communications in mathematical
match
and in computer chemistry

no. 46, October 2002
MATCDY (46) 199 - 214 (2002)
ISSN 0304-6253

Enumeration of Combinatorial Libraries obtained from Symmetrical Parent Compounds

Dedicated to E. Ruch

Werner Hässelbarth
Bundesanstalt für Materialforschung
Richard-Willstätter-Straße 11
D-12489 Berlin

Adalbert Kerber
Lehrstuhl II für Mathematik
Universität Bayreuth
D-95440 Bayreuth

Abstract

This contribution to the special issue on tables of marks in chemistry describes applications of enumeration under finite group actions to a particular class of combinatorial libraries. They arise from a symmetric parent compound via reactions with a given set of building blocks. Such libraries were described in particular in [2] and [3], their enumeration by weight is described in [1], using Pólya's theory of enumeration. The enumeration by symmetry group is given in the paper [7], where tables of marks are used. In the present paper we give a refinement of both these methods, enumerating such libraries by weight *and* symmetry group. The mathematical background can be found in the review article [9] contained in the this issue of MATCH and in the book [8].

1 Isomerism

Molecules are described in an *approximative* way. The lowest level of approximation is the *arithmetic level,* where just the multiplicities of the involved elements are given by the *molecular formula.* For example, in the formula C_6H_6 of benzene.

The next level, the *topological level,* uses the *structural formula.* It is an *interaction model* since it emphasizes in addition that certain atoms in the molecule are supposed to interact. The mathematical concept for such interaction models is the notion of *multigraph,* where the vertices are colored by atom names and the edges, the *covalent bonds,* indicate the interaction. Multiple edges can occur, but no loops. For example, to the molecular formula C_6H_6 there correspond 217 mathematically possible structural formulae, the *connectivity isomers.* Nowadays generators are available which give the complete system of mathematically possible structural formulae quickly and redundancy-free (no doublettes show up). They use the information from the arithmetic level (the molecular formula) and accept a lot of optional further conditions on the structural formula, for example ring sizes, hydrogen distribution, hybridization etc. An example is MOLGEN[1] (see e.g. [6]).

The third approximation is the *geometric level,* where structural formulae are placed in space, using, say, an energy model and optimization methods that give an idea how the molecule might "look like" if it is not disturbed by the presence of other molecules or something else. These isomers are called *stereoisomers,* and it is not clear yet, how all the stereoisomers corresponding a structural formulae obtained from a molecular formula can be constructed in general. The reason is that there are usually very many energetic minima, and there are no decent methods yet that allow us to evaluate them easily and in the general case, to classify the corresponding conformations and to evaluate a complete system of representatives of these classes.

A method that allows to bridge *part* of the gap between the topological and the geometric level was suggested very early, namely in the papers of A. C. Lunn and J. K. Senior ([14]) and the seminal paper by G. Pólya ([15]). Lunn and Senior express their approach ([14], p. 1030) to this problem in the following way:

> If the structural (connexity) formula of a compound be written out in full, it will be seen that the molecule can be thought of as a skeleton carrying a

[1] http://www.mathe2.uni-bayreuth.de/molgen4/

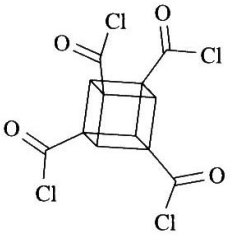
certain number of univalent substituents.

Of course they mention that there may be ambiguities, but once a skeleton and a set of substituents is chosen, it is clear how to proceed using the *symmetry group of the skeleton* and the substituents in order to enumerate the complete set of mathematically possible molecules obtained from the various distributions of the substituents over the *sites* of the skeleton. These molecules are called *permutational* or *substitutional isomers* of the skeleton and the given set of substituents.

This situation has been discussed in various papers many years ago, permutational isomers have been counted, and it was shown that the set of different permutational isomers can bijectively be mapped onto a set of double cosets (see [16], the review article [17] and [10], [11]), which is very helpful once we are interested in the structural formulae of permutational isomers. But we are not going to describe this, since we can refer to the review article [9] in the present issue of MATCH. Moreover, we should like to emphasize the *enumeration by symmetry type and we want to apply these methods to a particular case of combinatorial chemistry,* where libraries of molecules are considered which come from a symmetrical parent compound, as it is described, for example, in [2] and [3].

2 Pólya's Ansatz

Pólya gave a very clear description of a mathematical method that can be used in order to evaluate the *number* of these substitutional isomers as well as the size of libraries coming from symmetric parent compounds. In order to describe this *Ansatz,* we consider an example, taken from [3]: Assume a cubane derivative of the following form:



and suppose that the central cubane molecule forms a *regular cube* in space. The four active sites of this derivative react with amino acids. We assume that there are up to

21 of them. The cubane derivative can react with amino acids, the chemical reaction is well known. The question arises how we can enumerate the corresponding library of molecules, and how we can construct the corresponding structural formulae. In order to describe how this problem can be solved we recall Pólya's method from [9]:

2.1 Pólya's Ansatz

- *Denote the set of sites of the symmetric parent compound by X and the set of (different) substitutents by Y. (Hence, in our present cubane example, $X = \{x_1, x_2, x_3, x_4\}$ and $Y = \{y_1, \ldots, y_{21}\}$, consisting of the active sites and of the admissible different amino acids, respectively.)*

- *Consider an attachment of building blocks (amino acids) to the active sites as a mapping f from the set X of sites into the set Y of substituents:*

$$f: X \to Y.$$

- *Choose the symmetry group G in accordance with the problem in question. (In the cubane case it consists — since the amino acids are chiral and therefore no reflections can be allowed — of the 12 proper rotations of the cube.)*

- *The symmetry group G of the skeleton acts on the set X of sites, i.e. we are given a mapping*

$$G \times X \to X: (g, x) \mapsto gx,$$

such that $g(g'x) = (gg')x$ and $1x = x$, for all $x \in X$ and every $g, g' \in G$.

- *This action of the symmetry group G on the set X of sites induces an action of G on the set*

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

of all the colorations f of X :

$$G \times Y^X \to Y^X: (g, f) \mapsto gf,$$

where gf is defined by

$$gf(x) := f(g^{-1}x).$$

- *The set of orbits*

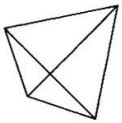
$$G\backslash\backslash Y^X = \{G(f) \mid f \in Y^X\},\ G(f) := \{gf \mid g \in G\},$$

is called the set of symmetry classes *of mappings. Each set of representatives of these classes is a complete set of essentially different molecules in our situation where X is the set of sites of the skeleton and Y the set of (different) admissible substituents, if the symmetry group G was correctly chosen.*

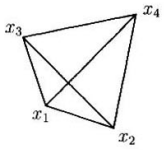
3 Application to Combinatorial Libraries

3.1 The Symmetry Group

In order to apply Pólya's Ansatz we begin with a discussion of the symmetry group of the central cubane derivative. It consists of 12 proper rotations, in mathematical terms it is the *alternating group* A_4. The reason is that the active sites of the cubane form the vertices of a regular *tetrahedron*



In terms of the usual chemical notation for point groups, it is denoted by **T**. The elements of **T** are easily described after numbering the vertices:



The group **T** contains the identity element, which leaves every vertex fixed, together with three double transpositions,

$$(x_1, x_2)(x_3, x_4), (x_1, x_3)(x_2, x_4), (x_1, x_4)(x_2, x_3),$$

each of which consists of two cyclic factors of length 2. Moreover, **T** contains 8 3-cycles,

$$(x_1, x_2, x_3)(x_4), (x_2, x_1, x_3)(x_4), (x_1, x_2, x_4)(x_3), (x_2, x_1, x_4)(x_3),$$

$$(x_1, x_3, x_4)(x_2), (x_3, x_1, x_4)(x_2), (x_2, x_3, x_4)(x_1), (x_3, x_2, x_4)(x_1).$$

(We use the mathematical notation of permutations as product of cyclic factors instead of the standard chemical notation for symmetry elements since it is more suitable for the formulae to come.)

3.2 The Size of the Library

The first step in the enumeration of the corresponding combinatorial library is the evaluation of the size of the library in terms of the number $|Y|$ of admissible amino acids. For this purpose we use the Lemma of Cauchy–Frobenius which says that the number of symmetry classes of mappings is equal to

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

where $c(g)$ means the number of cyclic factors of $g \in G$ on the set of vertices. In our particular case we obtain, since $G = \mathbf{T}$, the following expression for the size of the library:

$$\frac{1}{12} \left(1 \cdot |Y|^4 + 3 \cdot |Y|^2 + 8 \cdot |Y|^2\right). \tag{1}$$

For example, if $|Y| = 20$, we obtain 13700 different molecules that form the library, which means that the $20^4 = 160000$ different attachments of 20 admissible amino acids fall into exactly 13700 symmetry classes of mappings! Here is the sequence of library sizes $|\mathbf{T} \backslash\backslash Y^4|$:

$\lvert Y\rvert$:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$\lvert \mathbf{T} \backslash\backslash Y^4\rvert$:	1	5	15	36	75	141	245	400	621	925	1331	1860	2535	3381

$\lvert Y\rvert$:	15	16	17	18	19	20	21
$\lvert \mathbf{T} \backslash\backslash Y^4\rvert$:	4425	5696	7225	9045	11191	1370	16611

3.3 Enumeration by Weight

The next step is the enumeration by weight, i.e. by prescribed multiplicities for the occurrence of amino acids. According to the result of Pólya we have (for the notation and the definition of multiplicative weight see [9]):

3.1 Theorem *The generating function for the numbers of orbits by weight is*

$$\frac{1}{|G|} \sum_{g \in G} \prod_{i=1}^{|Y|} \left(y_1 + y_2 + \ldots + y_{|Y|}\right)^{a_i(g)},$$

where $a_i(g)$ is the number of cyclic factors of length i in the cycle decomposition of the permutation induced by g on the set X.

For our example library we get

$$\frac{1}{12}\left((y_1+\ldots+y_{|Y|})^4+3(y_1^2+\ldots+y_{|Y|}^2)^2+8(y_1^3+\ldots+y_{|Y|}^3)(y_1+\ldots+y_{|Y|})\right). \quad (2)$$

This gives, for $|Y|=2$ the generating function

$$y_1^4+y_1^3y_2^1+y_1^2y_2^2+y_1^1y_2^3+y_2^4,$$

where, for example, the presence of the summand $y_1^3y_2^1$ (with its coefficient 1) says that there is exactly one element in the library that contains 3 amino acids y_1 and one amino acid y_2. In the case when $|Y|=3$ we obtain

$$y_1^4+y_2^4+y_3^4+y_1^3y_2^1+y_1^3y_3^1+y_1^1y_2^3+y_1^1y_3^3$$
$$+y_2^3y_3^1+y_2^1y_3^3+y_1^2y_2^2+y_1^2y_3^2+y_2^2y_3^2+y_1^2y_2^1y_3^1+y_1^1y_2^2y_3^1+y_1^1y_2^1y_3^2.$$

This generating functions is often called the *group reduction function* and it should be mentioned that these functions can be evaluated online[2]. Moreover it should be said that its coefficients are not always equal to 1, for example, if $|Y|=4$, the group reduction function contains the summand $2y_1y_2y_3y_4$ which means that there are 2 essentially different molecules that are obtained by attaching 4 different amino acids.

3.4 Enumeration by Symmetry Group

The following step is the enumeration by stabilizer class, or, in terms of sciences, *the enumeration by symmetry group.* Burnside's Lemma (see [9] for details) implies

3.2 Theorem *The number of orbits of G on Y^X which have the elements of the conjugacy class $\widetilde{U}_i$ as stabilizers ($\widetilde{U}_i$ an element of the ith conjugacy class of subgroups of G) is the entry in the ith row of the vector*

$$B(G)\cdot\begin{pmatrix}\vdots\\ |Y|^{|U_i\backslash\backslash X|}\\ \vdots\end{pmatrix}.$$

$B(G)$ is the Burnside matrix of G, the inverse of the table of marks.

[2] http://www.mathe2.uni-bayreuth.de/axel/grf_engl.html
or
http://www-ang.kfunigraz.ac.at/~fripert/fga/k1polya.html

In order to apply this consequence of Burnside's Lemma to the cubane library, we take the *table of marks* of the symmetry group $\mathbf{T}$ from [9]: A transversal of the conjugacy classes of subgroups is

$$\begin{array}{lcll} U_1 & = & \langle 1 \rangle & \equiv C_1 = \{I\} \\ U_2 & = & \langle (13)(24) \rangle & \equiv C_2 = \{I, C_2\} \\ U_3 & = & \langle (123) \rangle & \equiv C_3 = \{I, C_3, C_3^2\} \\ U_4 & = & \langle (13)(24), (14)(23) \rangle & \equiv D_2 = \{I, C_2, C_2', C_2''\} \\ U_5 & = & \langle (123), (142) \rangle & \equiv \mathbf{T} \end{array} .$$

The table of marks is

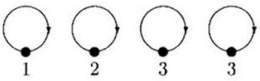
$$M(\mathbf{T}) = M(A_4) = \begin{pmatrix} 12 & 6 & 4 & 3 & 1 \\ 0 & 2 & 0 & 3 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} .$$

Its inverse, the Burnside matrix, looks as follows:

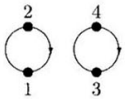
$$B(\mathbf{T}) = B(A_4) = \begin{pmatrix} 1/12 & -1/4 & -1/3 & 1/6 & 1/3 \\ 0 & 1/2 & 0 & -1/2 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1/3 & -1/3 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} .$$

It is easy to see from the cyclic factors of the respective generators that the subgroups U_i, $i = 1, 2, 3, 4, 5$, have $4, 2, 2, 1, 1$ orbits on X :

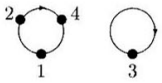
$U_1 = \langle 1 = (1)(2)(3)(4) \rangle$, and hence it has the orbits $\{1\}, \{2\}, \{3\}$, and $\{4\}$:



$U_2 = \langle (13)(24) \rangle$ has the orbits $\{1, 2\}$ and $\{3, 4\}$:



$U_3 = \langle (124) = (124)(3) \rangle$ has the orbits $\{1, 2, 4\}$ and $\{3\}$:



and so on with U_4 and U_5 which both have a single orbit. Hence the desired vector of numbers of molecules by symmetry type in the library is

$$\begin{pmatrix} 1/12 & -1/4 & -1/3 & 1/6 & 1/3 \\ 0 & 1/2 & 0 & -1/2 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1/3 & -1/3 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} |Y|^4 \\ |Y|^2 \\ |Y|^2 \\ |Y|^1 \\ |Y|^1 \end{pmatrix}$$

$$= \begin{pmatrix} (1/12)|Y|^4 - (1/4)|Y|^2 - (1/3)|Y|^2 + (1/6)|Y|^1 + (1/3)|Y|^1 \\ (1/2)|Y|^2 - (1/2)|Y|^1 \\ |Y|^2 - |Y|^1 \\ (1/3)|Y|^1 - (1/3)|Y|^1 \\ |Y|^1 \end{pmatrix}$$

$$= \begin{pmatrix} (1/12)|Y|^4 - (7/12)|Y|^2 + (1/2)|Y|^1 \\ (1/2)|Y|^2 - (1/2)|Y|^1 \\ |Y|^2 - |Y|^1 \\ 0 \\ |Y|^1 \end{pmatrix}.$$

Here is a table of values for the smallest values of $|Y|$:

$\lvert Y\rvert$:	1	2	3	4	5
$\lvert T\backslash\backslash_{\widetilde{U}_1} Y^4\rvert$:	0	0	3	14	40
$\lvert T\backslash\backslash_{\widetilde{U}_2} Y^4\rvert$:	0	1	3	6	10
$\lvert T\backslash\backslash_{\widetilde{U}_3} Y^4\rvert$:	0	2	6	12	20
$\lvert T\backslash\backslash_{\widetilde{U}_4} Y^4\rvert$:	0	0	0	0	0
$\lvert T\backslash\backslash_{\widetilde{U}_5} Y^4\rvert$:	1	2	3	4	5

This table shows that there are exactly $3, 14, 40, \ldots$ molecules with trivial (rotational) symmetry group $\widetilde{U}_1 = \{1\}$ in the library if $|Y| = 3, 4, 5, \ldots$

3.5 Enumeration by Weight and Symmetry Group

There is also the *weighted form of Burnside's Lemma* (see [9]). Its application to the enumeration of symmetry classes of mappings gives the following result:

3.3 Theorem *The generating function for the enumeration of G–classes on Y^X of type $\widetilde{U}_j$ by weight $w\colon f \mapsto \prod f(x) \in \mathbb{Q}[Y]$ is the j-th row of the following one column matrix:*

$$B(G) \cdot \begin{pmatrix} \vdots \\ \prod_{\nu \in |U_i \backslash X|} \sum_y y^{l_\nu(U_i)} \\ \vdots \end{pmatrix},$$

where $l_\nu(U_i)$ denotes the length of the ν-th orbit of U_i on X.

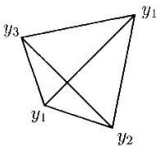
Since it is easy to see from their generators that the subgroups U_i, $i = 1, 2, 3, 4, 5$, have $4, 2, 2, 1, 1$ orbits on X, the desired vector of generating functions for the enumeration of the molecules by symmetry type and weight in the library arising from the central cubane derivative is

$$\begin{pmatrix} 1/12 & -1/4 & -1/3 & 1/6 & 1/3 \\ 0 & 1/2 & 0 & -1/2 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1/3 & -1/3 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \left(\sum_{y \in Y} y\right)^4 \\ \left(\sum_{y \in Y} y^2\right)^2 \\ \left(\sum_{y \in Y} y^3\right)\left(\sum_{y \in Y} y\right) \\ \left(\sum_{y \in Y} y^4\right) \\ \left(\sum_{y \in Y} y^4\right) \end{pmatrix}.$$

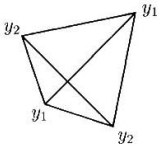
Here is a table of generating functions:

$\lvert Y\rvert$:	1	2	3
$\widetilde{U}_1$:	0	0	$y_1^2y_2y_3 + y_1y_2^2y_3 + y_1y_2y_3^2$
$\widetilde{U}_2$:	0	$y_1^2y_2^2$	$y_1^2y_2^2 + y_1^2y_3^2 + y_2^2y_3^2$
$\widetilde{U}_3$:	0	$y_1^3y_2 + y_1y_2^3$	$y_1^3y_2 + y_1^3y_3 + y_1y_2^3 + y_1y_3^3 + y_2^3y_3 + y_2y_3^3$
$\widetilde{U}_4$:	0	0	0
$\widetilde{U}_5$:	y_1^4	$y_1^4 + y_2^4$	$y_1^4 + y_2^4 + y_3^4$

For example,



is asymmetric, i.e. it has $U_1 = \{1\}$ as (rotational) symmetry group, it corresponds to the summand $y_1^2y_2y_3$ in the first row and the last column of the foregoing table, while



is of type $\widetilde{U}_2 = C_2$ and corresponds to the summand $y_1^2y_2^2$ in the second row and the last but one column.

3.6 Construction by Symmetry Group

A generator for combinatorial libraries, an extension of MOLGEN, is in progress, it will be described elsewhere. Besides the generation of combinatorial libraries it allows statistical

examinations of the library since it comprises an interface to powerful statistical packages like R or S-Plus.

Besides a construction of the total library we can restrict attention to elements with given symmetry group. For this purpose we can use Laue's Lemma ([12] and [9]), which is in fact a *constructive version* of Burnside's Lemma. It is based on the observation that the normalizer $N_G(U)$ of U in G acts on the set $(Y^X)_U$ of mappings fixed by U, and implies that every transversal of the orbits of the normalizer $N_G(U)$ on the set of fixed points

$$\left(\hat{Y^X}\right)_U = \left(Y^X\right)_U - \bigcup_{V:\ U\ max.\ in\ V} \left(Y^X\right)_V$$

is a transversal of the set of orbits of G on Y^X which are of type $\widetilde{U}$. More formally, each

$$T \in \mathcal{T}\left(N_G(U)/U \backslash\backslash \left(\hat{Y^X}\right)_U\right)$$

is a transversal of the orbits of type $\widetilde{U}$ of G on X :

$$T \in \mathcal{T}\left(G \backslash\backslash_{\widetilde{U}} Y^X\right).$$

(We use the notation $\mathcal{T}(H \backslash\backslash M)$ for the total set of transversals of the set of orbits $H \backslash\backslash M$.) Moreover, all these orbits of $N_G(U)/U$ are of the same size $|N_G(U)/U|$, and hence it is easy to generate elements of them uniformly at random.

For example, if $X = \{1,2,3,4\}$, $G = \mathbf{T} = A_4$ and $Y = \{y_1, y_2\}$, then we can see from [9], where the table of marks of $\mathbf{T}$ is given, that it has subgroups $U_2 = \{1, (13)(24)\}$ and $U_4 = \{1, (12)(34), (13)(24), (14)(23)\}$. Moreover, U_2 is maximal in U_4. Their orbit sets on X are

$$U_2 \backslash\backslash Y^X = \{\{1,3\},\{24\}\},\ U_4 \backslash\backslash Y^X = \{\{1,2,3,4\}\}.$$

Hence the set of mappings $f = (f(1), f(2), f(3), f(4))$ which are fixed under U_2 is

$$\left(Y^X\right)_{U_2} = \{(y_1,y_1,y_1,y_1), (y_2,y_2,y_2,y_2), (y_1,y_2,y_1,y_2), (y_2,y_1,y_2,y_1)\}.$$

Since U_4 is transitive, we have to subtract from this set the constant mappings, obtaining

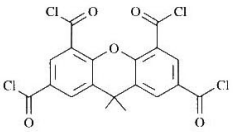
$$(\hat{Y^X})_{U_2} = \{(y_1,y_2,y_1,y_2), (y_2,y_1,y_2,y_1)\}.$$

Since U_4 is transitive and contained in the normalizer of U_2, this set $(\hat{Y^X})_{U_2}$ is an orbit of $N_{A_4}(U_2)$ and we find a transversal T of the type U_2-orbits of $\mathbf{T}$:

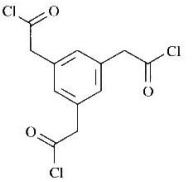
$$T = \{(y_1,y_2,y_1,y_2)\} \in \mathcal{T}\left(\mathbf{T} \backslash\backslash_{\widetilde{U_2}} \{y_1,y_2\}^X\right).$$

4 Two Further Libraries

In the papers [2] and [3] there are two further libraries contained which we should also consider under enumerative aspects. One of them has a xanthene as central molecule



and the other one a benzene triacid chlorine as its symmetric parent compound:



4.1 The library arising from xanthene

In the xanthene case, the symmetry group $G = C_2$ is of order 2 and there are 4 active sites. Hence the symmetry group consists of the identity element and a double transposition $(x_1x_3)(x_2x_4)$. Therefore, the formula for the size of the library turns out to be

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

Here is the table of the corresponding sizes of the libraries:

$\lvert Y\rvert$:	1	2	3	4	5	6	7	8	9	10	11	12
$\lvert \mathbf{T}\backslash\backslash Y^4\rvert$:	1	10	45	136	325	666	1225	2080	3321	5050	7381	10440

$\lvert Y\rvert$:	13	14	15	16	17	18	19	20	21
$\lvert \mathbf{T}\backslash\backslash Y^4\rvert$:	14365	19306	25425	32896	41905	52650	65341	80200	97461

The enumeration by weight gives, according to the theorem of Pólya, in the xanthene case the formula

$$\frac{1}{2}\left(\left(\sum_{y\in Y} y\right)^4 + \left(\sum_{y\in Y} y^2\right)^2\right),$$

for example we obtain y_1^4, if $|Y| = 1$, while for $|Y| = 2$ we get

$$y_1^4 + 2y_1^3y_2 + 4y_1^2y_2^2 + 2y_1y_2^3 + y_2^4.$$

Burnside's Lemma gives the numbers of molecules with prescribed symmetry group. In the xanthene case there are two subgroups only, $U_1 = \{1\}$ and $U_2 = \{(x_1x_3)(x_2x_4)\}$, and the formula is

$$\begin{pmatrix} 1/2 & -1/2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} |Y|^4 \\ |Y|^2 \end{pmatrix},$$

here are the values for several $|Y|$:

$\lvert Y\rvert$:	1	2	3	4	5	6	7
$\lvert C_2 \backslash\backslash_{\tilde{U}_1} Y^4\rvert$:	0	6	36	120	300	630	1176
$\lvert C_2 \backslash\backslash_{\tilde{U}_2} Y^4\rvert$:	1	4	9	16	25	36	49

For the enumeration of molecules by weight and symmetry group we find the formula

$$\begin{pmatrix} 1/2 & -1/2 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \left(\sum_{y\in Y} y\right)^4 \\ \left(\sum_{y\in Y} y^2\right)^2 \end{pmatrix},$$

here is the resulting vector fo the case $|Y| = 2$:

$$\begin{pmatrix} 2y_1^3y_2 + 2y_1^2y_2^2 + 2y_1y_2^3 \\ y_1^4 + 2y_1^2y_2^2 + y_2^4 \end{pmatrix}.$$

4.2 The library arising from the benzene triacid chlorine

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

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

The table of sizes is

$\lvert Y\rvert$:	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$\lvert \mathbf{T} \backslash\backslash Y^4\rvert$:	1	4	10	20	35	56	84	120	165	220	286	364	455	560

$\lvert Y\rvert$:	15	16	17	18	19	20	21
$\lvert \mathbf{T} \backslash\backslash Y^4\rvert$:	680	816	969	1140	1330	1540	1771

The enumeration by weight gives, according to the theorem of Pólya,

$$\frac{1}{6}\left(\left(\sum_{y\in Y} y\right)^3 + 3 \cdot \left(\sum_{y\in Y} y\right)\left(\sum_{y\in Y} y^2\right) + 2 \cdot \left(\sum_{y\in Y} y^3\right)\right).$$

Burnside's Lemma gives the numbers of molecules with prescribed symmetry group $G = C_3$: A transversal of the conjugacy classes of subgroups is

$$U_1 = \langle 1\rangle, U_2 = \langle (12)\rangle, U_3 = \langle (123)\rangle, U_4 = \langle (123), (12)\rangle = D_3.$$

The corresponding formula for the numbers of molecules by symmetry group turns out to be

$$\begin{pmatrix} 1/6 & -1/2 & -1/6 & 1/2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1/2 & -1/2 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} |Y|^3 \\ |Y|^2 \\ |Y| \end{pmatrix},$$

which gives the following values, for example,

$\lvert Y\rvert:$	1	2	3	4	5	6	7
$\lvert C_3 \backslash\backslash_{\tilde{U}_1} Y^4\rvert:$	0	0	1	4	10	20	35
$\lvert C_3 \backslash\backslash_{\tilde{U}_2} Y^4\rvert:$	0	2	6	12	20	30	42
$\lvert C_3 \backslash\backslash_{\tilde{U}_3} Y^4\rvert:$	0	0	0	0	0	0	0
$\lvert C_3 \backslash\backslash_{\tilde{U}_4} Y^4\rvert:$	1	2	3	4	5	6	7

For the enumeration of molecules by weight and symmetry group we find the formula

$$\begin{pmatrix} 1/6 & -1/2 & -1/6 & 1/2 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1/2 & -1/2 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \left(\sum_{y\in Y} y\right)^3 \\ \left(\sum_{y\in Y} y\right)\left(\sum_{y\in Y} y^2\right) \\ \left(\sum_{y\in Y} y^3\right) \\ \left(\sum_{y\in Y} y^3\right) \end{pmatrix},$$

here is the result for $|Y| = 2$:

$$\begin{pmatrix} 0 \\ y_1^2 y_2 + y_1 y_2^2 \\ 0 \\ y_1^3 + y_2^3 \end{pmatrix}.$$

References

[1] W. van Almsick, H. Dolhaine, H. Hönig: Efficient algorithms to enumerate isomers and diamutamers with more than one type of substituents. *J. Chem. Inf. Comput.*, **40**, 956-966, 2000.

[2] Th. Carell, E. A. Wintner, A. Bashir-Hashemi, A., J. Rebek Jr.: Neuartiges Verfahren zur Herstellung von Bibliotheken kleiner organischer Moleküle. Angewandte Chemie 106, 2159-2161 (1994).

[3] Carell, T., 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**, S. 171–183, 1995.

[4] S. EL–BASIL: Combinatorial Organic Chemistry: An Educational Approach. Nova Science Publishers Inc.,2000.

[5] S. FUJITA: Symmetry and Combinatorial Enumeration in Chemistry. Springer-Verlag, Berlin 1991.

[6] R. GRUND, A. KERBER, AND R. LAUE: MOLGEN, ein Computeralgebra-System für die Konstruktion molekularer Graphen. *MATCH*, **27**, 87-131, 1992.

[7] W. HÄSSELBARTH: Substitution symmetry. *Theor. Chim. Acta (Berl.)*, **67**, 339-367 (1985).

[8] A. KERBER: Applied Finite Group Actions. *Springer-Verlag 1999*

[9] A. KERBER: Enumeration under Finite Group Action, Basic Tools, Results and Methods. *MATCH* **46**, 151-198 (2002).

[10] A. KERBER, R. LAUE: Group actions, double cosets, and homomorphisms: unifying concepts for the constructive theory of discrete structures. *Acta Applicandae Mathematicae,* **52**, 63-90 (1998).

[11] A. KERBER, R. LAUE, TH. WIELAND: Discrete Mathematics for Combinatorial Chemistry. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science,* **51**, 225-234 (2000).

[12] R. LAUE: Eine konstruktive Version des Lemmas von Burnside. *Bayreuther Mathematische Schriften*, **28**, 111-125, 1989.

[13] R. LAUE: Construction of combinatorial objects – a tutorial. *Bayreuther Mathematische Schriften*, **43**, 53-96, 1993.

[14] A. C. LUNN, J. K. SENIOR: Isomerism and Configuration. *J. Phys. Chem.* 33 (1929), 1027 - 1079.

[15] G. PÓLYA: Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. *Acta mathematica*, **68**, 145-253, 1937.

[16] E. RUCH, W. HÄSSELBARTH, B. RICHTER: Doppelnebenklassen als Klassenbegriff und Nomenklaturprinzip für Isomere und ihre Abzählung. *Theoretica Chimica Acta*, **19** (1970), 288-300.

[17] E. RUCH, D. J. KLEIN: Double cosets in chemistry and physics. *Theoretica Chimica Acta*, **63** (1983), 447-472.