

# Systematic Search for New Types of Chemical Interconversions: Mathematical Models and Some Applications<sup>†</sup>

Serge S. Tratch\* and Nikolai S. Zefirov

Chemical Department of the Moscow State University, Leninskiye Gory, Moscow 119899, Russia

Received June 30, 1996

According to *Formal–Logical Approach to Organic Reactions*, the systematical search for new, unprecedented types of chemical interconversions needs the complete sets of *symbolic*, *reaction*, *structural*, and *skeletal* equations to be generated. In this paper, the generation problems are shown to be reduced to graph labeling problems and, more precisely, to constructive enumeration problems for orbit representatives of induced permutation groups acting on certain sets of vertex or edge labelings of well-defined graphs. The suggested combinatorial models make it possible to take into account *formal charges and unshared electrons* (needed to describe ionic, radical, and oxidation–reduction processes) and *unchanged bonds forming cycles* (included in cyclic structures of educts and products and needed to describe, e.g., degenerate rearrangements). The basic ideas of the effective generating algorithms, some results of the corresponding analytical enumeration techniques, and several new reaction types found by computer programs are also discussed.

During the last 25 years, an interest in formal description of chemical interconversions was significantly increased. In the most advanced approaches started in 1971–1973 (I. Ugi and co-workers<sup>1</sup>), in 1975–1976 (N. S. Zefirov, S. S. Tratch<sup>2</sup>), in 1983–1984 (V. Kvasnička and co-workers<sup>3</sup>), and in 1986–1987 (S. Fujita<sup>4</sup>), different concepts and mathematical tools were used in order to unequivocally represent structural changes in the course of organic reactions. In our Formal–Logical Approach to Organic Reactions, several types of specific equations were suggested<sup>2</sup> and are still successfully used in consideration of known as well as unknown types of chemical interconversions. These equations are very similar to traditional equations of organic chemistry, and this fact allows treatment of our approach as a unified language which makes it possible to discuss specific “structural” features of various organic interconversions.

On the other hand, the main purpose of the Formal–Logical Approach was and still is associated with reaction design problems, i.e., with computer-assisted systematic search for new, unprecedented types of chemical reactions. In order to successfully solve the design problems, several rigorous mathematical models, generation algorithms, enumeration techniques, and classification schemes were elaborated (initially for the simplest processes with cyclic electron transfer,<sup>2</sup> then for all possible interconversions of neutral systems,<sup>5,6</sup> and then for a special class of ionic and radicalic reactions<sup>7</sup>). In most published papers, “chemical aspects” of the Formal–Logical Approach (such as introduced terminology, classification problems, lists of generated equations, examples of real and hypothetical processes) were mainly discussed; the mathematical formalization of reaction design problems was treated only in papers,<sup>6,7b</sup> short abstracts,<sup>8</sup> and the dissertation thesis<sup>9</sup> of the first of the present authors.

During the last few years, several new applications of the Formal–Logical Approach were clearly recognized. These applications are associated with (a) new objects (i.e., skeletal and structural equations) to be generated; these objects more closely correspond to what the organic chemist should want to operate with when searching for new types of organic reactions; (b) more general mathematical models to be elaborated; these models are based on new notions (i.e., signed centers, fragments, and topology identifiers) which make it possible to uniformly treat generation problems related to neutral, charged, radical, or ion–radical systems; (c) several specific generation problems to be solved; the systematic search for new degenerate and highly degenerate rearrangements seems to be one of the most interesting and attractive problems.

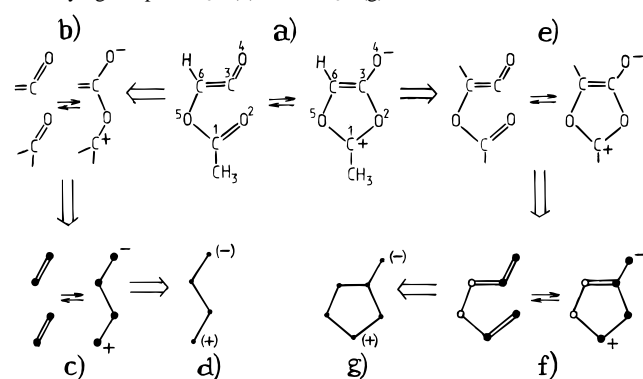
This paper may be considered as an attempt to overview the modern state of the Formal–Logical Approach in a very brief, illustrative manner. The main attention will be focused here on graph-theoretical and combinatorial models for various generation problems; the modern state of classification procedures will be discussed in the following paper in this issue.<sup>10</sup>

## 1. MAIN NOTIONS OF THE FORMAL–LOGICAL APPROACH

In the Formal–Logical Approach, organic reactions are treated as *bond redistributions*<sup>11</sup> between initial and final systems; all molecules (ions, radicals, or ion–radicals) are thought to be represented by definite resonance structures. The redistributed lone pairs and unshared electrons are not explicitly considered; they are taken into account by means of signs “+”, “.”, or “–” on corresponding atoms which are called *signed reaction centers* (or SRCs, for short). All other atoms which change the number and/or multiplicity of their bonds are denoted as unsigned reaction centers or, for brevity, simply *reaction centers* (RCs). It is evident that

<sup>†</sup> This material was presented by the first author at the Fourth International Conference on Chemical Structures. Noordwijkerhout, The Netherlands; June 2–6, 1996.

**Chart 1.** Interrelations between Chemical (a), Reaction (b), Symbolic (c), Skeletal (e), and Structural (f) Equations Are Illustrated for  $\alpha$ -Acetoxyketene Tautomerization Process and Corresponding Topology Identifying Graphs  $G_{TOP}$  (d) and  $G_{ETOP}$  (g)



this representation allows one to describe and classify practically all chemical reactions, but in order to get the reasonable number of results (specific equations generated by computer programs, cf. below), some additional restrictions are needed. For this purpose, only the SRCs with a single sign (i.e., “+”, “-”, or “.” but not “+.”<sup>12</sup>) appearing either in an initial or in a final system (but not in both!) are considered in this paper. The total number of such centers is typically equal to 0 or 2; it can be shown<sup>13</sup> that this number must be necessarily even. In the case of two or more signed centers, the overall transformation may also be considered as *bond-and-sign redistribution*. In our final introductory comment, we stress that *direct and opposite bond (or bond-and-sign) redistributions are not differentiated in the Formal-Logical Approach*. For this reason, only one of two possible equations (related to any nondegenerate process) is actually generated by a computer.

It is evident that the systematic search for new reactions needs not all atoms and bonds to be taken into account. For that reason, all unchanged bonds and “non-RC atoms” or only those of them which form (simple or complex) substituents may be deleted from traditional chemical equations. In the first case, *reaction equations* and, in the second case, *skeletal equations* are produced. In skeletal equations, the pairs of (unsigned or signed) reaction centers can be connected either by “isolated” bonds of unchanged multiplicity or by chains formed out of unchanged bonds and *structural centers*. The structural centers are formally not involved in the bond redistribution but typically form cycles in educt and/or product molecules and hence significantly help an organic chemist to recognize what he would like to consider as a “new type of reaction”. The examples of chemical, reaction, and skeletal equations related to tautomerization of  $\alpha$ -acetoxyketenes<sup>14a</sup> are depicted in Chart 1a,b,e, respectively (for other examples, see ref 10). The atoms 1, 4 are evidently signed centers, 2, 3—(unsigned) reaction centers, and 5, 6—structural centers.

The systematical search for synthetically significant bond redistributions (or bond-and-sign redistributions) evidently does not need the nature of atoms (C, H, N, O, etc.) to be explicitly known. This fact allows one to consider *symbolic and structural equations* in which atom symbols (taken from reaction and skeletal equations, respectively) are substituted by some abstract symbols. In the above example (cf. Chart 1c,f), structural centers are denoted by small circles, while

RCs and SRCs—by heavy points (other designations will be discussed in section 2).

In this paper, we do not consider all details related to structural and skeletal equations. In particular, some applications of these equations do not need all cycles to be explicitly taken into account. This is a case, e.g., for structural and skeletal equations which are used<sup>10</sup> as attributes in our hierarchical classification scheme for chemical reactions; in subsections 4.4 and 4.5 of ref 10, we demonstrate that it is reasonable to disregard some “chemically insignificant” cycles in the equations under consideration. The other important note is that the actual direction of an organic reaction is never reflected in symbolic and structural equations; in these equations, the interconversion of initial and final systems is always treated as a reversible process (cf. Charts 1c,f, 2c,e, etc.). The difference between irreversible reaction, reversible reaction, and resonance transformation is, however, often visualized in other equations in order to facilitate the perception of the corresponding process. The reaction equations of Charts 3g, 5, 6, 9, and 10 (and also skeletal and chemical equations represented in the following paper in this issue<sup>10</sup>) may serve as examples.

All the above objects of the Formal-Logical Approach are actually *equations* but not *graphs*. The graphs corresponding to the equations under discussion can, however, be easily constructed by superposition of left and right parts of symbolic and structural equations. In the “superposed” graphs, two vertices are joined by an edge if a corresponding bond of nonzero multiplicity exists in one or both parts of the equation under consideration. The graphs corresponding to symbolic equations (*topology identifying graphs*,  $G_{TOP}$ ) and to structural equations (*expanded topology identifying graphs*,  $G_{ETOP}$ ) are pictorially represented in Chart 1d,g; these graphs seem to have no analogs in other formal approaches to reaction design problems. Note that in both topology identifiers, the disposition of signs “+”, “-”, and/or “.” in initial and final systems must be explicitly reflected. In the example under discussion, the brackets mean that positive and negative charges appear in the final system.

At this step, four main types of generation problems can be formulated as follows: (a) to construct all possible *signed* topology or expanded topology identifiers ( $G_{TOP}$  or  $G_{ETOP}$ ) for a given unsigned graph  $G$ ; (b) to construct all possible symbolic (or structural) equations corresponding to a given graph  $G_{TOP}$  (or  $G_{ETOP}$ ); (c) to construct all possible reaction (or skeletal) equations corresponding to a given symbolic (or structural) equation; (d) to construct all possible chemical equations corresponding to a given reaction (or skeletal) equation; this problem (which evidently needs all simple as well as complex substituents to be preselected) will not be specifically discussed in this paper.

It should be noted that only two of the mentioned problems (i.e., generation of symbolic equations for a given  $G_{TOP}$  and generation of reaction equations for a given symbolic equation) were needed to be solved within the scope of the previous versions of the Formal-Logical Approach. On the other hand, the solution of all generation problems needs some selection criteria (making possible to disregard chemically incorrect or uninteresting results) to be formulated; most of these criteria are based on numerical and structural characteristics introduced in the next two sections.

## 2. NUMERICAL CHARACTERISTICS AND DESIGNATIONS OF UNSIGNED AND SIGNED REACTION CENTERS

Several numerical characteristics are introduced in the Formal–Logical Approach for RCs and SRCs. The first of these characteristics are *reaction and skeletal numbers*,  $RN$  and  $SN$ , which are equal to formal valencies of the center under discussion in the reaction (or symbolic) and skeletal (or structural) equations, respectively. Reaction and skeletal numbers of the same center must not be necessarily identical in initial and final systems. For this reason, the values of  $RN_{ini}$  and  $RN_{fin}$  ( $SN_{ini}$  and  $SN_{fin}$ ) are always differentiated.

The similar characteristic of any structural center (denoted by a small circle in a structural equation) is just the same for initial and final systems; this characteristic can be considered as “reduced valency” of the center under discussion. The other possible numerical characteristic of the structural center is its “reduced coordination number”, i.e., the number of its bonds in a skeletal or structural equation. The multiplicities of bonds are not taken into account in this case.

The next characteristics of RCs and SRCs, i.e., their degree, rank, and specificity, relate to the both systems simultaneously. The *degree*  $\rho$  of any reaction center denotes the number of its bonds which change their multiplicity in the course of the reaction. The *rank*  $R$  is equal to the total change of multiplicities for all bonds of the center under discussion. It was shown<sup>5a,9</sup> that the values of  $R$  are always positive and are *even for all unsigned centers and odd for all signed centers* (for which a sign “+”, “•”, or “–” appears either in an initial or in final system, cf. above). The final characteristic of a reaction center, its specificity  $r$ , is defined by expression  $r = |RN_{ini} - RN_{fin}|$  (or  $r = |SN_{ini} - SN_{fin}|$ ); just the values of  $r$  (which parities always coincide with parities of rank values) make it possible to classify reaction centers into several types.

The unsigned centers with  $r = 0$  are called *ordinary* and are denoted by heavy points. The centers with specificity values equal to 2, 4, 6, ... are called *specific*, *dispecific* (or doubly specific), *trispecific*, etc., and are designated by  $X$ ,  $X'$ ,  $X''$ , ..., respectively. The situation is somewhat more complex for signed centers ( $r = 1, 3, 5, \dots$ ) because two subtypes are possible for any given value of  $r = 2k + 1$ ,  $k = 0, 1, \dots$ . If in an initial or final system, the SRC with minimal value of  $RN$  is signed, then this center is called *ordinary* (if  $r = 1$ ), *specific* (if  $r = 3$ ), *dispecific* (if  $r = 5$ ), etc., and is denoted by the same symbol as for an unsigned center with specificity equal to  $r - 1$  (i.e., by heavy point if  $r = 1$ , by  $X$  if  $r = 3$ , by  $X'$  if  $r = 5$ , etc.). On the contrary, if the signed center with maximal value of  $RN$  is signed, then it is called *pseudospecific*, *pseudodispecific*, etc., and is designated by the underlined symbol of the corresponding unsigned center with specificity value equal to  $r + 1$  (i.e., by  $\underline{X}$  if  $r = 1$ , by  $\underline{X'}$  if  $r = 3$ , etc.).

The numerical characteristics and designations of centers are exemplified (see the tables in Chart 2) in relation to two equations which describe just the same oxirane formation process.<sup>14b</sup> These equations completely characterize cyclic structures of educt and product molecules and hence are considered here as skeletal equations.<sup>15</sup> The equation of Chart 2a corresponds to interconversion of nonpolar (initial and final) systems. In the other equation, that of Chart 2b,

**Chart 2.** Numerical Characteristics and Designations Are Tabulated for Centers Involved in the Oxirane Formation Process<sup>a</sup>

No.	$RN_{ini}$	$RN_{fin}$	$\rho$	$R$	$r$	Type
1	2	0	1	2	2	$X$
2	2	2	3	4	0	•
3	2	2	2	2	0	•
4	2	2	2	2	0	•
5–9	–	–	–	–	–	○

No.	$RN_{ini}$	$RN_{fin}$	$\rho$	$R$	$r$	Type
1	1	0	1	1	1	$\underline{X}$
2	1	2	3	3	1	•
3	2	2	2	2	0	•
4	2	2	2	2	0	•
5–9	–	–	–	–	–	○

<sup>a</sup> The transformations of nonpolar and dipolar systems are represented by skeletal (a and b) and symbolic (c and e) equations; the corresponding unsigned and signed topology identifiers  $G_{TOP}$  are characterized by the same (unsigned) linear–cyclic contour (d). The reduced coordination numbers and valencies of structural centers 5–9 (all equal to 2) are not explicitly shown in the tables of the chart.

the transformation of a dipolar (due to the dipolar resonance structure of the sulfonium methylenide) system into a nonpolar one is represented. Of the above-mentioned numerical characteristics, only the values of  $SN_{ini}$  and  $SN_{fin}$  are not tabulated in Chart 2; these values are identical with  $RN_{ini}$  and  $RN_{fin}$  for all unsigned and signed centers except  $C^4$  ( $SN_{ini} = SN_{fin} = 4$  for this center). The designations of ordinary ( $C^2$ ,  $O^3$ , and  $C^4$ ;  $r = 0$ ) and specific ( $S^1$ ;  $r = 2$ ) RCs and corresponding designations of ordinary ( $O^3$  and  $C^4$ ;  $r = 0$ ) RCs, ordinary SRC ( $C^2$ ;  $r = 1$ ), and pseudospecific SRC ( $S^1$ ;  $r = 1$ ) are used in symbolic equations Chart 2c,e. In these equations and in topology identifying graphs  $G_{TOP}$  of Chart 2d, all structural centers are evidently omitted.

## 3. STRUCTURAL CHARACTERISTICS OF BOND REDISTRIBUTIONS

Numerous structural characteristics are used in the Formal–Logical Approach in order to form the specific “language” making it possible to qualitatively describe the main features of the structural changes which take place in the course of organic reactions. Of these characteristics, only one, *canonicity*, explicitly depends on the presence of signs “+”, “•”, or “–” in the *symbolic or reaction* equation. All other structural characteristics relate to bonds which change their multiplicity in the course of reactions and hence characterize bond (rather than bond-and-sign) redistributions.

Any bond redistribution is called *canonical* if no signed reaction centers are involved in a process. It is important that all canonical bond redistributions describe neutral processes, while all ionic, radicalic, ion–radicalic, and oxidation–reduction processes (cf. classification scheme in ref 10) are characterized by noncanonical bond redistributions. Similarly, the bond redistribution is called *basic* if no triple bond is present in corresponding *symbolic* (or *reaction*) equation. All nonbasic bond redistributions can

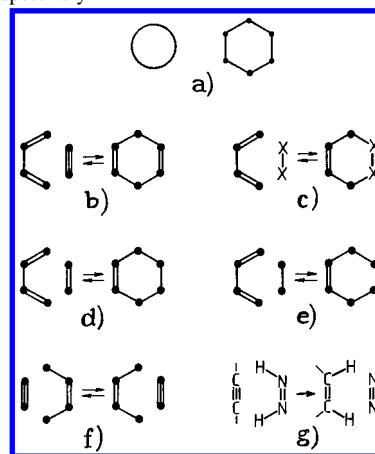
evidently be transformed into the basic ones if in structural formulas of all educts and products, the triple bonds are substituted by ordinary (single) bonds. A simple example is depicted in Chart 3b,c by two symbolic equations which describe just the same [2+4] cycloaddition–cyclofragmentation processes (e.g., forming and fragmentation of cyclohexa-1,4-diene ring system). In the second equation, an acetylene component is represented by  $\alpha$ -dicarbene structure ( $-\dot{C}-\dot{C}-$ ), and this fact results in transformation of a nonbasic process (with no specific RCs, cf. Chart 3b) into a corresponding basic process with two specific centers X.

The bond redistribution is called *trivial*<sup>16</sup> if all bonds in a symbolic (or reaction) equation are formed from zero or completely destroyed and *principal* if no bonds change their multiplicities from 2 to 3 (or vice versa). The nontrivial bond redistributions evidently contain one or more “valence lines” which formally remain unchanged in the course of reaction, while all processes related to derived (nonprincipal) bond redistributions always belong to the same interconversion type as corresponding “principal” processes do. For example, one can easily see that both the derived and principal bond redistributions represented by symbolic equations of Chart 3b,d describe [2+4] cycloaddition–cyclofragmentation reactions. In contrast, the very similar equation of Chart 3e describes six-centered addition–fragmentation processes (such as 1,4-addition of hydrogen or bromine to conjugated dienes). The specific equations which unambiguously characterize interconversion types—such as cycloaddition–cyclofragmentation or addition–fragmentation reactions—are discussed in detail in ref 10 (see section 2 and subsection 4.3 of that paper).

The next important characteristic of the bond redistribution is its *mode*. In a *single* (double or triple) *mode* bond redistribution, all bonds change their multiplicity by 1 (2 or 3) units, while in a *composite mode* bond redistribution some bonds change their multiplicity by 1 and some by 2 units. The *complex mode* redistributions (for which one or more bonds change their multiplicity by 3 units, while all others—by 1 or 2 units) are quite rare in organic chemistry. It should be noted here that canonical, single mode bond redistributions related to cyclic topology identifiers (see the graph of Chart 3a for an example) describe most of known concerted reactions; the corresponding *multicentered processes with cyclic electron transfer* were separately investigated in our early papers.<sup>2</sup> Similarly, the *processes with linear electron transfer*<sup>7</sup> are characterized by noncanonical, single mode bond redistributions related to linear graphs (such as that of Chart 4a); these bond redistributions describe the most important elementary steps of complex chemical interconversions.

Two structural characteristics are associated with degeneracy of bond redistributions and with presence of specific centers. Thus, the bond redistribution is *degenerate* if left and right parts of corresponding symbolic equation are isomorphic in graph-theoretical sense. Note that the notion of degeneracy can also be applied to reaction, structural, skeletal, and chemical equations, but the results must not be necessarily identical (this means that degeneracy also depends on distributions of atom symbols, unchanged bonds, and substituents). The symbolic (cf. Chart 3f) and reaction (cf. Chart 3g) equations can serve as typical examples. It is easy to see that the degeneracy disappears in the reaction process

**Chart 3.** Several Structural Characteristics Are Exemplified for Cyclic, Single Mode Six-Centered Processes Being Characterized by the Same Topology Contour and Identifier (a): Nonbasic Isodesmic (b) and Basic Allodesmic (c) Bond Redistributions Represent the Same Process; Derived (b) and Principal (d) Bond Redistributions Represent Different Processes of the Same Type; Bond Redistributions (b) and (e) Belong to Different Interconversion Types; the Degenerate and Nondegenerate Bond Redistributions Are Exemplified by Symbolic (f) and Reaction (g) Equations, Respectively



of Chart 3g due to an unsymmetrical disposition of symbols C and N in the reaction equation.

In *isodesmic bond redistributions*, only unsigned centers with  $r = 0$  (i.e., ordinary RCs) and signed centers with  $r = 1$  (i.e., ordinary and pseudospecific SRCs) are allowed to be present. It is important that to most of the known (one-step or multistep) organic reactions, isodesmic bond redistributions correspond. The *allodesmic* bond redistributions (i.e., those in which at least one unsigned or signed center with  $r > 1$  is involved) are relatively rare. Of the above-mentioned bond redistributions, only those represented by equations of Charts 2a,c and 3c are allodesmic.

The remaining structural characteristics of the bond redistribution relate to its *topology*, i.e., to the (unsigned or signed) contour formed by bonds which change their multiplicity in the course of the reaction. The contours under consideration define reaction classes (cf. subsection 4.2 in the following paper in this issue<sup>10</sup>) and are naturally classified into linear, bilinear, cyclic, linear–cyclic, etc. It should be noted that all introduced topology characteristics can also be recognized from any graph  $G_{TOP}$  which corresponds to a given topology contour.

The bond redistribution is called *connected* if the contour (and the topology identifying graph  $G_{TOP}$ ) is connected, *fundamental* if no more than three lines start at any point of the contour ( $G_{TOP}$  contains no vertices of degree 4 or more in this case), and *elementary* if canonical, single mode bond redistributions are possible for the topology under discussion.<sup>17</sup> Additionally, the topology identifying contours are classified into *disclosed* (no cycles in the contour and in the corresponding graph, cf., e.g., Chart 4a), *closed* (no endpoints in the contour and vertices of degree 1 in the graph  $G_{TOP}$ , cf. Chart 3a), and *mixed*. In the last case, cycles as well as endpoints are present in the contour under discussion (for an example, see linear–cyclic contour and corresponding unsigned and signed topology identifiers in Chart 2d).

In order to additionally illustrate the application of introduced structural characteristics, let us consider the symbolic and reaction equations of Chart 3f,g; these equa-

tions describe the well-known irreversible reaction of acetylenes with diimide. It is easy to see that the process under discussion is characterized by canonical, nonbasic, nontrivial, derived, isodesmic, single mode bond redistribution. The examination of the corresponding topology identifying contour and graph (cf. Chart 3a) shows that cyclic topology is, in fact, connected, closed, fundamental, and elementary.

In principle, all above structural characteristics are dependent on resonance structures chosen. However, for a particular process (in which some educts and/or products are depicted by different resonance structures), several characteristics can be identical. As a typical example, one can compare symbolic equations of Chart 2c,e (related to nonpolar and dipolar structures of ylid molecule, cf. Chart 2a,b). The first of the equations, that of Chart 2c, evidently represents canonical, allodesmic, composite mode bond redistribution, while the second equation (cf. Chart 2e) represents noncanonical, isodesmic, single mode bond redistribution. On the other hand, both the bond redistributions are basic, nontrivial, principal, and nondegenerate; the characteristics of topology also coincide (linear-cyclic topology is connected, mixed, fundamental, and nonelementary).

At the end of this section, several important theorems of the Formal-Logical Approach are mentioned. These theorems were shown<sup>9</sup> to be valid for *canonical bond redistributions*: (1) only double mode bond redistributions are possible for disclosed topologies; (2) only the vertices of even degree can be present in graphs  $G_{TOP}$  corresponding to elementary topologies; (3) only the cyclic topology is simultaneously connected, fundamental, and elementary; and (4) only the even number of ordinary centers is allowed for cyclic, single mode bond redistributions.

The very simple proofs of the above theorems are based on the fact that only (unsigned) RCs with even rank values are possible for canonical bond redistributions (cf. section 2). Thus, for all disclosed topologies, the multiplicity of any "terminal" bond is increased or decreased by two units because only the value  $R = 2$  is allowed for one of RCs. The fact that all "endpoint" vertices (corresponding to centers of degree 1) can be deleted from acyclic graph  $G_{TOP}$ , proves theorem (1). In order to prove theorem (2), one can note that in the case of canonical, single mode bond redistributions, the odd degree of any vertex in a topology identifying graph results in the (forbidden) odd rank value for a corresponding unsigned reaction center. Similarly, only even degree values which do not exceed 3 are allowed for all vertices of those graphs  $G_{TOP}$  which represent elementary and simultaneously fundamental topologies; the class of connected graphs with all vertices of degree 2 evidently consists of cyclic graphs similar to that of Chart 3a (theorem (3)).

In order to outline the proof of theorem (4), the supplementary graphs should be constructed; the edges of these graphs uniquely correspond to bonds which *increase their multiplicity in the course of any possible reaction*. In the case of canonical, cyclic, single mode bond redistributions, all ordinary RCs evidently relate to vertices of degree 1, while specific RCs—to vertices of degree 0 or 2 in the supplementary graph. But in any graph, the number of odd

degree vertices is always even, and this fact proves theorem (4).

#### 4. GRAPH MODELS OF REACTION DESIGN PROBLEMS

In this section, the main types of generation problems (mentioned in section 1) are shown to be reduced to multistep *graph labeling* problems, i.e., to sequential distribution of suitable labels between vertices or edges of a given graph. It should be noted that graph labelings are commonly used (see, e.g., ref 18) in order to describe generation and enumeration procedures for various derivatives, e.g., heteroanalogues, homologs, or other substitution products of a given organic structure.

**Step 1.** Let the connected graph  $G$  with the vertex set  $V$ ,  $|V| = m$ , and edge set  $Y$ ,  $|Y| = n$ , be given; this graph can be considered as the "parent" in relation to all generation procedures of steps 1–3. The selection of proper graphs  $G$  actually depends on some structural characteristics of symbolic or reaction equations to be generated. For example, this graph must not contain vertices of degree exceeding 3 (or vertices of odd degree) if only processes corresponding to fundamental topologies (or only canonical processes corresponding to elementary topologies) are searched for. Similarly, if symbolic or reaction equations related to closed (or disclosed) topologies are generated, then  $G$  is required to contain no vertices of degree 1 (or cycles).

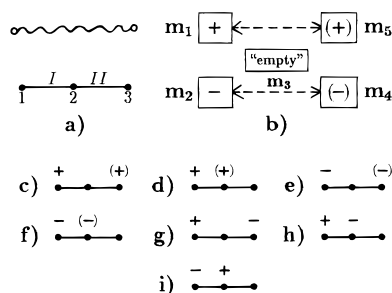
The vertices of the set  $V$  are associated with unsigned reaction centers in all symbolic and reaction equations which can be constructed starting from the graph  $G$ . Thus,  $G \equiv G_{TOP}$  if canonical processes (which contain no SRCs) are searched for. Otherwise, two of  $m$  vertices need to be labeled by symbol "+", "•", or "—" (if the signed center must be present in initial but not final system) or "(+)", "(•)", "(–)" (if the SRC must be present in final but not initial system). It is easy to see that in the general case, there are seven labels (the seventh, "empty" label corresponds to unsigned reaction centers), and the complete list of allowed labelings consists of  $6^2 \cdot \binom{m}{2}$  labeled graphs with  $m - 2$  vertices remaining unlabeled.

In practice, however, the labels chosen depend on the category and subcategory of processes to which generated equations correspond (cf. subsection 4.1 in the following paper in this issue<sup>10</sup>). Thus, in the case of ionic reactions, two vertices of  $G$  corresponding to SRCs must be labeled by "+" and "(+)" (c/c-processes), by "–" and "(–)" (a/a-processes), and by "+" and "–" (ca/n-processes); two pairs of required labels (characterizing positive and negative charges, respectively) are schematically represented in Chart 4b together with an "empty" label. All nonequivalent labelings (corresponding to unsigned linear topology contour and graph  $G$  of Chart 4a) are explicitly demonstrated by signed graphs  $G_{TOP}$  of Chart 4c–i. It is important that the represented list of (signed) linear topology identifiers<sup>19</sup> makes it possible to construct all symbolic and reaction equations related to three-centered ionic processes.

The signed graphs  $G_{ETOP}$  are constructed from a "parent" graph  $G$  in the same manner. In this case, two labeled vertices of the resulting graph represent SRCs, while the remaining  $m - 2$  vertices of  $G_{ETOP}$  correspond to (still indistinguished) unsigned reaction centers and structural centers.



**Chart 4.** Complete List (c–i) of Signed Graphs  $G_{TOP}$  (Corresponding to Three-Centered Ionic Processes) Is Constructed Starting from an Unsigned Graph Which Represents Linear Topology of Bond Redistribution (a) and Two Pairs of “Interchangeable” Labels Plus an “Empty” Label (b)



**Step 2.** Let the graph  $G_{TOP}$  and its symmetry be known. In order to construct the complete list of corresponding symbolic equations, the *edge labelings* of  $G_{TOP}$  need to be sequentially generated. In this case, the labels reflect 12 possible changes of bond multiplicity (from 0 to 1, 2, or 3; from 1 to 0, 2, 3; etc.). For this reason, there exists 12<sup>n</sup> formally possible labelings ( $n$  denotes the number of edges in  $G_{TOP}$ ) of which many—but not all!—labelings correspond to symbolic equations to be generated. In order to select “proper” edge labelings (i.e., those related to symbolic equations), one must take into account the possible values of specificity  $r$  for RCs and SRCs. The values of  $r$  must necessarily be even for unsigned centers and odd for signed centers (cf. section 2); additionally, in the case of isodesmic processes, only the specificity values  $r = 0$  and  $r = 1$  are allowed for all RCs and SRCs in corresponding symbolic equations. The values of  $r$  can be calculated from an expression

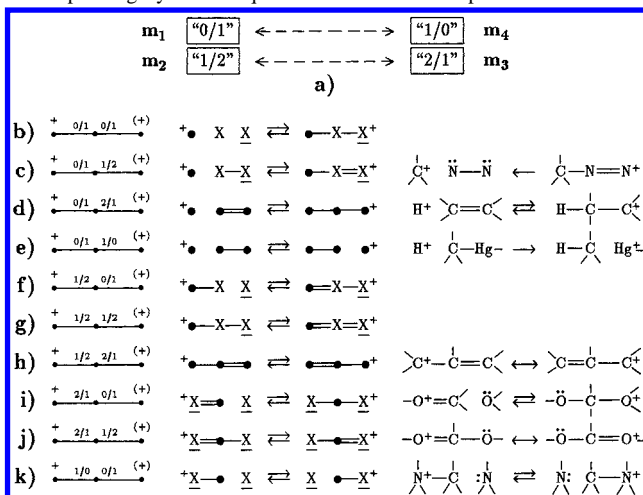
$$r = |RN_{ini} - RN_{fin}| = \left| \sum_{k=1}^{\rho} \nu_{ini}^{(k)} - \sum_{k=1}^{\rho} \nu_{fin}^{(k)} \right|$$

where  $\nu_{ini}^{(k)}$  and  $\nu_{fin}^{(k)}$  denote multiplicities of a  $k$ th bond of a given center in initial and final systems and  $\rho$ —the degree of corresponding vertex in  $G_{TOP}$ .

In practice, however, all 12 edge labels must not be necessarily considered. Let us code labels by “double” designations “0/1”, “0/2”, “0/3”, “1/2”, “1/3”, “2/3” (related to bonds which increase their multiplicity) and reversed designations “3/2”, “3/1”, “2/1”, “3/0”, “2/0”, “1/0” (related to bonds with decreased multiplicities). It is easy to see that in order to construct symbolic equations which correspond to basic bond redistributions, only the labels “0/1”, “0/2”, “1/2” (and also “reversed” labels) must be taken into account. Similarly, for trivial or principal bond redistributions, the labels “1/2”, “1/3”, “2/3” or only the label “2/3” (together with the “reversed” labels) are disallowed. In some instances, the list of used labels can be very short: thus, only two labels (“0/1” and “1/0”) are needed for generation of symbolic equations corresponding to trivial, single mode bond redistributions.

As an example, two pairs of labels related to principal, single mode bond redistributions are schematically represented in Chart 5a; note that each pair consists of the label (e.g., “0/1”) together with its counterpart (“1/0”) which denotes the change of bond multiplicity in the course of the reverse process. The complete list<sup>20</sup> of edge-labeled graphs (constructed starting from the signed graph  $G_{TOP}$  of Chart

**Chart 5.** Complete List (b–k) of Edge-Labeled Graphs Related to Signed Topology Identifier of Chart 4c and Two Pairs of “Interchangeable” Labels (a) Is Represented Together with Corresponding Symbolic Equations and Some Examples



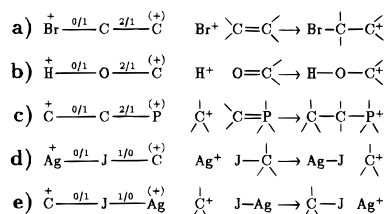
4c) is demonstrated in Chart 5b–k together with corresponding symbolic equations.

The reaction equations describing several well-known processes of organic chemistry are also shown in Chart 5b–k for clarity; these processes can be considered as three-centered reactions with linear electron transfer.<sup>7</sup> It should be mentioned that 4 of 10 edge-labeled graphs, and symbolic equations of Chart 5b,c,f,g would be deleted from this list if only isodesmic (i.e., containing no specific centers) bond redistributions were taken into account.

The graph model used for generation of structural equations (starting from  $G_{ETOP}$ ) is very similar to that described above. In this case, the list of edge labels must contain three additional labels, denoted “1”, “2”, “3” or “1/1”, “2/2”, “3/3” and related to bonds which do not change their multiplicities in the course of reaction. Thus, the total number of labels is equal to 15. Additionally, some new selection criteria for this generation problem must be elaborated. These criteria are needed in order to select those labelings in which unaffected bonds (together with or without bonds which actually change their multiplicities) form cycles of the prescribed size in an initial and/or final system. It is important to note that in edge-labeled graphs having been constructed starting from (unsigned or signed) graph  $G_{ETOP}$ , the structural centers can easily be recognized; all edges incident to any of corresponding vertices are always evaluated by labels “1/1”, “2/2”, or “3/3”.

**Step 3.** Let the edge-labeled graph (constructed from  $G_{TOP}$ ) be known. It is evident that the generation problem for reaction equations (corresponding to the fixed symbolic equation) is reduced to additional *vertex labeling problem*. The solution of this problem actually needs all  $m$  abstract symbols of RCs or SRCs to be substituted by vertex labels, i.e., appropriate symbols of elements-organogens. The set of labels consists of symbols  $C, H, N, O, P, S, Cl$ , etc., and the main selection criterion claims that reaction numbers of substituted abstract RCs or SRCs must not exceed formal valencies of corresponding chemical elements. The above formulated labeling problem is evidently very similar to labeling problems for substituted derivatives of a given structure (cf. ref 18); the resulting reaction equations were,

**Chart 6.** Three Vertex-Labeled Graphs (Constructed Starting from the Edge-Labeled Graph of Chart 5d) Are Represented Together with Reaction Equations (a–c) and the Hidden Symmetry of the Other Edge-Labeled Graph, i.e., That of Chart 5e, Is Illustrated by Two Equivalent Vertex-Labeled Graphs and Corresponding Equations (d and e)



in fact, considered by several investigators<sup>21</sup> in relation to specific classes of chemical interconversions.

As an example, three vertex-labeled graphs (constructed starting from the edge-labeled graph of Chart 5d) are shown in Chart 6a–c together with corresponding reaction equations; these equations describe the well-known electrophilic addition processes, i.e., bromination of olefines, protonation of a carbonyl group, and alkylation of phosphonium ylid, respectively.

It is easy to see that the generation procedure for reaction equations can also be used in order to sequentially construct all possible skeletal equations which correspond to a given structural equation. The values of skeletal numbers for RCs and SRCs and, additionally, the values of reduced valencies for all structural centers must be taken into account in this case.

The above consideration shows that graph-theoretical interpretations of generation problems related to graphs  $G_{TOP}$  and corresponding (i.e., symbolic and reaction) equations and to graphs  $G_{ETOP}$  and corresponding (i.e., structural and skeletal) equations are, in fact, very similar. That is why in the following discussion (see section 6), mainly the problems associated with nonduplicate construction of signed graphs  $G_{TOP}$ , symbolic equations, and reaction equations will be treated in detail.

An additional note is needed, however, in order to completely explain graph models related to the generation problem for structural equations. Up to this moment, we supposed that the vertices of unsigned or signed graphs  $G_{ETOP}$  were not partitioned into those corresponding to reaction and structural centers; the edges related to modified and unchanged bonds were also not differentiated (an example of such graph  $G_{ETOP}$  can be found in Chart 1g). The alternative representation of expanded topology identifiers arises if graphs  $G_{ETOP}$  are, in turn, constructed starting from the given (unsigned or signed) topology identifier  $G_{TOP}$ . The reaction and structural centers—and also the edges corresponding to modified and unaffected bonds—are necessarily differentiated in this case. In order to perceive the difference between two modifications of graphs  $G_{ETOP}$ , one can compare, e.g., two graphs of Chart 11b,c; in the second of these graphs, the reaction centers and modified bonds are distinguished by heavy points and solid lines, respectively. The similar graphs  $G_{ETOP}$  are used in the following paper in this issue, see Charts 2b and 13c in ref 10 for examples.

Although important for practical purposes, the generation procedure for expanded topology identifiers (starting from the graph  $G_{TOP}$ ) cannot be reduced to construction of vertex or edge labelings and hence is not discussed in this paper.

In contrast, the generation problem for structural equations (starting from any given  $G_{ETOP}$  with two sorts of vertices and edges being differentiated) can easily be represented as the labeling problem. It is easy to see that in order to construct any structural equation, the labels “0/1”, “0/2”, etc. should be distributed between those edges of the expanded topology identifier which correspond to bonds changing their multiplicity; the remaining labels (i.e., “1/1”, “2/2”, and/or “3/3”) must be assigned to all other edges of the graph  $G_{ETOP}$ . This consideration shows that the modified generation problem for structural equations is really covered by the graph model which was briefly outlined in step 2 of the above description; the application of quite complex selection criteria (associated with recognition of cycles in the resulting structural equations) is typically not needed in this case.

Summarizing the above discussion, we can state that for any given graph  $G$  (and some selection criteria, cf. above), the complete lists of multiply labeled graphs can be sequentially constructed. From these lists, the graphs (and related symbolic, structural, reaction, or skeletal equations) which describe actually new, unprecedented types of organic reactions should then be extracted.

The above description was, however, incomplete because all sequentially generated labelings of a graph  $G$  must be nonequivalent, otherwise duplicates of corresponding equations will necessarily appear. The equivalence of (multiply) labeled graphs may, in turn, be caused not only by obvious symmetry of the original (possibly labeled) graph but also by the fact that labels used in steps 1 and 2 are in some sense “interchangeable”. This means that interconversion of labels “+” and “(+)”, “·” and “(·)”, “–” and “(–)” and/or simultaneous interconversion of labels “0/1” and “1/0”, “1/2” and “2/1”, etc. results in graphs (cf. notes 19, 20 for examples) which always relate to reverse processes.

The additional difficulties in steps 2 and 3 of the generation procedure are associated with recognition of symmetry for some original graphs (i.e., those constructed in steps 1 and 2, respectively). For example, no obvious symmetry is observed at first sight for an edge-labeled graph of Chart 5e, and this fact can lead to (erroneous) conclusion that the two vertex-labeled graphs represented in Chart 6d,e are not equivalent; note that labels (Ag, J, and C) are not “interchangeable” in this case. The examination of corresponding reaction equations (cf. Chart 6d,e) shows, however, that both equations relate to opposite directions of just the same process, and hence only one of them must be actually generated. The equivalence of the vertex-labeled graphs (i.e., those of Chart 6d,e) evidently follows from the fact that symbolic equation of Chart 5e describes *degenerate bond-and-sign redistribution*. This degeneracy results in appearance of some “hidden” symmetry in corresponding graph; in order to recognize this symmetry, one must take into account the “interchangeability” of labels “+”, “(+)” and “0/1”, “1/0” in the edge-labeled graph of Chart 5e.

The last examples show that simple graph theory based models cannot completely and rigorously describe some problems associated with symmetry and “interchangeability” factors. For that purpose, more powerful (and more complicated!) combinatorial models of all above-mentioned reaction design problems are used in section 6; these models, in fact, generalize our previous results<sup>6b,7b,9</sup> related to one-

step generation of symbolic equations (starting from a given unsigned or signed graph  $G_{TOP}$ ). In the next section, the general combinatorial model is introduced in a very brief and abstract manner. All mathematical notions appeared in that section should not be immediately associated with chemical objects.

## 5. THE GENERAL MODEL: ORBITS OF THE GROUPS $S_2^H$ AND THEIR SUBGROUPS ON THE SETS OF FUNCTIONS $F = M^W$

Let two finite, nonempty sets, i.e., *site set*  $W = \{w_1, w_2, \dots, w_p\}$  and *label set*  $M = \{m_1, m_2, \dots, m_q\}$  be given. The label set is considered here as a union of three disjoint subsets:  $M^- = \{m_1, m_2, \dots, m_r\}$ ,  $M^0 = \{m_{r+1}, m_{r+2}, \dots, m_{q-r}\}$ , and  $M^+ = \{m_{q-r+1}, m_{q-r+2}, \dots, m_q\}$ . The elements of the subsets  $M^-$  and  $M^+$  are thought to be in a one-to-one correspondence; this means that for any label  $m_j \in M^-$  a single "opposite" label  $m_{q-j+1} \in M^+$  corresponds. No "opposite" labels are associated with labels of the subset  $M^0$ . Additionally, this subset (in contrast to subsets  $M^-$  and  $M^+$ ) is allowed to be empty;  $r = q/2$  in this case.

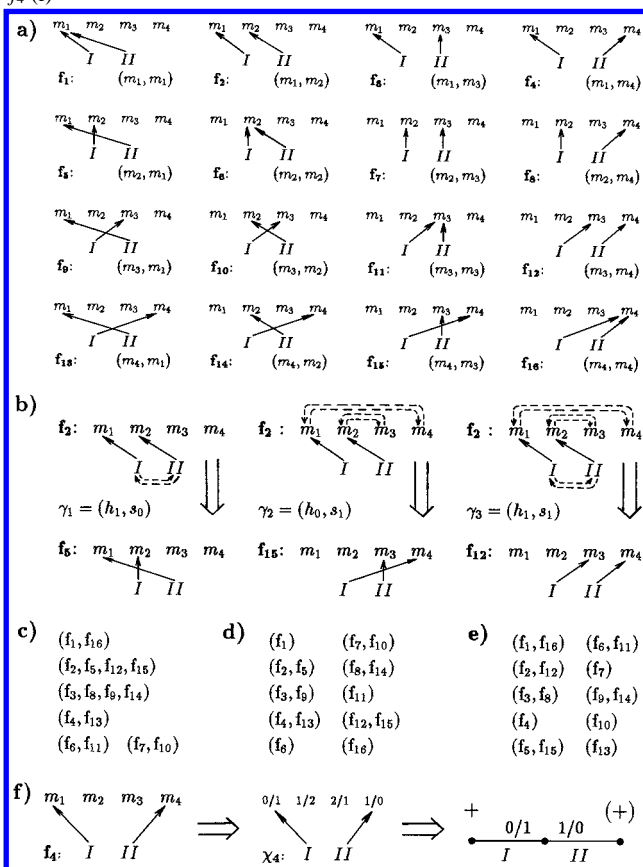
From the sets  $W$  and  $M$ , the complete set  $F = M^W$  of functions (or mappings)  $f = W \rightarrow M$  is then constructed; this set evidently consists of  $|F| = |M|^{|W|} = q^p$  functions. Each function  $f$  of  $F$  is uniquely characterized by its *linear code*, i.e., by sequence of labels from  $M$  corresponding to sites  $w_1, w_2, \dots, w_p$ , respectively. In an example of Chart 7a, all possible functions  $f$  from the two-element set  $W$  (with sites being denoted by roman numerals  $I$  and  $II$ ) into the four-element label set  $M = \{m_1, m_2, m_3, m_4\}$  are represented by arrow schemes and corresponding codes:  $(m_1, m_1), (m_1, m_2), \dots, (m_4, m_4)$ . Note that the construction of the complete set  $F$  of functions does not need the partition of the set  $M$  into three subsets to be actually known.

Two nonidentical functions  $f$  and  $\bar{f}$  are called *antipodes* if for all sites  $w_i, i = 1, 2, \dots, p$ , the labels  $f(w_i)$  and  $\bar{f}(w_i)$  either belong to the subset  $M^0$  and are identical, or are "opposite" (in this case, either  $f(w_i) \in M^-$  and  $\bar{f}(w_i) \in M^+$  or vice versa). Note that antipode function  $\bar{f}$  cannot be constructed if all values of  $f(w_i)$  belong to the set  $M^0$ . In an example of Chart 7a, the labels  $m_4$  and  $m_3$  are considered as "opposite" to labels  $m_1$  and  $m_2$ , respectively ( $M^- = \{m_1, m_2\}$ ,  $M^0 = \emptyset$ ,  $M^+ = \{m_3, m_4\}$ ), and hence for each function  $f$ , the unique antipode function can be recognized. The examination of Chart 7a shows that pairs of functions  $f_k$  and  $f_{17-k}$ ,  $k = 1, 2, \dots$ , are actually antipodes.

Additionally, some subsets  $\tilde{F} = \{f \in F : P(f)\}$  are considered in the general model. The subsets  $\tilde{F}$  consist only of those functions  $f$  which satisfy some *predicate system*  $P(f)$ ; the predicates are thought to reflect various restrictions (or selection criteria) associated with admissible functions  $f$  of  $F$ . One of the most typical predicates  $P(f)$  can be formulated as follows: the function  $f$  belongs to the set  $\tilde{F}$  if its code contains the specific number of some labels belonging either to the subset  $M^0$  or to two remaining subsets (i.e.,  $M^-$  and  $M^+$ ).

Let the arbitrary permutation group  $H = \{h_0, h_1, h_2, \dots, h_{n-1}\}$  (of order  $n$ ) act on the set  $W$ . This action evidently partitions  $W$  into disjoint subsets consisting of those sites which can be interconverted by one or more permutations  $h_k$  of  $H$ . Two sites  $w_i$  and  $w_j$  belong to the same subset (or

**Chart 7.** Pictorial Representations and Codes for All Functions  $f_k, k = 1, 2, \dots, 16$ , from the Set  $W = \{I, II\}$  into the Set  $M = \{m_1, m_2, m_3, m_4\}$  (a); Conversion of the Function  $f_2$  into Equivalent Functions  $f_5, f_{15}$ , and  $f_{12}$  (b); Orbits of the Group  $\Gamma = S_2^H$  (c) and Its Subgroups  $\Gamma'$  (d) and  $\Gamma''$  (e) on the Set  $F = M^W$ , and the Construction of Edge-Labeled Graph (cf. Chart 5e) from the Arrow Scheme Which Corresponds to Function  $f_4$  (f)



*equivalency class*, or orbit of  $H$  on  $W$ ) if there exists at least one permutation  $h_k \in H$  such as  $h_k(w_i) = w_j$ . Similarly, the symmetrical group  $S_2 = \{s_0, s_1\}$  of order 2 is thought here to act on the label set  $M$ ; the nonidentity permutation  $s_1$  is postulated<sup>22</sup> to convert any label of  $M^0$  into itself and any label  $m_j \in M^-$  (or  $m_j \in M^+$ ) into corresponding "opposite" label  $m_{q-j+1} \in M^+$  (or  $m_{q-j+1} \in M^-$ ).

In the above example (cf. Chart 7a), the permutation  $s_1$  interconverts "opposite" labels  $m_1$  and  $m_4$  (and also  $m_2$  and  $m_3$ ); the identity permutation  $s_0$  evidently converts all labels into themselves. In this example, the group  $H$  can be considered as consisting of two permutations:  $h_0 = (I)(II)$  (identity permutation) and  $h_1 = (I, II)$ . The second permutation evidently interconverts two sites of the set  $W$ .

From the groups  $H$  and  $S_2$ , the new, *induced* permutation group  $\Gamma = S_2^H$  is now constructed; this group consists of  $|S_2| \cdot |H|$ , i.e.,  $2n$  permutations, and acts on the set  $F = M^W$  (and its subset  $\tilde{F}$ ; note that all predicates  $P(f)$  must be invariant with respect to  $\Gamma$ ). The permutations  $\gamma \in \Gamma$  may be written in form  $\gamma = (h_k, s_0)$  or  $\gamma = (h_k, s_1)$ ,  $k = 0, 1, 2, \dots, n-1$ ; in both cases  $\gamma$  converts any function  $f$  into the equivalent function  $f'$  also belonging to  $F$

$$\gamma(f) = (h_k, s_l)(f) = s_l f h_k^{-1} = f'$$

for each  $k = 0, 1, 2, \dots, n-1$  and  $l = 0, 1$ . The last expression means that for any  $w_i \in W$ , the corresponding



value of the resulting function  $f'$  may be found as follows:  $f'(w_i) = s_i(f(h_k^{-1}(w_i)))$ . Similarly to the action of the groups  $H$  and  $S_2$  (on the sets  $W$  and  $M$ , respectively), the induced action of the group  $\Gamma = S_2^H$  also partitions the sets  $F = M^W$  and  $\tilde{F} \subseteq F$  into disjoint subsets (*equivalence classes* or *orbits*). Every orbit of  $\Gamma$  on  $F$  (or  $\tilde{F}$ ) consists of those functions  $f$  which can be interconverted by one or more permutations  $\gamma \in \Gamma$ .

The simplest way to visualize interconversions of the functions  $f$  (under action of permutations  $\gamma$ ) consists in examination of arrow schemes. Thus, in order to transform the arrow scheme corresponding to any function  $f$  into the scheme corresponding to the resulting function  $f' = \gamma(f)$ , the lower ends of all arrows must be shifted in accordance with permutations  $h_k^{-1}$ ,  $k = 1, 2, \dots, n - 1$  (because only nonidentity permutations really interconvert sites of the set  $W$ ), and the upper ends of arrows must be shifted in accordance with permutation  $s_1$  (because just this permutation interconverts labels of the set  $M$ ).


In an example of Chart 7b, the dashed arrows (related to permutations  $h_1 = (I, II)$  and  $s_1 = (m_1, m_4)(m_2, m_3)$ ) indicate the shifts which transform the arrow scheme of the function  $f_2$  (cf. Chart 7a) into the schemes corresponding to functions  $f_5, f_{15}$ , and  $f_{12}$ . This example shows that only three functions  $f_5 = (h_1, s_0)(f_2) = \gamma_1(f_2)$ ,  $f_{15} = (h_0, s_1)(f_2) = \gamma_2(f_2)$ , and  $f_{12} = (h_1, s_1)(f_2) = \gamma_3(f_2)$  are equivalent to the function  $f_2$  and (together with the original function  $f_2 = (h_0, s_0)(f_2) = \gamma_0(f_2)$ ) form an equivalence class or orbit of the group  $\Gamma = \{\gamma_0, \gamma_1, \gamma_2, \gamma_3\}$  on the set  $F$ . The other orbits can be found similarly; the resulting partition of the set  $F = M^W$  into orbits of the group  $\Gamma = S_2^H$  is represented in Chart 7c.

It should be noted here that  $\Gamma$  is, in fact, a particular case of the more general group-theoretical construction (because an arbitrary group can act on the set  $M$  instead of symmetrical group  $S_2$ ). This general construction was introduced by F. Harary<sup>23a</sup> and is known to mathematicians as *the power group*. The orbits of the power group acting on the set of functions from one finite set to another were called<sup>23b</sup> *combinatorial objects* in the narrow sense of this notion. In the broad sense, any construction from finite sets may be regarded as a combinatorial object because combinatorics is, in fact, the theory of finite sets.<sup>23c</sup>

From the above consideration, it is evident that the solution of any generation problem for combinatorial objects (in the narrow sense) needs some *representative system* or *transversal* for all orbits of the power group  $\Gamma = S_2^H$  on the set  $F$  (or its subset  $\tilde{F}$ ) to be found.<sup>24</sup> The functions  $f$  with extremal, i.e., minimal or maximal, codes are typically chosen as representatives of orbits. Strictly speaking, the choice between two possibilities is not important; in this paper, we prefer the functions with *minimal codes* to represent orbits of the group  $\Gamma$ . Thus, in the above example, the functions  $f_1$ – $f_4$ ,  $f_6$ , and  $f_7$  (with minimal codes, cf. Chart 7a,c) are regarded as representatives for six orbits of the group  $\Gamma$  on the set  $F$ .

In our second example, the solution of the generation problem for nonequivalent functions  $f = W \rightarrow M$  belonging to a specific subset  $\tilde{F} \subseteq F$  is outlined without details. In this example, the set  $W$  consists of three sites 1, 2, and 3, the label set<sup>26</sup>  $M = \{a, b, c, b^*, a^*\}$  is partitioned into three subsets  $M^- = \{a, b\}$ ,  $M^0 = \{c\}$ ,  $M^+ = \{b^*, a^*\}$ , and the

**Chart 8.** Orbits of the Group  $\Gamma = S_2^H$ ,  $H = \{(1)(2)(3), (1,3)(2)\}$ ,  $S_2 = \{(a)(b)(c)(b^*)(a^*), (a, a^*)(b, b^*)(c)\}$ , on the Set  $\tilde{F} \subseteq F = M^W$ ,  $W = \{1, 2, 3\}$ ,  $M = \{a, b, c, b^*, a^*\}$ , Are Represented by (Minimal) Codes of Functions  $f_1$ – $f_{14}$  (a) and Are Visualized by Labelings of Three Sequentially Numbered Points (b)

<b>a)</b>		
$f_1: (a, a, c)$	$f_2: (a, b, c)$	
$f_3: (a, c, a)$	$f_4: (a, c, b)$	$f_5: (a, c, b^*)$
$f_6: (a, c, a^*)$	$f_7: (a, b^*, c)$	$f_8: (a, a^*, c)$
$f_9: (b, a, c)$	$f_{10}: (b, b, c)$	$f_{11}: (b, c, b)$
$f_{12}: (b, c, b^*)$	$f_{13}: (b, b^*, c)$	$f_{14}: (b, a^*, c)$
<b>b)</b>		
	$f_1: \begin{matrix} a & a & c \\ \bullet & \bullet & \bullet \end{matrix}$	$f_2: \begin{matrix} a & b & c \\ \bullet & \bullet & \bullet \end{matrix}$
$f_3: \begin{matrix} a & c & a \\ \bullet & \bullet & \bullet \end{matrix}$	$f_4: \begin{matrix} a & c & b \\ \bullet & \bullet & \bullet \end{matrix}$	$f_5: \begin{matrix} a & c & b^* \\ \bullet & \bullet & \bullet \end{matrix}$
$f_6: \begin{matrix} a & c & a^* \\ \bullet & \bullet & \bullet \end{matrix}$	$f_7: \begin{matrix} a & b^* & c \\ \bullet & \bullet & \bullet \end{matrix}$	$f_8: \begin{matrix} a & a^* & c \\ \bullet & \bullet & \bullet \end{matrix}$
$f_9: \begin{matrix} b & a & c \\ \bullet & \bullet & \bullet \end{matrix}$	$f_{10}: \begin{matrix} b & b & c \\ \bullet & \bullet & \bullet \end{matrix}$	$f_{11}: \begin{matrix} b & c & b \\ \bullet & \bullet & \bullet \end{matrix}$
$f_{12}: \begin{matrix} b & c & b^* \\ \bullet & \bullet & \bullet \end{matrix}$	$f_{13}: \begin{matrix} b & b^* & c \\ \bullet & \bullet & \bullet \end{matrix}$	$f_{14}: \begin{matrix} b & a^* & c \\ \bullet & \bullet & \bullet \end{matrix}$

function  $f$  belongs to the set  $\tilde{F}$  if its code contains exactly one label (i.e., “c”) of the set  $M^0$ . The induced group  $\Gamma = S_2^H$ ,  $H = \{h_0, h_1\} = \{(1)(2)(3), (1, 3)(2)\}$ ,  $S_2 = \{s_0, s_1\} = \{(a)(b)(c)(b^*)(a^*), (a, a^*)(b, b^*)(c)\}$ , partitions the set  $\tilde{F}$  (consisting of 48 functions) into 14 orbits. These orbits are represented in Chart 8a by minimal (in relation to priority order  $a < b < c < b^* < a^*$ ) codes of corresponding functions  $f_1$ – $f_{14}$ .

In the group  $\Gamma = S_2^H$ , there are three subgroups of which two are needed to rigorously formulate some reaction design problems.<sup>27</sup> One of them is the power group,  $\Gamma' = E^H$ , which consists only of permutations  $\gamma = (h_k, s_0)$ ,  $h_k \in H$  ( $s_0$  is a single permutation of an identity group  $E$ ). The “action” of the group  $E$  on the set  $M$  means that no “opposite” labels are actually permuted. For this reason, the partition of the set  $M$  into three subsets  $M^-, M^0, M^+$  is not required; all labels can be considered as belonging to the set  $M \equiv M^0$  in this special case.

The other subgroup,  $\Gamma''$ , is not the power group; for its construction, the group  $H$  must contain a subgroup  $\tilde{H}$  of index 2 (this means that  $n/2$  permutations from  $H$  belong to this subgroup). The subgroup  $\Gamma''$  consists of  $n/2$  permutations  $\gamma = (h_k, s_0)$ ,  $h_k \in \tilde{H}$ , and  $n/2$  permutations  $\gamma = (h_k, s_1)$ ,  $h_k \in H \setminus \tilde{H}$ , and also acts on the set of functions  $f = W \rightarrow M$ . Both subgroups form orbits on the set  $F = M^W$  (and on the set  $\tilde{F}$ ) and representatives of these orbits uniquely correspond to solutions of some reaction design problems.

In order to illustrate the construction of orbits for subgroups  $\Gamma'$  and  $\Gamma''$ , let us turn to an example in Chart 7. In the above discussion, we have demonstrated the group  $\Gamma = S_2^H$  to consist of four permutations:  $\gamma_0 = (h_0, s_0)$ ,  $\gamma_1 = (h_1, s_0)$ ,  $\gamma_2 = (h_0, s_1)$ ,  $\gamma_3 = (h_1, s_1)$ , and this fact immediately shows that  $\Gamma' = E^H$  consists of two permutations, i.e.,  $\gamma_0$  and  $\gamma_1$ . On the other hand, it is easy to see that the single subgroup  $\tilde{H}$  of the group  $H = \{h_0, h_1\}$  is an identity group; for this reason, the subgroup  $\Gamma''$  of order 2 must be formulated as consisting of permutations  $\gamma_0 = (h_0, s_0)$  and  $\gamma_3 = (h_1, s_1)$ . Thus, for any given function, the transformation of the arrow scheme (cf. Chart 7b) in accordance either with permutation  $\gamma_1$  or with permutation  $\gamma_3$  results in an arrow scheme which corresponds to an equivalent function  $f'$ . In the example under consideration (cf. Chart 7b), the

function  $f_5$  is equivalent to  $f_2$  due to the action of permutation  $\gamma_1 \in \Gamma'$ . Similarly, the function  $f_{12}$  is equivalent to  $f_2$  due to the action of permutation  $\gamma_3 \in \Gamma''$ . The orbits of the subgroups  $\Gamma'$  and  $\Gamma''$  thus obtained are explicitly shown in Charts 7d and 7e, respectively.

The last important topic is associated with certain subgroups of the group  $H$  which describe the symmetry properties of any given combinatorial object (i.e., function  $f$ ). In the simplest situation (related to the action of the group  $\Gamma' = E^H$ ), the automorphism group  $Aut(f)$  of the particular function  $f$  is formulated as follows:

$$Aut(f) = \{h_k \in H : s_0 f h_k^{-1} = f\}$$

This notation means that for any  $h_k \in Aut(f)$  (and identity permutation  $s_0$  acting on  $M$ ), the induced action converts a given function  $f$  into itself. Surely, for any  $f$ , its automorphism group is a subgroup of  $H$ .

In the case of the group  $\Gamma$  and its subgroup  $\Gamma''$ , the other, expanded automorphism group  $Aut[f]$  is defined<sup>9</sup> as follows:

$$Aut[f] = \{h_k \in H : s_0 f h_k^{-1} = f \text{ or } s_0 f h_k^{-1} = \bar{f}\}$$

This notation means that the expanded group consists not only of “normal” or (+)-automorphisms (which convert  $f$  into itself and form the group  $Aut(f)$ ) but also of additional permutations, (−)-automorphisms, which convert  $f$  into its antipode function  $\bar{f}$ . It should be noted that coincidence of the specially constructed “normal” and expanded automorphism groups was considered<sup>9,25a</sup> as a chirality criterion for spatial (point, labeled, graph, and molecular) configurations; the application of the suggested criteria made it possible to elaborate<sup>9,25b</sup> the unambiguous classification system for chiral molecular objects.

In general, just the expanded automorphism groups,  $Aut[f]$ , are needed in order to rigorously formulate multistep reaction design problems. In practice, however, these groups are often identical to “normal” groups  $Aut(f)$ ; this is a case for most of above-generated combinatorial objects. For example, for only 2 of 14 functions  $f$  (represented in Chart 8a by their codes), “normal” and expanded automorphism groups do not coincide. In order to be convinced, one must verify that only two functions, i.e.,  $f_6$  and  $f_{12}$ , are converted into their antipodes by permutation  $h_1 = h_1^{-1} = (1, 3)(2)$ . In the example under consideration, the codes of functions  $f_6$  and  $f_{12}$  (i.e.,  $(a, c, a^*)$  and  $(b, c, b^*)$ ) should be read from the right to the left; the resulting codes  $(a^*, c, a)$  and  $(b^*, c, b)$  evidently relate to antipode functions  $f_6$  and  $f_{12}$ , respectively.

Some specific factors which result in coincidence of “normal” and expanded automorphism groups will be additionally considered below.

## 6. COMBINATORIAL MODELS OF REACTION DESIGN PROBLEMS

In this section, the three-step generation procedure of section 4 is reformulated in terms of combinatorial objects. For this purpose, three functions, i.e.,  $\varphi$ ,  $\chi$ ,  $\psi$ , their equivalence classes (with respect to appropriate groups  $\Gamma$ ,  $\Gamma'$ , or  $\Gamma''$ ), and corresponding automorphism groups are discussed in detail. The main applications of suggested combinatorial models are exemplified by construction of the

above-considered lists of signed graphs (cf. Chart 4c–i) and symbolic equations (cf. Chart 5b–k). The brief comments on other applications (related to possible reaction subcategories and two-centered processes with linear electron transfer) can be found in notes 31 and 32 in the following paper in this issue.<sup>10</sup>

**6.1. Generation of Signed Topology Identifiers.** The formal model of this generation step (i.e., step 1 in section 4) is needed only for *noncanonical bond redistributions*. Otherwise,  $G_{TOP} \equiv G$  (or  $G_{ETOP} \equiv G$ ) and the automorphism group  $Aut(G)$  of the “parent” graph  $G$  can be used for a solution of the next generation problem (see below).

Thus, the generation problem for signed graphs  $G_{TOP}$  or  $G_{ETOP}$  is now considered. Let the site set  $W_1$  be represented by  $m$  numbered vertices of a given graph  $G$ ; these vertices are thought to correspond to all unsigned and signed reaction centers. Let the label set  $M_1$  consist of an “empty” label (forming the subset  $M_1^0$ ) and of even number of labels corresponding to signs (such as “+” and “(+)”); note that one of these “interchangeable” labels must belong to the subset  $M_1^-$ , while the other—to the subset  $M_1^+$ . Let  $\Phi$  be a set consisting of all possible functions  $\varphi = W_1 \rightarrow M_1$ . The group  $H_1$  is a group which acts on the set  $W_1$ ; typically, this group is a vertex automorphism group of a graph  $G$ ,  $H_1 \equiv Aut(G)$ . The second group,  $S_2 = \{s_0, s_1\}$ , is defined as a group which acts on the label set. The nonidentity permutation  $s_1$  of this group interconverts labels “+” and “(+)”, “−” and “(−)”, etc. (which can be regarded as “opposite”) but converts into itself the single “empty” label. In this case, the orbits of the power group  $\Gamma_1 = S_2^{H_1}$  on the set  $\Phi = M_1^{W_1}$  represent all possible graphs with 0, 1, 2, ...,  $m$  signed vertices. Note that  $\Gamma_1$  is isomorphic to the first of three permutation groups (i.e., to  $\Gamma = S_2^H$ ) discussed in the preceding section.

In order to take into account selection criteria associated with topology identifiers (cf. step 1 in section 4), an appropriate predicate system  $P(\varphi)$  should be formulated. One of these predicates is required to reflect the necessary restriction: only two vertices in the proper topology identifying graph are allowed to be labeled by symbols of charge or unshared electron. The orbits of the power group  $\Gamma_1 = S_2^{H_1}$  on the subset  $\tilde{\Phi} = \{\varphi : P(\varphi)\}$  (consisting of functions which satisfy all predicates) must be searched for in this case. Thus, the generation problem for topology identifiers  $G_{TOP}$  (or  $G_{ETOP}$ ) is reduced to *construction of representatives for all orbits of the group  $\Gamma_1$  on the set  $\tilde{\Phi}$* . The “normal”  $Aut(\varphi)$  and expanded  $Aut[\varphi]$  groups of any particular function  $\varphi$  describe obvious and all (i.e., including “hidden”) symmetries of the corresponding signed graph.

As a typical example, let us consider the generation problem for signed topology identifiers  $G_{TOP}$  corresponding to the “parent” graph  $G$  of Chart 4a; note that two vertices of this graph are required to be labeled by symbols of positive and/or negative charges. In this case,  $W_1 = \{1, 2, 3\}$ ,  $M_1 = \{“+”, “−”, “empty” \text{ label}, “(−)”, “(+)”\}$  (cf. Chart 4b),  $\Phi = M_1^{W_1}$ , and the subset  $\tilde{\Phi} \subseteq \Phi$  consists of all functions  $\varphi$  which contain exactly one “empty” label in their codes. The groups  $H_1$ ,  $S_2$ , and  $\Gamma_1$  are formulated as described above; note that  $H_1 = \{(1)(2)(3), (1, 3)(2)\}$  is an automorphism group of a graph represented in Chart 4a.

An attentive reader will immediately see that the example under discussion is, in fact, identical to the second example of the preceding section (cf. Chart 8). One can easily verify that there is a one-to-one correspondence between elements of the sets  $W$  and  $W_1$ ,  $M$  and  $M_1$ ,  $\tilde{F}$  and  $\tilde{\Phi}$ ; the groups acting on these sets are also isomorphic. This fact shows that the functions  $f_1-f_{14}$  (representing orbits of the group  $\Gamma = S_2^H$  on the set  $\tilde{F}$ ) uniquely correspond to functions  $\varphi_1-\varphi_{14}$  and hence to 14 topology identifying graphs. In order to be convinced, one should substitute “abstract” labels “a”, “a\*”, “b”, “b\*”, and “c” in pictorial representations of Chart 8b by signs of charge “+”, “(+)”, “-”, “(-)”, and “empty” label, respectively. Additionally, two edges (connecting vertices 1 and 2, 2 and 3, cf. Chart 4a) should be explicitly shown. It is important that 7 of 14 resulting signed graphs<sup>28</sup> (corresponding to functions  $f_k \equiv \varphi_k$ ,  $k = 6, 8, 12, 13, 4, 2$ , and 9) are identical to graphs of Chart 4c-i and hence represent all possible topology identifiers related to three-centered processes with linear electron transfer.

In section 5, we have demonstrated that only for two functions (i.e.,  $f_6$  and  $f_{12}$ , cf. Chart 8a,b), “normal” and expanded automorphism groups were not identical. Keeping in mind the similarity between functions  $f$  and  $\varphi$  in this example, we can deduce that for all functions  $\varphi_k$ ,  $k = 1-5, 7-11, 13, 14$ , the groups  $Aut(\varphi_k)$  are isomorphic to groups  $Aut[\varphi_k]$ . This fact, in turn, leads to a conclusion that the groups  $Aut(\varphi_k) = Aut[\varphi_k]$  can be immediately observed by inspection of corresponding signed graphs. For example, the visual examination of unsymmetrical graphs represented in Chart 4d,f-i shows that both automorphism groups consist of a single identity permutation. In contrast, the “hidden” symmetry of two remaining graphs (cf. Chart 4c,e) is associated with (-)-automorphism (1, 3)(2) of the corresponding expanded group  $Aut[\varphi_k]$ ,  $k = 6, 12$ .

The above consideration convincingly shows that it is actually important to recognize whether the “normal” and expanded automorphism groups of a given function  $\varphi$  are identical. The following self-evident rules are valid for functions in which the labels of exactly two sites of the set  $W_1$  belong to the sets  $M_1^-$  and  $M_1^+$ ; in corresponding graphs, two vertices are evidently labeled by symbols of charge and/or unshared electron.

Thus, the groups  $Aut(\varphi)$  and  $Aut[\varphi]$  coincide if in the labeled graph: (1) two labels corresponding to signs are neither identical nor “opposite” (cf. graphs of Chart 4g-i as examples), (2) two signed vertices belong to different orbits of the group  $Aut(G)$  (cf. graphs of Chart 4d,f,h,i as examples), and (3) two vertices of the same orbit are labeled by the same symbol of sign; the graph of Chart 4g with the symbol “+” substituted by “-” (or vice versa) can serve as an example.<sup>29</sup>

As a result, we can state that “hidden” symmetry in signed topology identifiers exists in the single situation, i.e., if two vertices with “opposite” labels belong to the same orbit of a group  $Aut(G)$ .

## 6.2. Generation of Symbolic and Structural Equations.

In this generation step (corresponding to step 2 of section 4), all three types of induced groups—i.e., those similar to  $\Gamma$ ,  $\Gamma'$ , and  $\Gamma''$ —are used in formal models of the edge-labeling procedures.

Thus, the generation problems for edge-labeled (unsigned or signed) topology identifying graphs are now considered. Let  $W_2$  be an edge set of the graph  $G_{TOP}$  or  $G_{ETOP}$  and  $M_2$  an appropriate set of edge labels “0/1”, “1/0”, “0/2”, “2/0”, etc. Note that one of the “opposite” labels (such as “0/1” and “1/0”) must belong to the subset  $M_2^-$ , while the other—to the subset  $M_2^+$ ; the subset  $M_2^0$  is nonempty only if structural equations (which can contain bonds with unchanged multiplicity) are constructed. Let the set  $X$  be a set consisting of all possible functions  $\chi = W_2 \rightarrow M_2$  and the subset  $\tilde{X} \in X$  to consist of those functions  $\chi$  which satisfy some predicate system  $P(\chi)$ . The predicates can be associated with any possible selection criteria but one of them—i.e., that ensuring even specificity values for all unsigned RCs and odd specificity values for all SRCs (cf. sections 2 and 4)—must be necessarily taken into account.

Let  $H_2$  be an edge-group acting on the set  $W_2$ . This action is, in fact, an induced action because  $H_2$  interconverts edges (corresponding to some pairs of vertices) but not vertices of the “parent” graph  $G$ . The other group,  $S_2 = \{s_0, s_1\}$ , acts on the set  $M_2$ ; the nonidentity permutation  $s_1$  of this group is thought to interconvert “opposite” labels of the subsets  $M_2^-$  and  $M_2^+$ . In order to properly define the group  $\Gamma_2$  (acting on the sets  $X$  and  $\tilde{X}$ ) one must take into account several possible situations. Note that in all cases  $\Gamma_2$  is constructed out of the induced group  $H_2$  (and of the group  $S_2$ ) and hence can be considered as doubly induced permutation group.

In the case of unsigned topology identifiers, permutations of the group  $H_2$  correspond to (vertex) automorphisms of the group  $Aut(G)$  of the “parent” graph  $G$ . In this case, the group  $\Gamma_2$  (acting on  $X$  and  $\tilde{X}$ ) is, in fact, the power group of the type  $\Gamma$ . Thus, orbits of the group  $\Gamma_2 = S_2^{H_2}$  are needed in order to correctly generate all symbolic or structural equations related to canonical bond redistributions.

In the case of noncanonical bond redistributions, the full symmetry of the signed topology identifier is to be taken into account. For this reason, the permutations of the group  $H_2$  must necessarily correspond to all automorphisms of the expanded group  $Aut[\varphi]$ . In this case, two situations are, however, possible: (a) if “normal” and expanded automorphism groups of the particular function  $\varphi$  coincide,  $Aut(\varphi) = Aut[\varphi]$ , then  $\Gamma_2$  is a power group of the type  $\Gamma'$  (see section 5) and hence can be represented as  $\Gamma_2 = E^{H_2}$ ; (b) if the groups  $Aut(\varphi)$  and  $Aut[\varphi]$  are not identical, then the group of the type  $\Gamma''$  must be constructed. In this case, the subgroup  $\tilde{H}_2$  of a group  $H_2$  consists of permutations induced by all (+)-automorphisms of “normal” group  $Aut(\varphi)$ . The other permutations, i.e., those forming the set  $H_2 \setminus \tilde{H}_2$ , correspond to all (-)-automorphisms of the group  $Aut[\varphi]$ . The resulting doubly induced group  $\Gamma_2$  consists of permutations  $(h_k, s_0)$ ,  $h_k \in \tilde{H}_2$  and permutations  $(h_k, s_1)$ ,  $h_k \in H_2 \setminus \tilde{H}_2$ ; the order of this group is equal to  $|Aut[\varphi]|$ .

In all cases, orbits of the properly formulated groups  $\Gamma_2$  on the sets  $X$  and  $\tilde{X}$  correspond to nonequivalent edge-labeled graphs, and for this reason, any generation procedure for symbolic or structural equations needs the representatives of these orbits to be sequentially constructed.

The application of the above formal models is illustrated for one signed and one unsigned three-centered linear topology identifiers  $G_{TOP}$ . In both examples, set  $W_2$  consists

of two edges (*I* and *II*, cf. Chart 4a), set  $M_2$  consists of two pairs of appropriate “opposite” labels (thus, the subset  $M_2^0$  is empty), and set  $X$  consists of  $4^2 = 16$  functions  $\chi$ . Note that the subset  $\tilde{X}$  is considered here to be identical to set  $X$ . For this purpose, the preselected labels (of the set  $M_2$ ) must guarantee proper specificity values for all three reaction centers in the resulting symbolic equations.

The construction of the induced group  $H_2$  evidently needs all permutations of some vertex group (i.e.,  $Aut(G)$  or  $Aut[\varphi]$ ) to be explicitly known. The inspection of the unsigned topology identifier (cf. Chart 4a) shows that two permutations—(1)(2)(3) and (1, 3)(2)—of its vertex automorphism group uniquely correspond to permutations  $h_0 = (I)(II)$  and  $h_1 = (I, II)$  of the edge-group. Thus, the group  $H_2 = \{h_0, h_1\}$  (or its subgroup consisting of an identity permutation) acts on the set  $W_2$  in all possible situations.

In our first example, the signed topology identifier of Chart 4c is considered. The expanded  $Aut[\varphi]$  and “normal”  $Aut(\varphi)$  automorphism groups associated with this signed graph were demonstrated (cf. preceding subsection) to be nonidentical. This fact shows that  $H_2 = \{h_0, h_1\}$ ,  $\tilde{H}_2 = \{h_0\}$ , and orbits of the  $\Gamma''$ -like doubly induced group correspond to edge-labeled graphs and three-centered symbolic equations.

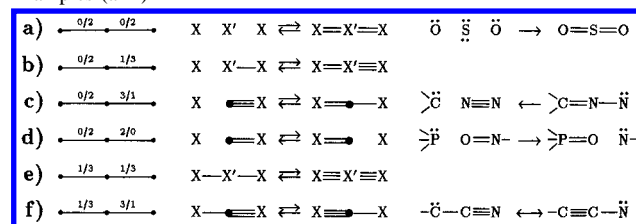
An attentive reader can recall that all possible functions  $f$  from a two-element set  $W$  into four-element set  $M$  were just constructed (cf. Chart 7a), and orbits of the group  $\Gamma'' \subseteq \Gamma = S_2^H$  on the set of  $4^2 = 16$  functions were found, cf. Chart 7e. Thus, the solution of the generation problem needs to transform the functions  $f_1$ – $f_7$ ,  $f_9$ ,  $f_{10}$ , and  $f_{13}$  (which were chosen to represent orbits of Chart 7e) into functions  $\chi = W_2 \rightarrow M_2$  and then into corresponding edge-labeled graphs.

Let the label set  $M_2$  to consist of symbols “0/1”, “1/2”, “2/1”, and “1/0” (cf. Chart 5a); these labels are evidently needed in order to construct equations which describe basic, principal, single mode processes. It is easy to verify that the selected labels ensure the even specificity value of the “central” unsigned reaction center and odd specificity values of two SRCs in all generated symbolic equations. The substitution of “abstract” labels ( $m_1, m_2, m_3, m_4$ ) by the selected labels in any arrow scheme of Chart 7a evidently produces the arrow scheme which corresponds to the function  $\chi$ . The resulting arrow schemes can, in turn, be transformed into corresponding edge-labeled graphs. Two steps of the transformation procedure are explicitly shown in Chart 7f for a function  $f_4$ .

The edge-labeled graph of Chart 7f is identical to the graph of Chart 5e and hence represents the symbolic equation which describes all possible  $S_E2$  electrophilic substitution processes. In a similar manner, other functions  $f_1$ – $f_3$ ,  $f_5$ – $f_7$ ,  $f_9$ ,  $f_{10}$ , and  $f_{13}$  (cf. Chart 7a,e) can be transformed into symbolic equations of Chart 5b–d,f–h,i,j,k. Thus, the complete list of three-centered symbolic equations related to signed topology identifying graph of Chart 4c has been successfully constructed.

In our second example, the symbolic equations related to canonical, principal, double mode bond redistributions are constructed starting from unsigned topology identifier of Chart 4a. In this case, the set  $M_2$  is defined to consist of four labels “0/2”, “1/3”, “3/1”, and “2/0”. This set of labels evidently guarantees even specificity values for all RCs in corresponding symbolic equations. Note that labels “0/2”

**Chart 9.** Three-Centered Double Mode Bond Redistributions with Linear Topology: Edge-Labeled Graphs, Symbolic Equations, and Examples (a–f)<sup>a</sup>



<sup>a</sup> Note that the initial system in the decomposition process of Chart 9c is represented by the nonpolar resonance structure of diazomethane.

and “2/0” (and also “1/3” and “3/1”) are regarded as “opposite”; just these labels are interconverted by nonidentity permutation  $s_1$  of the group  $S_2 = \{s_0, s_1\}$ .

The set  $X$  of all possible functions,  $\chi = W_2 \rightarrow M_2$ ,  $W_2 = \{I, II\}$ ,  $M_2 = \{“0/2”, “1/3”, “3/1”, “2/0”\}$ , consists of  $4^2 = 16$  functions which uniquely correspond to functions  $f$  of Chart 7a. The orbits of the group  $\Gamma = S_2^H$ ,  $H = H_2 \cong \{(1)(2)(3), (1, 3)(2)\}$ , were also recognized in section 5, cf. Chart 7c. For this reason, the transformation procedure (similar to that of Chart 7f) can be immediately applied<sup>30</sup> to arrow schemes for functions  $f_1$ – $f_4$ ,  $f_6$ , and  $f_7$  of Chart 7a. The resulting edge-labeled graphs and symbolic equations are depicted in Chart 9a–f.

The examination of the presented list of three-centered symbolic equations convincingly shows that at least four equations describe chemically reasonable canonical, double mode bond redistributions. The corresponding reaction equations can be found in Chart 9a,c,d,f.

The “normal”  $Aut(\chi)$  and expanded  $Aut[\chi]$  groups of a particular function  $\chi$  are subgroups of the corresponding groups  $Aut(\varphi)$  and  $Aut[\varphi]$  (or of the group  $Aut(G)$  in the case of canonical bond redistributions). All (+)-automorphisms of an arbitrary function  $\chi$  (which are present in the both automorphism groups) convert this function into itself. In contrast, (–)-automorphisms of  $\chi$  (which are present only in an expanded group) convert this function into its antipode  $\bar{\chi}$ . Note that in an antipode function, all edge labels are obviously converted into the “opposite” labels (i.e., the label “0/1” is substituted by “1/0” and vice versa).

The automorphism groups related to edge-labeled graphs of Chart 5d,e can serve as examples. The “normal” groups  $Aut(\chi)$  (of the both functions  $\chi$ ) evidently coincide with the identity group  $Aut(\varphi)$  of the signed topology identifier, cf. Chart 4c. In contrast, the group  $Aut[\chi]$  contains no (–)-automorphisms in the case of graph represented in Chart 5d but contains one (–)-automorphism (1, 3)(2) in the case of graph represented in Chart 5e. This (–)-automorphism interconverts “opposite” edge labels “0/1” and “1/0” as well as signs “+” and “(+)”. Thus, the groups  $Aut(\chi)$  and  $Aut[\chi]$  are identical for the edge-labeled graph of Chart 5d but are nonidentical for the edge-labeled graph of Chart 5e.

It is easy to see that symbolic equations of Chart 5d and Chart 5e describe nondegenerate and degenerate bond-and-sign redistributions. The other examples of degeneracy can be observed in Chart 5h,j,k and in Chart 9d,f, and in all cases the groups  $Aut(\chi)$  and  $Aut[\chi]$  do not coincide. In the general case, it can easily be proved that symbolic equations represent *degenerate bond or bond-and-sign redistribution* if “normal” and expanded automorphism groups of corresponding func-

tions  $\chi$  are nonidentical. Thus, the nondegeneracy criteria for bond redistributions are formally similar to the chirality criteria recently suggested for point, labeled, graph, and molecular configurations.<sup>9,25a,b</sup> This fact makes it possible to consider some analogy between chirality/achirality relation for spatial objects and nondegeneracy/degeneracy relation for processes which describe structural transformations of chemical objects.

### 6.3. Generation of Reaction and Skeletal Equations.

This is the simplest of three generation steps; the corresponding labeling procedures (cf. step 3 in section 4) are very similar to those used for substituted derivatives of a given structure.

Let  $W_3$  be a vertex set of the graph  $G_{TOP}$  or  $G_{ETOP}$  and the set  $M_3$  to consist of appropriate labels  $C, H, N, O, P, S$ , etc. Let  $\Psi$  be a set of all possible functions  $\psi = W_3 \rightarrow M_3$  and the subset  $\tilde{\Psi} = \{\psi \in \Psi : P(\psi)\}$  to consist of those functions  $\psi$  which satisfy some predicate system  $P(\psi)$ . Although the "main" selection criterion needs only formal valencies of abstract centers and corresponding atoms to be compared, other restrictions (e.g., those associated with disallowed bond multiplicities between specific pairs of atoms, with presence or absence of specific fragments, and even with energy characteristics of the overall process) can also be taken into account by means of appropriate predicates  $P(\psi)$ .

The group  $H_3$  (acting on  $W_3$ ) is defined to be identical to the expanded<sup>31</sup> automorphism group  $Aut[\chi]$  and the group acting on the set  $M_3$  is an identity group  $E$ . Thus, the group  $\Gamma_3 = E^{H_3}$  acts on the sets  $\Psi$  and  $\tilde{\Psi}$ ; this group is an induced group of the type  $\Gamma'$  (see section 5). The orbits of  $\Gamma_3$  on the sets  $\Psi$  and  $\tilde{\Psi}$  evidently represent all formally possible vertex-labeled graphs and allow the graphs which correspond to reaction or skeletal equations to be generated. The trivial examples (related to signed graph of Chart 4c and edge-labeled graphs of Chart 5d,e) can be found in Chart 6.

For any particular function  $\psi$ , its "normal" and expanded automorphism groups,  $Aut(\psi)$  and  $Aut[\psi]$ , are formulated to consist of those permutations (of the groups  $Aut(\chi)$  and  $Aut[\chi]$ , respectively) which convert  $\psi$  into itself. The application of these groups makes it possible to easily formulate the degeneracy criterion for processes represented by reaction or skeletal equations. More explicitly, the process is degenerate if  $Aut(\psi) \neq Aut[\psi]$ . For example, the loss of degeneracy in any reaction equation of Chart 6d,e results from the fact that a single  $(-)$ -automorphism of the group  $Aut[\chi]$  (related to edge-labeled graph of Chart 5e, cf. preceding subsection) does not convert the corresponding function  $\psi$  into itself.

It should be mentioned here that mathematical formulations of the generation problems related to skeletal and reaction equations are slightly different. (Some specific features of combinatorial models corresponding to signed graphs  $G_{ETOP}$  and structural equations could also be detected but were not explicitly discussed in the above text.) In particular, it is not necessary to consider just the same labels of  $M_3$  to be allowable for those vertices of the set  $W_3$  which correspond to reaction centers and to structural centers, respectively. The other difference is associated with the fact that in *skeletal equations*, the unusual valence numbers of heteroatoms are typically specified. As a result, different labels of the set  $M_3$  must be associated with several valence states (e.g.,  $S^{II}$ ,  $S^{IV}$ , and  $S^{VI}$ ) of the same atom.

It should be finally noted that three-step generation of nonequivalent signed topology identifiers (functions  $\varphi$ ), then edge-labeled graphs (function  $\chi$ ), and then vertex-labeled graphs (function  $\psi$ ) is formally very similar to stepwise generation of functions which represent molecular formulas, constitutional formulas, and stereoformulas. For this reason, the same mathematical construction—*The Ladder of Combinatorial Objects*<sup>9,25a</sup>—can be used to formalize the both problems. This fact makes it possible to denote the formalism used in this section as the *Ladder of Reaction Objects*.

## 7. ENUMERATION TECHNIQUES AND GENERATION ALGORITHMS

Although a great variety of enumeration problems is discussed in modern literature, most of them actually relate to orbits of the induced groups  $\Gamma' = E^H$ . The corresponding enumeration techniques (based mainly on the famous Pólya's Theorem<sup>32a</sup>) make it possible to easily solve counting problems for numerous substituted derivatives of organic structures; the above-cited refs 18a,b,d,e can serve as typical examples. The very similar "Pólya-like" methods were also used by S. Fujita who has enumerated "reaction graphs"<sup>4b,c</sup> and "reaction-center graphs"<sup>21c-e</sup> (in other words, symbolic and reaction equations) starting from a given "basic reaction graph"; the last type of graphs uniquely corresponds to "trivial" symbolic equations in which all bonds are formed from zero or completely destroyed.

In our investigations, we are mainly interested in analytical enumeration of symbolic equations corresponding to a given (unsigned or signed) graph  $G_{TOP}$ ; in this case, all above combinatorial models must be taken into account. The applications are, however, quite limited (cf. refs 6b, 7b, 8d,e, and 9) because commonly used enumeration tools—Burnside's<sup>32b</sup> (or Cauchy–Frobenius–Burnside's<sup>32c</sup>) Lemma, Pólya's Theorem,<sup>32a</sup> and Harary's Power Group Enumeration Theorem<sup>23a</sup>—cannot take into account most of the restrictions associated with various predicates  $P$ . For this reason, all predicates are disregarded in our enumeration techniques which are presented below in the most generalized form.<sup>33</sup>

Three main enumeration formulas, 1–3, make it possible to easily count the number  $N$  of orbits for arbitrary groups  $\Gamma$ ,  $\Gamma'$ , and  $\Gamma''$  acting on the sets  $F = M^W$ . The simplest formula 2 (related to the group  $\Gamma' = E^H$ ) is, in fact, well-known<sup>32c</sup> and directly follows from Burnside's Lemma. Two other, original<sup>8d,9</sup> formulas, 1 and 3, can also be considered as applications of Burnside's Lemma to specific power group  $\Gamma = S_2^H$  and to its subgroup  $\Gamma''$ , respectively. Note that permutations of the subgroup  $\tilde{H}$  are thought to be numbered by integers  $0, 1, \dots, |\tilde{H}| - 1$ , and permutations of the set  $H/\tilde{H}$ —by integers  $|\tilde{H}|, |\tilde{H}| + 1, \dots, |H| - 1$ .

$$N = \begin{cases} \frac{1}{2 \cdot |H|} \sum_{i=0}^{|\tilde{H}|-1} (q^{c(h_i)} + \delta(h_i) \cdot q^{c''(h_i)}) & \text{if } q = 2r \\ \frac{1}{2 \cdot |H|} \sum_{i=0}^{|\tilde{H}|-1} (q^{c(h_i)} + q^{c''(h_i)} \cdot (q-2r)^{c'(h_i)}) & \text{if } q > 2r \end{cases} \quad (1)$$

$$N = \frac{1}{|H|} \sum_{i=0}^{|\tilde{H}|-1} (q^{c(h_i)}) \quad (2)$$



$N =$

$$\begin{cases} \frac{1}{|H|} \left( \sum_{i=0}^{|\tilde{H}|-1} q^{c(h_i)} + \sum_{j=|\tilde{H}|}^{|\tilde{H}|-1} \delta(h_j) \cdot q^{c''(h_j)} \right) & \text{if } q = 2r \\ \frac{1}{|H|} \left( \sum_{i=0}^{|\tilde{H}|-1} q^{c(h_i)} + \sum_{j=|\tilde{H}|}^{|\tilde{H}|-1} q^{c''(h_j)} \cdot (q-2r)^{c'(h_j)} \right) & \text{if } q > 2r \end{cases} \quad (3)$$

In the above formulas,  $c$ ,  $c'$ , and  $c''$  denote total numbers of cycles, numbers of odd cycles, and numbers of even cycles in a permutation  $h_i$  or  $h_j$  acting on the set  $W$ . The values of  $\delta(h_i)$  and  $\delta(h_j)$  are equal to unity if the permutation contains no odd cycles (and are equal to zero otherwise). Note, that  $q$  is the number of labels in a set  $M$ , while  $r$  is the number of labels in a set  $M^-$  (and in a set  $M^+$ ).

In order to illustrate the applications of eqs 1–3, let us turn to the first example of section 5; in this case,  $W = \{I, II\}$ ,  $M = \{m_1, m_2, m_3, m_4\}$ ,  $\tilde{H} = \{h_0\} = \{(I)(II)\}$ ,  $H = \{h_0, h_1\} = \{(I)(II), (I, II)\}$  (cf. Chart 7a,b). It is easy to see that  $c(h_0) = c'(h_0) = 2$ ,  $c''(h_0) = 0$ ,  $\delta(h_0) = 0$  (for identity permutation of the groups  $\tilde{H}$  and  $H$ ) and  $c(h_1) = c''(h_1) = 1$ ,  $c'(h_1) = 0$ ,  $\delta(h_1) = 1$  (for nonidentity permutation of the group  $H$ ). Keeping in mind that  $M^- = \{m_1, m_2\}$ ,  $M^0 = \emptyset$ , and  $M^+ = \{m_3, m_4\}$ , we get  $q = 4$ ,  $r = 2$ , and the application of formulas 1 and 3 for  $q = 2r$  shows that numbers of orbits for the groups  $\Gamma$  and  $\Gamma''$  on the set  $F = M^W$  are equal to  $(4^2 + 0 \cdot 4^0 + 4^1 + 1 \cdot 4^1)/4 = 6$  and  $(4^2 + 1 \cdot 4^1)/2 = 10$ , respectively. The resulting values are evidently equal to the numbers of nonequivalent functions  $f$  (cf. Chart 7c,e) and also to the numbers of three-centered symbolic equations which are explicitly represented in Chart 9a–f and Chart 5b–k. In a similar manner, eq 2 can be used in order to compute the number of orbits,  $N = (4^2 + 4^1)/2 = 10$ , for the group  $\Gamma' = E^H$ ; the functions forming these orbits are listed in Chart 7d.

It should be noted that the above equations may be successfully applied to many other enumeration problems; an additional example (related to enumeration of possible reaction subcategories) is briefly commented in note 31 of ref 10. It is important that particular classes of (unsigned or signed) graphs and specific types of bond or bond-and-sign redistributions can also be calculated. Thus, the counting formula for (trivial or principal) canonical, cyclic, single mode bond redistributions may easily be obtained from eq 1 presuming that  $H$  is a dihedral group  $D_n$  and  $q = 2r$  is equal to 2 or 4. This formula allows one to directly calculate numbers of symbolic equations corresponding to  $n$ -centered processes with cyclic electron transfer, see ref 6b. The similar general formulas related to multicentered processes with linear electron transfer can be found in ref 7b.

The computer-assisted solutions of all generation problems (i.e., of those related to signed topology identifiers as well as to symbolic, structural, reaction, and skeletal equations) are also based on combinatorial models of section 6. Thus, the representatives of orbits for induced groups  $\Gamma$ ,  $\Gamma'$ , or  $\Gamma''$  (acting on appropriate subsets  $\tilde{\Phi}$ ,  $\tilde{X}$ , or  $\tilde{\Psi}$ ) must be sequentially constructed in all cases. It is important that in contrast to analytical enumeration (or counting) techniques, the proper constructive enumeration (or generation) algo-

gorithms need all predicates  $P(\varphi)$ ,  $P(\chi)$ , or  $P(\psi)$  to be explicitly taken into account.

The algorithms<sup>9</sup> which generate nonequivalent functions  $\varphi$ ,  $\chi$ , and  $\psi$  (in other words, signed, edge-labeled, and vertex-labeled graphs or corresponding equations) are quite complicated and cannot be discussed here in detail. For this reason, only the main principles of elaborated algorithms are listed below.

(a) The generation process is represented by multilevel *generation tree*; the nodes of this tree are associated with *partially (and also completely) constructed* signed, edge-labeled, or vertex-labeled graphs. In order to optimize the generation process, *strong selection criteria* for the nodes of the tree are required to be elaborated; these criteria make it possible to immediately cut off those branches of the tree which produce only duplicates or undesired combinatorial objects of any kind.

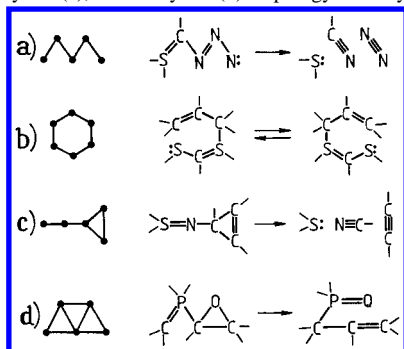
(b) All possible restrictions associated with appropriate labels (corresponding to each level of the generation tree) are compactly encoded by means of the special, *access matrix*; the rows of this matrix make it possible to quickly recognize the next of allowed labels.

(c) The partial labelings (i.e., partially constructed signed, edge-labeled, or vertex-labeled graphs) are classified into several types depending on the induced group which orbits are searched for. The original coding system for partial labelings makes it possible to formulate strong selection criteria and corresponding evaluation subprocedures; these subprocedures easily reject duplicates (and undesired labelings) at early steps of the generation process. As a result, *no special isomorphism checks are needed after the combinatorial object is completely constructed*.

## 8. COMPUTER IMPLEMENTATIONS AND PERSPECTIVES

The combinatorial models of section 6 are most general in the sense that they make it possible to formally describe all (one-step or multistep) generation procedures for computer-assisted nonduplicate construction of symbolic, reaction, structural, or skeletal equations starting from the given graph  $G_{TOP}$  or  $G_{ETOP}$ . The generation of chemical equations (corresponding to any given reaction or skeletal equation, cf. problem (d) in section 1) can also be covered by suggested models if only simple substituents—such as alkyl, aryl, or univalent functional groups—are taken into account. In this special case, the group  $Aut[\psi]$  produces all information needed to construct  $\Gamma'$ -like group acting on the set of appropriate functions (from the set of free valencies of reaction and structural centers into the set of preselected simple substituents). The orbits of this group uniquely correspond to traditional chemical equations.

Unlike computer-assisted solutions of synthesis design problems, the applications of computers to reaction design<sup>34</sup> (i.e., to systematical search for new types of chemical interconversions) are not numerous. The extensive investigations of the Munich group (based on Dugundji–Ugi's Model of Constitutional Chemistry<sup>1b</sup>) have resulted in multipurposed programs IGOR<sup>35b,c</sup> and IGOR2<sup>35d,e</sup> which made it possible to discover several new reaction types, cf. refs 35a–f for examples. Both IGOR programs were oriented to produce basis and hetero reactions (i.e., symbolic

**Chart 10.** Unusual Reaction Types Corresponding to Linear (a), Cyclic (b), Linear–Cyclic (c), and Tricyclic (d) Topology Identifying Graphs

and reaction equations, in our terminology) *starting from the given irreducible R-matrix*. These matrices correspond to “trivial” symbolic equations (such as those of Chart 5b,e,k and Chart 9a,d) with all bonds being formed from zero or completely destroyed.

In our group, the application of computers to reaction design problems also started in the early 1980s<sup>8b</sup> and resulted in two BASIC programs which generated symbolic (and reaction) equations *corresponding to a given topology identifying graph*. The fact that graphs  $G_{TOP}$  (but not irreducible  $R$ -matrices or their equivalents) have been constantly used in our approach allowed us to construct numerous symbolic equations with two or more specific centers  $X$ . The variety of processes with cyclic electron transfer can serve as an example: in addition to 6 three-centered, 6 four-centered, 20 five-centered, and 13 six-centered basis reactions (reported<sup>35c</sup> by the Munich group in 1985), more than 200 extra reactions—i.e., symbolic equations with 2, 3, 4, 5, or 6 centers  $X$ —were produced manually as early as in 1976.<sup>2c</sup> These “allodesmic” equations can also be generated by our computer program (SYMBEQ); some of them represent highly attractive new synthetic possibilities for ylids and imines.<sup>6b,9</sup> It should be noted here that most of hypothetical processes with two or more specific centers correspond to unprecedented “electron flow patterns” and hence represent the highest degree of novelty.<sup>35a</sup>

The first of our programs, SYMBEQ, has been worked out in collaboration with Baskin.<sup>5c,d,6c</sup> This program allows construction of (complete or restricted) lists of symbolic equations related to canonical 3–10 centered processes; additionally, it can produce the lists of reaction equations corresponding to any selected symbolic equation. All user-defined restrictions (associated with possible changes of bond multiplicity, with allowed numbers of specific RCs, with selected heteroatoms, etc.) as well as the structure of starting graph  $G_{TOP}$  can be fixed in the course of the friendly organized dialogue<sup>6c</sup> with the program.

Some results obtained by means of the program SYMBEQ (i.e., those related to linear and linear–cyclic topology identifiers) were systematically considered in separate papers.<sup>5c,d</sup> The most interesting reaction types detected by this program were briefly discussed,<sup>6b,c</sup> and at least one of new reactions was experimentally realized in the laboratory (cf. ref 6c). In this paper, we demonstrate only four unusual reaction types predicted by the program SYMBEQ; the corresponding reaction equations are represented in Chart 10a–d together with starting graphs  $G_{TOP}$ .

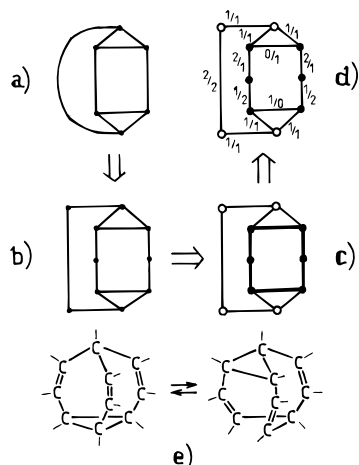
In contrast to SYMBEQ, the second program,<sup>7b,9</sup> ELSE-I, treats “linear” graphs  $G_{TOP}$  with 3–9 vertices and generates only symbolic equations describing (noncanonical) principal processes with linear electron transfer. The constructed equations belong to any of 12 subcategories which represent ionic, radical, monoredox, and diredox reactions, cf. subsection 4.1 in the following paper in this issue.<sup>10</sup> All equations or only the equations belonging to the selected “generalized” type<sup>5c,7c,9</sup> (cf. also section 1 in ref 10) are sequentially generated. For example, the above-discussed symbolic equations of Chart 5b–k represent three of six possible types (i.e., addition–fragmentation, substitution, and resonance in positively charged structures) for multicentered processes with linear topology. These equations were, in fact, produced by the program ELSE-I.

The present state of the Formal–Logical Approach (which includes all notions and mathematical models described above) allows for elaboration of a new generation of more powerful and flexible computer programs. These programs will probably make it possible to successfully solve many more intricate problems of reaction design. In particular, the following generation problems seem to be covered by the present state approach: (a) systematic search for new structural and skeletal equations which are much more informative to an organic chemist when compared with symbolic and reaction equations; (b) systematic search for all types of equations related to new ionic, radical, or oxidation–reduction processes starting from any given unsigned graph  $G$ ; (c) systematical search for new “global” types of organic interconversions, such as cyclizations, or condensations, or molecular rearrangements.

In many cases, the solution of the above problems evidently needs additional severe restrictions to be taken into account; these restrictions are anticipated to significantly decrease the extremely large number of possible results (combinatorial explosion!). The most promising restricting possibilities are thought to be associated with a generation of equations which correspond only to very limited but practically most important classes of bond redistributions (e.g., to isodesmic, principal, single mode and composite mode bond redistributions). The numbers of components in initial and final systems as well as the numbers and allowed sizes of cycles must also be regulated in order to ensure that reasonable numbers of equations be generated.

Finally, the special, extremely attractive problem of a systematic search for highly degenerate rearrangements will be mentioned. Although many details of this completely new problem must be theoretically investigated, the general strategy is quite clear. Thus, in order to construct “degenerate” structural and skeletal equations corresponding to canonical cyclic and noncanonical linear isodesmic, single mode bond redistributions (which describe practically all elementary organic reactions), a general scheme is suggested. The application of this scheme to (canonical) processes with cyclic electron transfer is illustrated by transformation of the regular graph (cf. Chart 11a) into its “homolog” (cf. Chart 11b), then into the graph of Chart 11c (in which the selected cycle is indicated by solid lines), and finally into the edge-labeled graph of Chart 11d. The resulting graph corresponds to the unique structural equation and also to the skeletal equation of Chart 11e. These equations evidently represent the highly degenerate Cope rearrangement of bullvalene.<sup>36</sup>

**Chart 11.** Stepwise Construction ( $a \Rightarrow b \Rightarrow c \Rightarrow d$ ) of the Edge-Labelled Graph Corresponding to Highly Degenerate Bullvalene Rearrangement (e)<sup>a</sup>



<sup>a</sup> Note that the both graphs (b) and (c) can be considered as unsigned graphs  $G_{ETOP}$ . In the second of expanded topology identifiers, the RCs and skeletal centers are, however, differentiated; the edges corresponding to modified bonds are also distinguished by solid lines.

It is interesting that highly degenerate ionic or radical rearrangement processes were not specifically discussed in modern chemical literature. The slight modification in the above sequence of transformations makes it, however, possible to systematically search for degeneracy in noncanonical, isodesmic, single mode bond redistributions. Thus, in order to construct degenerate processes with linear electron transfer, the nonequivalent chains (instead of cycles) must be found in graphs similar to that of Chart 11b; the endpoints of these chains evidently correspond to charged or radical centers. For both types of design problems (i.e., for those related to highly degenerate processes with cyclic and linear electron transfers), several supplementary procedures are worked out to date. Many other procedures<sup>38</sup> need, however, the special techniques to be additionally elaborated.

#### ACKNOWLEDGMENT

The partial financial support of this work from the Chemical Structure Association Trust and from the Russian Foundation for Basic Research (Grant No. 96-03-33003a) is gratefully acknowledged. The authors are also indebted to a referee for his valuable remarks and suggestions which helped us to significantly improve the initial version of our manuscript.

#### REFERENCES AND NOTES

- (1) (a) Ugi, I.; Gillespie, P. D. Chemistry and Logical Structures. 3. Representation of Chemical Systems and Interconversions by be Matrices and Their Transformation Properties. *Angew. Chem.* **1971**, *83*, 980–981. (b) Dugundji, J.; Ugi, I. An Algebraic Model of Constitutional Chemistry as a Basis for Chemical Computer Programs. *Topics Curr. Chem.* **1973**, *39*, 19–64.
- (2) (a) Zefirov, N. S.; Tratch, S. S. Rearrangements and Cyclizations. 10. Formal-Logical Approach to Synchronous Reactions. The Formal Tool to Derive the Complete Sets of Possible Types for Multicentered Processes with Cyclic Electron Transfer. *Zh. Org. Khimii* **1975**, *11*, 225–231 [in Russian]. (b) Zefirov, N. S.; Tratch, S. S. Rearrangements and Cyclizations. 11. Formal-Logical Approach to Synchronous Reactions. The Classification of Multicentered Processes with Cyclic Electron Transfer. *Zh. Org. Khimii* **1975**, *11*, 1785–1800 [in Russian]. (c) Zefirov, N. S.; Tratch, S. S. Rearrangements and Cyclizations. 12. Formal-Logical Approach to Synchronous Reactions. The Complete Sets of Multicentered Processes with Cyclic Electron Transfer on Three, Four, Five, and Six Centers. *Zh. Org. Khimii* **1976**, *12*, 7–18 [in Russian]. (d) Zefirov, N. S.; Tratch, S. S. Rearrangements and Cyclizations. 15. Tautomerism: General Problems, Classification, Search for New Topological and Reaction Types. *Zh. Org. Khimii* **1976**, *12*, 697–718 [in Russian].
- (3) (a) Kvasnička, V. Graph Theory Approach. *Coll. Czech. Chem. Commun.* **1983**, *48*, 2097–2117. (b) Kvasnička, V.; Kratochvil, M.; Koča, J. Reaction Graphs. *Coll. Czech. Chem. Commun.* **1983**, *48*, 2284–2304. (c) Kvasnička, V. Classification Scheme of Chemical Reactions. *Coll. Czech. Chem. Commun.* **1984**, *49*, 1090–1097. (d) Koča, J.; Kratochvil, M.; Kunz, M.; Kvasnička, V. Valence States of Atoms and Their Conversions. *Coll. Czech. Chem. Commun.* **1984**, *49*, 1247–1260.
- (4) (a) Fujita, S. Description of Organic Reactions Based on Imaginary Transition Structures. 1. Introduction of New Concepts. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 205–212. (b) Fujita, S. Description of Organic Reactions Based on Imaginary Transition Structures. 2. Classification of One-String Reactions Having an Even-Membered Cyclic Reaction Graph. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 212–223. (c) Fujita, S. Description of Organic Reactions Based on Imaginary Transition Structures. 3. Classification of One-String Reactions Having an Odd-Membered Cyclic Reaction Graph. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 224–230. (d) Fujita, S. Description of Organic Reactions Based on Imaginary Transition Structures. 8. Synthesis Space Attached by a Charge Space and Three-Dimensional Imaginary Transition Structures with Charges. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 111–115. (e) Fujita, S. Description of Organic Reactions Based on Imaginary Transition Structures. 9. Single-Access Perception of Rearrangement Reactions. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 115–120. (f) Fujita, S. "Structure-Reaction Type" Paradigm in the Conventional Methods of Describing Organic Reactions and the Concept of Imaginary Transition Structures Overcoming This Paradigm. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 120–126.
- (5) (a) Tratch, S. S.; Zefirov, N. S. Problems of Molecular Design and Computers. 7. Formal-Logical Approach to Organic Reactions. Main Notions and Terminology. *Zh. Org. Khimii* **1982**, *18*, 1561–1583 [in Russian]. (b) Zefirov, N. S.; Tratch, S. S. Problems of Molecular Design and Computers. 8. Reaction Fragments and Classification Equations in Formal-Logical Approach to Organic Reactions. *Zh. Org. Khimii* **1984**, *20*, 1121–1142 [in Russian]. (c) Tratch, S. S.; Baskin, I. I.; Zefirov, N. S. Problems of Molecular Design and Computers. 13. Systematic Analysis of Organic Processes Characterized by Disclosed Topologies of Bond Redistribution. *Zh. Org. Khimii* **1988**, *24*, 1121–1133 [in Russian]. (d) Tratch, S. S.; Baskin, I. I.; Zefirov, N. S. Problems of Molecular Design and Computers. 14. Systematic Analysis of Organic Processes Characterized by Linear-Cyclic Topology of Bond Redistribution. *Zh. Org. Khimii* **1989**, *25*, 1585–1606 [in Russian].
- (6) Zefirov, N. S.; Gordeeva, E. V.; Tratch, S. S. Problems of Molecular Design and the Computer. 11. The FLAMINGOES Program System for the Nonempirical Solution of Structural Problems of Organic Chemistry. The BASIC Program Oriented for Microcomputer. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 188–193. (b) Zefirov, N. S.; Tratch, S. S. Symbolic Equations and Their Application to Reaction Design. *Anal. Chim. Acta* **1990**, *235*, 115–134. (c) Zefirov, N. S.; Baskin, I. I.; Palyulin, V. A. SYMBEQ Program and Its Application in Computer-Assisted Reaction Design. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 994–999.
- (7) (a) Zefirov, N. S.; Tratch, S. S.; Gamziani, G. A. Problems of Molecular Design and Computers. 9. Formal-Logical Approach to Organic Reactions. General Description of the Processes with Linear Electron Transfer. *Zh. Org. Khimii* **1986**, *22*, 1341–1359 [in Russian]. (b) Tratch, S. S.; Gamziani, G. A.; Zefirov, N. S. Problems of Molecular Design and Computers. 10. Enumeration and Generation of Equations Characterizing Ionic, Radicalic, and Oxidation-Reduction Processes with Linear Electron Transfer. *Zh. Org. Khimii* **1987**, *23*, 2488–2507 [in Russian]. (c) Zefirov, N. S.; Tratch, S. S.; Gamziani, G. A. *Description of Organic Reactions in Frame of the Formal-Logical Approach*; Metsniereba Press: Tbilisi, 1993; 208 pp [in Russian].
- (8) (a) Tratch, S. S.; Podimova, E. V.; Zefirov, N. S. An Algorithm Generating Nonequivalent Labelings for Sets with Given Permutation Groups. *Solution Methods and Programs for Optimization Problems on Graphs and Networks* (Proceedings of the 2nd All-Union Conference); Novosibirsk, 1982; Part I, pp 210–213 [in Russian]. (b) Tratch, S. S.; Podimova, E. V.; Zefirov, N. S. An Algorithm Generating Complete Sets of Equations Which Describe Possible Reactions of Organic Compounds. *Application of Computers in Molecular Spectroscopy and Chemical Investigations* (Proceedings of the 6th All-Union Conference); Novosibirsk, 1983; pp 205–206 [in Russian]. (c) Tratch, S. S.; Zefirov, N. S. A Mathematical Model of Generation Problems for Structural Equations and Directed Search for Degenerate Rearrangements. *Applications of Computers in Molecular Spectroscopy*



