# Lawrence Berkeley National Laboratory Recent Work 

## Title

COMPUTER GENERATION OF ISOMERS

## Permalink

https://escholarship.org/uc/item/08r0f80m

## Author

Balasubramanian, K.
Publication Date
1981-12-01

# Lawrence Berkeley Laboratory UNIVERSITY OF CALIFORNIA <br> RECEIVED 

# Materials \& Molecular Research Division 

Submitted to Computers and Chemistry

## COMPUTER GENERATION OF ISOMERS

K. Balasubramanian

December 1981

## TWO-WEEK LOAN COPY

This is a Library Circulating Copy
which may be borrowed for two weeks.
For a personal retention copy, call Tech. Info. Division, Ext. 6782


Prepared for the U.S. Department of Energy under Contract W-7405-ENG-48

## DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

# Computer Generation of Isomers 

K. Balasubramanian<br>Department of Chemistry and<br>Lawrence Berkeley Laboratory University of California Berkeley, CA 94720

## Abstract

A computer program is developed for the enumeration of the isomers of polysubstituted compounds with $b_{1}$ substituents of the type $1, b_{2}$ substituents of the type $2, \ldots . b_{n}$ substituents of the type $n$. The procedure is illustrated with octahedral molecules containing 4 kinds of substituents (such as $\mathrm{F}, \mathrm{Cl}, \mathrm{Br}$ and I ) and the isomers of polysubstituted non-rigid pentane with 4 kinds of substituents.

This work was supported by the Director, Office of Energy Research, Office of Basic Energy Sciences, Chemical Sciences Division of the U. S. Department of Energy under Contract Number W-7405-ENG-48.

1. Introduction

Chemical applications of non-numerical computational methods are becoming quite important in recent years. (Randić (1975), Randić (1979), Randić (1980), Masinter et al. (1974a,b), Balasubramanian (1982a,b), Balasubramanian et al. (1980a), Dolhaine (1981)). Several of these papers concern developing algorithms for the generation of discrete combinatorial structures and their applications to chemistry. Recently, Dolhaine (1981) has developed a computer program for the enumeration of isomers of molecules containing 2 kinds of substituents. Balasubramanian (1981d) developed a combinatorial technique for nuclear spin statistics in molecular spectroscopy which was subsequently computerized (Balasubramanian (1982a,b)). In the present paper we develop computer programs and algorithms for the enumeration of isomers of polysubstituted compounds containing several kinds of substituents. The program that we use here is more general than the one given by Dolhaine (1981) in that this program can handle more than 2 substituents by way of multinomial expansions.

The history of isomer enumeration goes back to the last century. Several papers have appeared both in mathematical and chemical 1iterature. (Balaban (1976), Balaban, et al. (1976), Balaban (1975), Balasubramanian (1978, 1979a-c, 1981a-b), Dolhaine (1980), King (1972), Klein and Cowley (1978), Mislow (1976), Nourse (1979), Pólya (1937), Robinson (1970), and Ruch et al. (1970)). The topic has been reviewed by Rouvray (1974, 1975) and in the recent book by Balaban (1976). As pointed out by Dolhaine (1981) recently, even though these methods provide
for generators of isomers, yet, one needs to evaluate several polynomial generators and in general several polynomial products of multinomials. Consequently, an efficient and general computer program is warranted for polysubstituted compounds.

In the present paper we develop algorithms and computer programs which generate isomers of polysubstituted compounds with a minimal input. The required input for this program is just the cycle index and the information concerning the kinds of substituents. The program generates the total generating function wherein the coefficients of the various terms printed out give the number of isomers.

## 2. Theory

Let $G$ be a group acting on a discrete set $D$, which is the set of sites in the molecule that can accommodate substituents. Let $R$ be the set of substituents. In general let $|S|$ denote the number of elements in a set $S$. Consider the set $F$ of maps from D to R. Each element in this set is a map of sites to substituents and consequently, a way in which substituents can be placed on available sites. However, not all maps in $F$ are distinct because one map can be equivalent to another by the action of $G$ which permutes the sites and hence the maps. This action of $G$ on $D$ induces equivalence of maps in $F$. Two maps $f_{i}$ and $f_{j}$ are equivalent if there exists a $g \varepsilon G$ such that

$$
f_{i}(d)=f_{j}(g d), d \varepsilon D
$$

Thus maps in $F$ that are equivalent $c a n$ be grouped together in to equivalence classes. A representative in each class is an isomer
and the number of equivalence classes is exactly the number of isomers. In order to book-keep the number of different substituents in a map we introduce the concept of weight of a substituent. Let $w(x)$ be the weight of a substituent $r$ in $R$. Then with each map $f \varepsilon F$, we can associate a weight

$$
W(f)=\prod_{d \varepsilon D} W(f(d))
$$

For example, if $w_{1}, w_{2}, \ldots w_{\ell}$ are the weights of $\ell$ substituents in the set $R$ and if a structure contains $b_{1}$ substituents of the type $1, b_{2}$ substituents of the type $2, \ldots . b_{\ell}$ substituents of the type $\ell$, then the weight of this function would be ${ }^{w_{1}}{ }_{1} w_{2}$ $\ldots w_{\ell}$. Pólya (1937) proved a theorem now well-known as Pólya's theorem which gives a generating function for the equivalence classes of maps in $F$ from a group structure known as the cycle index of a group $G$. With each element $g \varepsilon G$ we can associate a cycle representation $x_{1} x_{2}{ }_{2} \ldots x_{n}$ if $g$ has $b_{1}$ cycles of length $1, b_{2}$ cycles of length $2, \ldots ., b_{n}$ cycles of length $n$. To illustrate, the permutation (12)(345)(67)(89) would have the cycle representation $x_{2}^{3} x_{3}$ since it has 3 cycles of length 2 and a cycle of length 3 . Then one can define a group structure known as the cycle index of a group $G, P_{G}$, defined as follows.

$$
\begin{equation*}
P_{G}\left(x_{1}, x_{2}, \ldots\right)=\frac{1}{|G|} \sum_{g \varepsilon G} x_{1}^{b_{1}}{ }_{x_{2}}^{b_{2}} \ldots x_{n}^{b_{n}} \tag{1}
\end{equation*}
$$

Polya (1937) proved that the generating function (G.F.) for isomers is obtained by replacing every $x_{k}$ by $\sum_{r \in R}(w(r))^{k}$. In symbols,

$$
\begin{equation*}
\text { G.F. }=P_{G}\left(x_{k} \rightarrow \sum_{r \& R}(w(r))^{k}\right) \tag{2}
\end{equation*}
$$

## 3. Algorithms and Programs

As noted by Dolhaine (1981), even though the above theorem is an elegant way to generate isomers, yet, one has to expand several multinomials (in general) and collect the coefficients in various terms. Thus a computer program is warranted for a general polysubstituent compound. Dolhaine's recent program can handle at most 2 substituents. We formulate here an algorithm and computer program, in general, for any number of substituents.

To illustrate, consider an octahedral molecule which con6 tains/substitutional sites. The cycle index of the rotational group 0 of this molecule is shown below.

$$
\begin{equation*}
P_{0}=\frac{1}{24}\left(x_{1}^{6}+6 x_{1}^{2} x_{4}+3 x_{1}^{2} x_{2}^{2}+6 x_{2}^{3}+8 x_{3}^{2}\right) \tag{3}
\end{equation*}
$$

If the set $R$ has 4 different substituents with the weights a, $b, c$, and $d$, then by Pólya's theorem the generating function is given by the following expression.
G.F. $=\frac{1}{24}\left[(a+b+c+d)^{6}+6(a+b+c+d)^{2}-\left(a^{4}+b^{4}+c^{4}+d^{4}\right)+3(a+b+c+d)^{2}\right.$ $\left.x\left(a^{2}+b^{2}+c^{3}+d^{2}\right)^{2}+6\left(a^{2}+b^{2}+c^{2}+d^{2}\right)^{3}+8\left(a^{3}+b^{3}+c^{3}+d^{3}\right)^{2}\right]$.

The above generating function has several multinomials which have to be expanded, multiplied and added together. The final generating function thus obtained contains 84 unique terms. Thus one can see that the problem addressed here is sufficiently complex that a computer program is warranted for this purpose.
A. Generation of Unique Terms in the Generating Function With each term ${ }^{w_{1}} 1_{w_{2}}{ }_{2} \ldots{ }^{b_{l}}{ }^{l}$ we can associate a vector $\left(b_{1}, b_{2}, \ldots, b_{\ell}\right)$ such that $\sum_{i} b_{i}$ equals the total number of substitutional sites. Let the total number of sites be |D|. Then the problem of generating the unique terms in the generating function amounts to generating all vectors ( $b_{1}, b_{2}, \ldots, b_{\ell}$ ), $b_{i} \geq 0$ such that $\sum_{i} b_{i}=|D|$. This indeed corresponds to the compositions of the integer $|D|$ into $|R|$ parts since $|R|=\ell$ is the number of substituents. This number can be found using the following logic. The procedure shown below is the one given by Nijennuis and Wilf (1975). Suppose we wish to distribute $|D|$ indistinguishable balls into $|R|$ labelled cells such that any cell contains $0,1,2, \ldots$ up to $|D|$ balls, then each such distribution yields a composition. This can be accomplished by constructing the cell boundaries as follows. Consider a cell containing 2 walls with $|D|+|R|-1$ spaces as shown below.

$$
\frac{1}{1} \overline{2}^{\prime} \cdots \cdots \cdot \frac{}{|D|+|R|} \frac{1}{|D|+|R|+1}
$$

The number of ways of distributing $|D|$ balls in the available $|D|+|R|-1$ spaces is

$$
\binom{|D|+|R|-1}{|D|}
$$

After distributing these balls cell boundaries can be constructed in the rest $|R|-1$ available spaces. Such an arrangement contains exactly $|R|$ cells among which we have distributed $|D|$ indistinguishable balls. Consequently, we have obtained a composition of $|D|$ into $|R|$ parts. The above procedure has
facilitated not only a method to count the number of compositions but also to construct them. The above procedure, in fact, amounts to finding all $|R|-1$ subsets from a $|D|+|R|-1$ set. There are standard combinatorial algorithms for this purpose, for example, see the algorithm NEXCOM given by Nijenhuis and Wilf (1975). The subroutine which is based on this algorithm generates all the unique terms in the generating function for isomers.
B. Generation of Coefficients (the number of isomers in the Generating Function)

A subroutine called "VEC" expands each multinomial in the generating function (the terms are given by the subroutine NEXCOM) . By expand we mean that it generates the coefficient for each unique term in a multinomial. Then it multiplies together several multinomials contained in each term of the cycle index. For example, the second term in the generating function for the isomers of polysubstituted octahedral compounds has two multinomials shown below.

$$
(a+b+c+d)^{2}\left(a^{4}+b^{4}+c^{4}+d^{4}\right) .
$$

VEC expands each multinomial separately then multiplies them together and returns the total coefficients and terms in this product. The main program multiplies these coefficients with the corresponding coefficient in the cycle index and adds to the previously generated vector and coefficients. In this manner the program scans through all the terms in the cycle index and generates the overall generating function for isomers.

## C. Input and Output Descriptions

The present program can actually handle the generalized character cycle indices introduced elsewhere (Balasubramanian (1981d). For the present purpose, however, we need to consider only the cycle index corresponding to the totally symmetric representation. The required input is essentially the coefficients and various terms in the cycle index. Note that to generate isomers, we restrict ourselves to the totally symmetric irreducible representation (which we denote by A1) of the rotational subgroup. The input description is shown in Table 1. We will expound further here on this table with examples. Let us consider the rotational group $D_{4}$ acting on 4 corners of a square. The cycle index is shown below.

$$
\begin{equation*}
P_{G}=\frac{1}{8}\left(x_{1}^{4}+3 x_{2}^{2}+2 x_{4}+2 x_{1}^{2} x_{2}\right) \tag{5}
\end{equation*}
$$

In particular we will concentrate on cards ensuing the fourth card. The last term $x_{1}^{2} x_{2}$ in the cycle index has 2 components (NPRO $=2$ ), the superfix and suffix of the first component is 2 and 1 (because it is $x_{1}^{2}$ ), while for the second component it is 1 and 2. Cards 4-9 are shown below for this cycle index.

| $\underline{\text { Card }}$ | 5 |  |  |  | (Number of terms in the cycle index) |
| :---: | :---: | :---: | :---: | :---: | :---: |
| 4 |  |  |  |  |  |
| 5 | 1 | 3 | 2 | 2 | (Coefficients in the same order as shown in (5)) |
| 6 | 1 | 4 | 1 |  | (The first term) |
| 7 | 1 | 2 | 2 |  | (The second term) |
| 8 | 1 | 1 | 4 |  | (The third term) |
| 9 | 2 | 2 | 1 | 1 | 2 (The fourth term). |

An example of a complete input is given in Table 2 for the isomers of polysubstituted octahedral compounds. The cycle index for this case is shown in Eq (3). We consider 4 different substituents. The output corresponding to this input is shown in Table 3. As one can see, Table 3 contains the total number of all isomers, the number of isomers with different types of substituents. The vector in each term of the generating function gives the number of times several different substituents occur. For example, a typical vector $\left(b_{1}, b_{2}, \ldots, b_{\ell}\right)$, stands for the isomers of a molecule containing $b_{1}$ substituents of the type 1 , $b_{2}$ substituents of the type $2, \ldots, b_{\ell}$ substituents of the type $\ell$. The coefficient of the corresponding vector gives the number of isomers of that kind. For example, the coefficient of the vector ( $4,2,0,0$ ) is 2 indicating there are 2 isomers for a compound of the type $\mathrm{Mx}_{4} \mathrm{Y}_{2}$, where M is the atom at the center. As a second example let us consider a non-rigid unbranched pentane with 4 different substituents. The cycle index is shown below with the rotational group being the generalized wreath product $C_{2}\left[C_{3}, E\right]$. For details of the cycle index of generalized wreath product, see Balasubramanian (1979a).

$$
\begin{equation*}
P\left(C_{2}\left[C_{3}, E\right]\right)=\frac{1}{18}\left[x_{1}^{12}+4 x_{1}^{9} x_{3}+4 x_{1}^{6} x_{3}^{2}+3 x_{2}^{6}+6 x_{2}^{3} x_{6}\right] \tag{6}
\end{equation*}
$$

The input and the output for this molecule are shown in Tables 4 and 5, respective 1 y.
D. Limitations and Error Messages

Arrays in the present program are dimensioned to sufficiently large numbers that most of the chemically interesting cases can be handled. The present program can handle generating functions of isomers with at most 1000 terms. Nevertheless, the program can be easily modified by a suitable expansion of arrays. The present version is restricted to cycle index which contains terms $x_{1}{ }^{b_{1}}{ }^{b_{2}} \ldots x_{n}{ }_{n}$ with $n \leq 5$. For $n \geq 6$, a message is printed out by the subroutine VEC specifying this limitation. This subroutine contains comment statements giving instructions as to where modifications are necessary.

This program can detect a number of inconsistent input errors. For example, it checks each term $x_{1}{ }_{1} x_{2}{ }_{2} \ldots x_{n}{ }_{n}$ in the cycle index to see if that term satisfies the following condition.

$$
\sum_{i=1}^{n} i b_{i}=|D|=N T .
$$

If this condition is not met then the program prints out an error message "Input error for this term check $N(I)$, I $\exp (I)$ ". The user should check the term just printed out and correct it. The second error message is based on the requirement that the coefficient of any term in the polynomial

$$
|G| P_{G}\left(x_{k} \rightarrow \sum_{r \varepsilon R}(w(r))^{k}\right)
$$

should be divisible by $|G|$. If not the program prints out an error message "ICO(J) is not divisible by MODG. Input error". The error is either in the set of coefficients in the cycle index or in the terms of the cycle index that could not otherwise be detected by the earlier criterion.

## References

Balaban, A. T. (1976) Chemical Applications of Graph Theory, Academic, N.Y.

Balaban, A. T., Palmer, E. M. and Harary, F. (1976) Rev. Roum. Chim. 22, 517.

Balaban, A. T. (1975) Rev. Roum. Chim. 20, 227.
Balasubramanian, K. (1978) Indian J. Chem. 16B, 1094.
Balasubramanian, K. (1979a) Theor. Chim. Acta. 51, 37.
Balasubramanian, K. (1979b) Theor. Chim. Acta. 53, 129.
Balasubramanian, K. (1979c) Annal. N.Y. Acad. Sci. 319, 33.
Balasubramanian, K., J. J. Kaufman, W. S. Koski and Balaban,
A. T. (1980a) J. Computational Chem. 1, 149.

Balasubramanian, K. (1980b) J. Chem. Phys. 72, 665.
Balasubramanian, K. (1981a) Theor. Chim. Acta. 59, 47.
Balasubramanian, K. (1981b) Theor. Chim. Acta. 59, 91.
Balasubramanian, K. (1981c) Theor. Chim. Acta. 59, 237.
Balasubramanian, K. (1981d) J. Chem. Phys. 74, 6824.
Balasubramanian, K. (1982a) J. Computational Chem. (in press).
Balasubramanian, K. (1982b) J. Computational Chem. (in press).
Dolhaine, H. (1980) Chemiker-Zeitung, 104, 287.
Do1haine, H. (1981) Computers and Chemistry, 5, 41.
King, R. B. (1972) Theor. Chim. Acta. 25, 309.
Klein, D. J. and Cowley, A. H. (1978) J. Am. Chem. Soc. 100, 2593.
Masinter, L. M., Sridharan, N. S., Lederberg, J., and Smith, D. H. (1974a), J. Am. Chem. Soc. 96, 7702.

Masinter, L. M., Sridharan, N. S., Carhart, R. E., and Smith, D. H. (1974b) J. Am. Chem. Soc. 96, 7714.

Mislow, K. (1976) Acc. Chem. Res. 9, 26.

Nijenhius, A., and Wilf, H. S. (1975) Combinatorial Algorithms, Academic, N.Y.

Nourse, J. G. (1979) J. Am. Chem. Soc. 101, 1216.
Pólya, G. (1937) Acta. Math. 38, 145.
Randić, M. (1979) Computers and Chemistry, 3, 5.
Randić, M. (1975) J. Am. Chem. Soc. 97, 6609.
Randić, M. (1980) J. Computational Chem. 1, 386.
Robinson, R. W. (1970) J. Combinatorial Theory, 9, 327.
Robinson, R. W., Harary, F. and Balaban, A. T. (1976), Tetrahedron, 32, 355.

Rouvray, D. H. (1974) Chem. Soc. Rev. 3, 355.
Rouvray, D. H. (1975) Endeavour, 34, 28.
Ruch, E., Hässelbarth, W., and Richter, B. (1970), Theor. Chim. Acta. 19, 288.

Table 1. Input for the Program Isomer

| Card | Format | Input Variables | Description |
| :---: | :---: | :---: | :---: |
| 1 | 10A8 | Title | Alphanumeric title |
| 2 | 16I5 | NGCI | Number of cycle indices. <br> Always 1 for generating isomers. |
|  |  | NSUBS | \|R|, number of different substituents. |
|  |  | NT | $\|D\|, ~ t o t a l ~ n u m b e r ~ o f ~ s u b s t i-~$ tutional sites |
|  |  | MODG | \|G|, number of elements in the group $G$. |
| 3 | A10 | SYM | Label of the irreducible representation. For the generation of Isomers it is always Al. |
| 4 | 1615 | NCI | Number of terms in the cycle index. |
| 5 | 1615 | $\begin{aligned} & I \operatorname{COCI}(2), \\ & I=1, N C I \end{aligned}$ | Coefficients of NCI terms in the cycle index. |
|  | For each $J=1$, NCI feed a card described as Card 6. |  |  |
| 6 | 16 I 5 | NPRO | Number of distinct components in each term of the cycle index |
|  |  | $\begin{aligned} & N(I, J), \\ & I=1, N P R O \end{aligned}$ | The superfixes of each component of a term in the cycle index. |
|  |  | $\begin{aligned} & I \exp (I, J), \\ & I=1, \quad \text { NPRO } \end{aligned}$ | The suffixes of each component of a term in the cycle index. |

Table 2. Input for the Isomers of Octahedral Molecules with Six Different Substituents

## Card

1 Isomers of polysubstituent octahedral molecules
$\begin{array}{lllll}2 & 1 & 4 & 6 & 24\end{array}$
3 A1
45

| 5 | 1 | 6 | 3 | 6 | 8 |
| :--- | :--- | :--- | :--- | :--- | :--- |

$\begin{array}{llll}6 & 1 & 6 & 1\end{array}$

| 7 | 2 | 2 | 1 | 1 | 4 |
| :--- | :--- | :--- | :--- | :--- | :--- |

$\begin{array}{llllll}8 & 2 & 2 & 2 & 1 & 2\end{array}$
$\begin{array}{llll}9 & 1 & 3\end{array}$
$\begin{array}{llll}10 & 1 & 2 & 3\end{array}$
ISOMERS CF PGLYSUBSTITUTED OCTAHEDRAL COMPOLNDS
41
PPRU,NIIIA=LNPRC $\quad 6$ IEXPS 1
NPRO,N(1),L=1,NPFO 221
 IEXPS 12
NPRENOIIII=1,NPRU 1
IEXPS
NPRO,N(II,L $=1, N P H C \quad 12$ EEXPS
TOTAL ALYBER OF ISCMERS THE GENERATING FUNCIION FOR ISCMERS
COEFFICIENT

VECTOR

Table 4. Input for the Isomers of Polysubstituted Unbranched Pentane with Four Kinds of Substituents

Card

| 1 | I somers of unbranded polysubstituted pentanes |  |  |  |  |
| ---: | :---: | :---: | :---: | :---: | :---: |
| 2 | 1 | 4 | 12 | 18 |  |
| 3 | A1 |  |  |  |  |
| 4 | 5 |  |  |  |  |
| 5 | 1 | 4 | 4 | 3 | 6 |
| 6 | 1 | 12 | 1 |  |  |
| 7 | 2 | 9 | 1 | 1 | 3 |
| 8 | 2 | 6 | 2 | 1 | 3 |
| 9 | 1 | 6 | 2 |  |  |
| 10 | 2 | 3 | 1 | 2 | 6 |

ISUMLRS OT UNERAACFEL PULYSUBSTITUTEO PENTANES
41

Table 5



Table 5
continued)

1450
$1 \zeta \in c$ 11128
7560
1956
1456
Scte
seit
1596
$136 C$
2066
1300
13 CO
5 C 4 $5 C 4$
88
162
564
1848
$14 \in 6$
1141
564 162
$2 \angle 12$
$+296$
$+256$
2272
$5 \mathrm{C4}$
:147
$42 \leq 6$
6342
$425 t$
1147
140 C
4256
4256
1460
1147
2672
1147
Si4
364
1 C2
88
$35 E$
748
$3!e$
88
356
456
260
1450
$3!6$
148
2200
$2<6 \mathrm{C}$
74 d
148
1496
148 356

[^0]Table 5 (continued)



OROONTROOOWNNHMMOOOONWNNNMFMOOOOOONA


This report was done with support from the Department of Energy. Any conclusions or opinions expressed in this report represent solely those of the author(s) and not necessarily those of The Regents of the University of California, the Lawrence Berkeley Laboratory or the Department of Energy.

Reference to a company or product name does not imply approval or recommendation of the product by the University of California or the U.S. Department of Energy to the exclusion of others that may be suitable.

## Ser


[^0]:    
    
     $\therefore$
    

