

On the Recognition of Composed Systems of Stereocenters in Molecular Graph Theory by Wreath Products

Bernd GÜNTHER* and Jochen ZÜGGE

Beilstein-Institut
Trakehner Straße 7-9, 60487 Frankfurt, Germany
*e-mail: bguenther@beilstein-institut.de

(Received: April 26, 2007; Accepted for publication: June 28, 2007; Published on Web: August 2, 2007)

We are going to demonstrate how group theory can be used for the detection, classification and validation of systems of related stereo centers in molecules. At the core of the formalism, which is suitable for computer programming and has been tested against ten million organic compounds, lies an analysis of the transformational properties of the molecule in terms of the action of those symmetry operations that conform to its graph theoretical structure while possibly modifying its stereochemical properties.

Keywords: Group actions, Wreath products, Molecular graphs, Stereochemistry, Graph automorphism groups, Graph canonisation

1 Introduction

Chemical structure processing is beyond pure graph theory, one of the most prominent reasons being the presence of stereochemical information. In organic chemistry, by far the most important stereocenters are of the tetrahedral or olefinic type, cf. Figure 1. Subjecting the four ligand atoms either of a tetrahedron or of an olefinic center to an odd permutation will switch its orientation. In the tetrahedral case the new orientation will be the mirror image of the original one (an instance of chirality), whereas in the olefinic case the two orientations are unrelated (the stereocenter is non chiral). More complex situations arise in crystallography or when several individual stereocenters interact. The key to the understanding is the action of the configuration symmetry group, i.e., the group of those operations that leave our stereo orientation invariant; precise definitions will be given in the next section. This has been established as a general principle in [1] and solves the case of individual stereocenters almost completely.

There is one reservation. The stereo information in Figure 1 is only genuine, if the four ligand atoms are “different”. Two ligands are indistinguishable if they can be exchanged by a symmetry operation which leaves the focus invariant, *unless this operation modifies the stereo configuration of some other center*. A familiar example for the latter effect occurs in spiro compounds [2](p.157) (cf. Figure 7), where three tetrahedral centers are located at opposite positions of two connected rings. A particular

challenge is presented by inositol [3](p.707) (cf. Figure 8), where each atom of a hexagonal ring is stereogenic: there are nine configurations, two of which are chiral opposites of each other. In order to distinguish the chiral stereoisomers it must be checked if a certain enumeration of the six ring atoms runs clockwise or counterclockwise. A possible way of dealing with composed systems of stereocenters would be to consider configuration symmetries on the union of all their ligand sets. However, the complexity of this approach is intimidating: in the case of inositol we would have to deal with six times four ligands, some of which overlap. In addition, most structure processing systems will fail to deduce the ring orientation necessary to identify the two chiral isomers.

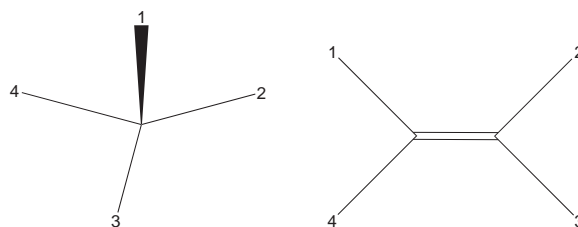


Figure 1. Tetrahedral (left) and olefinic stereocenter (right). Solid wedge bonds denote “up” bonds, while “down” bonds will be denoted by shaded wedge bonds (e.g. Figure 5). No other way to convey stereochemical information (such as line thickness, perspective drawing etc.) will be considered in our paper.

Fortunately, there is an easier approach, exploiting group theoretical methods to greater strength. A symmetry operation acting on our molecular graph has two effects on stereo information: it permutes the stereocenters and it modifies their orientation. The traces cut out by these actions are characteristic for certain combinations of stereocenters and can be used for their recognition. The action on n stereocenters will be described by a subgroup of the wreath product [4](§1.10) $S_n \wr \Gamma$ of the symmetric group S_n and the “group of configurations” (the group N/R in the notation of section 3) of the stereochemistry under consideration. In the tetrahedral or olefinic cases Γ is the two element group, because there are only two configurations. Moreover, the size of this subgroup will be small, about the size of the symmetry group of the considered configuration. For inositol this is the dihedral group $D(6)$ and hence contains twelve elements.

2 Stereocenters

A stereo descriptor consists of four pieces of data:

1. A focus, which may be either an atom or a bond.
2. A geometry type, e.g. tetrahedral, allenic, olefinic.
3. The list of atoms $[p_1, \dots, p_n]$, whose positions distinguish between different orientations of the geometry, and which we will call the *carrier* of the stereo descriptor.
4. The group of those permutations of the carrier that leave its orientation invariant (in the examples above precisely the even permutations), which will be called the rotation group of the corresponding geometry type. Thus a permutation $\pi \in S_n$ is a rotation if and only if the carriers $[p_1, \dots, p_n]$ and $[p_{\pi(1)}, \dots, p_{\pi(n)}]$ determine equivalent stereo descriptors.

For example, in Figure 2, the focus is the atom 1, the geometry type is “tetrahedral”, the carrier is the list of atoms $[2, 6, 11, 7]$ and the rotation group is the alternating group A_4 .

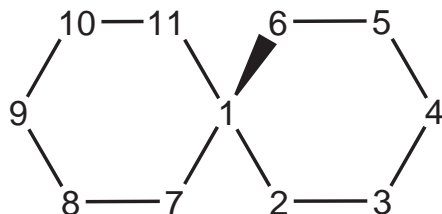


Figure 2. Spiro[5.5]undecane

In [1] the term configuration symmetry group is used for the group of rotations; see [5, 6] for bibliographies. The focus only serves for visualization and is otherwise dispensable. We will tacitly assume that any symmetry operation leaving the focus fixed also fixes the carrier setwise, or else disregard the focus and use the latter property as definition of the stabilizer group of a stereocenter. We note in passing that the full symmetric group S_n acts from the right on the stereo descriptors with given geometry and fixed focus: If D has the carrier $[p_1, \dots, p_n]$ and $\pi \in S_n$, then we denote by D^π the stereo descriptor with carrier $[p_{\pi(1)}, \dots, p_{\pi(n)}]$.

We assume that we have an efficient way of computing the automorphism group of a molecular graph in terms of its generators; these generators are customarily obtained as spinoff during graph canonisation [7, 8]. From these it is easily possible to derive generators of stabilizer subgroups for instance by the Schreier’s lemma [4](Thm.1.12).

Let G be a group of automorphisms acting from the left on our molecule. Then, evidently, G also acts from the left on the stereo descriptors of a given geometry: For a stereo descriptor D with focus p_0 and carrier $[p_1, \dots, p_n]$ and $g \in G$ we denote by gD the stereo descriptor with focus gp_0 and carrier $[gp_1, \dots, gp_n]$. The actions of G and S_n commute.

The stereocenter’s stabilizer group G_0 , to be understood as the setwise stabilizer of the carrier, requires more investigation. G_0 induces a permutation π of the sequence $[p_1, \dots, p_m]$ uniquely determined by the rule $gp_i = p_{\pi(i)}$, i.e. $gD = D^\pi$. Defining $\alpha(g) = \pi$ we obtain an antihomomorphism $\alpha: G_0 \rightarrow S_n$.

We can now test the validity of our stereo descriptor D as follows.

Definition 1 *The stereo descriptor D is called admissible with respect to the automorphism group G if the entire image $\alpha(G_0) \subseteq S_n$ is contained in the rotation group.*

In the example given in Figure 2, the stabilizer group G_0 is the eight element subgroup of the symmetric group S_{11} generated by the elements

$$\begin{aligned}
 h_1 &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 1 & 6 & 5 & 4 & 3 & 2 & 7 & 8 & 9 & 10 & 11 \end{pmatrix} \\
 h_2 &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 1 & 2 & 3 & 4 & 5 & 6 & 11 & 10 & 9 & 8 & 7 \end{pmatrix} \\
 h_3 &= \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 1 & 7 & 8 & 9 & 10 & 11 & 2 & 3 & 4 & 5 & 6 \end{pmatrix}
 \end{aligned}$$

Notice that h_1 turns the right hexagon once over, h_2 turns the left hexagon once over, and h_3 switches the two hexagons. Hence $\alpha(h_1)$ is the transposition $(2,6)$, $\alpha(h_2)$ is the transposition $(7,11)$, and $\alpha(h_3)$ is the permutation $(2, 7)(6, 11)$. On the other hand, the rotation group of the

tetrahedral center is the *alternating* group A_4 which does *not* contain any transpositions, hence the tetrahedral center is not admissible. Yet a chemical structure diagram is quite likely to be drawn that way in order to visualize the spatial arrangement of atoms: the right hexagon is perpendicular to the left with atom 2 lying above the paper plane. An automatic structure processing system must recognize the alleged stereo descriptor as fake and disregard it before registering the structure with a database.

3 Composed systems of stereo descriptors

There exist stereo descriptors which are not admissible in their own right but nevertheless exhibit stereogenic behavior in conjunction with other stereo descriptors. An example is provided in Figure 3. The carbon atoms 1 and 4 carry non admissible tetrahedral stereo descriptors D_1 and D_4 . The automorphism group of the molecule is Klein's four group generated by the axial reflection h_1 and the rotation h_2 with relations $h_1^2=h_2^2=h_1h_2h_1^{-1}h_2^{-1}=1$:

$$h_1 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 1 & 6 & 5 & 4 & 3 & 2 \end{pmatrix}$$

$$h_2 = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 4 & 5 & 6 & 1 & 2 & 3 \end{pmatrix}$$

Only ring atoms are listed. h_1 leaves both foci invariant but acts as transposition of the two ligand carbon atoms; since this is an odd permutation of the stereo descriptors' carrier they are not admissible. However, every automorphism either fixes both stereo descriptors' orientation or changes them coherently, so that the system consisting of both can be expected to be of stereochemical relevance and must be taken into account when registering the substance with a database. Note that one can conceive of a similar molecule with stereo descriptors located on three of the carbon atoms or on all six of them.

We set ourselves the task to recognize the presence of composed systems of stereocenters and describe their configuration. Let us suppose that after analyzing the admissibility of stereocenters in definition 1 some non admissible stereo descriptors remain, and let us consider a subset D_1, \dots, D_m of them, all of which shall be of the same geometry. The latter requirement only serves to

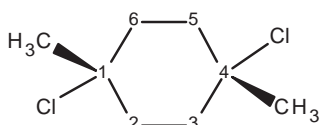


Figure 3. 1,4-dichloro-1,4-dimethylcyclohexane

simplify our presentation; more general cases such as the one depicted in Figure 4 can be treated by the same technique but with more cumbersome details. Suppose also that we are given a group H acting from the left on the index set $\{1, \dots, m\}$. In Figure 3 this would be Klein's four group, whereas for three resp. six stereo descriptors the dihedral groups $D(3)$ or $D(6)$ would occur. Furthermore, for each element $h \in H$ and every index i an isomorphism $\beta_{hi}: Z_i \approx Z_{h(i)}$ has to be specified, where Z_i denotes the carrier of stereo descriptor D_i . These isomorphisms describe the alteration of orientations of stereo descriptors affected by the operation of H . They are subject to the following conditions:

1. For all indices i the isomorphism $\beta_{1i}: Z_i \approx Z_i$ is a rotation.
2. For each element $h \in H$, every index i and every rotation ρ of the carrier Z_i of D_i the conjugate $\beta_{hi}\rho\beta_{hi}^{-1}$ is a rotation of the carrier $Z_{h(i)}$.
3. For any two elements $g, h \in H$ and each index i the composition $\beta_{gh,i}^{-1}\beta_{g,h(i)}\beta_{hi}$ is a rotation of Z_i .

We observe that this definition does not alter if the stereo descriptors D_i are replaced by rotated descriptors $D_i^{\rho_i}$; to see this one simply has to replace the isomorphisms β_{hi} by $\rho_{h(i)}^{-1}\beta_{hi}\rho_i$.

Thus equipped we may now define the rotation group of the system $D_1 \dots D_m$. A rotation consists of an element $h \in H$ and, for each index i , an isomorphism $\gamma_i: Z_i \approx Z_{h(i)}$ such that $\gamma_i \beta_{hi}^{-1}$ is a rotation for each i . Using conditions 1-3 it is straightforward though tedious to check that these rotations indeed constitute a group under the obvious law of composition.

A more concise description of all this may be given as follows. Let $R \subseteq S_n$ denote the rotation group of our stereo geometry and consider the normalizer $N = \{ \pi \in S_n \mid \pi^{-1} R \pi = R \}$ of R in S_n ; N is the largest subgroup of S_n such that R is a normal subgroup of N . There is a homomorphism $(\text{ad}, \beta): H \rightarrow S_m \times (N/R)^m$, where $\text{ad}: H \rightarrow S_m$ is the adjoint map corresponding to the operation of H on $\{1, \dots, m\}$ and the semidirect product $S_m \times (N/R)^m$ is formed with respect to the obvious operation on the exponent of the right factor. This construction is known as the *wreath product* [4](§1.10) of the groups S_m and N/R . Then the rotation group \mathbf{R} of our system of composed stereo descriptors is obtained as *pullback diagram* (also called *fibre product* [9](Ch.1 §4.8 p.46)):

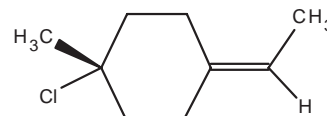


Figure 4. (1S)-1-chloro-4-ethylidene-1-methylcyclohexane

$$\begin{array}{ccc}
 \mathbf{R} & \longrightarrow & S_m \times N^m \\
 \downarrow & & \downarrow \\
 H & \xrightarrow{(\text{ad}, \beta)} & S_m \times (N/R)^m
 \end{array}$$

\mathbf{R} operates on $Z := \{1, \dots, m\} \times \{1, \dots, n\}$ by $(h, \pi_i)(i, j) := (h(i), \pi_i(j))$ and thus reveals itself as the rotation group of the abstract carrier Z . It contains $(\#H)(\#R)^m$ elements, but fortunately it is not necessary to write them down to check if a given permutation γ of Z is a rotation. Instead, one considers the induced permutation $\bar{\gamma} \in S_m$ of the index set $\{1, \dots, m\}$ and the switch of orientations $\gamma_i \in N/R$ from γD_i to $D_{\bar{\gamma}(i)}$ and checks if there exists an element $h \in H$ operating just that way, i.e., $\text{ad}(h) = \bar{\gamma}$ and $\beta(h) = (\gamma_i)$.

Figure 3 shall serve as illustration. Here $m=2$, H is Klein's four group $H = \mathbb{Z}_2 \times \mathbb{Z}_2$ and the rotations are the alternating group $R = A_4$. Since A_4 is a normal subgroup of the full symmetric group $N = S_4$ and $N/R = \mathbb{Z}_2$, S_2 acts on $(N/R)^2 = \mathbb{Z}_2 \times \mathbb{Z}_2$ by switching the factors, thus arises the semidirect product $S_2 \times (\mathbb{Z}_2 \times \mathbb{Z}_2)$. We define the homomorphism $(\text{ad}, \beta): \mathbb{Z}_2 \times \mathbb{Z}_2 \rightarrow S_2 \times (\mathbb{Z}_2 \times \mathbb{Z}_2)$ by specifying its values on the generators h_1 and h_2 used in section 3. $\text{ad}(g_i) \in S_2$ is the action of g_i on the index set $\{1, 4\}$, so $\text{ad}(h_1)$ is the identity and $\text{ad}(h_2)$ is the transposition. h_1 reverses the orientation of both stereo descriptors and h_2 preserves both orientations, so $\beta(h_1) = (1, 1) \in \mathbb{Z}_2 \times \mathbb{Z}_2$ and $\beta(h_2) = (0, 0)$.

$$\begin{array}{ccc}
 \mathbf{R} & \longrightarrow & S_2 \times (S_4 \times S_4) \\
 \downarrow & & \downarrow \\
 \mathbb{Z}_2 \times \mathbb{Z}_2 & \xrightarrow{(\text{ad}, \beta)} & S_2 \times (\mathbb{Z}_2 \times \mathbb{Z}_2)
 \end{array}$$

Identifying h_1 with $(1, 0) \in \mathbb{Z}_2 \times \mathbb{Z}_2$, h_2 with $(0, 1) \in \mathbb{Z}_2 \times \mathbb{Z}_2$ and denoting by $\tau \in S_2$ the transposition we can write $\text{ad}(m_1, m_2) = \tau^{m_2}$ and $\beta(m_1, m_2) = (m_1, m_1)$. Then \mathbf{R} becomes the set of all quintuples $(m_1, m_2, \sigma, \pi_1, \pi_2)$, $m_i \in \mathbb{Z}_2$, $\sigma \in S_2$, $\pi_i \in S_4$ such that $\sigma = \text{ad}(m_1, m_2) = \tau^{m_2}$ and $\text{sign}(\pi_1), \text{sign}(\pi_2) = \beta(m_1, m_2) = (m_1, m_1)$. Hence we can dispense with m_1 and m_2 and identify \mathbf{R} with the set of all triples (σ, π_1, π_2) such that $\text{sign}(\pi_1) = \text{sign}(\pi_2)$. We must observe that the group law in \mathbf{R} is *not simply coordinatewise multiplication*, but instead $(\bar{\sigma}, \bar{\pi}_1, \bar{\pi}_2) (\sigma, \pi_1, \pi_2) = (\bar{\sigma}\sigma, \bar{\pi}_{\sigma(1)}\pi_1, \bar{\pi}_{\sigma(2)}\pi_2)$.

In the example above the rotation group \mathbf{R} turned out to be a subgroup of $S_m \times N^m$. This is quite general: it holds whenever $(\text{ad}, \beta): H \rightarrow S_m \times (N/R)^m$ is a monomorphism, which may be assumed without loss of generality, because a kernel element would act trivially on the index set as well as on the stereo descriptor orientations and

hence could be considered non physical.

Our method has been tested against the Beilstein database with more than ten million organic compounds. We discuss here the most common structures encountered.

All individual stereocenters presented here are of the tetrahedral type. The number of carrier atoms is given by $n=4$, and the rotation group $R \subseteq S_4$ of orientation preserving carrier transformations is given by the alternating group $R = A_4$. Since A_4 is a normal subgroup of the full permutation group S_4 we have $N = S_4$, and the factor group $N/R = S_4/A_4$ is the cyclic group \mathbb{Z}_2 of order 2. It will be convenient to write \mathbb{Z}_2 as multiplicative group consisting of the two elements ± 1 .

The construction of the geometry of a composed system of m stereocenters requires us to specify a subgroup $H \subseteq S_m \times \mathbb{Z}_2^m$, which will be done by listing the elements. An element of H will be a sequence $(\pi, \varepsilon_1, \dots, \varepsilon_m)$ consisting of a permutation $\pi \in S_m$ and a sequence of numbers $\varepsilon_i = \pm 1$. We observe that the law of composition is given by the semidirect product rule $(\pi, \varepsilon_1, \dots, \varepsilon_m) (\rho, \delta_1, \dots, \delta_m) = (\pi\rho, \varepsilon_{\rho(1)}\delta_1, \dots, \varepsilon_{\rho(m)}\delta_m)$. A listing of such elements will then appear as a table containing in its first column the permutation π in disjoint cycle notation [10](p.40) and the elements $\varepsilon_1, \dots, \varepsilon_m$ in m further columns. The $(m+1)$ -tuple $(\pi, \varepsilon_1, \dots, \varepsilon_m)$ denotes a transformation such that the stereocenter with index i is shifted to position $\pi(i)$ and hereby preserves its orientation if and only if $\varepsilon_i = 1$. The transformation tables obtained hereby will serve *two* purposes: to recognize the geometry of composed systems of stereocenters and to distinguish their different configurations.

We emphasize that transformations are more general than symmetries: symmetries are precisely those transformations which preserve the orientation of all stereocenters and can easily be recognized by having $\varepsilon_i = 1$ for all i .

4 Pairs of stereocenters

The easiest situation is that of two tetrahedral stereocenters D_1 and D_2 located at opposite positions in a ring as in Figure 5. The group H consists of the four elements in Table 1. We observe that turning trans-1,4-dimethylcyclohexane into its opposite cis configuration would require the transformation $(\text{id}, -1, 1)$ which is *not* included in H . Hence our approach properly distinguishes different configurations. Reflection at the plane of drawing produces the transformation $(\text{id}, -1, -1)$ contained in H , therefore the molecules are non chiral.

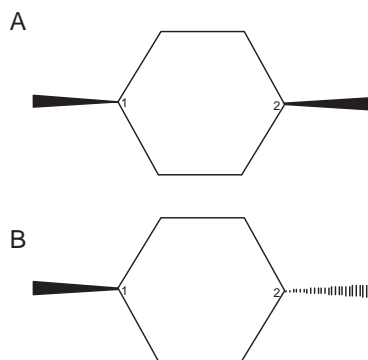


Figure 5. A: cis-1,4-dimethylcyclohexane and B: trans-1,4-dimethylcyclohexane.

Table 1. The group H for two interdependent stereo descriptors. The symbol id stands for the “identical” permutation leaving all elements fixed.

π	ε_1	ε_2
id	1	1
id	-1	-1
(1,2)	1	1
(1,2)	-1	-1

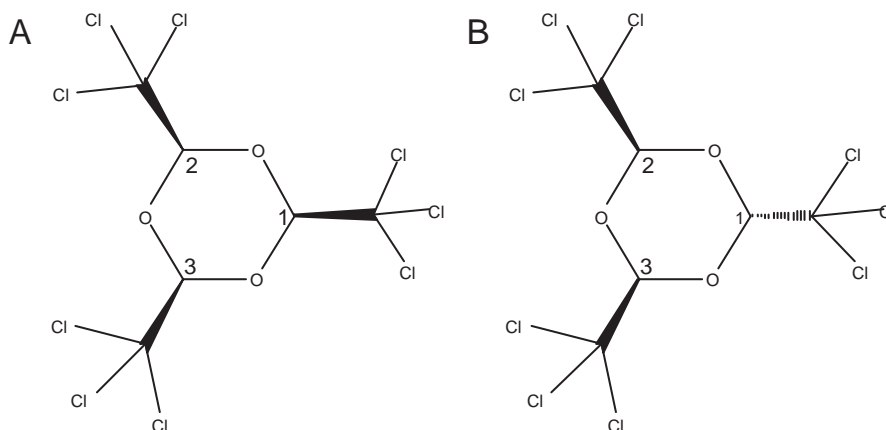


Figure 6. A: (2 α ,4 α ,6 α)-2,4,6-tris(trichloromethyl)-1,3,5-trioxane and B: (2,4 α ,6 β)-2,4,6-tris(trichloro-methyl)-1,3,5-trioxane.

Table 2. The groups H for the three cyclically interdependent stereo descriptors of Figure 6.

A	π	ε_1	ε_2	ε_3	
1	id	1	1	1	×
2	id	-1	-1	-1	
3	(1,2)	1	1	1	
4	(1,2)	-1	-1	-1	×
5	(1,3)	1	1	1	
6	(1,3)	-1	-1	-1	×
7	(2,3)	1	1	1	
8	(2,3)	-1	-1	-1	×
9	(1,2,3)	1	1	1	×
10	(1,2,3)	-1	-1	-1	
11	(1,3,2)	1	1	1	×
12	(1,3,2)	-1	-1	-1	

B	π	ε_1	ε_2	ε_3	
1	id	1	1	1	×
2	id	-1	-1	-1	
3	(1,2)	-1	-1	1	
4	(1,2)	1	1	-1	×
5	(1,3)	-1	1	-1	
6	(1,3)	1	-1	1	×
7	(2,3)	1	1	1	
8	(2,3)	-1	-1	-1	×
9	(1,2,3)	1	-1	1	
10	(1,2,3)	-1	1	-1	×
11	(1,3,2)	1	1	-1	
12	(1,3,2)	-1	-1	1	×

5 Three cyclically related stereocenters

Now the molecules in Figure 6 shall be investigated. In Table 2 we label the transformations actually occurring in the molecule by \times . In the left table pertaining to the *cis* configuration row 9 corresponds to the “vertical rotation” of the ring by $2\pi/3$, row 8 corresponds to the “horizontal rotation” around the axis connecting center 1 with the opposite oxygen by π and row 7 corresponds to the planar reflection at the vertical plane containing the same axis. These three elements generate the entire group: all other elements are obtained as compositions. We observe that no graph automorphism will ever realize the planar reflection, because the atom map that would affect this reflection reverses the sense of orientation of the ring and consequently of the stereocenters: the resulting transformation is the horizontal rotation mentioned above. Nevertheless, the planar reflection must be added to our transformation group to identify the configuration depicted in Figure 6 with its mirror image with all stereo bonds facing down. The transformation (id, -1, -1, -1) in row 2 is the composition of rows 7 and 8 and clearly shows that the molecule is not chiral.

There also is a purely algebraic reason for adding the “formal” transformation α_7 in row 7: denoting the group element in row i by α_i one shows $\alpha_7\alpha_9\alpha_7^{-1}=\alpha_{11}$ and $\alpha_7\alpha_8\alpha_7^{-1}=\alpha_8$. Since the elements α_9 and α_8 generate the subgroup of “actual transformations” this subgroup is normalized by α_7 , and if axiom 1 (cf. section 9) shall be satisfied α_7 must be included in the transformation table.

The right table (B) describes the *trans* configuration and is similar with rows 10, 8 and 7 containing vertical rotation, horizontal rotation and planar reflection. As above, axiom 1 tells us that row 7 must be included. We observe that the left table (A) is symmetric (cf. section 9) with respect to any index permutation while the right table (B) is symmetric with respect to the transposition (2, 3). In the latter case we cannot improve symmetry by adding more transformations because then H would encompass the entire group $S_3 \times \mathbb{Z}_2^3$.

6 Three linearly related stereocenters

Figure 7 shows (*R_a*)-spiro[5.5]undecane-3,9-diol and (*S_a*)-spiro[5.5]undecane-3,9-diol. In these molecules all three stereo centers are dependent of each other. Numbering the stereo centers from left to right results in the transformation table that is given in Table 3. We observe that it is possible to rotate each of the two rings separately and to switch the rings, giving a totality of eight

symmetry operations for this molecule (those marked by \times). The two possible configurations of this molecule are properly distinguished because their relation would have to be furnished by an operation like (id, 1, 1, -1) not contained in the Table 3. The molecule is chiral and therefore the transformation (id, -1, -1, -1) is not contained in H .

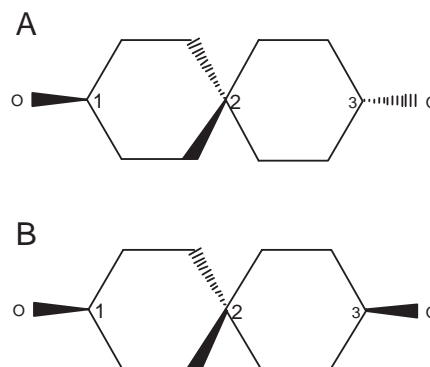


Figure 7. A: (*R_a*)-spiro[5.5]undecane-3,9-diol and B: (*S_a*)-spiro[5.5]undecane-3,9-diol.

Table 3. The groups H for the three stereo descriptors of Figure 7.

	π	ϵ_1	ϵ_2	ϵ_3	
1	id	-1	-1	1	\times
2	id	-1	1	-1	\times
3	id	1	-1	-1	\times
4	id	1	1	1	\times
5	(2,3)	-1	-1	1	
6	(2,3)	-1	1	-1	
7	(2,3)	1	-1	-1	
8	(2,3)	1	1	1	
9	(1,2)	-1	-1	1	
10	(1,2)	-1	1	-1	
11	(1,2)	1	-1	-1	
12	(1,2)	1	1	1	
13	(1,2,3)	-1	-1	1	
14	(1,2,3)	-1	1	-1	
15	(1,2,3)	1	-1	-1	
16	(1,2,3)	1	1	1	
17	(1,3,2)	-1	-1	1	
18	(1,3,2)	-1	1	-1	
19	(1,3,2)	1	-1	-1	
20	(1,3,2)	1	1	1	
21	(1,3)	-1	-1	1	\times
22	(1,3)	-1	1	-1	\times
23	(1,3)	1	-1	-1	\times
24	(1,3)	1	1	1	\times

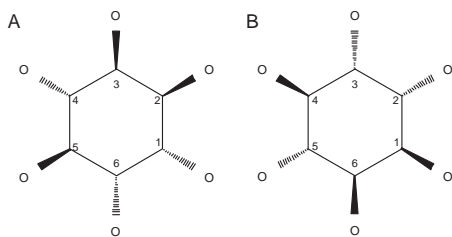


Figure 8. A: D-chiro-inositol and B: L-chiro-inositol.

Table 4. Pair and triple configurations on hexagons.

	inositol	1,4	2,5	3,6	1,3,5	2,4,6
1	cis	cis	cis	cis	cis	cis
2	epi	cis	cis	trans	cis	trans
3	allo	trans	trans	cis	trans	trans
4	myo	trans	cis	trans	trans	cis
5	muco	cis	cis	cis	trans	trans
6	neo	trans	trans	trans	trans	trans
7	scyllo	trans	trans	trans	cis	cis
8	D-chiro	cis	cis	trans	trans	trans
9	L-chiro	cis	trans	cis	trans	trans

Now let us suppose we break the symmetry of the molecules by substituting one of the OH-groups with a methyl-group. This will limit the “actual” transformations in our table to rows 1–4, but, paradoxically, we must enlarge the group if we want to comply with axiom 1 because this four element subgroup is normalized by the transformation $((1, 2, 3), 1, 1, 1)$ (row 16). Adding it creates a total of 16 new elements as compositions. The complete 24 element group now has the advantage of being symmetric with respect to any permutation of S_3 (cf. 9) and therefore simplifies the recognition test.

7 Hexagonal structures

There are nine different configurations of inositol [3](p.707). Inscribed in the hexagon we find three pairs of related stereocenters located on opposite atoms and two triples of stereocenters located on equidistant triangles. The diastereomers 1–7 are differentiated by these systems (cf. Table 4), while the two chiral configurations 8 and 9 (cf. Figure 8) are not. To distinguish between these we work out their transformation properties in Table 5.

The symmetry group of the hexagon is the dihedral group D_6 , that may be identified with the subgroup of the full symmetric group S_6 generated by the vertical rotation $(1, 2, 3, 4, 5, 6)$ and the horizontal rotation $(1, 2)(3, 6)(4, 5)$. The permutations obtained hereby constitute the first columns of the transformation tables. The two generators appear in rows 2 and 8, and the remaining elements are compositions.

Table 5. The groups H for the chiral configurations of inositol as depicted in Figure 8.

	π	ϵ_1	ϵ_2	ϵ_3	ϵ_4	ϵ_5	ϵ_6
1	id	1	1	1	1	1	1
2	$(1, 2, 3, 4, 5, 6)$	-1	1	-1	-1	-1	1
3	$(1, 3, 5)(2, 4, 6)$	-1	-1	1	1	-1	-1
4	$(1, 4)(2, 5)(3, 6)$	1	1	-1	1	1	-1
5	$(1, 5, 3)(2, 6, 4)$	-1	-1	-1	-1	1	1
6	$(1, 6, 5, 4, 3, 2)$	1	-1	1	-1	-1	-1
7	$(2, 6)(3, 5)$	-1	1	-1	-1	-1	1
8	$(1, 2)(3, 6)(4, 5)$	1	1	1	1	1	1
9	$(1, 3)(4, 6)$	1	-1	1	-1	-1	-1
10	$(1, 4)(2, 3)(5, 6)$	-1	-1	-1	-1	1	1
11	$(1, 5)(2, 4)$	1	1	-1	1	1	-1
12	$(1, 6)(2, 5)(3, 4)$	-1	-1	1	1	-1	-1

Both configurations have the same transformation properties. The distinction between them can be achieved by the sense of orientation of the hexagon: a polar arrow denoting the sense of rotation obtained by turning the hexagon following the numbered atoms increasingly will point in the same direction as the bond at atom 1 in the case of L-chiro-inositol and the opposite otherwise. Both configurations are chiral, in fact they are mirror images of one another, and consequently the planar reflection $(id, -1, -1, -1, -1, -1)$ is contained in neither transformation group.

8 The pentagonal structure

Our friends, the vertical and horizontal rotations and the planar reflection can be found in rows 7, 9 and 10 in the left table and rows 8, 5 and 6 in the right table. Both transformations encompass “formal” transformations in addition to the actual ones marked by \times , in particular the planar reflections in rows 10 (left table) and 6 (right table). As in section 5 it can be shown that their presence is enforced by axiom 1. Both configurations are non chiral.

Both transformation tables are conjugate subgroups of $S_5 \times \mathbb{Z}_2^5$, the conjugation being accomplished by the pentagram ordering $\sigma = ((1, 3, 4, 2), 1, \dots, 1)$. It suffices to check the conjugacy relation for the generating elements. Denoting the element in row i of the left table by α_i and the element in row j of the right table by β_j , we have $\sigma\alpha_7\sigma^{-1} = \beta_{15}$, $\sigma\alpha_9\sigma^{-1} = \beta_5$ and $\sigma\alpha_{10}\sigma^{-1} = \beta_6$. Consequently, when submitting one of the structures cyclopentane-pentol to the recognition test *both* transformation tables will match: one with the stereocenters in correct cyclic ordering and the other in pentagram ordering. An analysis of transformational properties alone will not reveal the correct match, instead a reference to the geometry has to be made to find out which stereocenters are connected by bonds.

Table 6. The groups H for the two pentagonal structures of Figure 9.

A	π	ε_1	ε_2	ε_3	ε_4	ε_5	
1	id	-1	-1	-1	-1	-1	
2	id	1	1	1	1	1	×
3	(2,5)(3,4)	-1	-1	1	1	-1	×
4	(2,5)(3,4)	1	1	-1	-1	1	
5	(1,2)(3,5)	-1	-1	-1	1	-1	
6	(1,2)(3,5)	1	1	1	-1	1	×
7	(1,2,3,4,5)	-1	-1	-1	1	-1	×
8	(1,2,3,4,5)	1	1	1	-1	1	
9	(1,3)(4,5)	-1	-1	-1	-1	-1	×
10	(1,3)(4,5)	1	1	1	1	1	
11	(1,3,5,2,4)	-1	-1	1	1	-1	
12	(1,3,5,2,4)	1	1	-1	-1	1	×
13	(1,4)(2,3)	-1	-1	-1	-1	1	
14	(1,4)(2,3)	1	1	1	1	-1	×
15	(1,4,2,5,3)	-1	1	1	1	-1	×
16	(1,4,2,5,3)	1	-1	-1	-1	1	
17	(1,5,4,3,2)	-1	-1	-1	-1	1	×
18	(1,5,4,3,2)	1	1	1	1	-1	
19	(1,5)(2,4)	-1	1	1	1	-1	
20	(1,5)(2,4)	1	-1	-1	-1	1	×

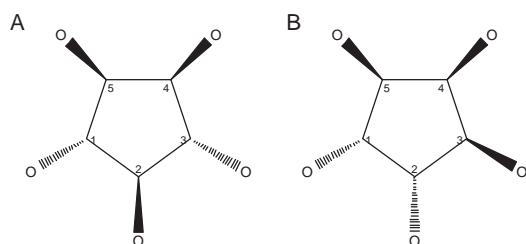


Figure 9. Two configurations of cyclopentane-pentol

9 Appendix: Symmetry in systems of stereocenters

Our analysis of putative systems of stereocenters D_1, \dots, D_m consists of observing their transformational behavior under symmetry operations and comparing it to a list of known transformation groups $H \subseteq S_m \times (N/R)^m$. Since the individual stereocenters may show up in arbitrary order we have to work out the effect of replacing the original sequence by a shuffled up sequence $(\bar{D}_1, \dots, \bar{D}_m) = (D_{\sigma(1)}, \dots, D_{\sigma(m)})$ for a permutation $\sigma \in S_m$. There arise two questions:

1. If the sequence D_1, \dots, D_m constitutes an admissible system of stereocenters, does the same hold for the permuted sequence $D_{\sigma(1)}, \dots, D_{\sigma(m)}$?
2. And if yes, do both determine the same configuration?

We consider a symmetry operation of our molecule and denote by $(h, \pi_i) \in S_m \times (N/R)^m$ its effect on the configuration of our system. This means that under this particular symmetry the system D_1, \dots, D_m corresponds to the system $D_{h(1)}^{\pi_1}, \dots, D_{h(m)}^{\pi_m}$. The shuffled up system $\bar{D}_1, \dots, \bar{D}_m$ is transformed into $\bar{D}_{\sigma^{-1}h\sigma(1)}^{\pi_{\sigma(1)}}, \dots, \bar{D}_{\sigma^{-1}h\sigma(m)}^{\pi_{\sigma(m)}}$, i.e. the original transformation is replaced by its conjugate $(\sigma, 1)^{-1}(h, \pi_i)(\sigma, 1)$. We conclude that question 1 has a positive answer if and only if for any occurring (h, π_i) its conjugate $(\sigma, 1)^{-1}(h, \pi_i)(\sigma, 1)$ belongs to our transformation group H .

Since the system $\bar{D}_1, \dots, \bar{D}_m$ is obtained from D_1, \dots, D_m by application of $(\sigma, 1)$ question 2 can be answered affirmatively if and only if $(\sigma, 1) \in H$.

We summarize:

Lemma 1 We suppose the geometry of our composed system of stereocenters is described by the transformation group $H \subseteq S_m \times (N/R)^m$, and the symmetry operations of a molecule under consideration act like a subgroup $H_0 \subseteq H$. Shuffling up our system by a permutation $\sigma \in S_m$ does not affect the admissibility of the system if and only if $(\sigma, 1)^{-1}H_0(\sigma, 1) \subseteq H$, and it does not alter the configuration of our system if and only if $(\sigma, 1) \in H$.

Manifestly, it must not happen that a permutation σ transforms an admissible system into another admissible system of different configuration. To avoid this anomaly

our transformation group H must satisfy the following axiom:

Axiom 1 If for a permutation $\sigma \in S_m$ ($\sigma, 1$) normalizes H , $(\sigma, 1)^{-1}H(\sigma, 1) \subseteq H$, then $(\sigma, 1)$ itself must belong to H .

Axiom 1 is satisfied in all our examples. We say that H is symmetric with respect to σ if the conclusion of axiom 1 is satisfied.

We explain the subtle distinction between the “formal” transformations constituting group H and the “actual” transformations lying in the subgroup $H_0 \subseteq H$ in the context of the three linearly connected stereocenters of section 6: Only transformations 1–4 and 21–24 are actually possible, all the others have been added for formal reasons. However, if the molecule’s symmetry is broken for instance by replacing one of the two hydroxy groups by, say, bromine, H_0 is further restricted to the four transformations 1–4. This is somewhat perilous, because this subgroup is normalized by $(\sigma, 1)$ for *any* permutation $\sigma \in S_m$, even allowing the central position to be placed at the end. To ensure that the configuration remains unaffected each $(\sigma, 1)$ must be included in H . Together with group elements obtained by composition we arrive at a total of 24 transformations, at most eight of which can be actually realized by symmetry operations.

The situation where H is symmetric with respect to *any* permutation is desirable because then the result of the recognition test is independent of the order in which our stereocenters D_1, \dots, D_m are submitted. In more general situations various orderings have to be considered, some causing the test to fail and others to accept, but in the case of acceptance always producing the same configuration provided axiom 1 is satisfied. Instead of trying all $m!$ permutations one should perform a preselection singling out impossible ones from the start. That can easily be done if our stereocenters constitute a polygon of size m , e.g. in section 7 with $m=6$, because then the dihedral group D_m will operate on the molecule. Among these symmetry operations there will be two generating rotations ρ by an angle $\pm 2\pi/m$ moving each vertex of the ring one step ahead. For $m=3, 4, 6$ these rotations are the only rotations of order m , so they can be recognized by purely algebraic properties. Pick any one atom of the ring, call it p and use the stereocenter located here as D_1 , the one on $\rho(p)$ as D_2 and the one on $\rho^{k-1}(p)$ as D_k . The ordering of stereocenters you arrive at will be cyclically correct. You may have to reverse the sense of orientation or pick a different starting point, but this leaves you with only $2m$ possibilities to try instead of the factorial $m!$.

For all other m different from 3, 4, 6 there is a pitfall. The case $m=5$ is conspicuous: *All four* non trivial rotations are of order five. This is the algebraic disguise of the geometric fact that the pentagon and an inscribed

pentagram (cf. Figure 10) have the same transformation group. Erroneously ordering the atoms using the rotation by $4\pi/5$ as a generator will lead to the pentagram ordering 0, 2, 4, 1, 3. We already noticed the role played by this permutation in section 8.

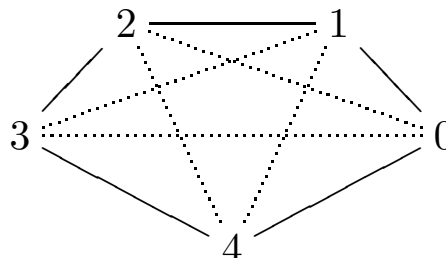


Figure 10. Pentagon and pentagram have the same transformation group.

References

- [1] J. G. Nourse, The Configuration Symmetry Group and Its Application to Stereoisomer Generation, Specification and Enumeration, *J. Am. Chem. Soc.*, **101**, 1210-1216 (1979).
- [2] T. Wieland, Erzeugung, Abzählung und Konstruktion von Stereoisomeren, *MATCH*, **31**, 153-203 (1994).
- [3] E. L. Eliel, S. H. Wilen, *Stereochemistry of Organic Compounds*, Wiley (1994).
- [4] P. J. Cameron, *Permutation Groups*, London Mathematical Society Student Texts 45, Cambridge (1999).
- [5] J. M. Barnard, A. P. F. Cook, B. Rohde, Storage and searching of stereochemistry in substructure search systems, *Chemical Information Systems - Beyond the Structure Diagram*, ed. by D. Bawden and E. Mitchell, New York (1990).
- [6] S. Fujita, *Symmetry and Combinatorial Enumeration in Chemistry*, Springer (1991).
- [7] B. D. McKay, *Practical graph isomorphism, Numerical mathematics and computing*, Proc. 10th Manitoba Conf., Winnipeg/Manitoba 1980, Congr. Numerantium 30, 45-87 (1981).
- [8] J. Braun, R. Gugisch, A. Kerber, R. Laue, M. Meringer, C. Rücker, MOLGEN-CID - A Canonizer for Molecules and Graphs Accessible through the Internet, *J. Chem. Inf. Comput. Sci.*, **44**, 542-548 (2004).

[9] Bourbaki, *Algebra I, Chapters 1-3*, engl. ed., Springer-Verlag, Berlin (1989).

[10] H. S. M. Coxeter, *Regular Polytopes*, 3rd ed., Dover (1973).