communications in mathematical
match
and in computer chemistry

no. 46, October 2002
MATCDY (46) 275 - 301 (2002)
ISSN 0304-6253

THE ARGENT PROGRAM SYSTEM: A SECOND-GENERATION TOOL AIMED AT COMBINATORIAL SEARCH FOR NEW TYPES OF ORGANIC REACTIONS. 2. MATHEMATICAL MODELS IN ARGENT–1

Serge S. Tratch,[†] Marina S. Molchanova,[‡] and Nikolai S. Zefirov[†]

[†] Department of Chemistry, Moscow State University;
Leninskie Gory, Moscow 119899, Russia
[‡] Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences;
Leninsky Prospekt 47, Moscow 117913, Russia

Abstract

The mathematical models of some reaction design problems are considered in detail; these models form the basis of the generation algorithm in ARGENT–1, the first of our new software systems for search and evaluation of unprecedented organic interconversions. The representation of graph symmetries by appropriate spatial symmetry operations makes it possible to explain the three-stage labeling process in ARGENT–1 in a clear and easily understandable manner; the nonequivalent vertex- and edge-labeled graphs are associated with orbits of the well-defined power group Γ or its two subgroups Γ' and Γ''. The use of the power group formalism for separate labeling stages resulted in presentation of mathematical models in a unified and very compact form. The applicability of the suggested models to some structural design problems is finally demonstrated.

1. INTRODUCTION

In the two papers presented in this issue, we demonstrate that the solution of typical reaction design problems necessitates sequential construction of several graph representations of organic reactions. The actual sequence of generation stages in ARGENT–1, the first version of our ARGENT system for design of new types of organic reactions, is $G \Rightarrow G_{TOP} \Rightarrow G_{SEQ} \Rightarrow G_{REQ}$, where, as was noted in the preceding paper,[1]

- G is the unsigned topology identifying graph, which represents the relative disposition of bonds that change their multiplicity during a reaction;

- G_{TOP} is the signed topology identifying graph, which, in addition to the topology described by G, also provides information about signs (symbols of charge or free electron) possibly borne by some atoms in the initial and/or final system of the reaction;

- G_{SEQ} is the graph representing the symbolic equation, which, in addition to the information contained in G_{TOP}, also provides information on the (initial and final) multiplicities of all bonds that actually participate in the reaction;

- G_{REQ} is the graph representing the reaction equation, which, in addition to the information of graph G_{SEQ}, also provides information on the chemical nature of the atoms (C, H, N, O, ...) whose bonds change their multiplicity in the course of the reaction.

As a result, the three stages of the generation process in ARGENT–1 are (1) assignment of "sign labels" to vertices of graph G, (2) assignment of "bond labels" to edges of graph G_{TOP}, and (3) assignment of "atom labels" to vertices of graph G_{SEQ}. However, although all these stages are essentially based on graph labeling, their mathematical models are different. This difference stems from the fact that the resultant graphs at any stage must be nonequivalent, and the equivalence relations depend not only on the symmetry of the starting graph but also on the interconvertibility of "paired" labels, which are converted into their counterparts as the direction of the reaction is reversed. In fact, as was shown in the preceding paper,

- generation of graphs G_{TOP} from G is based on the use of both paired and unpaired labels: the labels "+", "−", "·" (which denote the sign appearing in the initial system) and "(+)", "(−)", "(·)" (corresponding to signs in the final system) are paired, whereas the blank label (indicating that the corresponding atom is unsigned) is unpaired;

- generation of graphs G_{SEQ} from G_{TOP} involves only the use of paired labels; for each edge label a/b (with a and b, $a \neq b$, being the multiplicities of some edge in the initial and final systems), the label b/a is the counterpart;

- generation of graphs G_{REQ} from G_{SEQ} involves only the use of unpaired labels, each of them uniquely representing the name of the atom — and also its actual valence, if different valence states of the same chemical element are to be distinguished.

The necessity to make allowance for the interconvertibility of paired labels is responsible for the main difference between the aforementioned *problems of reaction design* and most typical *structural design problems*. The parallelism and differences between both types of multistage labeling problems is briefly overviewed in the final section of this paper.

In general, the term "labeling"[2,3] denotes some assignment of labels from a preselected set to specified parts (referred to as *sites*) of an arbitrary object possessing a given symmetry. Two labelings are considered nonequivalent (that is, essentially distinct) if they

are not interconvertible by any of the symmetry operations associated with the object in question. Perhaps the first systematic treatment of various labeling problems (associated with structural design) was described in 1974 in a paper entitled "Labeling of Objects Having Symmetry".[4a] The suggested efficient algorithm was applicable, however, only to unpaired labels; it formed the central part of the enhanced technique elaborated by the Stanford research group[4b] for exhaustive generation of all isomeric organic structures with a given molecular formula.

The general strategy for solving typical labeling problems can be described as follows:

- For a given unlabeled object (e.g., molecular graph or its given embedding in the 2D or 3D space), one should specify: (a) the set of sites — graph vertices (atoms), edges (chemical bonds), or chains, rings, k-tuples of atoms, etc. — to which labels are to be assigned; (b) the set of unpaired and/or paired labels associated with any characteristic that should be assigned to all sites; (c) the set of permutations that represent the symmetry group of the given unlabeled object; and possibly also (d) the set of additional requirements (constraints) that the resultant labeled objects must satisfy — e.g., the minimal and maximal allowed contents of some labels in the constructed labelings.

- The next thing to do is to construct the complete list of labelings (satisfying all the constraints, if any) and partition it into *equivalence classes*; each class is defined as consisting of labelings that are interconverted by at least one operation from the symmetry group.

- From each equivalence class, one should choose exactly one *representative labeling*; the list of these labelings selected from all classes is the solution to the labeling problem.

In the case of reaction design, partition of labelings into equivalence classes is a more complicated problem compared to typical problems of structural design. The complication is due to the presence of paired labels, such as "−" and "(−)" at the first stage or 0/1 and 1/0 at the second stage of the generation process described above. So, to recognize equivalent labelings (i.e., equivalent resultant graphs G_{TOP}, G_{SEQ}, and G_{REQ}), one must consider not only the symmetries of the starting graphs (G, G_{TOP}, and G_{SEQ}, respectively) but also interconversions of paired labels used at the first and second generation stages. A detailed and rigorous description of the corresponding mathematical models can be found in sections 5 and 6 of ref 5; however, this essentially formal description may seem too involved for those who are not familiar with applications of the permutation group theory to computational problems.

In the following two sections of this paper, we will consider the symmetry properties of starting graphs at each stage of reaction generation and the equivalence relations between resultant graphs in a simplified way, illustrating this analysis by examples corresponding to all the three stages. A formulation of the exact mathematical models for all relevant labeling problems and the easily understandable compact representations of these models can be found in sections 4 and 5. The application of just the same models to solution of some well-known structural design problems is demonstrated in section 6 of this paper.

2. SYMMETRY OPERATIONS AND INTERCONVERTIBILITY OF LABELS

As is known, symmetries of spatial molecular models are traditionally described in terms of *symmetry operations*, such as rotations about axes, reflections in planes, etc. In contrast to them, symmetries of a graph are usually represented by those permutations of its vertices that preserve all their connectivities. These permutations are traditionally referred to as *graph automorphisms*.[6] In other words, any graph automorphism is associated with such a renumbering of graph vertices that each pair of vertices with definite numbers is connected by an (ordinary) edge in the renumbered graph if and only if it was connected by an edge in the initial one. In the case of graphs containing *unpaired* vertex or edge labels, automorphisms are regarded as permutations that additionally preserve these labels.

The complete set of automorphisms for a given "parent" graph G forms its (vertex) *automorphism group* $Aut(G)$.[6] In many cases, permutations from the graph automorphism group may be unequivocally represented by operations from an isomorphic point symmetry group[7] associated with some embedding of the graph in the 2D or 3D space. In such a representation, one can easily visualize the symmetry properties of unlabeled graphs as well as graphs containing unpaired labels. As is shown below, a similar representation is also applicable to graphs with paired labels, and therefore it will be repeatedly used in this paper.

For example, let us consider graph G in Fig. 1a: the four permutations from its automorphism group $Aut(G)$ are in a one-to-one correspondence with four operations from the group C_{2v} — identity E, two reflections σ' and σ'' in vertical planes, and one rotation C_2 about a twofold axis. (Obviously, these symmetry operations pertain only to the proper — i.e., most symmetrical — embeddings of the graph.)

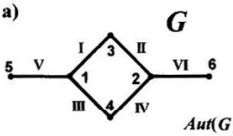
Any vertex automorphism group, just as the corresponding point symmetry group, partitions graph vertices into equivalence classes, or *orbits* of this group: vertices i and j belong to the same class if and only if vertex i can be moved to j (and vice versa) by at least one permutation from the group. Apparently, the vertex set of the graph in Fig. 1a is partitioned into three subsets consisting of equivalent vertices: $\{1,2\}$, $\{3,4\}$, and $\{5,6\}$. Considering the cyclic structures[8] of the permutations corresponding to the three nonidentity symmetry operations, one can see that, for any preselected pair of equivalent vertices, there are exactly two permutations in which these vertices belong to the same cycle.

Since the multistage generation process in ARGENT includes labeling of edges as well as of vertices, further analysis also requires explicit consideration of another permutation group, i.e., the *edge group*[6] of a graph. Permutations from this group are induced[9] by permutations from the vertex automorphism group but act on the set of graph edges rather than vertices. For an example of an edge group, see the rightmost part of Fig. 1a; examination of the four permutations clearly shows that the edge set {I, II, III, IV, V, VI} is partitioned into two equivalence classes — {I, II, III, IV} and {V, VI}. Note that the vertex and edge groups of the same graph are typically isomorphic,[10] and hence the same symmetry operations (i.e., E, σ', σ'', and C_2 in the example graph of Fig. 1a) may be associated with permutations from these two groups.

On the basis of the above considerations, it is easy to construct all labelings and select representative nonequivalent labelings in the case where *only unpaired labels* are used; the automorphism groups of resultant vertex-labeled or edge-labeled graphs may also be found without difficulty. For example, if there are two possible vertex labels (say,

Figure 1: Symmetry operations associated with an example graph G: (a) the cyclic notations of permutations that form the vertex (on the left) and edge (on the right) automorphism groups of G, (b) supplementary (overbarred) symmetry operations and the expanded group $Aut[G]$. The vertices and edges of the graph are indexed with Arabic and Roman numerals, respectively.

a)



$Aut(G) = \{E, \sigma', \sigma'', C_2\}$

E : (1)(2)(3)(4)(5)(6)	(I)(II)(III)(IV)(V)(VI)
σ' : (1)(2)(3 4)(5)(6)	(I III)(II IV)(V)(VI)
σ'': (1 2)(3)(4)(5 6)	(I II)(III IV)(V VI)
C_2: (1 2)(3 4)(5 6)	(I IV)(II III)(V VI)

b)

$$\left.\begin{array}{l} \overline{E} = E \\ \overline{\sigma}' = \sigma' \\ \overline{\sigma}'' = \sigma'' \\ \overline{C}_2 = C_2 \end{array}\right\} + \begin{array}{c}\text{simultaneous conversion} \\ \text{of all paired labels} \\ \text{into their counterparts}\end{array}$$

$$Aut[G] = \{E, \sigma', \sigma'', C_2, \overline{E}, \overline{\sigma}', \overline{\sigma}'', \overline{C}_2\}$$

A and B), the graph of Fig. 1a with vertices 3 and 5 assigned label A (and all other vertices assigned label B) is equivalent to three other graphs — i.e., those with vertices 4 and 5, 3 and 6, or 4 and 6 labeled with A. The automorphism group of any of these labeled graphs is evidently the identity group. Similarly, if one considers edge labelings with two labels as well (A′ and B′), the labeling where edges I and II are assigned label A′ (and all other edges assigned label B′) is equivalent only to the labeling where edges III and IV are assigned label A′. The automorphism groups of the two possible edge-labeled graphs[11] consist of two permutations, i.e., those corresponding to the symmetry operations E and σ''.

This example with unpaired labels was simple. However, as was noted above, graphs containing paired labels are also used in reaction design problems, since reversal of the reaction direction results in simultaneous conversion of all such labels into their counterparts. Recalling that direct and inverse processes are regarded as the same interconversion in the Formal–Logical Approach,[1,5] one can expect that the presence of paired labels would result in higher cardinalities of the sets consisting of equivalent labelings and a smaller total number of these sets. The orders of the symmetry groups whose operations convert each of graph labelings into themselves can also increase.

Let us examine some permutation representing the symmetry of a given unlabeled graph G. If we consider the corresponding interconversions of the vertex- and/or edge-labeled graphs to be generated (G_{TOP}, G_{SEQ}, G_{REQ}), we may distinguish two possibilities: either (a) vertices and edges of resultant graphs are just moved according to this permutation or (b) not only they are moved, but, in addition, all paired labels are simultaneously replaced by their counterparts, thus reflecting reversal of the reaction direction. If a (labeled or unlabeled) graph is converted into itself under a transformation of the first or second type, the permutation in question is called a (+)- or (−)-automorphism of this graph, respectively.

Actually, the above definition of a (+)-automorphism means just the same as the common definition of an automorphism; however, the notion of a (−)-automorphism is somewhat unusual and, in fact, rather artificial for graphs without paired labels. Indeed, if a graph contains no paired labels, then the action of all its (−)-automorphisms coincides with the action of its (+)-automorphisms. For example, in the case of graph G in Fig. 1a, its four (−)-automorphisms (corresponding to the symmetry operations marked with overbars in Fig. 1b) are indentical to the relevant (+)-automorphisms. However, as soon as any vertex or edge of this graph is assigned a paired label, the (+)- and (−)-automorphisms of the resultant labeled graph are immediately distinguished from each other.

It is important that, just as the set of all (+)-automorphisms forms a group, the whole set of (+)- and (−)-automorphisms of any labeled or unlabeled graph also forms a group. This latter group was called[5] the *expanded automorphism group* of this graph. According to the above paragraph, the expanded automorphism group of any graph without paired labels is actually an artificial construction, the so-called *action*[12] of the group. Thus, the expanded group $Aut[G]$ (represented in Fig. 1b by four nonoverbarred and four overbarred symmetry operations) moves vertices of graph G in just the same way as two copies of the "normal" group $Aut(G)$ do; the orbits (i.e., the equivalence classes of vertices) are also identical for both groups.

3. EQUIVALENCE RELATIONS BETWEEN LABELED GRAPHS

On the basis of the above considerations, we can directly draw three important conclusions concerning equivalence relations between vertex- and/or edge-labeled graphs:

- For each resultant graph at any generation stage, all labeled graphs equivalent to it should be recognized and then deleted from the complete list of resultant graphs. As a result, the remainder of the complete list will consist only of pairwise nonequivalent labeled graphs, which represent the solution to the labeling problem for this stage.
- To recognize the equivalence classes of resultant graphs (G_{TOP}, G_{SEQ}, or G_{REQ}), the expanded automorphism groups of the starting graphs (G, G_{TOP}, or G_{SEQ}, respectively) should be examined. These groups contain all information on the symmetry of starting graphs and also on the interconvertibility of labels used at previous generation stages.
- Formally representing (+)- and (−)-automorphisms from the expanded groups of starting graphs by operations from some symmetry groups (without and with overbars, see above), one can easily visualize interconversions of equivalent vertex- or edge-labeled resultant graphs. In this section, we will treat these symmetry operations as if they were actual (+)- and (−)-automorphisms. The purpose of this oversimplification is to make explanations as clear as possible — however, one should always remember that all graph automorphism groups are actually formed of permutations but not of corresponding operations of any kind.

It is evident that the expanded automorphism groups can be successively constructed for graphs G_{TOP}, then G_{SEQ}, and finally G_{REQ}. For this purpose, one should select those symmetry operations from the expanded group of the starting graph ($Aut[G]$ at the first stage, $Aut[G_{TOP}]$ at the second, and $Aut[G_{SEQ}]$ at the third) that convert the relevant resultant labeled graph into itself. Hence, the expanded group of any resultant graph is either a subgroup of the expanded automorphism group of the starting graph or identical to it.

The "normal" automorphism groups $Aut(G_{TOP})$, $Aut(G_{SEQ})$, and $Aut(G_{REQ})$ can be similarly constructed starting from $Aut(G)$; these groups consist of (+)-automorphisms and are subgroups of corresponding expanded groups or coincide with them.

Let us consider some specific situations and details for separate generation stages.

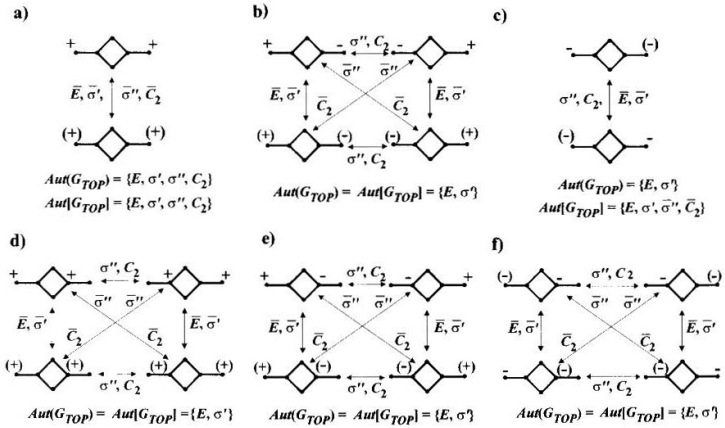
Signed topology identifiers. The vertex-labeled graphs G_{TOP} constructed from a given graph G must be nonequivalent with respect to operations from the expanded group $Aut[G]$. The paired labels actually assigned to the two signed centers can be identical (e.g., "+" and "+"); opposite (e.g., "−" and "(−)"); or different — that is, neither identical nor opposite (e.g., "+" and "−"). In addition, the vertices to which these labels are assigned can belong either to the same equivalence class or to different classes.

As a result, $3 \cdot 2 = 6$ situations are possible at the first generation stage; they are illustrated in Figs. 2a–f for some graphs G_{TOP} corresponding to the unsigned topology identifying graph G of Fig. 1a. Interconversions of all labeled graphs belonging to the same equivalence class (from which a single graph should be selected) are explicitly shown; the symmetry operations from group $Aut[G]$ that convert any resultant graph G_{TOP} into an equivalent one are specified in Figs. 2a–f at the relevant two-sided arrows. The remaining

operations (the same for all equivalent graphs) represent permutation from the "normal" and expanded automorphism groups of G_{TOP}. Evidently, in all the six possible cases, $Aut(G_{TOP}) \subseteq Aut[G_{TOP}]$, $Aut(G_{TOP}) \subseteq Aut(G)$, and $Aut[G_{TOP}] \subseteq Aut[G]$. It is also evident that the expanded groups of graphs G_{TOP} and G are identical if and only if G_{TOP} contains no signed vertices, that is, if $G_{TOP} \equiv G$.

The number of vertex labelings in each equivalence class can be obtained by dividing the order of group $Aut[G]$ (equal to 8 in our example, see Fig. 1b) by the order of the corresponding group $Aut[G_{TOP}]$; see note 11 for explanation. In the above examples, this number equals either 2 (Figs. 2a,c) or 4 (Figs. 2b,d–f).

Figure 2: The equivalence classes of six preselected graphs G_{TOP} and the corresponding automorphism groups. The signed vertices belong to (a, b, c) the same orbit and (d, e, f) different orbits of $Aut(G)$; labels at these vertices are (a, d) identical, (b, e) different, and (c, f) opposite.



Examination of Figs. 2a–f shows that $Aut(G_{TOP}) = Aut[G_{TOP}]$ in most cases — namely, if the two signed vertices belong to different orbits of $Aut(G)$ **or** belong to the same orbit but bear identical or different sign labels. In these cases, the expanded group $Aut[G_{TOP}]$ does not contain (−)-automorphisms associated with reversal of the reaction direction (see above). Such graphs G_{TOP} cannot produce degenerate bond redistributions[13] at subsequent stages of the generation process. On the contrary, one can easily see that nonidentity of groups $Aut(G_{TOP})$ and $Aut[G_{TOP}]$ can be observed if and only if the two signed vertices belong to the same orbit of $Aut(G)$ **and** are assigned opposite labels (see Fig. 2c). The total number of (−)-automorphisms is necessarily equal to the number of (+)-automorphisms in this case.[14]

Symbolic equations. The resultant edge-labeled graphs G_{SEQ} constructed from a given graph G_{TOP} must be nonequivalent with respect to operations from the expanded group

$Aut[G_{TOP}]$. Since paired labels (such as 0/1 and 1/0, 0/2 and 2/0, etc.) are assigned at this stage to all edges of the given topology identifying graph, unsigned or signed ($G_{TOP} \equiv G$ or $G_{TOP} \not\equiv G$, respectively), three situations are possible:

(a) $Aut(G_{TOP}) \subset Aut[G_{TOP}] = Aut[G]$
(unsigned topology identifier; there are (−)-automorphisms in its expanded group);

(b) $Aut(G_{TOP}) = Aut[G_{TOP}] \subset Aut[G]$
(signed topology identifier; there are no (−)-automorphisms in its expanded group);

(c) $Aut(G_{TOP}) \subset Aut[G_{TOP}] \subset Aut[G]$
(signed topology identifier; there are (−)-automorphisms in its expanded group).

These situations are exemplified in Figs. 3a,c,e for arbitrary edge-labeled graphs G_{SEQ} corresponding to the topology identifiers of Figs. 1a, 2b, and 2c, respectively. For each G_{SEQ}, the corresponding symbolic equation is explicitly shown (Figs. 3b,d,f), as well as all equivalent graphs obtained by action of relevant operations from $Aut[G_{TOP}]$. To solve the second-stage labeling problem, one should again, just as at the first stage, select a single representative graph G_{SEQ} (e.g., the leftmost one in each of Figs. 3a,c,e) from each equivalence class of edge-labeled graphs.

The symmetry operations from group $Aut[G_{TOP}]$ that convert each edge-labeled graph G_{SEQ} into itself represent its expanded automorphism group $Aut[G_{SEQ}]$, which reflects the full symmetry associated with the corresponding symbolic equation. The "normal" groups $Aut(G_{SEQ})$ can be similarly obtained from groups $Aut(G_{TOP})$. These "normal" groups are represented by those operations from the expanded groups $Aut[G_{SEQ}]$ that are not marked with overbars in our notation. Surely, $Aut(G_{SEQ}) \subseteq Aut(G_{TOP})$ and $Aut[G_{SEQ}] \subseteq Aut[G_{TOP}]$.

Groups $Aut(G_{SEQ})$ and $Aut[G_{SEQ}]$ can be identical or nonidentical if $Aut[G_{TOP}]$ contains (−)-automorphisms and are necessarily identical if $Aut(G_{TOP}) = Aut[G_{TOP}]$. Among the examples of Fig. 3, nonidentity of groups $Aut(G_{SEQ})$ and $Aut[G_{SEQ}]$ is observed only for the graph in Fig. 3a; the corresponding symbolic equation of Fig. 3b represents a *degenerate bond redistribution*. This fact is not accidental: the degeneracy criteria[5] (actually applied to symbolic equations in the ARGENT-1 program) are really based on comparing the two automorphism groups of graph labelings in question.

Reaction equations. The vertex-labeled graphs G_{REQ} constructed from a given edge-labeled graph G_{SEQ} must be nonequivalent with respect to operations from the expanded group $Aut[G_{SEQ}]$. Note that, although all labels (i.e., atom symbols) used at this generation stage are actually unpaired, the use of the "normal" automorphism group $Aut(G_{SEQ})$ can lead to erroneous results, because paired labels have necessarily been used at previous stages.

To illustrate the solution to this labeling problem, let us consider three vertex-labeled graphs G_{REQ} (Figs. 4a,c,e) that represent two orbits of group $Aut[G_{SEQ}]$ for the graph in Fig. 3a and one orbit of group $Aut[G_{SEQ}]$ for the graph in Fig. 3e.

The symmetry operations associated with the "normal" and expanded groups of graphs G_{REQ} are explicitly shown in Figs. 4a,c,e together with operations from the two automorphism groups of each starting graph G_{SEQ}. In general, the relationships between the groups are similar to those observed at previous stages: $Aut(G_{REQ}) \subseteq Aut(G_{SEQ})$ and $Aut[G_{REQ}] \subseteq Aut[G_{SEQ}]$. Actually, groups $Aut[G_{REQ}]$ and $Aut[G_{SEQ}]$ are identical

Figure 3: (a, c, e) The equivalence classes of three graphs G_{SEQ} and (b, d, f) the symbolic equations corresponding to the leftmost graphs of these equivalence classes. The operations of group $Aut[G_{TOP}]$ that convert an edge-labeled graph into itself and into other equivalent graphs are explicitly shown.

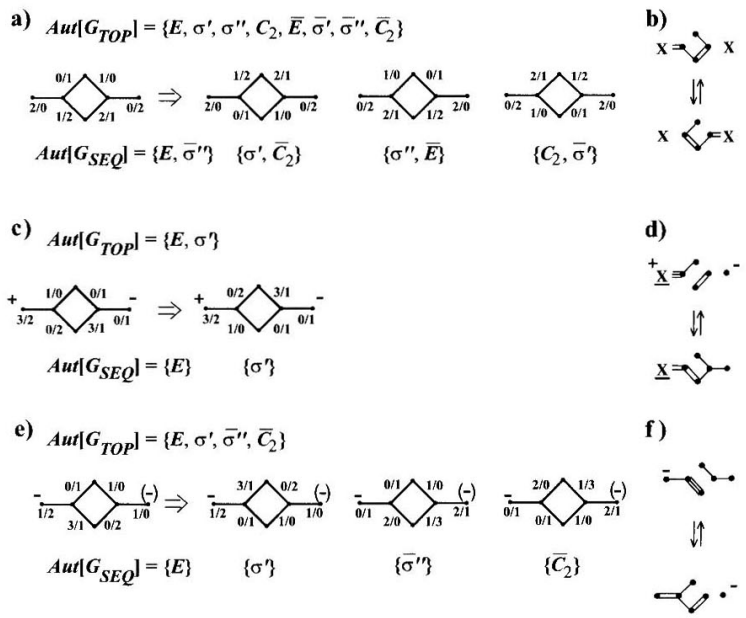
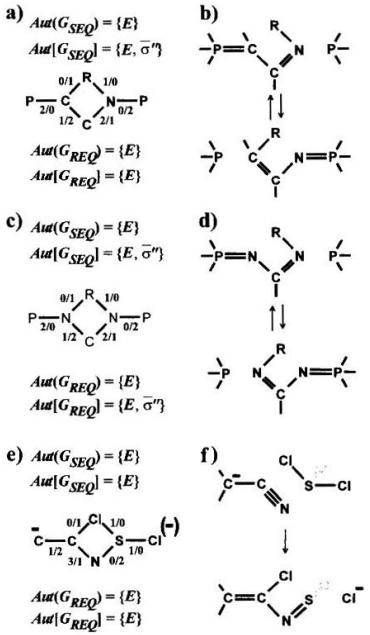


Figure 4: Selected vertex-labeled graphs G_{REQ} presented together with (a, c, e) groups $Aut(G_{SEQ})$, $Aut[G_{SEQ}]$, $Aut(G_{REQ})$, and $Aut[G_{REQ}]$ and with the (b, d, f) corresponding reaction equations, where R represents a carbon (C), hydrogen (H), or any other terminal center of a migrating group. The oxygen atom in part (f) of the figure is not a reaction center but a substituent shown to clarify the essence of the real process.



in two of the three examples under consideration; the only exception is the expanded group of the graph in Fig. 4a, nonidentical to that of the graph in Fig. 3a.

Also similarly to previous stages, the two automorphism groups of any graph G_{REQ} can be either identical (e.g., for the asymmetrically labeled graph in Fig. 4a) or nonidentical (e.g., for the symmetrically labeled graph in Fig. 4c). In the latter case, the presence of (−)-automorphisms in $Aut[G_{REQ}]$ leads to degeneracy[15] in the corresponding reaction equation (compare the hypothetical[16a] processes in Figs. 4b,d). On the contrary, identity of groups $Aut(G_{SEQ})$ and $Aut[G_{SEQ}]$ necessarily results in nondegeneracy of the corresponding interconversions: indeed, $Aut(G_{SEQ}) = Aut[G_{SEQ}]$ implies $Aut(G_{REQ}) = Aut[G_{REQ}]$. A reaction equation corresponding to a nondegenerate process actually investigated[16a,b] may be found in Fig. 4f.

The above description was essentially based on the simple idea to separately treat overbarred and nonoverbarred symmetry operations. Since any pair of such operations represents the same permutation from the group $Aut(G)$, the corresponding (+)- and (−)-automorphisms of vertex- and/or edge-labeled graphs must be somehow distinguished in the rigorous mathematical models of all labeling procedures. Some important notions required for this purpose are briefly considered in the next section; for their more detailed discussion and numerous examples, one can consult sections 5 and 6 of ref 5.

4. THE POWER GROUP FORMALISM

In the preceding section, we considered the symmetries of unlabeled and labeled graphs as permutations from automorphism groups and illustrated them by spatial symmetry operations. However, this description is not completely rigorous, although the reasons seem rather subtle. The problem is that, strictly speaking, automorphisms from the "normal" vertex group (or edge group) of a graph actually permute *graph vertices* (or *edges*) but not *graph labelings*. In fact, a labeling is a function, or, in other words, a mapping from the set of sites (vertices or edges) to the set of labels. That is why permutations from the "normal" automorphism group move sites but do not directly convert a function into another function or into itself. To rigorously consider interconversions of labelings, one should use the group induced[9] by the automorphism group but acting on the set of all possible labeled graphs rather than on the set of sites. Such a group enables one to make allowance not only for permutations of sites but also for interconversions of paired labels.

In general, a group acting on the set of functions is well-known to mathematicians: it is *the power group*[17a] introduced by Harary and Palmer[17b] in 1966.

In order to accurately apply the power group formalism, let us once more recall the succession of labeling stages in generation of graphs G_{REQ} from some preselected parent graph G with the vertex set V and edge set X. At the first and third generation stages, the sites to be labeled are vertices of graphs G and G_{SEQ}, respectively; at the second stage, these sites are edges of graph G_{TOP}.

Let us denote the set of sites to be labeled at some stage by W ($W = V$ or $W = X$, depending on the stage), and their number by p: $W = \{w_1, w_2, \ldots, w_p\}$. Let us also assume that $M = \{m_1, m_2, \ldots, m_q\}$ is the set of q labels that can be assigned to these sites: sign labels, "double" bond labels, or atom labels in the case of graph G, G_{TOP}, or G_{SEQ}, respectively.

Then, the set $F = M^W$ ($|F| = |M|^{|W|} = q^p$) of functions $f = W \rightarrow M$ represents all

labelings possible for the given sets of sites and labels at each stage.[5] For example, if one considers generation of graphs G_{TOP} from the unsigned topology identifying graph G in Fig. 1a, then $p = 6$ (the number of sites, i.e., vertices), $q = 7$ (the total number of sign labels, including the blank one), and $|F| = 7^6 = 117\ 649$. Note that here we disregard equivalence relations, as well as any possible built-in or user-specified selection criteria.

Let us also consider another example for the same topology identifier: in generation of graphs G_{SEQ} from graph $G_{TOP} \equiv G$, having $p = 6$ (the number of sites, i.e., edges) and $q = 12$ (the total number of edge labels used in ARGENT–1, such as 0/1, 1/0, 1/2, 2/1, etc.), one obtains $|F| = 12^6 = 2\ 985\ 984$.

However, as has already been noted, many of these q^p labelings are equivalent, and the numbers of nonequivalent labelings are much smaller. Therefore, one must define such an induced group acting on set F that each orbit of this group would form a class of equivalent labelings, and labelings belonging to different orbits (classes) would be nonequivalent. Then, the labeling process at each stage would be reduced to the combinatorial problem of finding a system of orbit representatives (a *transversal*) of this group on set F.

As was shown above, the equivalence of labelings depends on the symmetry group acting on set W and on the interconvertibility of labels in set M. To consider the first of these two aspects, let us assume that a permutation group $H = \{h_1, h_2, \ldots, h_n\}$ of order n acts on set W and describes the symmetry of the starting graph for the current generation stage. Different groups H appear at the first, second, and third generation stages, but permutations in all these groups can be associated with some symmetries of the parent graph G (see above). For definiteness, permutations from any group H should be numbered in a unified way; in this paper and in the ARGENT–1 program, the numbering system satisfies the following two rules: (1) the first permutation h_1 from any group H is the identity one, and (2) if (−)-automorphisms are present, they form the second "half" of the group; in this case, n is even, permutations $h_1, h_2, \ldots, h_{n/2}$ are (+)-automorphisms, and permutations $h_{n/2+1}, h_{n/2+2}, \ldots, h_n$ are (−)-automorphisms.

To make allowance for interconvertibility of labels, let us assume that the label set M is ordered in such a way that it may be subdivided into three disjoint subsets: $M^0 = \{m_1, m_2, \ldots, m_{q-2r}\}$, $M^- = \{m_{q-2r+1}, m_{q-2r+2}, \ldots, m_{q-r}\}$, and $M^+ = \{m_{q-r+1}, m_{q-r+2}, \ldots, m_q\}$, with M^0 consisting of $q - 2r$ unpaired labels, and M^- and M^+ ($|M^-| = |M^+|$) being the sets of r paired labels. Note that each label m_j ($q - 2r < j \leq q - r$) from M^- must uniquely correspond to its counterpart $m_{2q-2r-j+1}$ from M^+ and vice versa. In this situation, label interconversions can be formally described by introducing the symmetrical group $S_2 = \{s_1, s_2\}$ acting on M: in this group, s_1 is the identity permutation, and s_2 is the permutation that converts all labels from M^- into their counterparts from M^+ (and vice versa) and all labels from M^0 into themselves. This notation will be preserved throughout this paper.

Note that M cannot be empty but any of its subsets can: $|M^0| = 0$ at the second stage (all edge labels are paired), and $|M^-| = |M^+| = 0$ at the third stage (all atom symbols are unpaired labels).

After the two permutation groups (i.e., H and S_2) and the sets on which they act (W and M, respectively) are specified, the power group formalism[17a] can be actually applied to rigorous description of reaction design problems. For this purpose, let us construct the induced permutation group $\Gamma = S_2^H$ whose elements γ convert any function f from the set $F = M^W$ (and the corresponding graph labeling) into some function f' also belonging to F (into an equivalent labeling, respectively). The permutations γ from the power group

Γ can be written down as ordered pairs (h_k, s_l), where $h_k \in H$ $(k = 1, 2, \ldots, n)$ and $s_l \in S_2$ $(l = 1, 2)$. The action of any such pair on some function f means that all sites of set W are initially moved by the permutation h_k, and their labels from set M are then converted (into themselves or counterpart labels) by the permutation s_l.

The power group Γ of order $2n$ is evidently applicable to description of any one-stage generation problem involving paired labels. However, in the case of multistage problems, certain well-defined "smaller" groups (subgroups of the group Γ) must be used at some generation stages. Two such subgroups, denoted as Γ' and Γ'', are described in the next section of this paper; each of them consists of n pairs (h_k, s_l) out of $2n$ possible pairs.

5. MATHEMATICAL MODELS OF GENERATION STAGES

In order to describe the exact mathematical models of reaction design problems, let us recall some results of section 3. In that section, we demonstrated that equivalence relations between resultant vertex- and edge-labeled graphs actually depend on the presence or absence of (a) specific symmetry operations representing $(-)$-automorphisms of the starting graph in question and (b) paired labels at any generation stage. That is why the use of the power group formalism must be separately considered for each generation stage and for each possible situation.

The necessary comments to individual generation stages are given below. For reference, see Table 1, which contains the representation of the overall results in a compact unified form.

I. The first stage is generation of graphs G_{TOP} from G. Here, the site set is $W_1 = V$, and the label set is $M_1 = M_1^0 \cup M_1^- \cup M_1^+$ with $M_1^0 = \{\text{blank label}\}$, $M_1^- = \{+, -, \cdot\}$, and $M_1^+ = \{(+), (-), (\cdot)\}$ (see column I in Table 1). The group defining the symmetry of graph G is its "normal" vertex automorphism group: $H_1 = Aut(G)$, $|H_1| = n_1$. As is clear from the above discussion, two graphs G_{TOP} generated from the same graph G are equivalent if one of them can be converted into the other by a permutation γ that is represented by a pair (h, s_1) or (h, s_2), where $h \in H_1$ and $s_1, s_2 \in S_2$. Indeed, any permutation $\gamma = (h, s_1)$ corresponds to some nonoverbarred symmetry operation (such as σ'' or C_2 in Figs. 2b–f) which, in turn, may be regarded as a $(+)$-automorphism of G. Similarly, any pair (h, s_2) uniquely corresponds to an overbarred symmetry operation (e.g., $\overline{E}$ and $\overline{\sigma}'$ in Figs. 2a–f), or, in other words, to a $(-)$-automorphism of G. As a result, equivalent labelings uniquely correspond to orbits of the induced permutation group $\Gamma = S_2^{H_1}$ on the set $F_1 = M_1^V$, which consists of functions $f = V \to M_1$. The order of the power group Γ is evidently equal to $|S_2| \cdot |H_1| = 2n_1$.

II. The second stage is generation of graphs G_{SEQ} from G_{TOP}. The site set is $W_2 = X$ (the edge set of a given graph G_{TOP}); the label set is $M_2 = M_2^- \cup M_2^+$ with M_2^- being the set of "double" edge labels a/b $(a < b)$ and M_2^+ being the set of their counterparts b/a (see columns II(a)–II(c) in Table 1). Evidently, the symmetry of graph G_{TOP} at this stage is characterized by its edge group H_2 induced by the relevant vertex automorphism group. However, since the topology identifying graphs G_{TOP} can be unsigned or signed and the symmetry of signed topology identifiers can be characterized by two kinds of automorphism groups (see section 3), three cases must be considered separately.

Table 1: Mathematical models of reaction design problems. See text for explanation

Stage	I	II(a)	II(b)	II(c)	III
Site set	$W_1 = V$	$W_2 = X$			$W_3 = V$
Permut. group	$H_1 = Aut(G)$	H_2 induced by $Aut(G)$ ($\equiv Aut(G_{TOP})$)	H_2 induced by $Aut(G_{TOP})$	$\widetilde{H_2}$ and H_2 induced by $Aut(G_{TOP})$ & $Aut[G_{TOP}]$, respectively	$H_3 = Aut[G_{SEQ}]$
Label set	$M_1 = M_1^0 \cup M_1^- \cup M_1^+$, $M_1^0 = \{\text{blank label}\}$, $M_1^- = \{+, -, \cdot\}$, $M_1^+ = \{(+), (-), (\cdot)\}$	$M_2 = M_2^- \cup M_2^+$, $M_2^- = \{0/1, 1/2, 2/3, 0/2, 1/3, 0/3\}$ $M_2^+ = \{1/0, 2/1, 3/2, 2/0, 3/1, 3/0\}$			$M_3 = M_3^0 = \{\text{H, O, N, C, P, S, J, }\ldots\}$
Permut. group	$S_2 = \{s_1, s_2\}$	$S_2 = \{s_1, s_2\}$	$E = \{s_1\}$	$S_2 = \{s_1, s_2\}$	$E = \{s_1\}$
Set of functions	$F_1 = M_1^V$	$F_2 = M_2^X$			$F_3 = M_3^V$
Elements of the induced group	pairs (h, s_1), (h, s_2); $h \in H_1$	pairs (h, s_1), (h, s_2); $h \in H_2$	pairs (h, s_1), $h \in H_2$	pairs (h, s_1), $h \in \widetilde{H_2}$ and (h, s_2), $h \in H_2 \setminus \widetilde{H_2}$	pairs (h, s_1), $h \in H_3$
Group type	$\Gamma = S_2^{H_1}$	$\Gamma = S_2^{H_2}$	$\Gamma' = E^{H_2}$	Γ''	$\Gamma' = E^{H_3}$

II(a) Let us assume $G_{TOP} \equiv G$, that is, no signed vertices are present, see the graph in Fig. 1a. Then, the edge group H_2 ($|H_2| = n_2$) is induced by the automorphism group $Aut(G_{TOP}) = Aut(G)$ — see column II(a) in Table 1. As is apparent from Fig. 3a, the leftmost edge-labeled graph G_{SEQ} can be converted into any other equivalent edge-labeled graph either by a pair (h, s_1) corresponding to a nonoverbarred symmetry operation or by a pair (h, s_2) corresponding to an overbarred operation. Here, $h \in H_2$, and s_1, s_2 are the identity and nonidentity permutations from group S_2 acting on M_2. As a result, equivalent labelings uniquely correspond to orbits of the power group $\Gamma = S_2^{H_2}$ ($|\Gamma| = 2n_2$) on the set of functions $F_2 = M_2^X$.

II(b) Let us assume there are two signed vertices in graph G_{TOP} and $Aut(G_{TOP}) = Aut[G_{TOP}]$; that is, the sign labels at the two vertices are not opposite or these vertices belong to different orbits of the group $Aut(G)$ (see section 3 and Figs. 2a,b,d–f). Then, evidently, no overbarred symmetry operations exist and no permutations (h, s_2) convert resultant graphs G_{SEQ} into equivalent ones. Therefore, all permutations interconverting equivalent labelings of graph G_{TOP} are represented by pairs (h, s_1), $h \in H_2$, with H_2 being the edge group induced by $Aut(G_{TOP})$; an example of a symmetry operation corresponding to a (h, s_1) pair is σ' in Fig. 3c. As a result, equivalent labelings form orbits of another power group, $\Gamma' = E^{H_2}$, acting on the set of functions $F_2 = M_2^X$. This group is constructed from group H_2 acting on X and the identity group $E = \{s_1\}$ acting on the label set M_2 — see column II(b) of Table 1. Obviously, Γ' is a subgroup of $\Gamma = S_2^{H_2}$; its order is equal to $n_2 = |H_2|$.

II(c) Let us assume that $Aut(G_{TOP}) \subset Aut[G_{TOP}]$; that is, graph G_{TOP} contains two vertices bearing opposite sign labels and belonging to the same orbit of group $Aut(G)$ (see Fig. 2c). Then there are two different types of permutations that can convert an edge-labeled graph into an equivalent one. First, let us consider an edge permutation h induced by some permutation $h^* \in Aut(G_{TOP})$. Since the (+)-automorphism h^* leaves sign labels unchanged (cf. the nonoverbarred operation σ' in Fig. 3e), h will form a pair only with s_1, the permutation that also preserves bond labels. At the same time, any (−)-automorphism $h^* \in Aut[G_{TOP}] \setminus Aut(G_{TOP})$ converts all sign labels into opposite ones; therefore, the permutation h induced by h^* can produce an equivalent edge-labeled graph only when paired with s_2 (see operations $\overline{\sigma}''$ and $\overline{C}_2$ in Fig. 3e). As a result, any conversion of graph G_{SEQ} into an equivalent one ensures that sign and bond labels are either all preserved or all simultaneously converted into their counterparts.

For the sake of uniformity, let us now consider two edge groups of graph G_{TOP}: group H_2 is induced by $Aut[G_{TOP}]$ ($|H_2| = |Aut[G_{TOP}]| = n_2$), and its subgroup $\widetilde{H}_2$ is induced by $Aut(G_{TOP})$ ($|\widetilde{H}_2| = |Aut(G_{TOP})| = n_2/2$). In this case, the group[18] Γ'' defining the equivalence classes on set $F_2 = M_2^X$ can be represented as a union of two disjoint subsets: one consists of $n_2/2$ pairs (h, s_1) with $h \in \widetilde{H}_2$, whereas the other consists of $n_2/2$ pairs (h, s_2) with $h \in H_2 \setminus \widetilde{H}_2$ — see column II(c) of Table 1. Just as Γ', the Γ'' group is also a subgroup of Γ of the order n_2, but, in contrast to Γ', is not a power group.

III. The third stage is generation of graphs G_{REQ} from G_{SEQ}; examples of nonequivalent graphs G_{REQ} and the corresponding reaction equations can be found in Fig. 4.

At this stage, $W = V$, and $M_3 = \{\text{H, O, N, C, P, S, J, } \ldots\}$.[19] All atom labels are unpaired: $M_3^- = M_3^+ = \varnothing$. The group H_3 acting on the vertex set is always the expanded automorphism group $Aut[G_{SEQ}]$ (with the order $|Aut[G_{SEQ}]| = n_3$). Indeed, two vertex- and edge-labeled graphs G_{REQ} constructed from the same graph G_{SEQ} are equivalent if and only if some permutation $h \in H_3$ paired with the identity permutation s_1 converts one of the graphs into the other. In other words, equivalent labelings, corresponding to functions of the set $F_3 = M_3^V$, are converted into each other by pairs (h, s_1) with $h \in H_3$, see column III in Table 1. The power group consisting of these permutations is $\Gamma' = E^{H_3}$; its order is $|\Gamma'| = n_3$.

Analysis of Table 1 shows that mathematical models of all the generation problems under discussion have much in common: the induced permutation group Γ, as well as its subgroups Γ' and Γ'', partitions the set F of appropriate functions into equivalence classes, each of these classes representing a definite vertex- and/or edge-labeling of the parent graph G. The difference between the mathematical models associated with the three group types (Γ, Γ', and Γ'') is due to the natures of sites and labels used at different generation stages — see the second and fourth rows of Table 1.

Yet another conclusion from Table 1 is that the group H_i acting on set W_i at any generation stage ($i = 1, 2, 3$) coincides with or is induced by some automorphism group of the graph corresponding to the preceding stage. Recalling that any expanded automorphism group can include only permutations from the expanded group of the graph generated at the preceding stage, one can conclude that

$$Aut[G] \supseteq Aut[G_{TOP}] \supseteq Aut[G_{SEQ}] \supseteq Aut[G_{REQ}];$$

that is, the full symmetry of any resultant graph (G_{TOP}, G_{SEQ}, or G_{REQ}) cannot exceed the full symmetry of the corresponding starting graph (G, G_{TOP}, or G_{SEQ}, respectively). The similar relationship between "normal" automorphism groups

$$Aut(G) \supseteq Aut(G_{TOP}) \supseteq Aut(G_{SEQ}) \supseteq Aut(G_{REQ})$$

also holds; note that $Aut[G]$ actually consists of two copies of group $Aut(G)$, see section 2.

Finally, we can also refer to another terminology: graphs G, G_{TOP}, G_{SEQ}, and G_{REQ} may be regarded as "reaction objects" of the zeroth, first, second, and third levels, respectively. In this representation, the whole mathematical construction used here to describe the three-stage generation of all possible vertex- and edge-labeled graphs G_{REQ} starting from a given unlabeled graph G may be viewed as the *Ladder of Reaction Objects*. A very similar construction, i.e., the Ladder of Molecular Objects, was used[20] to describe successive generation of all molecular formulas, constitutional formulas, and stereochemical formulas starting from a given number of atoms involved; this approach also resulted in mathematical formalization of the three fundamental characteristics of organic molecules — their composition, connectivity, and configuration.[15c]

6. ON SOME SIMILARITY BETWEEN REACTION AND STRUCTURAL DESIGN PROBLEMS

In the last section of this paper, we want to demonstrate a parallelism between the above-considered reaction design problems and several structural design problems whose solution

is based on just the same mathematical models. To put it more precisely, we will briefly outline the applicability of the multistage labeling procedures (and of the induced groups Γ, Γ', and Γ'') to description of the generation problems for various "structurally similar" derivatives corresponding to the given "parent" compound.

First of all, many structural design problems require successive construction of more complicated molecular graphs or graph-like objects from less complicated ones; this process is typically terminated when complete sets of the relevant constitutional (structural) formulas are obtained. For example, the generation algorithm suggested in the fundamental paper[4a] on labeling problems starts from so-called vertex-graphs and involves successive production of "cyclic skeletons", then "ciliated skeletons", then "superatoms", and finally "chemical graphs" representing all possible constitutional formulas corresponding to a given molecular (gross) formula.[4b]

Balaban[21a] created his "adamantaneworld" in a similar manner. In his study, all four-vertex general cubic graphs (i.e., regular graphs of degree 3 that can contain multiple edges and/or loops) were constructed at the first stage, and the resulting adamantane isomers were obtained at the second stage by assignment of appropriate edge labels $(CH_2)_k$, $k = 1, 2, \ldots, 6$, to all or some edges of these general graphs. Note that heteroanalogs and/or substituted derivatives of the resultant structures can easily be produced at subsequent labeling stage(s) if supplementary vertex labels representing heteroatoms and/or preselected substituents are additionally used.

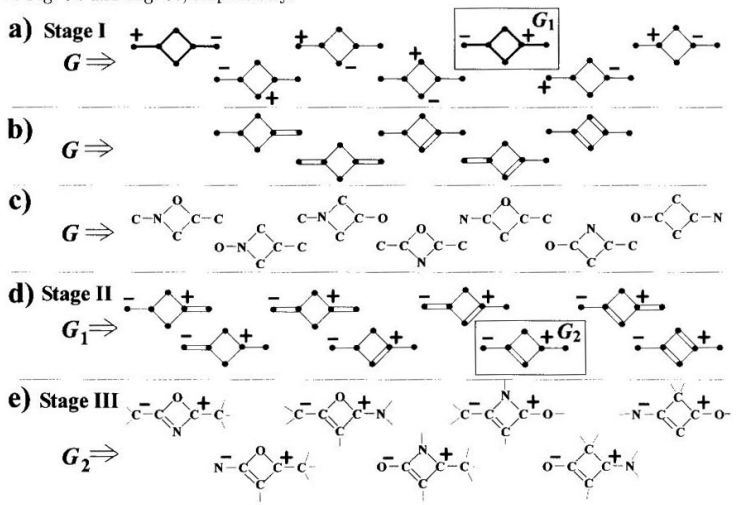
An even more specialized literature example[21b] is associated with multistage construction of complex acyclic fragments starting from a set of given "elementary fragments". Afterwards, at the final stage, the resulting substituents saturate the free valences of some "central fragment", thus producing complete sets of acyclic derivatives of the parent, typically cyclic, skeleton. Note that the structures thus obtained are not necessarily isomeric.

For a more detailed discussion, let us consider an extremely simple and somewhat artificial example of a structure generation problem involving the use of "sign labels", "bond multiplicity labels", and "atom labels" at the first, second, and third labeling stages, respectively. The labels actually chosen are (1) signs "+", "−", and the blank label; (2) double and ordinary bonds; and (3) divalent O, trivalent N, and tetravalent C atoms. Note that these labels are very similar to those used at the three stages of the above-discussed reaction design problems, except that sign and bond labels are unpaired now. Some nonequivalent signed graphs corresponding to dipolar structures, multigraphs containing one or two nonadjacent double edges, and vertex-labeled graphs representing O,N-heteroanalogs of the 1,3-dimethylcyclobutane skeleton can be found in Figs. 5a–c; all these graphs are constructed from the parent graph G of Fig. 1a.

The three-stage labeling procedure is illustrated by graph labelings of Figs. 5a,d,e; for the second and third generation stages, only the labelings obtained from selected graphs G_1 and G_2 are explicitly shown. The final results of Fig. 5e evidently represent hydrogen-depleted constitutional formulas of those unsaturated 1,4-dipolar species with the molecular formula C_4H_5NO which can be regarded as heteroderivatives of the parent 1,3-dimethylcyclobutane structure. (Note that valences of the charged O, N, and C atoms in all multiply labeled graphs of Fig. 5e are equal to 1, 2, and 3, respectively.)

The nonequivalent labeled graphs of Figs. 5a,d,e can evidently be associated with orbits of some induced groups that act on sets F_i ($i = 1, 2, 3$ is the number of the generation stage) consisting of functions from one finite set W_i into another set M_i. Here, W_i is the

Figure 5: Some labeled graphs generated from the parent graph G of Fig. 1a: (a) signed graphs corresponding to 1,3- 1,4-, and 1,5-dipolar structures, (b) multigraphs containing no adjacent double edges, and (c) vertex-labeled graphs that are consistent with the prescribed valences of O, N, C and contain exactly one oxygen, one nitrogen, and four carbons. The signed multigraphs constructed from the selected signed graph G_1 and the constitutional formulas constructed from the selected signed multigraph G_2 are depicted in Fig. 5d and Fig. 5e, respectively.



set of vertices, edges, and again vertices of the parent graph G for the first, second, and third stage, respectively, and M_i consists of sign labels, bond multiplicity labels, and atom labels for the same stages. The induced groups are, in turn, constructed from "normal" vertex or edge automorphism groups H_i of the appropriate starting graphs. Note that $Aut(G) \supset Aut(G_1) \supset Aut(G_2)$ in this example, and this is the main reason why the numbers of results in Figs. 5d and 5e are different from those in Figs. 5b and 5c, respectively.

However, as was mentioned above, all labels of sets M_i at the three labeling stages are unpaired, and this is the main difference between the above reaction design problems and the majority of typical structural design problems (surely, the additional constraints can also be different). As a result, all groups acting on sets M_i, $i = 1, 2, 3$, are identity groups consisting of a single permutation s_1, and all induced groups whose permutations convert the labeled graphs of Fig. 5a–e into equivalent graphs are power groups of the type $\Gamma' = E^H$; cf. similar groups represented in columns II(b) and III of Table 1.

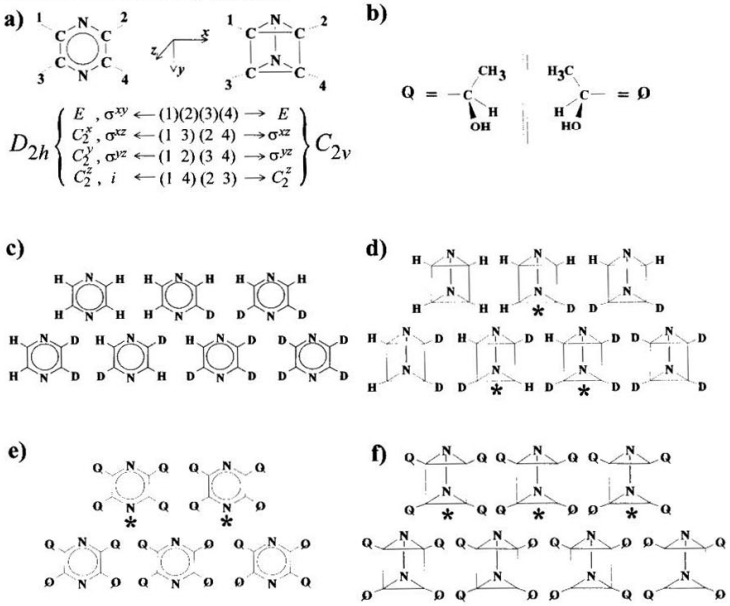
On the other hand, we can still demonstrate the applicability of groups Γ and Γ'' to some structural design problems. For this purpose, let us consider one-stage[22] labeling procedures aimed at generation of substituted derivatives of planar pyrazine and nonplanar diazaprismane (1,4-diazatetracyclo[2.2.0.0]^{2,6}.0^{3,5}]hexane) structures; both are represented in Fig. 6a. In contrast to all the above examples, let us make explicit allowance for the stereochemistry[23,24] of the parent as well as resultant structures. Consideration of stereochemistry means that the labeling problem in question is associated with some embedding of the parent pyrazine or diazaprismane graph in the 3D space rather then with the graph itself. In this situation, some symmetry group of the "spatial" (embedded) graph is responsible for equivalence relations between resultant labelings.

As is well known since Lunn & Senior[25a] and Pólya,[25b] application of the rotation group produces all *stereoisomeric derivatives* of the starting spatial graph, whereas application of the rotation–reflection group results in generation of all *achiral derivatives* plus representatives of all *pairs of chiral derivatives*; different enantiomers are not distinguished in this case. At the same time, as was mentioned above, the automorphism groups of graphs themselves (in the mathematical sense) yield labelings that represent constitutionally isomeric derivatives of the parent organic structure. For more recent investigations on the interrelationship between the above three groups, see ref 25c.

Here we do not distinguish between different enantiomers of substituted derivatives, and hence, we consider the rotation–reflection groups whose operations permute the four numbered free valences of the planar pyrazine and nonplanar diazaprismane skeletons. As is apparent from Fig. 6a, these groups consist of eight and four symmetry operations, respectively. On the other hand, the permutation group[26] associated with the graphs under discussion is the same in these two examples; all the four permutations from this group are also shown in Fig. 6a.

Let us consider the simplest labeling problem consisting in assignment of only two different labels (substituents) to the four sites (free valences).[27] Moreover, let us assume that the labels are either both unpaired (e.g., H and D) or both paired; in the latter case, the symbols Q and Ǫ denote the two enantiomeric forms of some chiral substituent. An example of such a substituent (–CH(OH)CH$_3$, the α-hydroxyethyl group) is represented in Fig. 6b; it is evident that the two enantiomeric forms of the α-hydroxyethyl group cannot be superimposed in the 3D space but are interconvertible under reflection in a mirror plane.

Figure 6: The one-stage generation problem for substituted derivatives of planar (pyrazine) and non-planar (diazaprismane) structures: (a) the starting graphs G with numbered "free valences" to be labeled; the four permutations from group H acting on the set of free valences and the corresponding spatial symmetries are also indicated; (b) an example of paired labels (Q and Ǫ) interconverted by improper symmetry operations; (c, d) the complete lists of substituted derivatives with two unpaired labels H and D — orbits of the induced group Γ'; (e, f) the complete lists of substituted derivatives with two paired labels Q and Ǫ — orbits of the induced groups Γ and Γ'', respectively. Chiral resultant structures are marked by asterisks.



The complete lists of deuterium-substituted pyrazines and diazaprismanes are presented in Figs. 6c,d; their examination reveals the existence of a one-to-one correspondence between the derivatives of these two parent structures. The corresponding complete lists of Q,Q-derivatives are shown in Figs. 6e,f, and, apparently, they represent just another situation: the total numbers of diastereomers (non-enantiomeric stereoisomers) are different, and the two extra substitution products of Fig. 6f — i.e., the third and seventh ones — have no analogs among the Q,Q-substituted pyrazine structures.

To explain these results, let us explicitly construct the mathematical models of the above labeling problems. Evidently, $|W| = 4$ and $|M| = 2$ in all cases under consideration. Hence, the set $F = M^W$ consists of $2^4 = 16$ functions $f = W \rightarrow M$ that uniquely correspond to all possible labelings. In the case of the unpaired labels H and D, the group acting on M is the identity group $E = \{s_1\}$, and therefore the induced group interconverting labelings (i.e., functions f) of Figs. 6c,d is the power group $\Gamma' = E^H$, with H being the group of four permutations shown in Fig. 6a. Note that, in the pyrazine case, two symmetry operations corresponding to the same permutation $h \in H$ (such as the identity operation E and σ^{xy}, C_2^x and σ^{xz}, etc.) can be associated with two identical pairs (h, s_1); surely, only one of these pairs is taken into account.

In the case of paired labels Q and Q, the group acting on the set $M = \{Q, Q\}$ is $S_2 = \{s_1, s_2\}$, where the nonidentity permutation s_2 interconverts Q and Q. In this situation, the four proper symmetry operations associated with the pyrazine skeleton graph (E, C_2^x, C_2^y, and C_2^z, see the left-hand part of Fig. 6a) correspond to pairs (h, s_1), and the four improper symmetry operations (σ^{xy}, σ^{xz}, σ^{yz}, and i) correspond to pairs (h, s_2). Thus, each permutation $h \in H$ forms two nonidentical permutations γ, and the resultant induced group is the power group $\Gamma = S_2^H$.

Finally, in the case of diazaprismane derivatives, two out of four permutations h (those corresponding to the proper symmetry operations E and C_2^z) are paired with s_1; these permutations form a subgroup $\widetilde{H}$ of group H. The other two permutations, i.e., those of the set $H \setminus \widetilde{H}$, evidently correspond to reflections in the σ^{xz} and σ^{yz} planes and hence form pairs (h, s_2). Accordingly, the seven nonequivalent labelings of Fig. 6f represent orbits of the subgroup Γ'' on set F.

Summarizing the above discussion, we can state that the generation problem for substituted derivatives of a given organic structure can be formalized by means of an induced group belonging to any of the above three types. The Γ' group appears if the label set consists only of unpaired labels, and the Γ and Γ'' groups are used if paired labels are also present;[28] these two groups correspond to planar and nonplanar spatial embeddings of the starting graph, respectively. Surely, mathematical models of some other structural design problems (although also reducible to labeling problems, see note 29) are different as to their essence. Consideration of these problems lies outside the scope of this paper.

Returning to the reaction design and ARGENT–1 program, we can state that the above mathematical models form the basis for solving two further problems: those of analytical and constructive enumeration.

- Analytical enumeration, or counting, consists in calculating the numbers of orbits that represent permissible labeled graphs G_{TOP}, G_{SEQ}, and G_{REQ}, with allowance for all the built-in selection criteria as formulated in section 5 of the preceding paper.[1] The corresponding techniques[30a] are actually incorporated into the ARGENT–1 software; they enable one to estimate the maximal numbers of possible resultant

graphs prior to their generation by the program.

- Constructive enumeration, or generation, consists in recognizing orbits of the relevant group (Γ, Γ', or Γ'') on the set of labelings and selecting a single representative from each orbit. An extremely efficient combinatorial algorithm was elaborated for solving this problem at any generation stage for both reaction and structural design problems.[30b]

A detailed consideration of the analytical and constructive enumeration techniques applicable to the above-formulated design problems is the subject of two forthcoming papers in this series.[30a,b] A thorough discussion of user-specified selection criteria and their implementation in the ARGENT–1 program are to be published afterwards.[30c]

ACKNOWLEDGMENT

Partial financial support of this work by the Russian Foundation for Basic Research (Grant No. 00-03-32778a) is gratefully acknowledged. The authors are very grateful to Prof. A.T. Balaban for valuable suggestions.

REFERENCES AND NOTES

1. Zefirov, N.S.; Tratch, S.S.; Molchanova, M.S. Previous paper in this issue.

2. Some authors (e.g., see ref 3a) use the term "coloring" in the same sense as "labeling". The notion of a combinatorial object is preferred by some mathematicians;[3b] a brief summary of applications of combinatorial objects to various problems of structural design, reaction design, and also of stereochemistry was presented[3c] by one of the authors as the lecture course at the 11th Dubrovnik International Conference.

3. (a) Balasubramanian, K. *Theor. Chim. Acta* **1988**, *74*, 111–122. (b) Faradgev, I.A. In *Algorithmic Investigations in Combinatorics*; Nauka Press: Moscow, 1978; pp 3–11. (c) Tratch, S.S. In *MATH/CHEM/COMP'96, Book of Abstracts*; Inter-University Centre: Dubrovnik, 1996.

4. (a) Masinter, L.M.; Sridharan, N.S.; et al. *J. Am. Chem. Soc.* **1974**, *96*, 7714–7723. (b) Masinter, L.M.; Sridharan, N.S.; et al. *Ibid.*; pp 7702–7714.

5. Tratch, S.S.; Zefirov, N.S. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 331–348.

6. (a) Harary, F. *Graph Theory*; Addison–Wesley: Reading, MA, 1969. (b) Harary, F.; Palmer, E.M. *Graphical Enumeration*; Academic Press: New York, 1973.

7. In the general case, there is no one-to-one correspondence between graph automorphisms and spatial symmetry operations. In some instances, graph automorphism groups are supergroups of all point groups corresponding to embeddings of a graph in the 2D or 3D space. (For example, the group of the ethane graph, with all hydrogens considered, consists of $2 \cdot 6^2 = 72$ permutations, whereas the point groups D_{3d} and D_{3h} of the most symmetrical ethane conformations consist of 12 operations.) On the other hand, for a planar but nonlinear molecule, any graph automorphism may

be put into correspondence with two symmetry operations in the 3D space. That is why automorphisms of the graph in Fig. 1a are associated here with operations from group C_{2v} rather than with those from its supergroup D_{2h}; the latter point symmetry group consists of all 8 operations in the 3D space that convert the planar embedding of the graph into itself.

8. Permutations from the automorphism group or other groups are commonly represented in the cyclic notation (e.g., see Fig. 1a). Cycles are disjoint subsets of the set of permuted elements; their union is the whole set. In any particular cycle of some permutation, each element but the last one moves to the next element of this cycle, and the last element moves to the first one. For example, the permutation $(1)(2)(3,4)(5)(6)$ of the vertex set means that vertex 3 moves to vertex 4 and vice versa; these two vertices form a cycle of length 2. A cycle of length 1 indicates that the corresponding vertex (1, 2, 5, or 6 in this example) remains unmoved. Cycles of length 1 are often omitted from the notation for the sake of brevity.

9. (a) The induced group is constructed from one or several other groups and acts on the set that is, in turn, constructed from the sets on which the original groups act. Various induced groups are used by mathematicians in formulation of many discrete structures; a high-level monograph[9b] on the theory and applications of such structures was recently published. (b) Kerber, A. *Algebraic Combinatorics Via Finite Group Actions.* BI-Wiss.-Verl.: Mannheim, 1991.

10. For any connected graph with three or more vertices, the vertex and edge automorphism groups are isomorphic. For other graphs, the necessary and sufficient conditions of isomorphism are defined by Theorem 14.1 in Harary's handbook.[6a]

11. It is well known[6] that the length of orbit Y for any permutation group A — in other words, the number of elements in the corresponding equivalence class — is equal to the index of the subgroup $A(y)$ that stabilizes (moves to itself) some element $y \in Y$: $|Y| = |A|/|A(y)|$. That is why the number of labeled graphs in each equivalence class considered here is equal to the order of the "large" group (which corresponds to some starting graph) divided by the order of the "small" automorphism group (i.e., the group of the resultant vertex- or edge-labeled graph).

12. (a) Some algebraists (e.g., see ref 12b) distinguish between usual permutation groups and actions of abstract groups on appropriate sets; the latter notion is applied in situations where two or more copies of the same group permute elements of the same set. (b) Kaluzhnin, L.A.; Sushtchanski, V.I.; Ustimenko, V.A. *Kibernetika* **1982**, 83–94.

13. (a) In our latest investigations,[13b] we found that three main types of degeneracy are theoretically possible: (1) the degeneracy that appears due to the presence of $(-)$-automorphisms in the expanded groups of edge-labeled graphs (*regular degeneracy*); (2) the degeneracy that is caused by some symmetries of graph G but cannot be associated with any $(-)$-automorphism (*semiregular degeneracy*); and (3) the degeneracy that is completely independent of the symmetry properties of the topology identifying graph G (*irregular degeneracy*). All results discussed in this paper, as well as all degenerate chemical interconversions (isomerizations and intermolecular

processes) actually investigated by organic chemists, refer only to regular degeneracy; that is why the adjective "regular" is not explicitly used in the text. The present version of the ARGENT–1 program makes it possible to perform specialized search for chemically feasible examples of two other, still unprecedented types of degenerate interconversions. Further, we are planning to publish a separate series of papers on the theoretical investigation of degeneracy and the actual results obtained. (b) Tratch, S.S.; Molchanova, M.S.; Zefirov, N.S. In *Molecular Modeling* (Proc. 2nd All-Russian Conf.); Moscow, 2001; p 18.

14. This intuitively clear fact may be briefly explained as follows: (+)-automorphisms of any labeled graph are its isomorphisms onto itself, and (−)-automorphisms are its isomorphisms onto its unique antipode (in other words, onto a labeled graph with all paired labels substituted by opposite ones). Thus, the number of (−)-automorphisms is either zero (if the antipode labeling is non-isomorphic to the original one) or coincides with the number of (+)-automorphisms (in the other case).

15. (a) The degeneracy/nondegeneracy criteria for reactions are quite similar to achirality/chirality criteria for chemical structures.[15b] This conclusion stems from the fact that (−)-automorphisms (responsible for the regular degeneracy) may be considered as analogs of improper symmetry operations. The corresponding combinatorial chirality criteria and their applications to classification of chiral molecules are discussed in refs 15c,d. (b) Tratch, S.S.; Zefirov, N.S. In *Molecular Modeling* (Proc. 1st All-Russian Conf.); Moscow, 1998; p U2. (c) Tratch, S.S. *Zh. Org. Khimii* **1995**, *31*, 1320–1351. (d) Tratch, S.S.; Zefirov, N.S. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 448–464.

16. (a) The intermolecular ylide–imine and imine–imine interconversions of Figs. 4b,d seem to be unprecedented. The α-chlorovinylsulfinylamine formation process of Fig. 4f has really been observed.[16b] The representation of this process depends on the choice of the resonance structure for the cyanomethide anion: the reaction equation of Fig. 4f and that of Chart 11f in ref 16c can both be used for describing this interesting process. (b) Zefirov, N.S.; Chapovskaya, N.K.; et al. *Zh. Org. Khimii* **1975**, *11*, 1981. (c) Tratch, S.S.; Zefirov, N.S. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 349–366.

17. (a) In the general case, the notion of a power group can be described as follows. Let A and B be two finite permutation groups acting on the finite sets X and Y (each consisting of two or more elements), respectively. Then the power group $\Gamma = B^A$ consists of all possible ordered pairs $\gamma = (\alpha, \beta)$, $\alpha \in A$, $\beta \in B$, and acts on the set $F = Y^X$ of all functions (or mappings) $f = X \to Y$ from set X into set Y. (b) Harary, F.; Palmer, E. *J. Combin. Theory* **1966**, *1*, 157–173.

18. To prove that Γ'' is actually a group, one can multiply any two permutations $\gamma_a = (h_a, s_a)$ and $\gamma_b = (h_b, s_b)$; the product, i.e., the permutation $(h_a h_b, s_a s_b)$, is a (+)-automorphism if $s_a = s_b$ ($s_a s_b = s_1$) and a (−)-automorphism if $s_a \neq s_b$ ($s_a s_b = s_2$). This conclusion directly follows from the fact that $h_a h_b$ belongs to $\widetilde{H_2}$ if both h_a and h_b simultaneously belong either to the group $\widetilde{H_2}$ or to the set $H_2 \setminus \widetilde{H_2}$.

19. These seven labels are regarded as the "main" atom labels in ARGENT–1 because they cover the whole range of maximal valences from one to seven; surely, the user may extend this set if needed.

20. (a) Tratch, S.S.; Zefirov, N.S. In *Principles of Symmetry and Systemology in Chemistry*; Moscow University Press: Moscow, 1987; pp 54–86. (b) Tratch, S.S. Doctoral Dissertation; Moscow, 1993; Vol. 2, pp 56–144.

21. (a) Balaban, A.T. *Rev. Roum. Chim.* **1986**, *31*, 795–810. (b) Tratch, S.S.; Lomova, O.A.; et al. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 130–139.

22. One-stage labeling problems are considered here only for the sake of brevity; surely, both starting graphs in Fig. 6a can themselves be produced via a two-stage procedure (skeleton graphs corresponding to benzene isomers and their diazaanalogs can be constructed at the first and second labeling stages, respectively). Note that the extremely strained and probably very unstable diazaprismane structure was chosen here only for convenience, so that the symmetry of both graphs in Fig. 6a would be characterized by the same permutation group.

23. Description of some general stereochemical problems associated with configuration of organic molecules[15c,d,20] is also based on the use of induced groups Γ and their subgroups. More specific formal models were suggested to enumerate geometrically isomeric unbranched polyenes (or ternary 2D chain configurations[20b,21a]) and stereoisomeric unbranched triangulanes (or binary 3D chain configurations[20b,21b]); some other applications are briefly mentioned in refs 24c,d.

24. (a) Tratch, S.S.; Devdariani, R.O.; Zefirov, N.S. *Zh. Org. Khimii* **1990**, *26*, 921–932. (b) Zefirov, N.S.; Kozhushkov, S.I.; et al. *J. Am. Chem. Soc.* **1990**, *112*, 7702–7707. (c) Tratch, S.S. In *MATH/CHEM/COMP'96, Book of Abstracts*; Inter-University Centre: Dubrovnik, 1996. (d) Zefirov, N.S.; Tratch, S.S. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 900–912.

25. (a) Lunn, A.C.; Senior, J.K. *J. Phys. Chem.* **1929**, *33*, 1027–1079. (b) Pólya, G. *Acta Math.* **1937**, *68*, 145–254. (c) Iliev, V.V. *MATCH* **1999**, *40*, 153–186.

26. Although the permutations presented in Fig. 6a actually permute free valences rather then vertices of both skeleton graphs, they uniquely correspond to vertex automorphisms of the graphs under consideration. For the correspondence between graph symmetries and spatial symmetry operations pertaining to planar embeddings of the graphs in the 3D space, see note 7.

27. More complicated labeling problems (e.g., involving one or more unpaired and an even number of paired labels) are based on similar mathematical models; the examples are to be explicitly considered in one of the forthcoming publications in this series. It is important that "saturation" of free valences by appropriate substituents means just the same as replacement of H in the pyrazine or diazaprismane structure by some unpaired or paired label. Note that H itself, if present, must be considered as one of unpaired labels.

28. (a) There are several papers on labeling problems with paired labels also represented by enantiomeric forms of chiral substituents. However, most authors prefer to discuss these problems in terms of double cosets[28b,c] or coset representations;[28d,e] consideration of both theories lies outside the scope of this paper. (b) Hässelbarth, W.; Ruch, E. *Isr. J. Chem.* **1976/77**, *15*, 112–115. (c) Ruch, E.; Klein, D.J. *Theor. Chim. Acta* **1983**, *63*, 447–472. (d) Fujita, S. *J. Math. Chem.* **1990**, *5*, 121–156. (e) Fujita, S. *Symmetry and Combinatorial Enumeration in Chemistry.* Springer-Verl.: Berlin, 1991.

29. (a) For example, generation of all nonisomorphic multigraphs[29b] corresponding to the skeleton of any hydrocarbon structure is reduced to construction of orbit representatives for the group $E^{S_p^{\{2\}}}$; in this case, the labels 0, 1, 2, and 3 are assigned to $p(p-1)/2$ vertex pairs, and $S_p^{\{2\}}$ is the pair group induced by the automorphism group S_p of the empty graph with p vertices. Another nontrivial structural design problem[29c] is associated with specific edge-substituted derivatives of a given organic structure; as far as we know, situations with some or all edge labels represented by *asymmetrical* achiral (such as $-CH_2O-$) or chiral (such as $-CH(CH_3)O-$) homological modules have not yet been analyzed in literature. (b) Kerber, A. *MATCH* **1975**, *1*, 5–10. (c) Tratch, S.S. Unpublished results.

30. (a) Tratch, S.S.; Molchanova, M.S.; Zefirov, N.S. Contribution 3 of this series. To be submitted. (b) Tratch, S.S.; Molchanova, M.S.; Zefirov, N.S. Contribution 4 of this series. To be submitted. (c) Molchanova, M.S.; Tratch, S.S.; Zefirov, N.S. Contributions 5 and 6. In preparation.