Journal of Mathematical Chemistry 21 (1997) 141–157

Combinatorics of combinatorial chemistry *

Thomas Wieland **

Lehrstuhl II für Mathematik, University of Bayreuth, 95440 Bayreuth, Germany

Received 20 November 1996

The hot topic among medicinal chemists today is a novel technique for chemical synthesis in drug research called *combinatorial chemistry*, where usually a core structure and some building-block molecules are given and all combinatorially possible combinations are produced. The resulting set of compounds (called a *library*) can afterwards be systematically screened for a desired biological activity. In this paper we discuss the applications of the mathematical discipline of combinatorics to this process, especially an algorithm for the exhaustive and redundancy-free generation of a combinatorial library as well as equations for the enumeration of library sizes.

1. Introduction

With the upcoming of new analysis automata, it became possible to examine several thousands of compounds a day for their biological activity. Together with the necessity of cost reduction in industrial research, this *high-throughput screening* has raised the desire for making very large numbers of novel molecules available. In the recent years a new technique is used for this purpose: *combinatorial chemistry* [6,12,13,32,35]. It does not aim at the classical objective of synthesizing *one* substance as pure as possible, but it deliberately utilizes the structural variety to produce a large number of compounds simultaneously.

Typically a set of *building-blocks* is taken that is systematically combined with a *core* structure in all combinatorially possible ways where the actual reations make use of chemical, biological or biosynthetical procedures. The set of all resulting molecules is called a *combinatorial library*.

Despite the high-throughput capacities, it is still reasonable to keep the sizes of the libraries small – especially to avoid one pharmacological class being tested over and over again. So in the preparation of combinatorial chemistry experiments, mathematical modelling is essential [23,25], and mathematical combinatorics can provide it.

In this paper we want to discuss an algorithm for the generation of combinatorial

* Supported by the BMBF under 03-KE7BAY-9

** Present address: German Aerospace Research Establishment (DLR), DFD, 82234 Weßling, Germany. E-mail: wieland@dfd.dlr.de

libraries which is exhaustive (i.e., it must generate all molecules corresponding to the given data) and redundancy-free (i.e., no molecules should occur twice). It can be compared with isomer generators [3,4,7,8,11,18,24,36], i.e., computer programs capable of calculating all possible isomers to a given empirical formula and further optional conditions, but is different in the task since here usually blocks cannot be attached by simply adding an edge.

We will also present general formulae to compute the sizes of combinatorial libraries, following the tradition of Cayley [9], Redfield [29] and Pólya [26,27] who already emphasized the benefits of combinatorial enumeration in chemistry.

2. Basic definitions

In this paper we consider graphs as labeled multi-graphs, i.e., as mappings

$$\gamma : \boldsymbol{p}^{[2]} \to \{0, \ldots, m-1\}, \quad \text{in short } \gamma \in m^{\boldsymbol{p}^{[2]}},$$

where $\boldsymbol{p}^{[2]}$ is the set of all 2-subsets of $\boldsymbol{p} := \{1, \ldots, p\}$, if the graph has p vertices, $\gamma(\{i,j\}) = k$ if there is an edge of degree k between the vertices i and j, and $\gamma(\{i,j\}) = 0$ if the two vertices are not connected.

For molecules we take the usual model dating back to Crum Brown (cf. [5]), identifying atoms as vertices and bonds as edges. The atomic types are defined by an additional mapping $\beta : \boldsymbol{p} \to \{E_1, E_2, \ldots\}$ with the E_i representing chemical elements, such that a molecular graph is a pair (γ, β) of a graph and a coloring of the vertices with atomic types.

Furthermore, we call

$$\eta \in m^{T^{[2]}} \quad \text{with } T \subseteq \boldsymbol{p},\ \forall i,j \in T\colon\ \eta(\{i,j\}) = \gamma(\{i,j\}),$$

a *subgraph* of γ, denoted by $\eta \subseteq \gamma$.

3. Group actions and orderly generation

We will introduce some basic definitions and notations from algebra which will be needed furtheron (for more details, see, e.g., [17]).

Definition 1. Let G be a group, and Ω a non-empty set. A mapping

$$G \times \Omega \to \Omega,\ (g, \omega) \mapsto g\omega$$

with $g'(g\omega) = (g'g)\omega\ \forall g, g' \in G,\ \forall \omega \in \Omega$ and $1\omega = \omega$ is called an *action of G on Ω*, abbreviated by ${}_G\Omega$.

A group action is called *finite* if both the group and the set are finite. In this paper we will only use finite actions.

Group actions give rise to several important sets:

Definition 2. Let G be a group acting finitely on Ω, $\omega \in \Omega$ and $\Delta \subseteq \Omega$.

- $G(\omega) := \{g\omega \mid g \in G\}$ is called *orbit* of ω.
- $G\backslash\backslash\Omega := \{G(\omega) \mid \omega \in \Omega\}$ is called *set of orbits*.
- $\mathcal{T}(G\backslash\backslash\Omega)$ is called *transversal of the orbits* with $\Omega = \dot{\bigcup}_{t\in\mathcal{T}} G(t)$, derived from the equivalence class property of the orbits.
- $\Omega_g := \{\omega \in \Omega \mid g\omega = \omega\}$ is called the *set of fixed points* of g.
- $C_G(\Delta) := \{g \in G \mid g\delta = \delta \ \forall\delta \in \Delta\}$ is called *centralizer* or *pointwise stabilizer* of Δ in G.
- $N_G(\Delta) := \{g \in G \mid g\delta \in \Delta \ \forall\delta \in \Delta\}$ is called *normalizer* or *setwise stabilizer* of Δ in G.

Let X and Y denote two finite non-empty sets. Then we set $Y^X := \{f \mid f : X \to Y\}$. If G acts on X, then G also acts on Y^X as

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

Typical sets for X and Y are sets of natural numbers like $\boldsymbol{n} := \{1, 2, \ldots, n\}$.

The type of group which we will use only is the *symmetric group* $S_n := \{\pi \in \boldsymbol{n}^{\boldsymbol{n}} \mid \pi \text{ bijective}\}$ and its subgroups, which we call *permutation groups*.

In the context of graphs, automorphism groups are important structures, since a group action on the points induces an action on the pairs of points:

Definition 3. The stabilizer of a labeled multi-graph $\gamma \in m^{\boldsymbol{p}^{[2]}}$ is called *automorphism group* of γ:

$$\mathrm{Aut}(\gamma) := C_{S_p}(\gamma) = \{\pi \in S_p \mid \pi\gamma = \gamma\}.$$

For molecular graphs (γ, β), we set $\mathrm{Aut}(\gamma, \beta) := C_{S_p}(\gamma, \beta)$.

We will now consider a subgraph $\eta \subseteq \gamma$, as defined above. The automorphisms $\pi \in A := \mathrm{Aut}(\gamma)$ that keep η fixed certainly hold: $\pi \in N_A(\eta)$. There may be some $\pi, \pi' \in N_A(\eta) \subseteq A$ with $\pi \downarrow \eta = \pi' \downarrow \eta$, but $\pi \neq \pi'$. So to obtain the automorphisms of the subgraph we have to consider cosets after the centralizer of η. This yields:

Proposition 4. The automorphisms of a subgraph $\eta \subseteq \gamma$ in a graph γ with automorphism group A induce on η a group isomorphic to $N_A(\eta)/C_A(\eta)$.

Proof. The embedding of $N_A(\eta)$ in $\mathrm{Aut}(\eta)$ has the kernel $C_A(\eta)$ such that the homorphism theorem immediately yields the assertion. □

For constructing transversals of orbits, the naive approach is to compare any new element with all previously calculated; but this is completely inappropriate for practical use. A helpful principle is the concept of *orderly generation*, a method that

was introduced by R.C. Read [28] and that can be refined considerably [15–17,22,33]. It is based on the fact that total orders on X and Y induce a canonic total order on Y^X, the lexicographic order, so that a *canonic transversal*

$$\mathcal{T}_>\big(G\backslash\backslash Y^X\big),$$

consisting of the biggest elements of the orbits, does exist.

Theorem 5 ([17,22,28]). Let ${}_G\Omega$ be a finite group action where Ω is assumed to be totally ordered by $\leqslant$ and to possess a disjoint decomposition

$$\Omega = \bigcup_{i=1}^{n} \Omega_i$$

in invariant and non-empty subsets Ω_i. Let A be an algorithm that produces for each $\omega \in \Omega$ either the empty set or a set $A(\omega) \subseteq \Omega$ in descending order such that the following conditions hold for the canonic transversals $\mathcal{T}_>^{(i)}$ of $G\backslash\backslash\Omega_i$ for all $i \in \{1,\dots,n-1\}$:

- $$\mathcal{T}_>^{(i+1)} \subseteq \bigcup_{\omega\in\mathcal{T}_>^{(i)}} A(\omega).$$
- For all $\omega_1,\omega_2 \in \mathcal{T}_>^{(i+1)}$ with $\omega_1 < \omega_2$ we have that $\omega_1 \in A(\omega_1')$ and $\omega_2 \in A(\omega_2')$ implies $\omega_1' < \omega_2'$.

Then the desired transversal of $G\backslash\backslash\Omega$ can be obtained by proceeding as follows:

1. Determine $\mathcal{T}_>^{(1)}$ and set $\mathcal{T}_> \leftarrow \mathcal{T}_>^{(1)}$.
2. For $i \in \{1,\dots,n-1\}$ determine all $A(\omega^{(i)})$ for $\omega^{(i)} \in \mathcal{T}_>^{(i)}$, calculate $\mathcal{T}_>^{(i+1)}$ by unifying the $A(\omega^{(i)})$ and eliminating all non-canonic elements and set
$$\mathcal{T}_> \leftarrow \mathcal{T}_> \cup \mathcal{T}_>^{(i+1)}.$$

Proof. Clear. □

The decisive result that can be derived from the general form reads:

Proposition 6. Let $f \in \mathcal{T}_>(G\backslash\backslash Y^X)$ and $f_1 \in Y^X$ be a starting piece of f, i.e., there exists a $t \leqslant n$ with

$$f_1(j) = \begin{cases} f(j), & \text{for } j < t,\\ 0, & \text{for } j \geqslant t.\end{cases}$$

Then $f_1 \in \mathcal{T}_>(G\backslash\backslash Y^X)$.

Proof. We perform an indirect proof and assume that $f_1 \notin \mathcal{T}_>(G\backslash\backslash Y^X)$. Then a $\pi \in G$ exists with $\pi f_1 > f_1$, i.e., there is also $\tilde{t} < n$, such that

$$\pi f_1(j) = f_1(j) \quad \text{for } j < \tilde{t} \quad \text{and} \quad \pi f_1(\tilde{t}) > f_1(\tilde{t}).$$

We have $\tilde{t} < t$, since otherweise due to $f_1(j) = 0$ for $j \geqslant t$ also πf_1 could not become larger than f_1 under the permutation π.

We differentiate after the impact of π:

1. π exchanges the places $k < \tilde{t}$ and $k' > \tilde{t}$ (i.e., $\pi^{-1}(k) = k'$) with $f_1(k) = f_1(k') = 0$, where $f(k') > 0$. (If there are more than one pair (k, k'), we choose the one with the smallest k.) According to our assumption we get $\pi f(j) = \pi f_1(j) = f_1(j) = f(j)$ for $j < k$ and $f(k) = f_1(k) = \pi f_1(k) = 0$, but $\pi f(k) = f(k') > 0$, leading to $f(k) < \pi f(k)$. This means $f < \pi f$, in contradiction to the assumed maximality of f.
2. In all other cases we have $\pi f(j) = f(j)\, \forall j < \tilde{t}$ and $\pi f(\tilde{t}) = \pi f_1(\tilde{t}) > f_1(\tilde{t}) = f(\tilde{t})$, which is again in contradiction to the maximality of f.

□

This proposition tells us how to use orderly generation algorithmically: it is sufficient to expand starting pieces lexicographically without having to re-test them on maximality. This means in the opposite that, if the starting piece is already not canonic, it cannot become a canonic representative by filling the remaining places. (For more details see [16,22,33].)

4. Reaction schemes

We would like to start the presentation of our method for library generation with a syntax to describe the underlying chemical reactions formally, especially the two-component synthesis like

$$\mathrm{A} + \mathrm{B} \rightarrow \mathrm{C}.$$

In most cases, subgraphs determine the course of the reaction.

Definition 7. Let (η_1, β_1) and (η_2, β_2) with $\eta_1 \in m^{\boldsymbol{r}^{[2]}}$ and $\eta_2 \in m^{\boldsymbol{s}^{[2]}}$ be molecular graphs. A *reaction scheme* is defined as the triple

$$\big((\eta_1, \beta_1),\ (\eta_2, \beta_2),\ \rho\big),$$

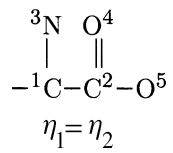
where $\rho : \boldsymbol{p} \times \boldsymbol{q} \rightarrow \mathbb{Z} \cup \{-\infty\}$ is a mapping with

$$\rho(i,j) = \begin{cases} k, & i \text{ and } j \text{ are to be connected by a bond of degree } k, \\ 0, & i \text{ and } j \text{ remain unconnected}, \\ -\infty, & \text{one of the atoms } i \text{ or } j \text{ is dropped.} \end{cases}$$

By means of this definition[1] many two component reactions can be described sufficiently. In fact, our main interest in such a reaction lies in the changes of the graphs, and not in the experimental aspects (like reaction conditions, catalysts or equilibria).

A corresponding algorithm could be formulated to link two graphs by a reaction scheme over all reacting subgraphs. As we will not explicitly need such a procedure for library generation, we omit a deeper discussion and just present an example:

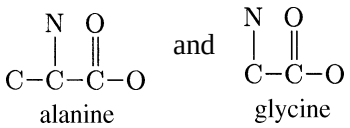
Example 1. Peptids are protein molecules built from at least two amino acids which play a central role in biochemistry. The joining of the single amino acids is performed by condensation of the acid group (COOH) and the amid group (NH_2). Thus the decisive reaction structure is the α amino acid group, which must be contained in both reaction partners:



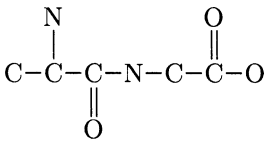
The condensation is represented by the mapping

$$\rho = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -\infty & -\infty & -\infty & -\infty & -\infty \end{pmatrix}.$$

We consider the amino acids



Obviously both contain the subgraph η_1. Despite the equality of the subgraphs in the reaction scheme, the order of the initial graphs is essential. Taking alanine as the first one, say, we obtain



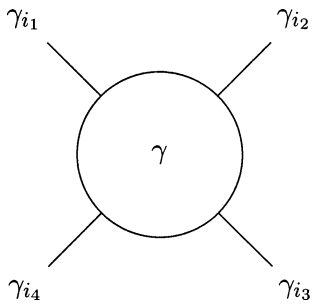
[1] This definition is a simplification of the situation and is only used for a formalization of the construction problem discussed below. For more sophisticated purposes more comprehensive approaches like the algebra of be- & r-matrices [10] are necessary.

5. Multiple attachments to a core structure

The evolvement of a combinatorial library includes a special type of reaction scheme that is given by a core structure with several reaction sites and a number of ligand compounds.

Let $((\eta_1, \chi_1),\ (\eta_2, \chi_2),\ \rho)$ denote a reaction scheme, (γ, β) a molecular graph containing k substructures isomorphic to (η_1, χ_1) with $k > 1$, and a set of molecular graphs $(\gamma_1, \beta_1), \ldots, (\gamma_n, \beta_n)$; for sake of simplicity we assume that each contains exactly one substructure isomorphic to (η_2, χ_2).

The first task is to determine all attachments of the ligands to the sites of the core, where the sites are given by the substructures of the reaction scheme. For $k = 4$, e.g., the situation is:



Topological equivalence of the atoms is expressed by the automorphism group $\mathrm{Aut}(\gamma, \beta)$. The permutation group $P_\gamma \leqslant S_k$ of the sites can be derived according to proposition 4 as an induced subgraph automorphism group of the subgraph which is determined by the first vertex of each ζ_i (see algorithm 9). So P_γ acts on the sites $\boldsymbol{k}$ which we want to assign with n different ligands. These arguments lead to:

Lemma 8. The essentially different possibilities to attach n ligand structures, which contain the corresponding subgraph of the given reaction scheme exactly once, to the k different reaction sites of a core structure (γ, β) correspond to a transversal of the action of P_γ on $\boldsymbol{n}^{\boldsymbol{k}}$:

$$\mathcal{T}\left(P_\gamma \backslash\backslash \boldsymbol{n}^{\boldsymbol{k}}\right).$$

So now we can state a strategy:

Algorithm 9 (Attachment of ligands to a core structure).

1. Determine the group P_γ as well as all subgraphs $\zeta_1, \ldots, \zeta_k \subseteq \gamma$ which are isomorphic to η_1.
2. Compute for $i \in \boldsymbol{n}$ the subgraphs $\zeta^{(i)} \subseteq \gamma_i$ which are isomorphic to η_2.
3. Use orderly generation to obtain the next representative $f \in \boldsymbol{n}^{\boldsymbol{k}}$ under the action of P_γ.

4. Determine the total graph which yields from the attachment of the ligands $\gamma_{f(1)}, \ldots, \gamma_{f(k)}$ to γ according to ρ, i.e., by combining the graphs, eliminating vertices which have to be dropped and adding the necessary edges.
5. If there are further orbit representatives, go to step 3.

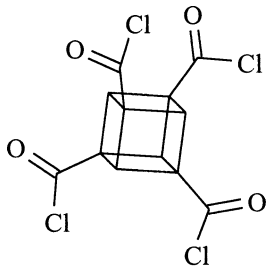
Due to orderly generation in step 3 and the uniqueness of the subgraphs of the ligands our requirements of exhaustiveness and irredundancy are fulfilled.

6. Generation of libraries

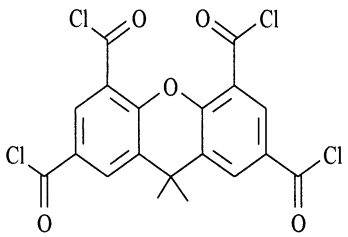
For the generation of a combinatorial library from given building-blocks algorithm 9 is perfectly suited, since the basic situation of combinatorial chemistry as described in section 1 is just that of this method.[2]

For practical use it is moreover relevant that the multiplicity of a certain building-block can be restricted, i.e., that a (γ_i, β_i) occurs in all compounds of the library at least u and at most v times. This can be reached by an additional test in algorithm 9 between step 3 and step 4. In laboratory, this restriction can be statisfied by an appropriate modification of the reaction conditions.

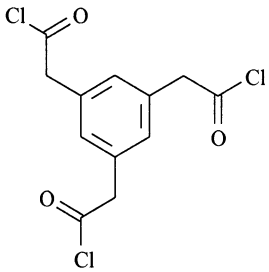
As an example we consider the combinatorial libraries from [6]. The authors used as building-blocks the twenty natural amino acids (cf. figure 1) and as core structures some acid chlorides:



I



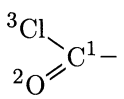
II

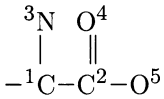


III

a cubane-derivative (structure **I**), xanthene (**II**) and a benzene triacid chloride (**III**).

The reaction scheme consists of the substructures





and the matrix

$$\rho = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ -\infty & -\infty & -\infty & -\infty & -\infty \end{pmatrix}.$$

[2] We assume that each building-block is admissible for each site. In the other case additional rules must be formulated.

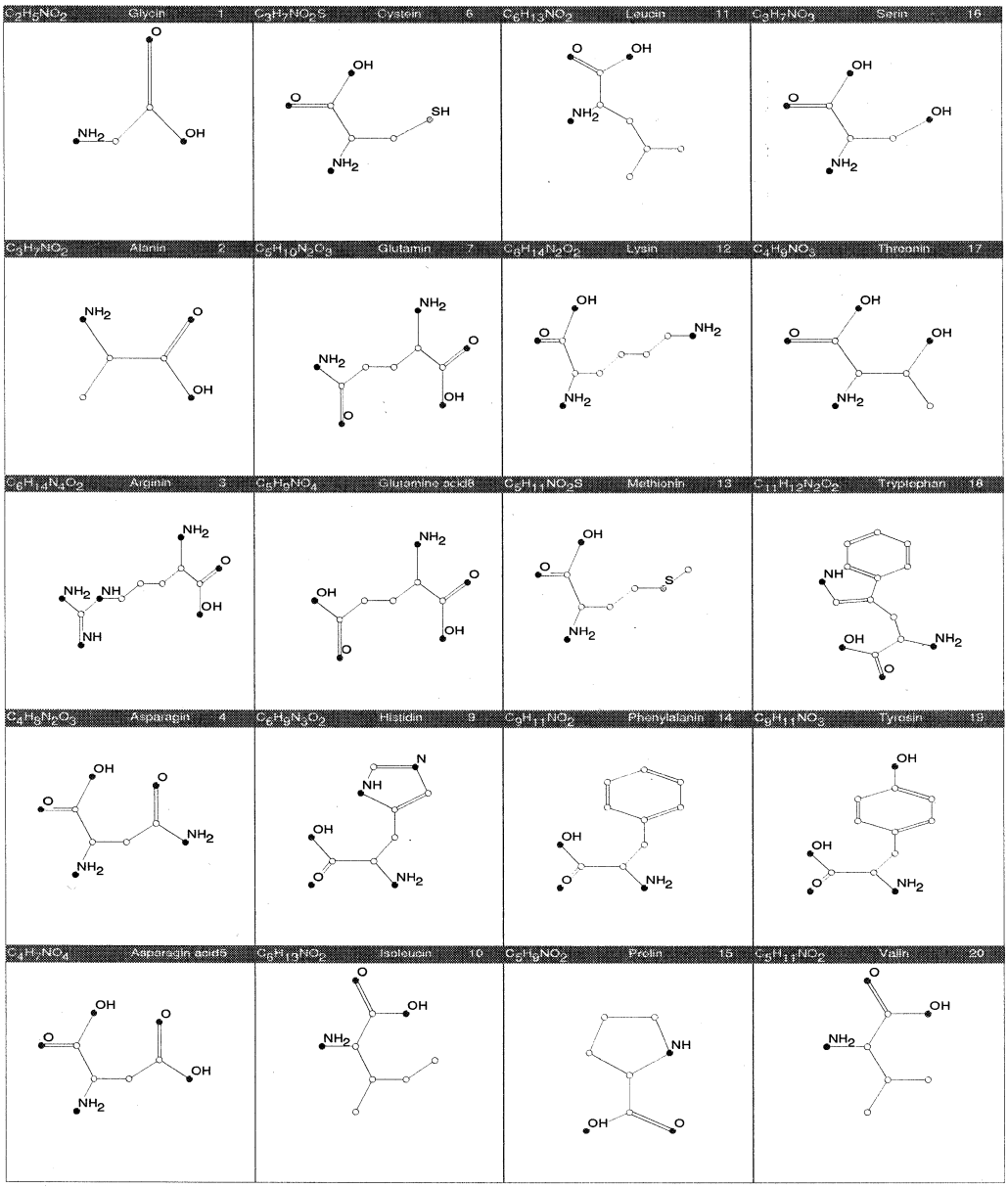


Figure 1. The 20 natural amino acids.

The cubane-derivative **I** is the structure with the highest symmetry, i.e., the largest automorphism group (derived from the symmetry group of the cube without inversions), which has 24 elements. Since each automorphism comprises a movement of the reacting substructures, we have $|P_\gamma| = 24$, too. As there are just four sites, it turns out that $P_\gamma = S_4$.

The xanthene **II** has an automorphism group with four elements. Two of them include the exchange of the methylene groups on the carbon bridge atom, such that just $|P_\gamma| = 2$. Besides the identity, this is the reflection of the rings on the vertical symmetry axis.

The benzene triacid chloride **III** has cyclic symmetry. Thus P_γ equals the cyclic group C_3, having three elements.

Even though the symmetry situation is a little complicated just for one of the three cores, the advantages of the mathematical concept behind algorithm 9 are obvious. The general *Ansatz* with an arbitrary permutation group and the efficient orderly generation (cf. [16,22,33]) allows a very rapid generation of the combinatorial libraries in all three cases. The computing speed is about 40 structures per second on a Pentium 90 MHz PC, writing all solutions to the hard disk.

Details about the sizes of the libraries are given in section 7 where we want to enumerate them.

Figure 2 shows six molecules from each of the three libraries.[3]

7. Enumeration of libaries

In this section we will present methods for the enumeration of the sizes of combinatorial libraries. A key tool for enumeration in algebraic combinatorics is the lemma of Cauchy–Frobenius:

Proposition 10 (Lemma of Cauchy–Frobenius). Let G be a finite permutation group acting on a finite set X.

- The number of orbits of this action is

$$|G\backslash\backslash X| = \frac{1}{|G|}\sum_{g\in G}|X_g|.$$

- G also acts on Y^X if Y denotes another finite set. For the number of orbits the following equation holds:

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

  where $c(g)$ is the number of cycles of the permutation g.

Proof. See, e.g., [17]. □

[3] The 2D placements were automatically calculated by the drawing module of MOLGEN [3,14,36]. These pictures reveal the current inacurracies of the employed placement algorithm [3,30] for combinatorial libraries.

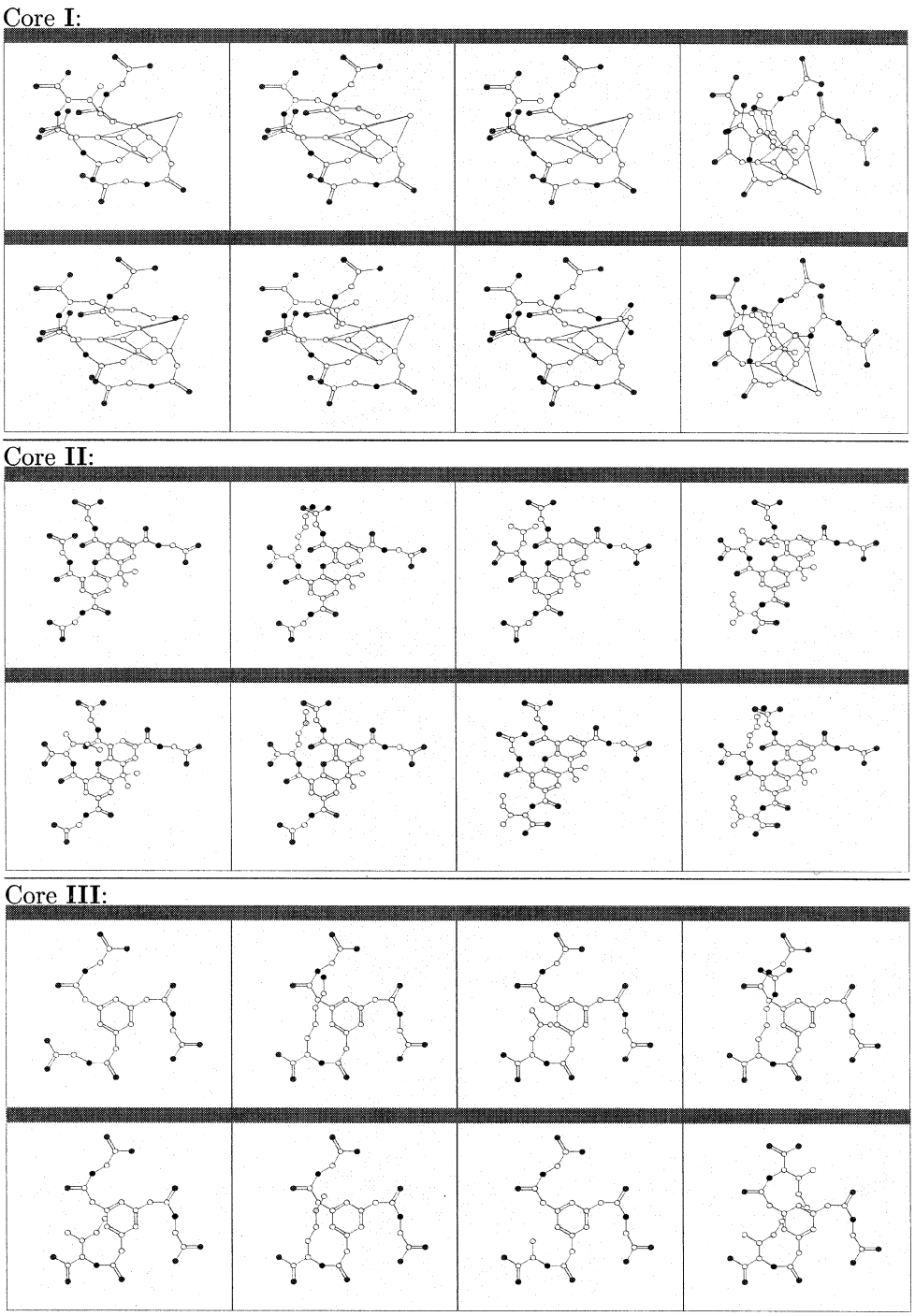


Figure 2. Extracts from the combinatorial libraries produced from the structures **I**, **II** and **III** and the natural amino acides from figure 1.

As we saw in lemma 8, the attachment of building-blocks to a core structure can be represented by a group action. So we can immediately state a result on the size of such libraries:

Theorem 11. Let (γ, β) be a core structure with k different reaction sites. Its automorphism group Aut(γ, β) may induce a permutation group $P_\gamma \leqslant S_k$ among the sites.

Furthermore a set of n different building-blocks may be given. Then the combinatorial library which can be built from the core structure and the building-blocks according to a corresponding reaction scheme has

$$\left|P_\gamma \backslash\backslash \boldsymbol{n}^{\boldsymbol{k}}\right| = \frac{1}{|P_\gamma|} \sum_{\pi \in P_\gamma} n^{c(\pi)}$$

elements.

Example 2. Again we consider the example from section 6 (taken from [6]) with the cores **I**, **II** and **III** and the amino acids (see figure 1) as building-blocks.

- The group P_γ for the cubane derivative (**I**) is the symmetric group S_4 with 24 elements and $k = 4$. Then we get by theorem 11

$$\left|P_\gamma \backslash\backslash \boldsymbol{n}^4\right| = \frac{1}{24}\left(n^4 + 6n^3 + 11n^2 + 6n\right).$$

- For xanthene (structure **II**), we have $k = 4$ and $P_\gamma = \{1, (12)(34)\}$. Here our formula yields

$$\left|P_\gamma \backslash\backslash \boldsymbol{n}^4\right| = \frac{1}{2}\left(n^4 + n^2\right).$$

- In case of the benzene triacid chloride (**III**) there is $k = 3$ and $P_\gamma = \{1, (123), (132)\}$. So the equation reads

$$\left|P_\gamma \backslash\backslash \boldsymbol{n}^3\right| = \frac{1}{3}\left(n^3 + 2n\right).$$

Table 1 provides an overview over the sizes of libraries depending on the number of building-blocks used.

Although there are equally many sites in **I** and **II**, the libraries with the first one are considerably smaller due to the higher symmetry of the core.

A combinatorial enumeration can also be obtained, if not all building-blocks shall be allowed for all possible multiplicities, i.e., if the frequencies shall be restricted. The tool for this task is called *weighted enumeration* (cf. [17], also for proofs).

Table 1
Size of libraries depending on numbers of building-blocks.

n	**I**	**II**	**III**
1	1	1	1
2	5	10	4
3	15	45	11
4	35	136	24
5	70	325	45
6	126	666	76
7	210	1225	119
8	330	2080	176
9	495	3321	249
10	715	5050	340
11	1001	7381	451
12	1365	10440	584
13	1820	14365	741
14	2380	19306	924
15	3060	25425	1135
16	3876	32896	1376
17	4845	41905	1649
18	5985	52650	1956
19	7315	65341	2299
20	8855	80200	2680

Proposition 12. 1. *Weighted form of the Cauchy–Frobenius lemma:* let ${}_GX$ be a finite group action and $W : X \to \mathbb{R}$ a weight function. If W is constant on the orbits of G on X, then for any transversal $\mathcal{T}$ of the orbits we have

$$\sum_{t\in\mathcal{T}} W(t) = \frac{1}{|G|}\sum_{g\in G}\sum_{x\in X} W(x).$$

2. Let $w : Y^X \to \mathbb{R},\ f \mapsto \prod_{x\in X} W(f(x))$ denote the *multiplicative weight* for the weight function $W : Y \to \mathbb{R}$. Then w is constant on the orbits of the permutation group G on Y^X and for any transversal $\mathcal{T}$ of the orbits we have

$$\sum_{t\in\mathcal{T}} w(t) = \frac{1}{|G|}\sum_{g\in G}\prod_{i=1}^{|X|}\Bigg(\sum_{y\in Y} W(y)^i\Bigg)^{a_i(g)},$$

where $a_i(g)$ denotes the number of cycles of length i in the permutation g.

3. Let $c(f, \_) : Y \to \mathbb{N},\ y \mapsto |f^{-1}(\{y\})|$ be the *content* of the mapping $f \in Y^X$, i.e., $c(f, y)$ denotes how often f takes the value y. Then the number of G-orbits on

Y^X, the elements of which have the same content as $f \in Y^X$, is equal to the coefficient of the monomial $\prod_y y^{c(f,y)}$ in the polynomial

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

Applied to combinatorial libraries we obtain:

Theorem 13. Let (γ,β) be a core structure with k different reaction sites. Its automorphism group Aut(γ,β) may induce a permutation group $P_\gamma \leqslant S_k$ among the sites.

Furthermore a set of n different building-blocks and a distribution $f \in \boldsymbol{n^k}$ of the blocks may be given.

Then the number of elements of the library, the distributions of which have the same content $c(f,\_)$ as f, is equal to the coefficient of the monom $\prod_r y_r^{c(f,y_r)}$ in the polynomial

$$\frac{1}{|P_\gamma|}\sum_{\pi\in P_\gamma}\prod_{i=1}^{k}\Big(\sum_{r=1}^{n} {y_r}^i\Big)^{a_i(\pi)}$$

over the unknowns $y_1,\dots,y_n$.

Example 3. As above, we consider the example from section 6 (taken from [6]) with the core structures **I**, **II** and **III** and the amino acids as building-blocks.

Then theorem 13 yields the following relations:

- The cubane derivative (**I**) has $P_\gamma = S_4$. We differentiate its elements according to their *cycle type* $a(\pi) = (a_1(\pi),\dots,a_4(\pi))$:

  1 element of type (4, 0, 0, 0), e.g., (1)(2)(3)(4);
  6 elements of type (2, 1, 0, 0), e.g., (12)(3)(4);
  3 elements of type (0, 2, 0, 0), e.g., (12)(34);
  8 elements of type (1, 0, 1, 0), e.g., (123)(4);
  6 elements of type (0, 0, 0, 1), e.g., (1234).

  So the polynomial is

$$\begin{aligned}\frac{1}{24}\big((y_1+\cdots+y_n)^4 &+ 6(y_1+\cdots+y_n)^2\big(y_1^2+\cdots+y_n^2\big)\\ &+ 3\big(y_1^2+\cdots+y_n^2\big)^2 + 8(y_1+\cdots+y_n)\big(y_1^3+\cdots+y_n^3\big)\\ &+ 6\big(y_1^4+\cdots+y_n^4\big)\big).\end{aligned}$$

  If we have four building-blocks, say, this term becomes

$$y_1^4+y_2^4+y_3^4+y_2y_3^2y_4+y_1^2y_2y_3+y_1y_2^2y_3+y_1y_2y_3^2+y_1^3y_2+y_1^3y_3$$

$$+ y_1^2y_2^2 + y_1^2y_3^2 + y_1y_2^3 + y_1y_3^3 + y_2^3y_3 + y_2^2y_3^2 + y_2y_3^3 + y_1y_2^2y_4 + y_1^2y_2y_4$$
$$+ y_1^2y_3y_4 + y_2^2y_3y_4 + y_1y_2y_4^2 + y_1y_3^2y_4 + y_1y_3y_4^2 + y_2y_3y_4^2 + y_4^4 + y_1^3y_4$$
$$+ y_1^2y_4^2 + y_1y_4^3 + y_2^3y_4 + y_2^2y_4^2 + y_2y_4^3 + y_3^3y_4 + y_3^2y_4^2 + y_3y_4^3 + y_1y_2y_3y_4.$$

- For xanthene (**II**) we obtain the polynomial

$$\frac{1}{2}\big((y_1 + \ldots + y_n)^4 + \big(y_1^2 + \ldots + y_n^2\big)^2\big).$$

In the case $n = 4$, this expands to the sum

$$6y_1y_2^2y_4 + 6y_1y_2y_4^2 + 6y_1y_2^2y_3 + y_1^4 + y_2^4 + y_3^4 + y_4^4 + 6y_1y_3^2y_4 + 6y_1^2y_2y_3$$
$$+ 6y_1^2y_2y_4 + 6y_1^2y_3y_4 + 6y_2^2y_3y_4 + 2y_1y_2^3 + 2y_1y_3^3 + 2y_1y_4^3 + 2y_1^3y_2$$
$$+ 2y_1^3y_3 + 2y_1^3y_4 + 4y_1^2y_2^2 + 4y_1^2y_3^2 + 4y_1^2y_4^2 + 2y_2y_3^3 + 2y_2y_4^3 + 2y_2^3y_3$$
$$+ 2y_2^3y_4 + 4y_2^2y_3^2 + 4y_2^2y_4^2 + 2y_3y_4^3 + 2y_3^3y_4 + 4y_3^2y_4^2 + 6y_1y_2y_3^2 + 6y_1y_3y_4^2$$
$$+ 12y_1y_2y_3y_4 + 6y_2y_3^2y_4 + 6y_2y_3y_4^2.$$

The condition that, for instance, only those library elements are of interest which contain the first building-block exactly once corresponds to the summands

$$6y_1y_2^2y_4 + 6y_1y_2y_4^2 + 6y_1y_2^2y_3 + 6y_1y_3^2y_4 + 2y_1y_2^3$$
$$+ 2y_1y_3^3 + 2y_1y_4^3 + 6y_1y_2y_3^2 + 6y_1y_3y_4^2 + 12y_1y_2y_3y_4.$$

So there are $6 + 6 + 6 + 6 + 2 + 2 + 2 + 6 + 6 + 12 = 54$ elements of that kind.

- For the benzene triacid chloride (**III**) the polynomial is

$$\frac{1}{3}\big((y_1 + \ldots + y_n)^3 + 2\big(y_1^3 + \ldots + y_n^3\big)\big).$$

Considering three building-blocks, say, this means

$$y_1^3 + y_1^2y_2 + y_1^2y_3 + y_1y_2^2 + 2y_1y_2y_3 + y_1y_3^2 + y_2^3 + y_2^2y_3 + y_2y_3^2 + y_3^3.$$

8. Conclusion

The methods and theorems presented above provide a number of tools for the analysis of combinatorial libraries. They should enable the researcher to understand and overview his libraries better and eventually to design his experiments more precisely. In combination with procedures for building-block selection [23,31] combinatorics will thus allow a deeper insight into the chemical and biological reactions. The following and last step in combinatorial chemistry, the screening of the library, is, however, a much more difficult task, but mathematical models (e.g., QSAR correlations [1,2,19, 20,34] or CoFMA [20,21]) are already in sight.

Acknowledgements

The author is grateful to the German Federal Ministry of Education, Science, Research and Technology (BMBF) for financial support and to the large number of researchers in the institute and on the Internet for useful hints and discussions.

References

[1] S.C. Basak, Use of molecular complexity indices in predictive pharmacology and toxicology: a QSAR approach, Med. Sci. Res. 15 (1987) 605–609.
[2] S.C. Basak, S. Bertelsen and G.D. Grunwald, Application of graph-theoretical parameters in quantifying molecular similarity and structure-activity studies, J. Chem. Inf. Comput. Sci. 34 (1994) 270–276.
[3] C. Benecke, R. Grund, R. Hohberger, A. Kerber, R. Laue and T. Wieland, MOLGEN+, a generator of connectivity isomers and stereoisomers for molecular structure elucidation, Anal. Chim. Act. 314 (1995) 141–147.
[4] C. Benecke, R. Grund, R. Hohberger, A. Kerber, R. Laue and T. Wieland, Chemical isomerism, a challenge for algebraic combinatorics and for computer science, in: *Applied Algebra, Algebraic Algorithms and Error-Correcting Codes*, eds. C. Cohen, M. Giusti and T. Mora, Lect. Not. Comput. Sci. (Springer, New York, 1995).
[5] N.L. Biggs, K.E. Lloyd and R.J. Wilson, *Graph Theory 1736–1936* (Clarendon Press, 1977).
[6] T. Carell, E.A. Wintner, A.J. Sutherland, J. Rebek, Jr., Y.M. Dunayevskiy and P. Vouros, New promise in combinatorial chemistry: synthesis, characterization, and screening of small-molecule libraries in solution, Chem. & Biol. 2 (1995) 171–183.
[7] R.E. Carhart, D.H. Smith, H. Brown and C. Djerassi, An approach to computer-assisted elucidation of molecular structure, J. Am. Chem. Soc. 97 (1975) 5755–5762.
[8] R.E. Carhart, D.H. Smith, N.A. Gray, J.G. Nourse and C. Djerassi, GENOA: a computer program for structure elucidation utilizing overlapping and alternative substructures, J. Org. Chem. 46 (1981) 1708–1718.
[9] A. Cayley, On the mathematical theory of isomers, Philosoph. Magaz. 74(4) (1874) 444–446.
[10] J. Dugundji and I.K. Ugi, An algebraic model of contitutional chemistry as a basis for chemical computer programs, Top. Curr. Chem. 39 (1973) 19–64.
[11] K. Funatsu, N. Miyabayaski and S. Sasaki, Further development of structure generation in the automated structure elucidation system CHEMICS, J. Chem. Inf. Comput. Sci. 28 (1988) 18–28.
[12] M.A. Gallop, R.W. Barrett, W.J. Dower, S.P.A. Fodor and E.M. Gordon, Applications of combinatorial technologies to drug discovery. 1. Background and peptide combinatorial libraries, J. Med. Chem. 37 (1994) 1233–1251.
[13] E.M. Gordon, R.W. Barrett, W.J. Dower, S.P.A. Fodor and M.A. Gallop, Applications of combinatorial technologies to drug discovery. 2. Combinatorial organic synthesis, library screening strategies, and future directions, J. Med. Chem. 37 (1994) 1385–1401.
[14] R. Grund, A. Kerber and R. Laue, MOLGEN, ein Computeralgebra-System für die Konstruktion molekularer Graphen, MATCH 27 (1992) 87–131.
[15] R. Grund, A. Kerber and R. Laue, Construction of discrete structures, especially isomers, Discr. Appl. Math. 67 (1996) 115–126.
[16] T. Grüner, R. Laue and M. Meringer, Applications for group actions applied to graph generation, in: *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, eds. L. Finkelstein, and C. Kantor (Providence, RI, 1995) (in press).
[17] A. Kerber, *Algebraic Combinatorics Via Finite Group Actions* (BI-Wissenschaftsverlag, Mannheim, Wien, Zürich, 1991).

[18] A. Kerber, ed., Generatoren für molekulare Graphen, Nomenklaturfragen, MATCH 27 (1992).
[19] L.B. Kier and L.H. Hall, *Molecular Connectivity in Structure-Activity Analysis* (Research Studies Press, Chichester, 1986).
[20] H. Kubinyi, ed., *3D QSAR in Drug Design. Theory, Methods and Applications* (ESCOM Science Publishers, Leiden, 1993).
[21] H. Kubinyi, Der Schlüssel zum Schloß. II. Hansch-Analyse, 3D-QSAR und De novo-Design, Pharmazie in unserer Zeit 23(5) (1994) 281–290.
[22] R. Laue, Construction of combinatorial objects – a tutorial, Bayreuther Mathem. Schr. 43 (1993) 53–96.
[23] E.J. Martin, J.M. Blaney, M.A. Siani, D.C. Spellmeyer, A.K. Wong and W.H. Moos, Measuring diversity: experimental design of combinatorial libraries for drug discovery, J. Med. Chem. 38 (1995) 1431–1436.
[24] L.M. Masinter, N.S. Sridharan, J. Lederberg and D.H. Smith, Applications of artificial intelligence for chemical inference. XII. Exhaustive generation of cyclic and acyclic isomers, J. Am. Chem. Soc. 96 (1974) 7702–7714.
[25] M.R. Pavia, The chemical generation of molecular diversity, Network Science (August 1995).
[26] G. Pólya, Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen, Acta mathematica 68 (1937) 145–253.
[27] G. Pólya and R.C. Read, *Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds* (Springer, Berlin, 1987).
[28] R.C. Read, Every-one a winner, Ann. Discr. Math. 2 (1978) 107–120.
[29] H.J. Redfield, The theory of group-reduced distributions, Am. J. Math. 49 (1927) 433–455.
[30] C.A. Shelley, Heuristic approach for displaying chemical structures, J. Chem. Inf. Comput. Sci. 23 (1983) 61–65.
[31] R.P. Sheridan and S.K. Kearsley, Using a genetic algorithm to suggest combinatorial libraries, J. Chem. Inf. Comput. Sci. 35 (1995) 310–320.
[32] I. Ugi, Fast and permanent changes in preparative and pharmaceutical chemistry through multicomponent reactions and their 'libraries', Proc. Eston. Acad. Sci. Chem. 44 (1995) 237–273.
[33] T. Wieland, Konstruktionsalgorithmen bei molekularen Graphen und deren Anwendung, Ph.D. thesis, University of Bayreuth (1996).
[34] T. Wieland, The use of structure generators in predictive pharmacology and toxicology, Arzneim.-Forsch./Drug Res. 46(I) (1996) 223–227.
[35] T. Wieland, Mathematical simulations in combinatorial chemistry, MATCH 34 (1996) 179–206.
[36] T. Wieland, A. Kerber and R. Laue, Principles of the generation of constitutional and configurational isomers, J. Chem. Inf. Comput. Sci. 36 (1996) 413–419.