

# Computers in Chemistry: Molecular Structure Generation and Applications\*

Adalbert Kerber

Department of Mathematics  
Universität Bayreuth  
Germany

\* joint book (in preparation) with R. Gugisch, R. Laue,  
M. Meringer, C. Rücker, E. Schymanski

# A particular case of “classification”

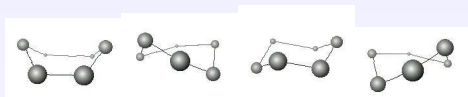
- In general, classification means counting and representing **classes** of basic structures of a theory, say vector spaces, graphs, groups, etc.
- In **Mathematical Chemistry**: structural formulas of molecules, libraries of molecules (for structure elucidation), combinatorial libraries (for predictions), patent libraries (patent violations), molecules with particular biological or medical properties (where geometry may come in), . . .

Three levels of precision in describing molecules:

- the *arithmetic* level: Molecular formula,
- the *topologic* level: Connectivity isomers,
- the *geometric* level: Stereo isomers.

# An example: Cyclohexane

- The molecular formula:  $C_6H_{12}$ ,
- has 25 connectivity isomers, e.g. a 6-membered ring,
- the ring, cyclohexane, possesses 4 stereo isomers:

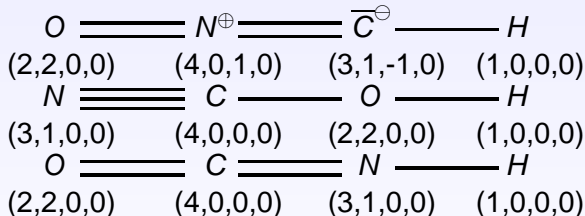


In this talk: The topologic level, connectivity isomers, and the geometric level, stereo isomers.

**I.e. interaction models on the topologic and even on the geometric level!**

# The usual topological model of molecule

- Isomerism *stated* by A. v. Humboldt in 1796, *verified* by J. v. Liebig and F. Wöhler in 1825
- First isomers: fulminic acid, cyanic acid and isocyanic acid



- these are multigraphs, nodes *colored* by element symbols and atom states  $(v_z, p_z, q_z, r_z)$ , *but unlabeled!*

# The basic (mathematical) problem

- Unlabeled graphs are labeled graphs, “up to relabeling”.
- We use *Constructive Classification*, describing the classes as orbits of groups according to the following definition
- $X$  a set (e.g. labeled graphs on  $n$  nodes),  $G$  a group (e.g. the symmetric group on the nodes of the graphs). An **action of  $G$  on  $X$**  is a mapping

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

such that  $g(g'x) = (gg')x$  and  $1x = x$ . It induces pairwise **disjoint orbits**, i.e. it yields a *classification* into sets

$$G(x) = \{gx \mid g \in G\} \subseteq X.$$

And vice versa: Each classification can be described this way!

# The main results on orbits

- **Lemma of Cauchy–Frobenius:** The total number of orbits is the average number of fixed points:

$$|G \backslash X| = \frac{1}{|G|} \sum_{g \in G} |X_g|, \text{ where } X_g = \{x \mid gx = x\}.$$

This gives the number of the classes!

- **The Fundamental Lemma:** If  $G_x = \{g \mid gx = x\}$ , the *stabilizer* of  $x$ , then the action of  $G$  on the orbit  $G(x)$  is essentially the same as the action of  $G$  on the set of left cosets  $G/G_x$ . Hence, the set of orbits of a subgroup  $U$  of  $G$  on  $G(x)$  is bijective to the set of orbits of  $U$  on  $G/G_x$ , which is  $U \backslash G/G_x = \{UgG_x \mid g \in G\}$ , a set of **double cosets**.

$$U \backslash G(x) \rightarrow U \backslash G/G_x: U(gx) \mapsto UgG_x.$$

# Constructive classification methods:

- **The algebraic way:** In order to construct a transversal of the orbits of  $U$  on  $X$ , choose a suitably bigger group  $G$  that extends the action of  $U$ , evaluate a transversal of the set of double cosets  $U \backslash G / G_x$  and retranslate into elements of  $X$ .
- **A combinatorial approach:** If  $X$  is totally ordered, then evaluate a canonic transversal, e.g. the smallest elements in their orbit, using R. Read's method, described in his paper *Every one a winner*, Ann. Discrete Math. 2, 107-120 (1978) (cf. also Faradzhev, same year)

Standard application: Construction of unlabeled graphs and of molecular graphs!



# Construction of a sample

In the case of an 'astronomic' number of classes: Generate a *sample*, a set of labeled structures, uniformly distributed over the classes, using

## The Dixon/Wilf Algorithm:

If  ${}_G X$  denotes a finite action, then we can generate orbit representatives uniformly at random in the following way:

- Choose a conjugacy class  $C^G$  of the finite group  $G$  (i.e. an orbit of the action  $g' \mapsto gg'g^{-1}$ ) with probability

$$p(C^G) = \frac{|C^G||X_g|}{|G||G \backslash X|}.$$

- pick an element  $g \in C^G$  and generate a fixed point  $x$  of  $g$ , uniformly at random.

Then the probability that  $x$  is an element of a particular orbit is

$$\frac{1}{|G \backslash X|}.$$

r

# Important generalization: Pólya's paper of 1937

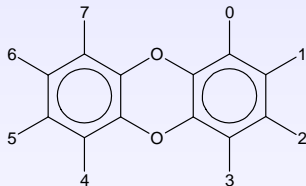
Use *two* actions: of  $G$  on  $X$  and of  $H$  on  $Y$  and consider the induced actions of  $G$ ,  $H$ ,  $H \times G$ ,  $H \wr G$  on the set of mappings

$$Y^X = \{f \mid f: X \rightarrow Y\}$$

- His idea:  $X$  the set of active sites of the skeleton of a molecule,  $G$  its point group,  $Y$  a set of substituents.  $H$  a “transposition” in case of chirality (see below).
- Enumerate the essentially different distributions of the substituents over the active sites, i.e. the **permutational isomers**.

# Example 1: Seveso dioxine

The skeleton is



$X = \{0, 1, 2, 3, 4, 5, 6, 7\}$ , the set of 'active sites'

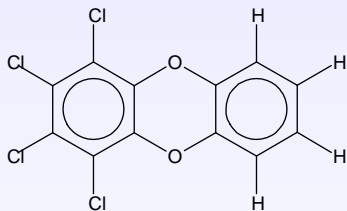
$Y = \{H, Cl\}$ , the set of 'types of substituents'

$$|Y^X| = |\{H, Cl\}| = 2^8 = 256$$

$G$  is generated by the horizontal and vertical reflection.

# apply the Fundamental Theorem

in order to evaluate the permutational isomers of Seveso dioxine



Its stabilizer is  $S_{(4,4)} = S_{\{0,1,2,3\}} S_{\{4,5,6,7\}}$ .

The use of double cosets: The bigger group is  $S_8$ ,

the set of cosets of the stabilizer:  $|S_8/S_{(4,4)}| = 70$

The double cosets can be represented by *tabloids*

$$\frac{\overline{i_0 \dots i_3}}{i_4 \dots i_7}, \{i_0, \dots, i_7\} = \{0, \dots, 7\} \text{ and } i_0 < \dots < i_3, i_4 < \dots < i_7$$

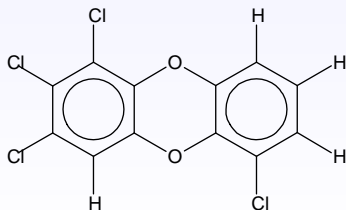
It suffices to consider the upper rows:

$$\overline{0123}, \overline{0124}, \overline{0125}, \overline{0126}, \overline{0127}, \overline{0134}, \overline{0135}, \dots, \overline{4567}.$$

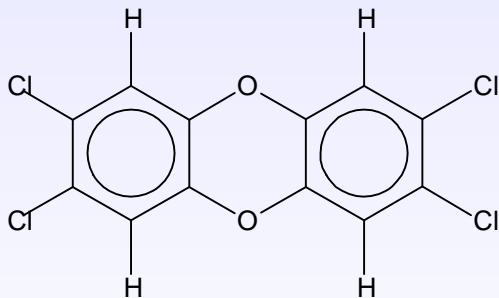
The symmetry group divides this set into 22 orbits

e.g.  $\overline{0123} \in \Omega_0$ ,  $\overline{0124} \in \Omega_1$ ,

$\Omega_1$  corresponds to



# The Seveso dioxine is



It is obtained from the orbit of the tabloid 0347.

# The generation of structural formulas

*A labeled molecular graph:*

$$\begin{aligned}(\varepsilon, \zeta, \gamma) &= (\text{sequence of elements, of atom states, a multigraph}) \\ &\in \mathcal{E}^n \times \mathcal{Z}_{\mathcal{E}}^n \times \mathcal{G}_{4,n}\end{aligned}$$

unlabeled molecular graphs correspond to orbits of the action

$$S_n \times (\mathcal{E}^n \times \mathcal{Z}_{\mathcal{E}}^n \times \mathcal{G}_{4,n}) \rightarrow (\mathcal{E}^n \times \mathcal{Z}_{\mathcal{E}}^n \times \mathcal{G}_{4,n})$$

$$\pi\varepsilon(i) = \varepsilon(\pi^{-1}i), \quad \pi\zeta(i) = \zeta(\pi^{-1}i), \quad \pi\gamma(\{i, j\}) = \gamma(\{\pi^{-1}i, \pi^{-1}j\}),$$

i.e.  $\pi \in S_n$  keeps elements, atom states, bond multiplicities

General 'philosophy':

**Use groups, avoid combinatorial explosion!**



# Generating molecular libraries

For various generators see MATCH 27

I refer in particular to MOLGEN 5.0

For details, requirements, manual and publications cf.

`http://www.molgen.de`

Play online with a reduced version of MOLGEN 5.0

# Its main advantage

MOLGEN 5.0 allows the use of *fuzzy molecular formula*,

for example, enter

C1-8H0-16N0-6O0-4

in

<http://www.molgen.de>

together with the condition 100 for the molecular mass

and obtain in ca. 10 seconds the *molecular library* containing 33,537 structural formulas that were *constructed*.

They correspond to 16 molecular formulas covered by the fuzzy formula.

# A table of further sizes of molecular libraries

mass	MF	MG	MGNAD	BS	MS
30	2	2	2	2	2
40	3	5	5	5	1
50	1	7	7	1	1
60	6	47	47	25	12
70	6	380	380	84	31
80	6	1,645	1,644	100	23
90	11	5,849	5,818	107	28
100	16	33,627	33,537	710	154

MF: number of molecular formulas,

MG: corresponding molecular graphs,

MGNAD: no. of structural formulas without aromatic duplicates.

BS: numbers of isomers contained in the Beilstein database

MS: refers to the NIST mass spectral library.

Molecular structure generation is crucial whenever the unknown chemical compound may not be contained in the available databases. For a detailed description of method using MOLGEN cf.

E. Schymanski: *Integrated analytical and computer tools for toxicant identification in effect-directed analysis*. PhD thesis 07/2011, Helmholtz Centre for Environmental Research – UFZ.

Contaminants in groundwater of Bitterfeld, Germany:  
150 spectra, 42 identified using NIST alone.  
32 of these confirmed using structure generation techniques. In addition, 20 further peaks were tentatively identified using structure generation techniques alone, resulting in a total of 62 tentative identifications.

There is also a reaction based generation

T. Wieland: *Konstruktionsalgorithmen bei molekularen Graphen und deren Anwendung*. MATCH Commun. Math. Comput. Chem., 36:7–157, 1997.

T. Wieland, A. Kerber, and R. Laue: *Principles of the generation of constitutional and configurational isomers*. J. Chem. Inf. Comput. Sci., 36:413–419, 1996.

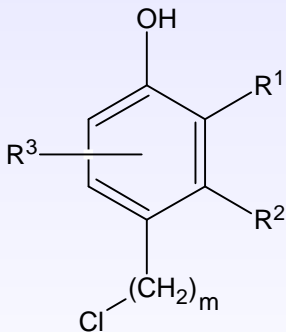
Use MOLGEN-COMB and MOLGEN-QSPR, in order to optimize an experiment of combinatorial chemistry *in advance*.

The basic problems:

- Generate the molecular library covered by the Markush formula
- Generate it in a *canonic way*, MOLGEN does this!

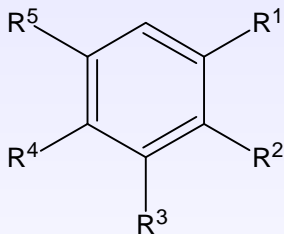
It is needed in order to **find patent violations!**

# An example



where R<sup>1</sup>: CH<sub>3</sub> or C<sub>2</sub>H<sub>5</sub>, R<sup>2</sup>: Alkyl (1–6 C atoms), R<sup>3</sup>: NH<sub>2</sub>,  
*m* : 1–3. This formula covers a library  $\mathcal{L}_1$  of altogether 396 structural formulas.

# compare with the patent



where R<sup>1</sup>: CH<sub>3</sub>, C<sub>2</sub>H<sub>5</sub>, OH, R<sup>2</sup>: Alkyl (1–6 C atoms),  
R<sup>3</sup>: OH, OCH<sub>3</sub>, OC<sub>2</sub>H<sub>5</sub>, CH<sub>3</sub>, C<sub>2</sub>H<sub>5</sub>, R<sup>4</sup>: OH, CH<sub>2</sub>Cl, NH<sub>2</sub>,  
R<sup>5</sup>: H, CH<sub>3</sub>, C<sub>2</sub>H<sub>5</sub>, NH<sub>2</sub>.

The patent library consists of  $|\mathcal{L}_2| = 5,939$  molecules



# there are patent violations:

Because of *canonic generation* we find in a fraction of a second

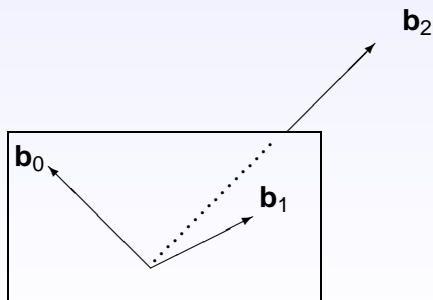
$$|\mathcal{L}_1 \cap \mathcal{L}_2| = 4$$

See

A. Kerber, R. Laue, and M. Meringer: *An application of the structure generator MOLGEN to patents in chemistry*. MATCH Commun. Math. Comput. Chem., 47:169–172, 2003.

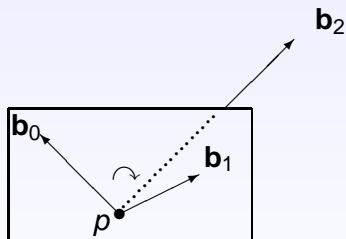
# Entering the geometric level: Orientation

- **Orientation** of the 3D-space  $\mathbb{R}^3$  means fixing a sequence  $(\mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2)$  of linearly independent vectors, i.e. the choice of an ordered basis of the space considered.
- Another ordered basis  $(\mathbf{c}_0, \mathbf{c}_1, \mathbf{c}_2)$  is said to have the *same orientation* if its determinant has the same sign:  
 $sign(\det(\mathbf{b}_0, \mathbf{b}_1, \mathbf{b}_2)) = sign(\det(\mathbf{c}_0, \mathbf{c}_1, \mathbf{c}_2))$ .



# Right-handed, left-handed

An orientation allows to introduce a right-handed (left-handed) screw or orientation by using *in addition* the standard convention of “clockwise rotation”:



# Point group, proper and improper rotations

- Consider a molecule as a rigid arrangement of points, barycentrically placed in 3D space. Symmetry operations are contained in the *orthogonal group*

$$\mathbf{O}_3 = \{\Phi \mid \Phi \text{ is a linear isometry}\}.$$

In particular,  $\det(\Phi) = \pm 1$ .

- We distinguish the following subgroup

$$\mathbf{SO}_3 = \{\Phi \mid \Phi \in \mathbf{O}_3, \det(\Phi) = +1\}.$$

The elements of this *special orthogonal group* are called *proper rotations*, the other elements of the orthogonal group are the *improper rotations*.

- The **point group**, the symmetry group of the molecule, is a subgroup of the orthogonal group, and chirality means in particular to distinguish the cases

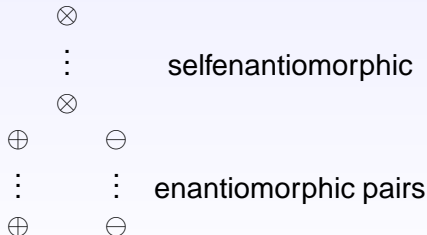
$$\mathbf{P} \subseteq \mathbf{SO}_3 \text{ and } \mathbf{P} \not\subseteq \mathbf{SO}_3.$$

- The molecule is equal to its mirror image if and only if  $\mathbf{P} \not\subseteq \mathbf{SO}_3$ . Such molecules are called **achiral**, the others **chiral**.
- The **rotation group**  $\mathbf{R}$  of the molecule is the intersection

$$\mathbf{R} = \mathbf{P} \cap \mathbf{SO}_3.$$

# A sketch of the situation

- Chirality is an equivalence relation on the set of all molecules. The molecules that form an equivalence class consisting of a single molecule are selfenantiomorphic (identical to their respective mirror images) and achiral. The others form enantiomorphic pairs and are chiral.

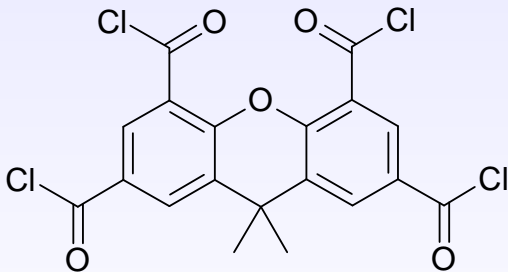


# Serious open problem

- How can we evaluate the point group “in silico” since we have to start from a barycentric placement of the molecule in space? A placement can be obtained using, say, MM2, but this ends in a local energy minimum that will not be “ideal”, e.g. cubane will (in silico) not be a cube!

## Example 2: An amidation

Consider a xanthenetetracarboxylic acid chloride assumed to be planar:



Each of the four acid chloride functional groups ( $\text{C}(=\text{O})\text{Cl}$ ) reacts with amino ( $\text{NH}_2$ ) groups of amino acids, producing an amide.

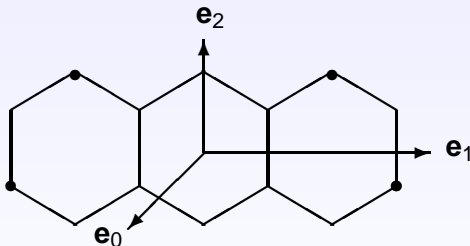


# Counting the amidations

$X$  the set of labels of the substitutable positions,

$Y$  a set of amino acids.

Use the following simplified version of the skeleton:

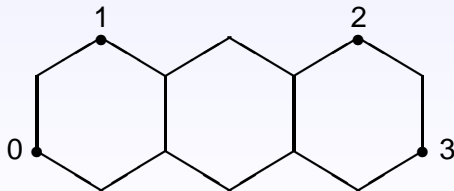


# The point group

is the set of linear isometries

$$\rho_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \rho_1 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \rho_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$
$$\rho_3 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

while the subgroup of proper rotations is  $\mathbf{R} = \{\rho_0, \rho_3\}$ , After numbering substitutable positions, say



we obtain the permutations induced by isometries  $\rho_i$



$$\bar{\rho}_0 = (0)(1)(2)(3), \bar{\rho}_1 = (0)(1)(2)(3), \bar{\rho}_2 = (03)(12), \bar{\rho}_3 = (03)(12).$$

We have

$$|\Delta(\mathbf{P} \times \mathbf{P}) \setminus Y^X| = \frac{1}{|\mathbf{P}|} \sum_{\rho \in \mathbf{R}} |Y|^{c(\bar{\rho})} + \frac{1}{|\mathbf{P}|} \sum_{\rho \in \mathbf{P} \setminus \mathbf{R}} |Y_{ac}|^{c_o(\bar{\rho})} \cdot |Y|^{c_e(\bar{\rho})}.$$

$$= \begin{cases} 25 & \text{if } |Y| = 3 \\ 166 & \text{if } |Y| = 5 \\ \vdots & \vdots \\ 579,121 & \text{if } |Y| = 39. \end{cases}$$

assuming that  $Y$  consists of glycine and further amino acids, together with their enantiomorphs.

For the construction: Cf. W. Hässelbarth

# Stereo isomers

- **Constitution** describes the nature of atoms and their connectivity, disregarding their spatial array.
- **Stereoisomers** are isomers of identical constitution that differ in the arrangement of their atoms in space.
- **Configuration** describes aspects of steric arrangement that are long-lived on a laboratory time scale.
- **Conformation** describes details of atom arrangement that are short-lived on a laboratory time scale.
- **Conformer** is a conformation that is a (global or local) potential energy minimum.

# An approach using matroids

- Dreiding and Dress used chirotopes (strongly related to oriented matroids) as a tool for describing conformations. Cf. Crippen and Havel in chapter 2 of their book. Similar considerations: Klin, Tratch, Zefirov and Kuz'min.
- Implemented by R. Gugisch, cf.

*R. Gugisch. Konstruktion von Isomorphieklassen orientierter Matroide. Bayreuther Mathematische Schriften, 72:1–129, 2005.*

*R. Gugisch and C. Rücker. Unified generation of conformations, conformers, and stereoisomers: A discrete mathematics-based approach. MATCH Commun. Math. Comput. Chem., 61:117–148, 2009.*

# The main notions

- *Treat a given conformation of a molecule of  $n$  atoms in space as a collection of tetrahedra or tripods of atoms.* I.e. consider sets of four atoms, and each of the  $\binom{n}{4}$  sets of order 4 may be considered. The main problem will be to reduce the number of tetrahedra to be taken into account.
- Each tetrahedron's volume is the determinant of a matrix obtained from the atom coordinates, divided by 6: From a set of 4 atoms we build a sequence  $(\mathbf{a}_0, \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3)$  of coordinate vectors

$$\mathbf{a}_i = \begin{pmatrix} a_{i0} \\ a_{i1} \\ a_{i2} \end{pmatrix},$$

they form the volume determinant

$$V(\mathbf{a}_0, \mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3) = \det(\mathbf{a}_1 - \mathbf{a}_0, \mathbf{a}_2 - \mathbf{a}_0, \mathbf{a}_3 - \mathbf{a}_0)$$

- The sign of the volume depends on the chosen orientation of the 3D space considered, i.e. of the chosen basis. We define the orientation function  $\chi$  as follows:

$$\begin{aligned}\chi(a_0, a_1, a_2, a_3) &= \text{sign}(V(a_0, a_1, a_2, a_3)) \\ &= \text{sign}(\det(a_1 - a_0, a_2 - a_0, a_3 - a_0)).\end{aligned}$$

**is called the tetrahedron's or the atom quadruple's orientation.**

- Thus, **a given conformation of a multiatom molecule can roughly be described by its orientation function  $\chi$ , a sequence of plus and minus signs and zeros that are the orientations of all quadruples!**

# From quadruples of atoms to 4-subsets of atoms

Once we know  $\chi(a_0, a_1, a_2, a_3)$ , we also know that

$$\begin{aligned}\chi(a_0, a_1, a_2, a_3) &= \chi(a_0, a_2, a_3, a_1) = \chi(a_0, a_3, a_1, a_2) = \\ \chi(a_1, a_0, a_3, a_2) &= \chi(a_1, a_2, a_0, a_3) = \chi(a_1, a_3, a_2, a_0) = \\ \chi(a_2, a_1, a_3, a_0) &= \chi(a_2, a_0, a_1, a_3) = \chi(a_2, a_3, a_0, a_1) = \\ \chi(a_3, a_2, a_1, a_0) &= \chi(a_3, a_0, a_2, a_1) = \chi(a_3, a_1, a_0, a_2).\end{aligned}$$

The twelve remaining permutations of  $(a_0, a_1, a_2, a_3)$  all have the opposite orientation  $-\chi(a_0, a_1, a_2, a_3)$ . The orientation function is constant on the orbits of the *alternating group*

$$A_4 = \langle (012), (013), (023), (123) \rangle.$$

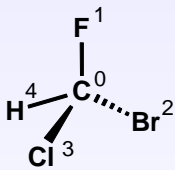
Thus, instead of  $n^4$  we only need to specify

$$\binom{n}{4} = \frac{n(n-1)(n-2)(n-3)}{24}$$

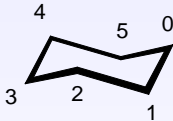
orientations of four *distinct* atoms. For example, for  $n = 10$ , we have  $10^4 = 10,000$  while  $\binom{10}{4} = 210$ .



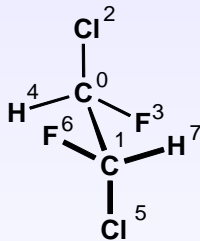
# Some conformations:



a



b



c

# Their orientation functions

$$\begin{array}{ccccc} 0123 & 0124 & 0134 & 0234 & 1234 \\ \chi_{\mathbf{a}} = & - & + & - & + & + \end{array}$$

$$\begin{array}{cccccccccccccccc} 0123 & 0124 & 0125 & 0134 & 0135 & 0145 & 0234 & 0235 & 0245 & 0345 & 1234 & 1235 & 1245 & 1345 & 2345 \\ \chi_{\mathbf{b}} = & + & + & + & \mathbf{0} & + & + & - & \mathbf{0} & + & + & - & - & \mathbf{0} & + & + \end{array}$$

$$\begin{array}{cccccccccccccccc} 0123 & 0124 & 0125 & 0126 & 0127 & 0134 & 0135 & 0136 & 0137 & 0145 & 0146 & 0147 & 0156 & 0157 & 0167 \\ \chi_{\mathbf{c}} = & - & + & \mathbf{0} & + & - & - & - & \mathbf{0} & + & + & - & \mathbf{0} & - & + & - \end{array}$$

... (altogether 70 quadruples) ... ?

- Orientation functions obtained from each other by a graph automorphic renumbering are called *isomorphic*. A conformation is *achiral* if and only if its orientation function is isomorphic to its negative.
- A graph automorphism that does not change the orientation function is called an *orientational* automorphism. The set of all orientational automorphisms is called the orientational *automorphism group*. It is a subgroup of the graph automorphism group.
- geometric  $\subseteq$  orientational  $\subseteq$  graph automorphism group.

# Canonic representatives

- When generating conformations, we need to introduce *canonic orientation functions*.
- The particular choice is not essential. Essential is the ability to test for canonicity *efficiently*.
- Minimality can be tested by calculating and comparing the orientation functions of all graph automorphic renumberings. For small graph automorphism groups this is practicable. For a large graph automorphism group, a more sophisticated algorithm is available.

# Partial orientation functions

- We generally restrict our considerations by selecting a subset  $S$  of the set of  $n^4$  atom quadruples,  
 $\chi: S \rightarrow \{-, +, 0\}$ , a **partial orientation function (POF)**.
- We request  $S$  to be closed under permutations of the quadruples and under application of graph automorphisms of the molecule.
- We can introduce canonic forms for POFs with respect to  $S$ , and generate canonic representatives only, avoiding the generation of isomorphic conformations.

# Tests for realizability

A conformer is a realization of a POF  $\chi$  if the abstract POF describes the orientations of that conformer. We distinguish three kinds of abstract POFs:

- $\chi$  is *not realizable*: There is no embedding of the atoms in space such that  $\chi$  is the POF of this embedding.
- $\chi$  is *geometrically realizable*: There exists an embedding of the atoms into space having  $\chi$  as POF. However, bond lengths, bond angles and torsion angles are not considered.
- $\chi$  is *chemically realizable*: There is (at least) one conformer having  $\chi$  as POF.

It is crucial to provide a set of combinatorial tests to exclude as many as possible non-realizable and not chemically realizable abstract POFs.

# We distinguish two kinds of tests:

- Tests excluding non-realizable abstract POFs only may be used without any concern. Their validity can be proven in the mathematical sense.
- Tests excluding geometrically realizable but not chemically realizable abstract POFs have to be justified chemically.

The computer program **origen**, a generator of alternating functions, was developed by R. Gugisch that accepts as input among other parameters a set of tuples as relevant function domain

# Example: cyclohexane

$C_6H_{12}$ . We generate *chirotopes* of rank  $k = 4$  over  $n = 6$  points, i.e. alternating, not constantly zero functions  $\chi: n^4 \rightarrow \{+, -, 0\}$  fulfilling the binary *Grassmann–Plücker relations*

$$\chi(a_0, a_1, a_2, a_3) = \chi(b_0, b_1, b_2, b_3) \neq 0 \implies$$

$$\exists i: \chi(b_i, a_1, a_2, a_3) = \chi(b_0, \dots, \underbrace{a_0}_{\text{position } i}, \dots, b_3) \neq 0.$$

Thus the orientation function of any sequence of atoms (not all in one plane) is a *chirotope*. The realizable chirotopes correspond to the 3D point configurations of six atoms.



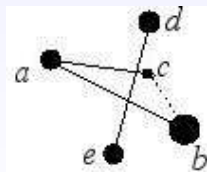
# The procedure:

- We generate chirotopes up to relabeling under the (dihedral) automorphism group  $D_6$  of order 12.
- We restrict attention to chirotopes without zero function values.
- Generating all isomorphism classes of such chirotopes results in 386 solutions. The first five are listed below:

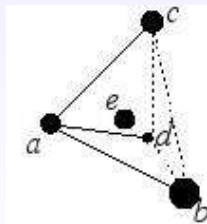
$0123$	$0124$	$0134$	$0234$	$1234$	$0125$	$0135$	$0235$	$1235$	$0145$	$0245$	$1245$	$0345$	$1345$	$2345$
+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
+	+	+	+	+	+	+	+	+	+	+	+	+	+	-
+	+	+	+	+	+	+	+	+	+	+	+	+	+	-
+	+	+	+	+	+	+	+	+	+	+	+	+	-	+
+	+	+	+	+	+	+	+	+	+	+	+	-	-	+
⋮														

# Further reductions

- We can exclude conformations where a bond intersects with such a triangle, as the steric energy of this conformation would be very high. Thus we can exclude all chirotopes containing a circuit of the form  $(+a, +b, +c, -d, -e)$  where atoms  $a$ ,  $b$ , and  $c$  are consecutively connected as well as atoms  $d$  and  $e$ .



(a)



(b)

# Further reductions

This reduces the number of solutions to 162. Focusing on a set  $R$  of relevant quadruples of atoms, for example the four neighbors of each stereocenter,

$$R = \{0123, 1234, 0125, 0145, 0345, 2345\},$$

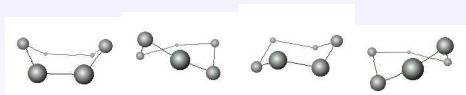
we obtain 13 partial chirotopes shown in

$0123$	$0124$	$0134$	$0234$	$1234$	$0125$	$0135$	$0235$	$1235$	$0145$	$0245$	$1245$	$0345$	$1345$	$2345$
+			+	+				+		+		+	+	
+			+	+				+		+		+	-	
+			+	+				+		-		-	+	
+			+	+				+		-		-	-	
+			+	+				-		+		+	+	
+			+	+				-		+		+	-	
+			+	+				-		-		-	+	
+			+	-				+		-		-	+	
+			-	+				+		+		+	+	
+			-	+				+		+		+	+	

# The resulting stereoisomers:

Using constrained optimization, the expected conformations are

- For only 4 of the given partial chirotopes, we found a stable realization:



+++++-

++++--

+-----

-----

Note that the first of the conformations, known as boat form, is indeed an unstable conformation having zero gradient (i.e. a saddle point of the energy function). Though it is interesting that this was found by our algorithms, its unstability can easily be recognized by considering the Hesse matrix of the energy function.

The remaining three conformations represent exactly the usual classification of all possible conformations of cyclohexane, i.e. the chair form (the third conformation) as well as two enantiomorphic twist forms (second and fourth conformation).

# Thanks for your attention!

My email addresses:

[adalbert-kerber@t-online.de](mailto:adalbert-kerber@t-online.de)

also

[kerber@uni-bayreuth.de](mailto:kerber@uni-bayreuth.de)

URL of MOLGEN-ONLINE:

<http://www.molgen.de>