communications in mathematical
match
and in computer chemistry

no. 38, October 1998

ISSN 0340-6253

MATCDY (38) 163-172 (1998)

Online applications of SYMMETRICA to the enumeration of permutational isomers and to the enumeration of certain combinatorial libraries

Adalbert Kerber and Axel Kohnert

University of Bayreuth, Department of Mathematics

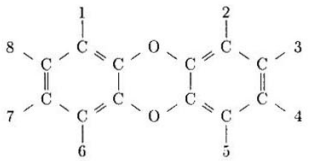
D-95440 Bayreuth, Germany

Abstract

We should like to demonstrate, how the computer algebra package SYMMETRICA can be applied in order to evaluate the size of certain combinatorial libraries of molecules, and, moreover, in order to evaluate a generating function for counting the elements of such libraries by weight. The methods used are the very same ones that can be applied for the enumeration of permutational isomers. They are based on the work of Pólya, and refined using results from the representation theory of symmetric groups.

1 Permutational isomers

In the case when we want to obtain *permutational isomers,* say the isomers of dioxin, we are given the *skeleton* of a molecule, for example, the skeleton of dioxin, which is (after numbering the eight sites from 1 to 8)



Moreover, we are given a set of *ligands* that we have to distribute over the set Y, say, of *sites* of the skeleton. In the case of dioxin, there are altogether 8 sites, and

the set of ligands X is the set $\{H, Cl\}$ consisting of hydrogen and chlorine only. The question is, how many different molecules are there which arise from the skeleton by attaching 4 hydrogen atoms and 4 chlorine atoms to the skeleton. It is well known that there are exactly 22 of them. We should like to describe how this number can be obtained with the help of SYMMETRICA.

Pólya introduced a mathematical concept that allows to treat this case in a lucid way ([1]) which also allows to cover the case of particular combinatorial libraries in a completely analogous way as we shall show down below. He considered sets

$$Y^X := \{f: X \to Y\}$$

of mappings together with actions of a group G on X and the corresponding action of G on Y^X :

$$G \times Y^X \to Y^X : (g, f) \mapsto f \circ g^{-1}.$$

In the case of permutational isomers, X is the set of sites, while Y is the set of ligands — we used these notations already above — and G means the symmetry group of the skeleton. The set of mappings is the set of all the mathematically possible attachments of ligands, while the classes of equivalent attachments are the *orbits* $G(f)$ of the symmetry group on the set of attachments. Hence the total set of orbits

$$G \backslash\backslash Y^X$$

of G on Y^X consists of the classes of essentially different molecules. The desired set of *permutational isomers* is just a complete set of representatives of these classes, a so-called *transversal.*

The total number of these classes or permutational isomers can be evaluated using the well-known Lemma of Cauchy-Frobenius which says that

$$|G \backslash\backslash Y^X| = \frac{1}{|G|} \sum_{g \in G} |Y|^{c(g)},$$

where $c(g)$ means the number of cyclic factors of g on X.

In the dioxin case it is very easy to apply this formula, since the symmetry group G consists just of the following four permutations of the eight sites:

$$1, (16)(25)(34)(78), (12)(38)(47)(56), (15)(26)(37)(48).$$

The number of cyclic factors is 8,4,4,4, respectively. Hence, depending on the number $|Y|$ of ligands, the total number of permutational isomers with this particular skeleton is

$$\frac{1}{4}(|Y|^8 + 3|Y|^4).$$

If $|Y| = 2$ this gives 76 essentially different attachments of hydrogen or chlorine atoms the the dioxin skeleton. The permutational isomers of dioxin form a subset of these since we have to restrict attention to the attachments that contain exactly 4 hydrogen and 4 chlorine atoms.

We therefore go one step further, we enumerate the attachments by weight, i.e. by number of hydrogen and chlorine atoms. Pólya gave a solution for this more difficult problem, too. He proved that the generating function for the classes of attachments by weight is the so-called *group reduction function,* a polynomial that is obtained from the *cycle index* of the action of G on X :

$$Cyc(G,X) := \frac{1}{|G|}\sum_{g\in G}\prod_i z_i^{a_i(g)},$$

where $a_i(g)$ means the number of i-cycles of g on X. The action of the symmetry group of the dioxin skeleton has the following cycle index:

$$\frac{1}{4}(z_1^8 + 3z_2^4).$$

The group reduction function arises from $Cyc(G,X)$ by *Pólya-substitution,* which means by replacing z_i by the polynomial $\sum_{y\in Y} y^i$:

$$Grf(G,X,Y) := \frac{1}{|G|}\sum_{g\in G}\prod_i(\sum_{y\in Y} y^i)^{a_i(g)}.$$

For our example this is the polynomial

$$\frac{1}{4}((y_1+y_2)^8 + 3(y_1^2+y_2^2)^4)$$

$$= y_1^8 + 2y_1^7y_2 + 10y_1^6y_2^2 + 14y_1^5y_2^3 + 22y_1^4y_2^4 + 14y_1^3y_2^5 + 10y_1^2y_2^6 + 2y_1y_2^7 + y_2^8.$$

Replacing y_1 by the symbol H and y_2 by the symbol Cl and exponents by indices, we obtain the generating function of the different attachments:

$$H_8 + 2H_7Cl_2 + 10H_6Cl_2 + 14H_5Cl_3 + 22H_4Cl_4 + 14H_3Cl_5 + 10H_2Cl_6 + 2H_1Cl_7 + Cl_8.$$

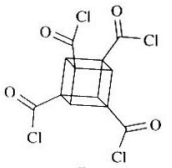
The summand $22H_4Cl_4$ shows that there are exactly 22 essentially different attachments of four hydrogen and four chlorine atoms. These are the permutational isomers of dioxin.

2 Combinatorial chemistry

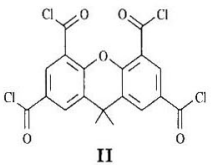
It is interesting to see as it was shown by T. Wieland in MATCH **36** ([2]) — that the very same mathematical tools can be used in order to construct the libraries of molecules described in famous contributions to chemistry, for example in

> T. Carell, E. A. Wintner, A. Bashir-Hashemi, J. Rebek Jr.: Neuartiges Verfahren zur Herstellung von Bibliotheken kleiner organischer Moleküle. *Angewandte Chemie* **1994**, 106, Nr. 20, 2159-2161.
>
> T. Carell, E. A. Wintner, A. J. Sutherland, J. Rebek Jr., Y. M. Dunayevskiy, P. Vouros: New promise in combinatorial chemistry: synthesis, characterization, and screening of small-molecule libraries in solution. *Chem. & Biol.* **2** (1995), 171-183.

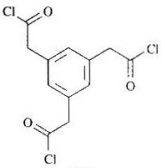
The authors' method is to start from a *central molecule* and then to add *bulding blocks* that react with the central molecule in a well defined way. They mention in particular the following central molecules which react with up to 21 different amino acids:



I



II



III

The first is a cubane derivative, the second one a xanthene, and the third one is a benzene triacid chlorine.

We are now in a position to apply the methods that were used for permutational isomers: The set X now consists of the *active sites* of the central molecule, while the set Y consists of amino acids. The set Y^X of mappings can be identified with all the possible reactions with amino acids, the group G is the symmetry group of the

central molecule, and the orbits of G on Y^X are the classes of equivalent molecules that arise from these reactions. Hence the set

$$G\backslash\backslash Y^X$$

of orbits can be identified with the desired library of molecules consisting of all the mathematically possible molecules that arise from the central molecule by reactions of all the active sites with amino acids taken from Y.

This gives, to begin with, that the *size of the library* is equal to (see above)

$$|G\backslash\backslash Y^X| = \frac{1}{|G|}\sum_{g\in G}|Y|^{c(g)}.$$

This is the cycle index polynomial setting all the the z_i equal $|Y|$. For example in the case of the cubane the symmetry group T consists of 12 rotations of the cube, and so we obtain for the number of orbits the expression [2]

$$\frac{1}{12}\left(|Y|^4 + 11\cdot|Y|^2\right).$$

In the xanthene case, the symmetry group is of order 2, and so the formula for the size of the library turns out to be

$$\frac{1}{2}(|Y|^4 + |Y|^2).$$

In the case of the benzene triacid chloride, the symmetry group is of order 6 on the set of three active sites, and so the size of the library is

$$\frac{1}{6}(|Y|^3 + 3\cdot|Y|^2 + 2\cdot|Y|).$$

Here is a table that provides the sizes of libraries depending on the number of building-blocks used, the first column belongs to the cubane, and the symmetric group as symmetry group, the second column is based on the xanthene core, the third one on the triacid derivative:

n	cubane	xanthen	triacid
1	1	1	1
2	5	10	4
3	15	45	10
4	36	136	20
5	75	325	35
6	141	666	56
7	245	1225	84
8	400	2080	120
9	621	3321	165
10	925	5050	220
11	1331	7381	286
12	1860	10440	364
13	2535	14365	455
14	3381	19306	560
15	4425	25425	680
16	5696	32896	816
17	7225	41905	969
18	9045	52650	1140
19	11191	65341	1330
20	13700	80200	1540
21	16611	97461	1771

Moreover, the polynomial

$$Grf(G,X,Y) := \frac{1}{|G|}\sum_{g\in G}\prod_{i}(\sum_{y\in Y} y^i)^{a_i(g)}.$$

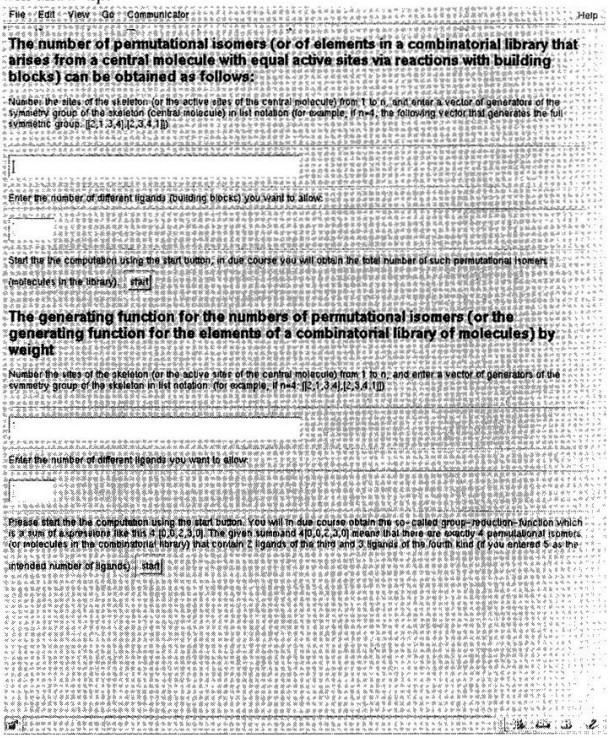
describes the combinatorial library by weight. And so it is time now to see what SYMMETRICA can do in this case.

3 The application of SYMMETRICA

The computer algebra package SYMMETRICA is a collection of routines, written in the programming language C, and so it can be run on nearly every computer platform. The user has to write a few lines of code only, obtaining after compilation a fast executable program. For people, who just want to try small examples without having to write a program there are also *online versions* of example programs written in SYMMETRICA. Let us discuss those of them which apply to the problems described above. The are meanwhile incorporated into the home pages of MOLGEN4.0, the address is

`http://www.mathe2.uni-bayreuth.de/molgen4/`

The interested user needs only to push the button "Online Calc.". The following page will show up:



It shows that you are asked for the generators of the symmetry group in question. Please enter them in *list notation*. For example, in the dioxin case, where we already numbered the sites of the skeleton from 1 to 8, the following two reflections are obviously generators of the symmetry group, and so we enter the following vector

$$[[6,5,4,3,2,1,8,7],[2,1,8,7,6,5,4,3]]$$

Then we enter the number 2 for the number of the two different ligands hydrogen H and chlorine Cl, that we want to consider.

In the first case, when we want only to obtain the *total number* of permutational isomers, then we will obtain the result

```
The input was
[[6,5,4,3,2,1,8,7],[2,1,8,7,6,5,4,3]]
and
2
The result of the computation is
76
the computation was finished after 0.00 seconds on a pentium 133 MHz
```

In the second case, when we want to obtain the *generating function* for the permutaional isomers, we will obtain the following result:

```
The input was
[[6,5,4,3,2,1,8,7],[2,1,8,7,6,5,4,3]]
and
2
The result of the computation is
1 [0,8]  2 [1,7]  10 [2,6]  14 [3,5]  22 [4,4]  14 [5,3]  10 [6,2]
2 [7,1]  1 [8]
the computation was finished after 0.02 seconds on a pentium 133 MHz
```

It shows, for example, that there are exactly 22 permutational isomers of type $[4,4]$, i.e. which contain exactly 4 hydrogen and exactly 4 chlorine atoms. These are the well-known 22 isomers of dioxin. If you want them displayed on the screen, you can use MOLGEN.

In the case of the combinatorial library obtained from cubane by reactions with amino acids we enter a vector of generators of the symmetric group on the four active sites, for example the vector

$$[[2,3,1,4],[1,3,4,2]]$$

and then we enter the number of different amino acids that we want to take into account. Lets say, 5 amino acids. We obtain for the size of the corresponding molecular library the following result:

```
The input was
[[2,3,1,4],[1,3,4,2]]
and
4
The result of the computation is
36
the computation was finished after 0.00 seconds on a pentium 133 MHz
```

in accordance with the table given above. The output of the generating function for the xanthene looks as follows:

```
The input was
[[2,1,4,3]]
and
4
The result of the computation is
1 [0,0,0,4]  2 [0,0,1,3]  4 [0,0,2,2]  2 [0,0,3,1]  1 [0,0,4]
2 [0,1,0,3]  6 [0,1,1,2]  6 [0,1,2,1]  2 [0,1,3]  4 [0,2,0,2]
6 [0,2,1,1]  4 [0,2,2]  2 [0,3,0,1]  2 [0,3,1]  1 [0,4]
2 [1,0,0,3]  6 [1,0,1,2]  6 [1,0,2,1]  2 [1,0,3]  6 [1,1,0,2]
12 [1,1,1,1]  6 [1,1,2]  6 [1,2,0,1]  6 [1,2,1]  2 [1,3]
4 [2,0,0,2]  6 [2,0,1,1]  4 [2,0,2]  6 [2,1,0,1]  6 [2,1,1]
4 [2,2]  2 [3,0,0,1]  2 [3,0,1]  2 [3,1]  1 [4]
```

You may now want to evaluate the other cases, the sizes of the libraries with central xanthene or benzene triacid chloride. You may even want to go deeper into detail and to look for further tools provided by SYMMETRICA. Here is the corresponding internet address:

```
http://www.mathe2.uni-bayreuth.de/axel/symneu_engl.html
```

References

[1] G. PÓLYA. Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. *Acta mathematica*, **68**, pp. 145–253, 1937.

[2] T. WIELAND. Konstruktionsalgorithmen bei molekularen Graphen und deren Anwendung. *MATCH*, **36**, pp. 5–155, 1997.