k) then k4:=k-k2-k3 Else k4:=k2;
      Goto 1; End
    Else begin Dec(k3); Inc(k4); End;
    Goto 1;
9:  End;
  For N:=1 to size Do f32[N]:=0;
  For N:=1 to size-1 Do For j:=1 to N Do f32[N+1]:= f32[N+1]+b[j]*b[N-j+1];
  For N:=1 to size Do
    capf[N]:= g[N] - (f32[N] - b[N div 2]*((N+1)mod 2))/2;
  For N:=1 to (size div 10) Do writeln(capf[N*10],'(',N*10:3,')'); End.
```

Each rooted tree is distinctively counted with no redundancy. Therefore, the process could be used to identify rooted trees. The root is configured through a Young diagram. In so doing, a tree is divided into one to four branches, each resulting in a subtree with a new root, from

⊗ Abstract published in *Advance ACS Abstracts*, September 1, 1995.

0095-2338/95/1635-0912$09.00/0

Table 2. Numbers of C_NH_{2N+2} Constitutional Isomers

N	isomers
10	7.50000000000000E+0001
20	3.66319000000000E+0005
30	4.11184676300000E+0009
40	6.24818011473410E+0013
50	1.11774365174695E+0018
60	2.21587345357704E+0022
70	4.71484798515330E+0026
80	1.05644769069467E+0031
90	2.46245150242821E+0035
100	5.92107203812581E+0039

which a Young diagram of lesser degree is constructed. The process continues until it reaches all the tips of the tree. The nested Young diagram thus constructed identifies the tree. If two branches are of the same size, they are put in reverse lexicographic order. However, this process of identification through all branches in a tree, similar to the N-tuple method,[6,7] is not needed for enumeration since all subtrees have already been counted during the recursion.

The rest of the problem is to "unlabel" the rooted trees in $g(x)$, for which the Otter formula[8] is used

$$F(x) = g(x) - \frac{1}{2}[f_3^2(x) - f_3(x^2)]$$

$F(x)$ yields the constitutional isomer count for alkanes.

The above algorithm is coded in Turbo Pascal and listed in Table 1. Variables are compiled in extended double precision with 64 bit mantissa. For such a precision, 19–20 significant digits can be accommodated; the truncation or rounding error begins to show up for $N \geq 47$. Results of the isomer count of alkanes agree with ref 9 where values for $N \leq 100$ are listed. Isomer counts of C_NH_{2N+2} for tens of N are tabulated in Table 2. Isomer counts of single-^{13}C labeled C_NH_{2N+2} for $N \leq 47$ have not been published before and are tabulated in Table 3. Other outputs can be obtained by rephrasing the last statement in Table 1. Results of $N >$ 100 can be obtained by increasing the constant "size", with rounding error to be tolerated.

In summary, coefficients of $f_3(x)$, $g(x)$, and $F(x)$ are the constitutional isomer counts for monosubstituted, single C-atom isotopically labeled, and unlabeled alkanes, respectively.

Table 3. Numbers of Single-^{13}C Labeled C_NH_{2N+2} Constitutional Isomers

N	isomers	N	isomers
1	1	26	2084521232
2	1	27	5549613097
3	2	28	14804572332
4	4	29	39568107511
5	9	30	105938822149
6	18	31	284103144805
7	42	32	763067158047
8	96	33	2052459438451
9	229	34	5528079077194
10	549	35	14908290599141
11	1347	36	40253559153599
12	3326	37	108811562245870
13	8330	38	294451568992057
14	21000	39	797620980258275
15	53407	40	2162722316575299
16	136639	41	5869562580635247
17	351757	42	15943815991204954
18	909962	43	43345348850358244
19	2365146	44	117934205327801553
20	6172068	45	321120920694833012
21	16166991	46	875012884317194451
22	42488077	47	2385963304276811920
23	112004630		
24	296080425		
25	784688263		

REFERENCES AND NOTES

(1) Cayley, A. On the analytical forms called trees. *Am. J. Math.* **1881**, *4*, 266–268.
(2) Hermann, F. On the problem of evaluating the numbers of isomeric paraffins of the formula C_nH_{2n+2}. *Ber. Dtsch. Chem. Ges.* **1880**, *13*, 792.
(3) Henze H. R.; Blair, C. M. The number of isomeric hydrocarbons of the methane series. *J. Am. Chem. Soc.* **1931**, *53*, 3077–3085.
(4) Davis, C. C.; Cross, K.; Ebel, M. Computer calculation of isomers. *J. Chem. Ed.* **1971**, *48*, 675.
(5) Robinson, R. W.; Harary, F.; Balaban, A. T. The numbers of chiral and achiral alkanes and monosubstituted alkanes. *Tetrahedron* **1976**, *32*, 355–361.
(6) Knop, J. V.; Müller, W. R.; Jeričević, Ž.; Trinajstić, N. Computer enumeration and generation of trees and rooted trees. *J. Chem. Inf. Comput. Sci.* **1981**, *21*, 91–99.
(7) Contreras, M. L.; Valdivia, R.; Rozas, R. Exhaustive generation of organic isomers. 1. acyclic structures. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 323–330.
(8) Otter, R. The number of trees. *Annals Math.* **1948**, *49*, 583–599.
(9) Trinajstić, N.; Nikolić, S.; Knop, J. V.; Müller, W. R.; Szymanski, K. *Computational Chemical Graph Theory: Characterization, Enumeration, and Generation of Chemical Structures by Computer Methods*; Ellis Horwood, Chichester, 1991; pp 140–149.

CI950056Q