

A Hierarchical Classification Scheme for Chemical Reactions[†]

Serge S. Tratch* and Nikolai S. Zefirov

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

Received June 30, 1996

Several known approaches to classification of structural changes in the course of chemical reactions are briefly discussed. The hierarchical five-level classification scheme is suggested for organic interconversions in which the bond redistributions (summarizing all required steps of the reaction process) are exactly known. The main hierarchy levels, i.e., *reaction category*, *reaction class*, *interconversion type*, *topological type*, and *reaction type*—and also corresponding sublevels—are illustrated by numerous examples. The first two levels together with reaction subcategories and subclasses relate to the nature of educt and product species (neutral or dipolar molecules, ions, radicals, etc.) and to contours and graphs formed by bonds which change their multiplicity during the reaction. The interconversion, topological, and reaction types are uniquely characterized by classification, symbolic, and skeletal equations, respectively; the subtypes are also associated with specific kinds of equations. The central hierarchy attribute, i.e., *classification equation*, is shown to closely correspond to what organic chemists typically use in order to systematize simple processes with cyclic and linear electron transfers. Finally, the application of the suggested classification scheme to construction of hierarchically organized codes for organic reactions is finally demonstrated in the paper.

The main purpose of the Formal–Logical Approach to Organic Reactions is a solution of various reaction design problems associated with the computer-assisted search for new, unprecedented types of chemical interconversions, cf. the preceding paper.¹ The recognized (real or hypothetical) reaction types need, however, to be systematized, and the classification principle must be general enough in order to be applicable to all known as well as generated chemical reactions. The fact that the Formal–Logical Approach is based on classical structural theory leads to a conclusion that corresponding classification scheme must cover all possible interconversions of covalently bonded neutral molecules, ions, radicals, and ion–radicals.

Although many classification schemes for organic reactions have been suggested in literature, none of them can satisfy all purposes of an organic chemist.² In contrast to traditional classification attributes (such as nature or structural characteristics of the reagent, or substrate, or product of the reaction), new attributes associated with formal structural changes in the course of chemical processes were shown^{2b} to be more applicable to computerized information retrieval systems. In the present paper, several hierarchical schemes (related exclusively to overall bond redistributions during organic reactions³) are briefly reviewed.

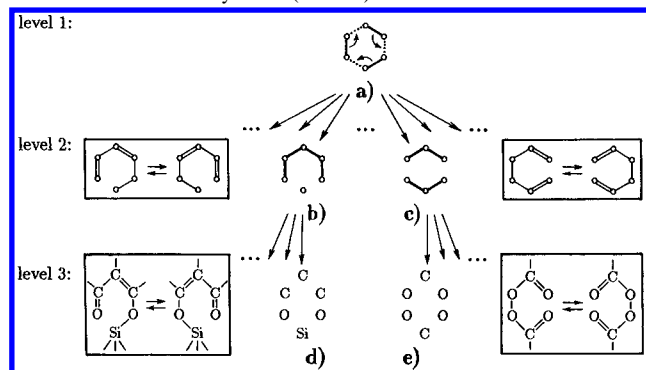
The main levels of the most widely known classification systems are based on essentially the same classification attributes which are illustrated here in terms of the Munich group approach.⁵ All processes which are characterized by the same *valence electron shift pattern* (and by corresponding irreducible R-matrix in Dugundji–Ugi's Algebraic Model of Constitutional Chemistry^{5a}) form **reaction or R-category**.^{5b–d} The selected R-category—also denoted as r-class^{5e}—is sche-

matically represented in Chart 1a by electron redistribution scheme. The corresponding processes are then partitioned into several **RB-classes**^{5b}—or rb-subclasses in terminology of ref 5e—in accordance with invariant σ -frames^{5d} (or with intact BE-matrices^{5b,e} of Dugundji–Ugi's matrix model). These frames are, in fact, connected or disconnected graphs with the edges representing only the valence lines which are unaffected in the course of the reaction. Two examples of σ -frames are shown in Chart 1b,c; it is easy to see that the superposition of any electron flow pattern (with the bonds appearing either in educt or in product molecules) with any σ -frame results in the specific scheme which is referred to as *basis*^{5c} or *basic*^{5d,e} *reaction*. The basis reactions can evidently correspond either to unknown or to more or less common reaction classes; two examples, i.e., those of [1,5] and [3,3] sigmatropic rearrangements, are represented (in boxes) in the second level of Chart 1.

The processes belonging to the same RB-class are then classified into **RA-classes**^{5b}—also denoted as ra-subclasses^{5e}—in accordance with the so-called *atom vectors*^{5b} which unambiguously describe how the symbols of chemical elements are distributed between atoms forming the reaction core (see Chart 1d,e for examples). In the corresponding *hetero reactions*^{5d} (exemplified by the degenerate silyl group migration process^{6a} and by the peroxide automerization process,^{6b} cf. level 3 of Chart 1), all substituents are not specified. The atoms which do not belong to the reaction core and hence do not directly participate in the reaction process can also be taken into account by means of additional classification levels, such as **R1-classes**, **R2-classes**, etc.^{5b} Each of these classes is formed out of processes with the first, second, etc. neighbors of reaction centers being identical. It should be noted that an additional (*zeroth* with respect to Chart 1) classification level, that is **CD-class**,^{5c,e} is characterized in Ugi's approach by the minimal number

[†] 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. Three Main Levels of I. Ugi's Classification System Are Represented by Six-Centered Six-Electron Redistribution Pattern (a), Two of 13 σ -Frames (b and c), and Two of Numerous Possible Distributions of Atom Symbols (d and e)^a



^a The examples of basis and hetero reactions are shown in boxes near corresponding classification attributes.

of valence electrons being redistributed in the course of the reaction process; in the example of Chart 1, this number is evidently equal to 6.

The very similar classification schemes were suggested by Kvasnička^{7a} and Fujita;^{7b,c} for other publications of these authors, see refs 3, 4, and 21c,d in the preceding paper.¹ Thus, in the graph theory based Kvasnička classification system,^{7a} three levels of Chart 1 are represented by signed **reaction graph** R , **intact molecular subgraph** G , and **evaluation** φ of vertices from G by appropriate atomic symbols. In this approach, the ordered pair (R, G) and ordered triple (R, G, φ) (which is considered as reaction core) uniquely correspond to the above-mentioned basis and hetero reactions.

Three graphs corresponding to Kvasnička's graph R , pair (R, G) , and triple (R, G, φ) can also be recognized as subgraphs of Fujita's Imaginary Transition Structures; these graphs were denoted^{7b} as **Basic Reaction Graph**, **Reaction Graph**, and **Reaction Center Graph**. Two last graphs—also considered^{7c} as reaction center graphs of levels 0 and 1—unambiguously describe basis and hetero reactions (e.g., those explicitly shown in Chart 1).

Several "levels of mechanistic information" were considered by Roberts^{7d} in 1978; in his classification scheme, the first level was represented by left and right parts of **concerted process skeleton** (corresponding, e.g., to electron redistribution pattern of Chart 1a), while the second and third levels do not have special names. We shall not discuss here many other notions which have been suggested by several investigators^{8a-g} in order to introduce a very similar hierarchy in description of organic reactions. Our only additional comment is that in many early approaches (cf., e.g., refs 8a,b,e), all classification attributes were applied to an important but a very limited class of organic transformations, namely to pericyclic reactions with neutral educt and product molecules.

In our Formal-Logical Approach, the classification principles were also initially applied to "neutral" pericyclic reactions (or to processes with cyclic electron transfer, in the other terminology). Thus, the complete lists of symbolic equations (representing just the same information as basis reactions of Chart 1) were partitioned^{9a,b} into "generalized" and interconversion types as early as in 1976; the main tools of the classification procedure were demonstrated to also be

applicable to tautomerization reactions.^{9c,d} In our subsequent papers, the less traditional "neutral" processes, i.e., those with linear^{10a} and linear-cyclic^{10b} topologies of electron transfer, were classified into "generalized" and interconversion types. The similar treatment^{10c,d} of multicentered "linear" processes in charged and radicalic systems has shown, however, that the distribution of charge and/or unshared electron symbols must be explicitly taken into account in the classification scheme.

Although the levels of our old classification scheme—i.e., **topology of electron transfer**, **interconversion type**, **topological type**, and **reaction type**—have been declared and extensively used in the early publication,^{9c} the corresponding classification attributes, i.e., *topology identifying graphs*, *classification*, *symbolic*, and *reaction equations*, were completely and systematically described in the subsequent, more fundamental papers.^{11a,b} The applications of supplementary "generalized" types (characterized by schematic pictures and closely related to interconversion types, cf. below) were additionally discussed in ref 11c. It should be noted that two of the main classification levels, i.e., topological and reaction types, were actually identical with RB- and RA-classes of the Munich group, while two other levels—topology of electron transfer and interconversion type—seem to have no analogs in other classification systems.

In our recent investigations,^{11d} we have clearly understood that several additional classification levels must be introduced in order to reflect the modern state of the Formal-Logical Approach. These new levels are needed to cover the newly formulated generation problems (cf. section 1 in ref 1) and are associated with the nature and cyclic structure of the educt and product molecules. Keeping in mind that our classification principles were never considered as a whole in English-language publications, we regard this paper as a survey in which we try to overview and exemplify all hierarchy levels of the modern state classification procedure.

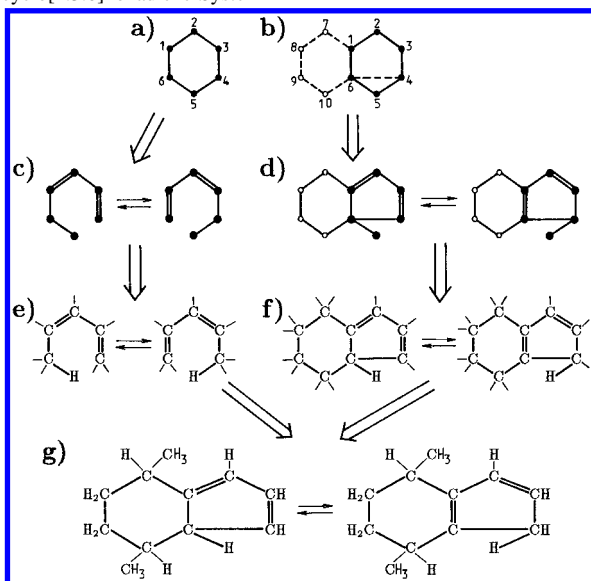
In section 1, the main notions of the Formal-Logical Approach are visualized, and the main differences of our hierarchy levels (from those suggested by other authors, cf. above) are briefly discussed. In section 2, the specific notions of reaction fragments and classification equations are considered in detail; these notions are needed to completely understand the central and most nontraditional level (i.e., interconversion type) in our classification scheme. The overall hierarchy is exposed in section 3; the main levels of this new hierarchy—**reaction category**, **reaction class**, **interconversion type**, **topological type**, and **reaction type**—are considered (together with corresponding subcategories, subclasses, and subtypes) in section 4 with more details. In the final section of this paper, some perspective applications (associated with the so-called Hierarchical Coding System for organic reactions) are outlined.

1. SOME PRELIMINARY NOTES

In the modern state Formal-Logical Approach, all chemical reactions are represented by *bond redistributions* (or *bond-and-sign redistributions*, cf. section 1 in ref 1) between initial and final systems. The main notions of this approach are visualized in Chart 2 by means of the following:

(a) two graphs, i.e., *topology identifying graph* G_{TOP} (cf. Chart 2a) and *expanded topology identifying graph* G_{ETOP}

Chart 2. The Main Notions of the Formal–Logical Approach Are Illustrated by Topology (a) and Expanded Topology (b) Identifying Graphs, Symbolic (c), Structural (d), Reaction (e), Skeletal (f) and Chemical (g) Equations Representing [1,5] Hydrogen Shift in the Bicyclo[4.3.0]nonadiene System^a



^a The graph (a) and equations (c, d, f, and g) are actual attributes in the modern state classification procedure.

(cf. Chart 2b). The edges and vertices of G_{TOP} correspond to all bonds which change their multiplicities in the course of the reaction and to all reaction centers (RCs), respectively. The graphs G_{ETOP} can contain additional vertices (corresponding to structural centers, cf. small circles in Chart 2b) and also additional edges corresponding to bonds which are unaffected in the course of reaction. The “isolated” additional edges (such as the dashed edge 4–6 in the graph of Chart 2b) and the chains formed from such edges (cf. the “dashed” chain 1–7–8–9–10–6) are required to connect reaction centers, e.g., RCs 4, 6 and 1, 6 in the example of Chart 2b.

(b) two equations, i.e., symbolic (Chart 2c) and structural (Chart 2d) equations, in which all bonds changing their multiplicities—and also additional bonds in the case of structural equations—are explicitly shown. In these equations, the reaction centers are represented by some abstract symbols (by heavy points in Chart 2c,d), while the structural centers—corresponding to vertices 7–10 of Chart 2b—by small circles.

(c) two equations, i.e., reaction (Chart 2e) and skeletal (Chart 2f) equations, in which the chemical nature of all centers is specified but simple as well as complex substituents are still unknown.

(d) the traditional chemical equation with all substituents (and hydrogen atoms) of the educt and product molecules being taken into account. In Chart 2g, the reversible sigmatropic rearrangement^{12a} of 2,5-dimethyl substituted bicyclo[4.3.0]nonadienes is shown as an example.

It is evident that two graphs and four equations of Chart 2a–f can be easily extracted from the traditional chemical equation (the procedure is exemplified in section 1 of ref 1). On the other hand, the above-mentioned graphs and equations can obviously be used as classification attributes. Let us initially analyze the left branch ($a \Rightarrow c \Rightarrow e \Rightarrow g$) of Chart 2 which visualizes the “old”^{9c} levels of our classifica-

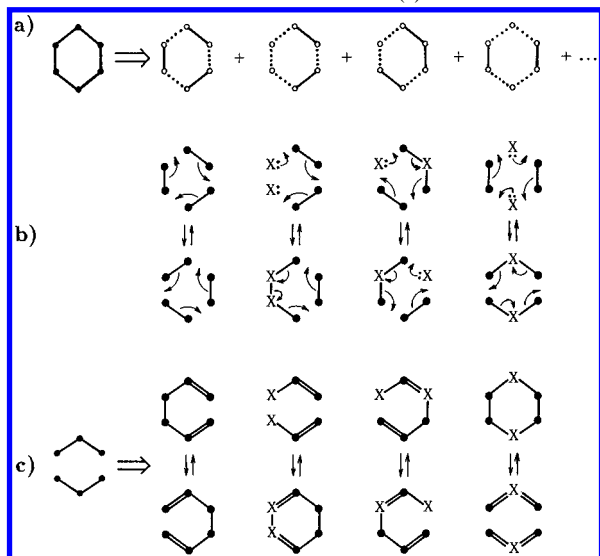
tion scheme. At this step, the reader can be convinced that symbolic and reaction equations represent just the same information as basis and hetero reactions in the Munich group approach (for that purpose, compare Chart 2c,e with the equations represented at levels 2 and 3 of Chart 1). As a result, two of our classification levels, i.e., **topological and reaction types**,^{9c} actually correspond to Ugi’s^{5b,e} RB- and RA-classes, to Fujita’s^{7b,c} Reaction Graphs and Reaction Center Graphs, etc.

The situation is completely different for the first level of Chart 1—no object directly corresponding to “electron redistribution scheme” can be found in Chart 2. Instead, the cyclic graph G_{TOP} of Chart 2a (which identifies the topology of electron transfer or, more strictly speaking, the topology of bond redistribution) appears as a classification attribute in our approach. The edges of the topology identifying graph specify only the bonds which are affected in the course of the reaction, and, for this reason, many “electron redistribution patterns” (similar to that of Chart 1a) can correspond to the same graph G_{TOP} . As an example, four of eight “electron redistribution patterns” (which are possible if only single bonds are allowed to be formed or broken, cf. ref 9b) are explicitly shown in Chart 3a. These patterns correspond to “trivial” symbolic equations, i.e., to the symbolic equations with all bonds being formed from zero or completely destroyed. Note that in three (“eight-electron”) equations of Chart 3b, two specific RCs—i.e., centers which increase or decrease their valency by two units—are designated by X; for other possible designations of RCs see section 2 in ref 1.

It is important that the above-mentioned σ -frames^{5d} or intact molecular subgraphs^{7a} can be successfully used as classification attributes (cf. Chart 1b,c) if and only if (a) the particular “electron redistribution pattern” is preselected and (b) this pattern is highly symmetric (otherwise several nonidentical superpositions are possible). In Chart 3c, we explicitly demonstrate that the superposition of one and the same σ -frame—i.e., that of Chart 1c—with “trivial” symbolic equations depicted in Chart 3b (or with corresponding electron redistribution patterns of Chart 3a) results in “nontrivial” symbolic equations of which the first and the third ones describe different sorts of sigmatropic rearrangements (i.e., [3,3] and [3,3 $\alpha\gamma$] shifts,¹³ respectively), while the others—cycloaddition–cyclofragmentation processes ([3 α +3 α] cycloaddition and [3 β +3 β] cyclofragmentation,¹³ respectively). That is why just symbolic equations (such as those of Chart 3b,c) but not their invariants are used as attributes in our classification system.

The other significant feature of our approach is that the specific level which reflects the perception of structural changes by an organic chemist (rather than by a computer) always was and still is an important part of the classification scheme. The idea of this level is based on the fact that many different symbolic equations (with not necessarily equal numbers of RCs) can be collected in one “generalized” type. The six-centered symbolic equations representing [1,5], [3,3], and [3,3 $\alpha\gamma$] sigmatropic shifts (cf. Charts 1 and 3c) can serve as examples; the above-mentioned—and additionally, all possible [1,2], [1,3], [2,3], etc.—shifts can evidently be declared to form the “generalized” type of sigmatropic rearrangements.

Chart 3. Four (of Eight Possible) “Electron Redistribution Patterns” Corresponding to Cyclic Graph G_{TOP} (a), Four “Trivial” Symbolic Equations (with 0 and 2 Specific RCs) Describing Six-Centered Processes with Cyclic Electron Transfer (b), and Four “Nontrivial” Symbolic Equations Corresponding to the Same Invariant σ -Frame but to Different “Electron Redistribution Patterns” (c)

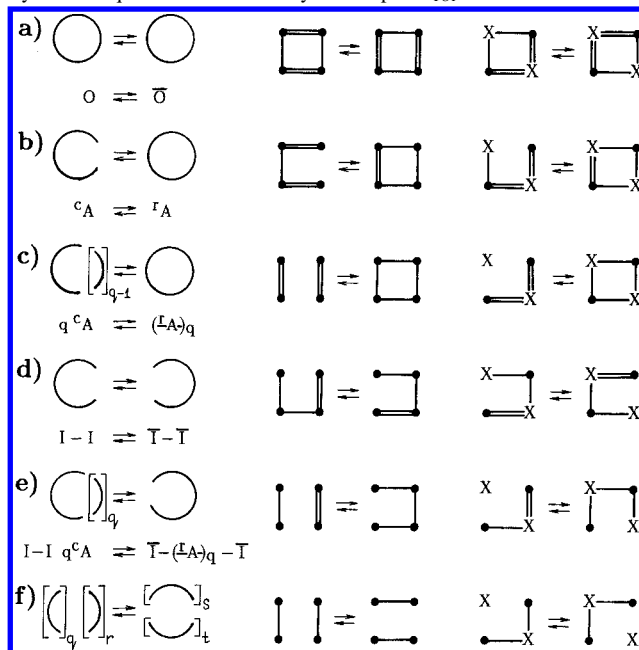


Two methods are used in the Formal–Logical Approach in order to collect the processes belonging to several different types (and hence characterized by nonidentical symbolic equations, cf. above) into the more general type. One of them is based on schematic pictures^{9d,11c} which reflect the types (and also numbers) of components in the left and right part of any symbolic equation. Thus, for the processes with cyclic topology of bond redistribution, the initial or final system can consist (1) of one cyclic component (“ring”), (2) of one linear component (“chain”), or (3) of two or more linear components. For this reason, only six “generalized” types, i.e., “ring–ring”, “ring–chain”, “ring–several components”, “chain–chain”, “chain–several components”, and “several components–several components”, are possible for pericyclic reactions with neutral educt and product molecules.

The above-mentioned six types are schematically represented in Chart 4a–f by interconversions of closed and disclosed contours. It is evident that Chart 4a relates to resonance of nonpolar structures (or to nonresonance “ring \rightleftharpoons ring” transformations such as those for substituted cyclobutadienes^{12b}), Chart 4b—to typical electrocyclic (“chain \rightleftharpoons ring”) reactions, Chart 4c—to q -component, $q \geq 2$, cycloadditions and cyclofragmentations, Chart 4d—to all sigmatropic (“chain \rightleftharpoons chain”) rearrangements, Chart 4e—to various (oligo)addition–(oligo)fragmentation¹⁴ processes, and Chart 4f—to various “cyclodismutation” reactions^{9a,b} in which both the initial and final systems consist of two or more components. The examples of Chart 4a–f represent the complete list of the very well-known^{5c,d,15} four-electron *isodesmic* processes (i.e., reactions with no specific RCs) and the incomplete list of much less investigated^{15b,c} six-electron *allodesmic* processes; in the symbolic equations related to allodesmic reactions, two of four centers (X) change their formal valencies by two units.

The very similar considerations result in classification of neutral processes with linear topology into six “generalized” types:^{10a,11d} (1) linear resonance, (2) linear (oligo)addition and (oligo)fragmentation, (3) linear (oligo)substitution, (4)

Chart 4. Schematic Representations of Six “Generalized” Types, Corresponding Classification Equations, and Examples of Four-Centered Symbolic Equations Related to Cyclic Graphs G_{TOP} ^a



^a Four-centered “ring–ring” transformations (a), [4] and $[4\alpha\gamma]$ electrocyclic reactions (b), $[2+2]$ and $[1+3\beta]$ cycloaddition–cyclofragmentation processes (c), $[1,3]$ and $[2\alpha,2\beta]$ sigmatropic rearrangements (d), $[2+(1,1)]$ and $[1+(1,2\alpha)]$ addition–fragmentation processes (e), $[(1,1)+(1,1)]$ and $[1+(1,1,1)]$ cyclodismutation processes (f) are visualized by symbolic equations which contain no specific RCs and two “nonadjacent” specific centers X.

linear (oligo)metathesis, (5) linear (oligo)substitutive (oligo)–addition and (oligo)fragmentation, and (6) linear (oligo)–dismutation. In the names of the above types, the optional prefix “oligo” relates to reactions with the enlarged numbers of components (for example, any addition process with $q > 2$ components can be regarded as an “oligo” process).

Although the schematic representations similar to those of Chart 4 can be easily drawn in simple cases, their application to reactions with more complex topologies of bond redistribution seems to be quite cumbersome. Thus, for the processes with linear–cyclic topology, the “generalized” types were, in fact, characterized^{10b,11d} by two symbols of which the first one related to the linear “part” and the second—to the cyclic “part” of the bond redistribution. For this reason, the more generally applicable (although somewhat more complicated) classification attributes were and still are used in the hierarchical system instead of schematic representations of Chart 4. The corresponding hierarchy level—**interconversion type**—needs, however, two specific notions, i.e., *reaction fragments* and *classification equations*, to be initially introduced, cf. the next section.

At this point, we only state that in many simple cases, there is a one-to-one correspondence between “generalized” types and classification equations (see, e.g., classification equations presented in the left part of Chart 4a–e). In the more complex cases, many classification equations can, however, correspond to a simple schematic picture. This is the case, for example, for cyclodismutation processes; two reactions of Chart 4f are characterized by different classification equations and hence belong to different interconversion types.

The next specific feature of the modern state classification scheme is associated with some new notions of the Formal–Logical Approach. Let us turn to the right branch of Chart 2 ($b \Rightarrow d \Rightarrow f \Rightarrow g$). Two equations of that branch—i.e., structural equation of Chart 2d and skeletal equation of Chart 2f—represent more detailed information on the reaction process of Chart 2g when compared with corresponding symbolic and reaction equations (cf. Chart 2c,e, respectively). In particular, it is obvious that the reaction under discussion is the [1,5] sigmatropic rearrangement process in the bicyclic system; none of the cycles is, however, opened or closed during this reaction. In the additional examples (see subsections 4.4 and 4.5), it will be shown that newly formed and disclosed cycles and also cycles which exist only in transition structures can be reflected in structural as well as skeletal equations. That is why the special hierarchy (sub)levels are associated with these equations in the modern classification scheme.

Thus, all processes related to a given structural equation (e.g., that of Chart 2d) are considered as belonging to the same **topological subtype**, and all processes related to a given skeletal equation (e.g., that of Chart 2f) are considered as belonging to the same **reaction type**. It is important to note that any topological subtype is actually a sublevel which corresponds to the definite topological type, e.g., to topological type of [1,5] sigmatropic rearrangement in the example under consideration.

In our last introductory comment, we mention that unshared electrons and electron pairs are not directly taken into account in the Formal–Logical Approach—the signed reaction centers (SRCs) are considered instead, cf. ref 1. The nature of chemical species (i.e., neutral molecules, ions, radicals, etc.) and the number of “extra” electrons (added to or removed from the initial system in the course of oxidation–reduction processes) are, in our opinion, very significant factors which need to be explicitly taken into account in the detailed classification scheme. That is why, in contrast to other classification systems, the specific attributes characterizing **category and subcategory** of the process reflect this information and represent a separate level and sublevel in the modern state classification scheme.

2. REACTION FRAGMENTS AND CLASSIFICATION EQUATIONS

Reaction fragments (RFs) can be observed in any reaction or symbolic equation if all newly formed and completely destroyed bonds are deleted from that equation. The simple procedure which makes it possible to recognize RFs is illustrated in Chart 5a for an ene/retroene reaction; in this example, there are exactly three “blocks” (i.e., hydrogen atom, ethylene component, and allylic group) from which initial and final systems are constructed. The RCs of the same fragment are necessarily connected by bonds of nonzero multiplicity (see bonds 2–3 and 4–5, 5–6 in the symbolic equation of Chart 5a), and these bonds can be regarded as “inner” bonds of the fragment under consideration. All other, “outer” bonds (represented by dashed lines in the symbolic equations of Chart 5a–c) exist either in the initial or the final system; these bonds of multiplicity 1, 2, or 3 typically connect RCs belonging to different reaction fragments.

In many situations, the outer bonds connect, however, two centers of the same RF; an example can be found in Chart

Chart 5. Recognition of RFs in Reaction and Symbolic Equations Is Illustrated for Ene/Retroene Reaction (a), 1,1-Dibromocyclopropane Debromination (b), and Nucleophilic Substitution (c) Processes^a

	Designations of reaction fragments	RN_{ini}^*	RN_{fin}^*	ρ^*	R^*	r^*
1	I_1	T_1	1	1	2	2
2	I_1	T_1	1	0	1	1
3	I_3	T_3	1	1	2	0
4	C_{A_1}	r_{A_1}	0	2	2	2
5	C_{A_1}	$r_{A_1}^+$	0	1	1	1
6	C_{A_2}	r_{A_2}	0	2	2	2
7	r_{A_3}	C_{A_3}	4	0	4	4

^a The designations and numerical characteristics (RN_{ini}^* , RN_{fin}^* , ρ^* , R^* , and r^*) of all fragments are listed in the table below this chart.

5b which represents the known dehalogenation method resulting in formation of allene and its derivatives. An intrafragment outer bond 1–2 (or *loop* 1–2) in both initial systems of Chart 5b evidently appears due to saturation of two “external” valencies starting at centers 1 and 2 of the fragment under consideration.

In the general case, the reaction fragment of order n consists of $n \geq 1$ reaction centers¹⁶ and exists in two forms corresponding to initial and final systems. These forms can evidently be considered as “large” RCs for which only external valencies (corresponding to outer bonds) should be taken into account. This fact shows that certain numerical characteristics of RCs (cf. section 2 in ref 1) are also applicable to RFs. There are four such characteristics of which one, i.e., *reaction number*, relates to a particular form of the reaction fragment. More precisely, the value of RN_{ini}^* (RN_{fin}^*) denotes the total multiplicity of all outer bonds in an initial (or final) form of the fragment under consideration.

Three other numerical characteristics, i.e., ρ^* , R^* , and r^* , are very similar to local characteristics of RCs (ρ , R , and r ; cf. section 2 in ref 1) but relate to the fragment as a whole. Thus, the *degree* ρ^* denotes the total number of external valencies in both forms of any RF; the multiplicities of corresponding outer bonds are not taken into account in this case. The rank R^* value is equal to the total change of multiplicities for outer bonds; the very simple considerations (based on the fact that outer bonds of any RF change their multiplicities either from zero to 1, 2, 3 or from 1, 2, 3 to zero) show that $R^* = RN_{ini}^* + RN_{fin}^*$ for all reaction fragments. The last numerical characteristic of the RF, i.e., its *specificity* r^* , is defined by the expression $r^* = |RN_{ini}^* - RN_{fin}^*|$; just the values of r^* make it possible to classify reaction fragments into several types, cf. below.

The numerical characteristics of one-, two-, and three-centered RFs having been recognized in reaction and symbolic equations of Chart 5a,b are listed in rows 1, 3, 4, 6, and 7 (see the table below Chart 5). Only the values in

the seventh row of this table should be additionally commented. At first sight, the values of RN_{ini}^* , ρ^* , R^* , and r^* for three-centered fragment of Chart 5b should be equal to 3 because three outer bonds (see dashed lines) are broken in both equations representing 1,1-dibromocyclopropane dehalogenation reaction. One of these bonds (i.e., that connecting centers 1 and 2 in an initial system) is, however, a loop which corresponds to two external valencies of the reaction fragment. For this reason, the correct values of RN_{ini}^* , ρ^* , R^* , and r^* for the fragment under discussion are all equal to 4.

The fact that any fragment really behaves as a "block" or enlarged reaction center (with the bonds being either formed from zero or completely destroyed) makes it possible to partition all RFs into several types in accordance with corresponding values of r^* ; the very similar classification of RCs—i.e., in accordance with specificity values r —was discussed in section 2 of the preceding paper.¹ It should be mentioned here that in neutral processes—i.e., in processes with no charged or radical centers being involved—only reaction centers (and hence only RFs) with even specificity values are formally possible.¹ For this reason, the fragments which contain no charged or radical centers will be initially considered.

Thus, the fragments with $r^* = 0$ (presuming that $RN_{ini}^* = RN_{fin}^* \neq 0$) are analogs of ordinary reaction centers. These RFs are called *inversion*¹⁷ or *I-fragments* and are designated by symbol *I* (or I_n if the order is explicitly specified). The univalent hydrogen atom and allylic group in Chart 5a are typical examples of I_1 - and I_3 -fragments with degree and rank values equal to 2. Note that two forms of the same *I*-fragment need not be identical. If these forms need to be differentiated, then the designation of one (traditionally, final) form is overlined; compare designations *I* and \bar{I} or I_n and \bar{I}_n .

The fragments with $r^* = 2$ (A - or A_n -fragments), $r^* = 4$ (A' - or A'_n -fragments), $r^* = 6$ (A'' or A''_n -fragments), etc. are analogs of specific (X), dispecific (X'), trispecific X'' , etc. reaction centers which specificity values are also equal to 2, 4, 6, ..., respectively. All these fragments are considered as *addition fragments*¹⁸ (in the broad sense of this word) and are partitioned into strictly addition fragments (A), doubly addition fragments (A'), etc. It is important that two forms of addition fragments can always be differentiated by their reaction numbers RN^* . The forms with minimal values of RN^* are called *chain* or *c-forms* (cA_n , $^cA'_n$, $^cA''_n$, ...); the typical example is represented by an ethylene unit which forms the two-atom chain in initial systems of Chart 5a. The forms with maximal values of RN^* are considered as *ring* or *r-forms* (rA_n , $^rA'_n$, $^rA''_n$, ...) of the fragments under discussion. In this case, the outer bonds of rA_n , $^rA'_n$, and other addition fragments either form rings (for example, three-membered ring in initial systems of Chart 5b) or connect RCs of the given fragment with the centers belonging to other RFs (e.g., with hydrogen atom and carbon atom of the inverted allylic group; see the example in Chart 5a).

An additional type of RFs is represented by fragments with $RN_{ini}^* = RN_{fin}^* = r^* = 0$. These fragments are called *O*- or *O_n-fragments* (the letter "O" resembles zero which is, in turn, associated with the number of outer bonds in *O*-fragments). It is evident that in both forms, i.e., O_n and \bar{O}_n , the fragments under discussion contain only inner bonds, and, for this reason, they are associated either with nonpolar resonance

structures (similar to Kekulé structures of the benzene skeleton) or with initial and final systems which interconversion needs no bonds to be formed from zero or completely destroyed. The trivial examples are illustrated by symbolic equations in Chart 4a.

It is important to note that all above-mentioned designations, i.e., I_n , \bar{I}_n , cA_n , rA_n , $^cA'_n$, $^rA'_n$, ..., O_n , \bar{O}_n identify type, particular form, and order (n) of any RF but contain no information on number, multiplicity, and relative disposition of outer (and surely, inner) bonds in an n -centered fragment. The more informative coding designations or *codes*¹⁹ for particular forms of *I*-, *A*-, and *O*-fragments were suggested in ref 11b; these codes make it possible to additionally recognize multiplicities of ρ^* outer bonds in the both forms of any reaction fragment.

Up to this moment, only interconversions of neutral systems were considered. In many reactions and key steps of reactions, an even number of charged and radical RCs (signed centers, cf. sections 1 and 2 in ref 1) can, however, participate. These centers are characterized by *odd specificity (r) values*¹ and are classified into types in accordance with these values and with disposition of signs "+", "•", or "−". Thus, if the RC with the minimal reaction number (or formal valency) RN is signed, then this signed center is called ordinary (if $r = 1$), specific (if $r = 3$), dispecific (if $r = 5$), etc.; a trivial example of an ordinary center—negatively signed in the final system—is represented by a bromine atom in the nucleophilic substitution process of Chart 5c. In contrast, the signed centers with $r = 1, 3, 5, \dots$ are pseudospecific, pseudodispecific, pseudotrispecific, etc. if the maximal value of RN and the sign appear in the same (initial or final) system. In order to differentiate "pseudo" specific RCs, the standard designations (X , X' , X'' , etc.) are underlined; an example of pseudospecific center \underline{X} (corresponding to nitrogen atom in the nucleophilic substitution process) can be found in Chart 5c.

The *signed reaction fragments* can be considered as analogs of either unsigned or signed RCs. Typically, the signed fragments contain²⁰ (1) a single signed RC in one form and no signed RCs in the other form, (2) two signed RCs in one form (and no signed RCs in the other form), or (3) one signed RC in the both forms of the fragment.

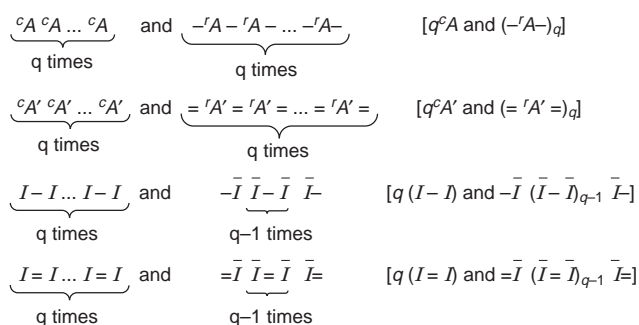
In the last two cases, the signed RFs are characterized by even specificity values,²¹ $r^* = 0, 2, 4, \dots$, and hence can be considered as analogs of unsigned RCs. These signed fragments are classified into several types (*I*, *A*, *A'*, ...) in accordance with specificity values only. The signs "+", "•", "−" must not be taken into account in the classification procedure but need to be explicitly presented in designations of one (in case 2) or both (in case 3) forms of RFs. Note that *O*-fragments are possible in both situations. The RF representing the ethylene unit and its dipolar form (O_2 and \bar{O}_2^{+-} , cf. Chart 9h) as well as two identical forms (O_3^- and \bar{O}_3^-) of the fragment corresponding to resonance structures of allyl-anion can serve as simple examples of the "no bond" signed fragments.

In the first of the above-mentioned cases, the signed reaction fragments can be proven (cf. note 21) to be characterized by odd specificity values. The classification of these fragments is very similar to the classification of signed RCs. Thus, the signed n -centered RF with $r^* = 1$,

3, 5, etc. is called inversion (I or I_n), addition (A or A_n), double addition (A' or A'_n), etc. if the form with the minimal RN^* value (i.e., c -form) is signed. In contrast, the “pseudo” addition fragments (\bar{A} or \bar{A}_n , \bar{A}' or \bar{A}'_n , \bar{A}'' or \bar{A}''_n , etc.) correspond to specificity values $r^* = 1, 3, 5, \dots$ if r -forms of the RFs under discussion are signed. Note that in designations, the needed symbols “+”, “•”, and “−” as well as symbols c and r (for chain and ring forms of all addition and pseudo addition fragments) must be explicitly presented. The examples of the simplest, one-centered inversion and pseudoaddition signed RFs are represented by bromine and nitrogen atoms in Chart 5c; the designations of both forms and numerical characteristics RN^*_{ini} , RN^*_{fin} , ρ^* , R^* , and r^* for these fragments can be found in the second and fifth rows of the table below Chart 5.

If in initial and final systems (of symbolic or reaction equations) all unsigned as well as signed fragments are substituted by their designations, then the *fragment systems* are obtained. In these systems, only the outer bonds (connecting different RFs) and also loops²² (corresponding to bonds which connect two centers of the same RF) are explicitly presented. The initial and final fragment systems ${}^cA_2 I_1 - I_3$ and $\bar{I}_1 - {}^rA_2 - I_3$ (cf. with reaction and symbolic equations of Chart 5a) and also ${}^cA_1 I_1 - I_1$ and ${}^rA_1 + - \bar{I}_1 \bar{I}_1$ (cf. with equations of Chart 5c) can serve as typical examples.

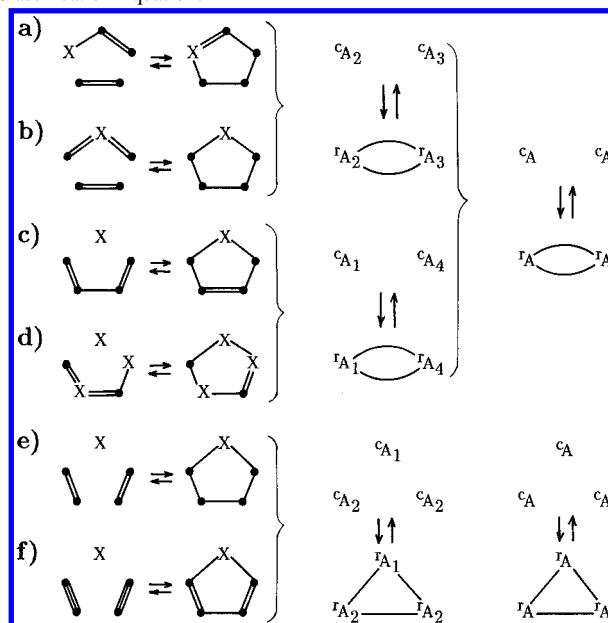
The equations which describe interconversions of fragment systems are called *classification equations*; the number of RFs in any of these equations evidently cannot exceed the number of RCs in the corresponding symbolic or reaction equation. There are three types of classification equations. In *particular* (and *standard*) equations, the orders of all fragments are explicitly indicated (or omitted, respectively). In *generalized* classification equations, the sequences of similar RFs of degree 2 are “joined together” by means of repetition coefficients. The following sequences of similar fragments are typical:



The application of classification equations is illustrated in Chart 6 for five widely known²³ and one unknown²⁵ types of five-centered cycloaddition–cyclofragmentation reactions. The direct processes corresponding to symbolic equations of Chart 6a,b and Chart 6c–f evidently represent 1,3-dipolar and cheletropic (two-component and three-component) additions, respectively.

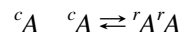
The examples of Chart 6 convincingly demonstrate that all processes differed only by internal structures of one or more n -centered, $n > 1$, fragments necessarily relate to just the same particular classification equation. Three equations in the second column of Chart 6 actually define intercon-

Chart 6. Five-Centered $[2+3\alpha]$ (a), $[2+3\beta]$ (b), $[1+4]$ (c), $[1+4\alpha\gamma]$ (d), $[1+2+2]$ (e), and $[1+2'+2']$ (f) Cycloaddition–Cyclofragmentation Processes Are Represented by Symbolic Equations and Corresponding Particular (in the Second Column) and Standard (in the Third Column) Classification Equations



version subtypes in the modern state classification system.

In standard classification equations, depicted in the third column of Chart 6, the numbers of centers in RFs are ignored. For this reason, not only five-centered processes of Chart 6a–d but the well-known^{24a,b} three-centered $[1+2]$, four-centered $[2+2]$, and six-centered $[2+4]$ as well as many other two-component cycloaddition–cyclofragmentation reactions are covered by the “two-fragment” equation



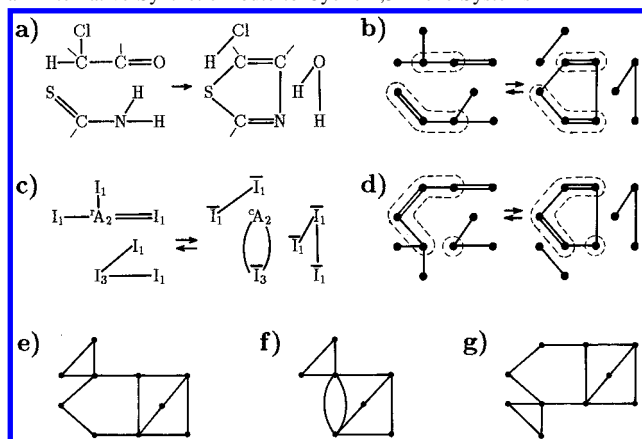
In the general case, it is possible to collect all related processes with equal numbers of similar RFs into the same interconversion type.

If the numbers of similar fragments are not directly taken into account, then the next generalization level can be reached. Thus, in all two- and multicomponent cycloaddition–cyclofragmentation reactions of Chart 6, the unique equation $q {}^cA \rightleftharpoons ({}^rA -)_q$ evidently corresponds, and this fact shows that just generalized classification equations must be considered as typical analogs of the schematic pictures in the left part of Chart 4.

It is easy to see that any classification equation (which contains at least one interfragment bond in an initial or final system) can be regarded as an analog of some “trivial” symbolic or reaction equation.²⁶ For this reason, the structural characteristics (having been introduced in section 3 of ref 1 for bond redistributions in reaction systems) are also applicable to bond redistributions in the fragment systems. This fact does not mean, however, that particular characteristics related to some symbolic (or reaction) equation and to a corresponding classification equation must necessarily coincide.

Let us consider, for example, the five-centered symbolic equation of Chart 6f—it evidently relates to a canonical,

Chart 7. The Hantzsch Thiazole Synthesis is Represented by Reaction (a), Symbolic (b), and Classification (c) Equations, The Substitution of Original Two-Centered Addition and Three-Centered Inversion Fragments (in the Symbolic Equation of Hantzsch Reaction) for Four-Centered Addition and One-Centered Inversion Fragments Is Shown To Produce the New Symbolic Equation (d) Which Represents an Alternative Synthetic Route to Cyclic 1,3-Diene Systems^a



^a Topology identifying graphs (e–g) corresponding to equations of Chart 7b–d are demonstrated to be nonisomorphic.

nonbasic, nontrivial, derived (nonprincipal), allodesmic, single mode bond redistribution. The corresponding three-fragment particular and standard classification equations also represent canonical, allodesmic (because A-fragments are analogs of centers X), single mode bond redistributions which must, however, be considered as basic, trivial, and principal.²⁷

The characteristics associated with topologies of bond redistribution can also be different; this conclusion directly follows from the fact that topology identifying graphs (G_{TOP} , cf. section 1) related to symbolic and classification equations need not be necessarily isomorphic or even similar. Three nonisomorphic graphs G_{TOP} corresponding to symbolic equations of Chart 7b,d and to the classification equation of Chart 7c are visualized in Chart 7e,g and Chart 7f, respectively.

Our last comment on reaction fragments and classification equations is associated with their applications to reaction design problems. The simple consideration shows that there is no sense to construct complete sets of symbolic equations *via* intermediate classification equations corresponding to a given topology identifying graph. The main arguments are based on the fact that the resulting symbolic equations can correspond to the very different topology identifiers (which must not be necessarily isomorphic to the initial graph G_{TOP} ; cf. three graphs of Chart 7e–g). The additional difficulties are associated with symmetry recognition in individual fragments and also in topology identifying graphs which contain parallel edges or loops²⁸ (the topology identifier of Chart 7f can serve as an example of such graphs). The above reasons lead to a conclusion that the application of reaction fragments and classification equations to design problems needs more specific techniques to be elaborated. The main ideas of quite trivial “analogy based” approach are briefly considered below.

In the simplest situation, one or more RFs from a given symbolic or reaction equation can be substituted by new fragments with the same numerical characteristics (RN_{ini}^* , RN_{fin}^* , ρ^* , R^* , and r^*). The resulting symbolic equations must not necessarily be similar to the starting one. In Chart

7, this fact is exemplified by transformation of starting symbolic equation (representing the Hantzsch thiazole synthesis,^{24c} cf. equations of Chart 7a,b) into the symbolic equation of Chart 7d. For this purpose, the two-centered addition fragment (with $RN_{ini}^* = 4$, $RN_{fin}^* = 2$, $\rho^* = 5$, $R^* = 6$, and $r^* = 2$) and three-centered inversion fragment (with $RN_{ini}^* = RN_{fin}^* = 2$, $\rho^* = 4$, $R^* = 4$, $r^* = 0$) are substituted by new fragments of order 4 and 1



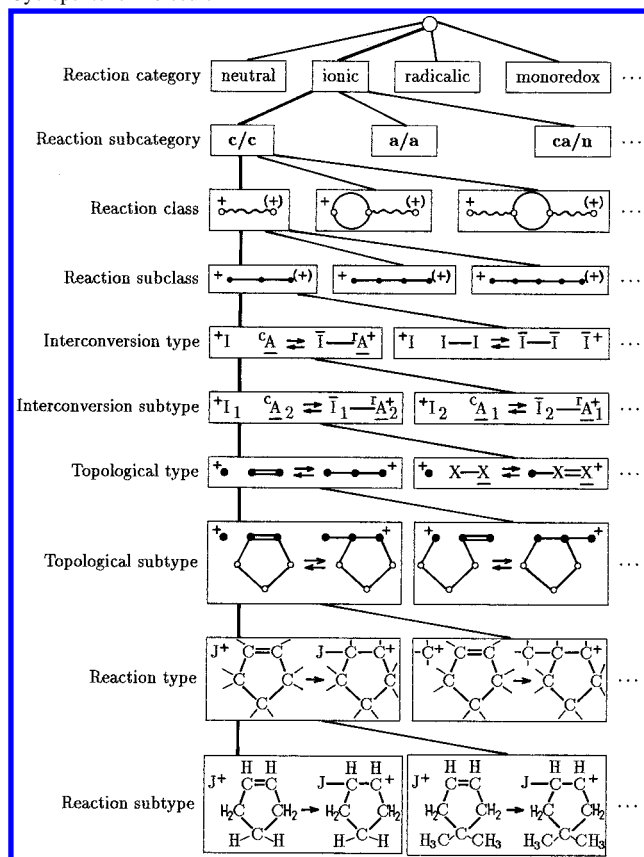
with the same numerical characteristics. The resulting symbolic equation of Chart 7d evidently represents the perspective synthetic route to five-membered carbo- or heterocyclic systems (e.g., to cyclopentadienes starting from γ -chlorosubstituted α,β -unsaturated ketones and compounds with an active methylene group).

In the general case, new symbolic or reaction equations can be obtained from the given particular or standard classification equations (e.g., from the equation of Chart 7c) if all symbols of RFs are substituted by appropriate “synthetically significant” fragments of the preselected set. In order to avoid a combinatorial explosion, it seems reasonable to consider only the RFs which participate in *real organic transformations*; these fragments can be extracted, e.g., from the reaction data base. The computer implementation of the approach under consideration needs, however, several theoretical subproblems to be initially solved. These subproblems are associated with the effective coding system for multi-centered fragments, with a description of their symmetry, and with a formulation of specific rules making it possible to recognize the allowed combinations of RFs in symbolic or reaction equations. The elaboration of rigorous mathematical models for the “analogy based” approach to reaction design problems is also planned, and the results will be published in our future papers.

3. HIERARCHY LEVELS ILLUSTRATED

At this point, the modern state classification scheme can at last be considered as a whole. All hierarchy levels and corresponding sublevels are exposed in Chart 8 in reference to the typical process with linear electron transfer,^{10c,d} i.e., to electrophilic addition of an iodine cation (I^+) to a neutral cyclopentene molecule. The first hierarchy level and sublevel—reaction category and corresponding subcategory—reflect the fact that in an example process, only ions (together with neutral cyclopentene structure) are involved. The designations “c/c”, “a/a”, and “ca/n” denote the presence of positively (“c” = cation) or negatively (“a” = anion) charged centers in initial and final systems; the symbol “n” corresponds to the neutral final system. The next level and sublevel—reaction class and subclass—are represented by topology identifying contours and graphs; note that in the case of non-neutral processes, these contours and graphs necessarily afford information on relative disposition of charged and/or radicalic centers. In the “c/c” example under consideration, the symbols “+” and “(+)” show that a positively charged atom appears in an initial or final system, respectively. All other hierarchy levels and sublevels are characterized in Chart 8 by the above-discussed equations, i.e., by standard and particular classification equations, by

Chart 8. The Tree-like Structure of the Classification Scheme Is Illustrated for Electrophilic Addition of Iodine Cation to Neutral Cyclopentene Molecule^a



^a The hierarchy attributes corresponding to this process are linked by fatty lines.

symbolic and structural equations, and by skeletal and chemical equations. Of these equations, only one uniquely corresponds to the 2-iodocyclopentylcation formation process; all structural (but not stereochemical) features of this process are evidently associated with its reaction subtype; cf. the lowest row in Chart 8.

The main feature of the suggested scheme is that it can obviously be applied to chemical interconversions of *any level of complexity*. This means that elementary reaction steps, "normal" one-step processes with no isolable intermediates, and complex multistep transformations are treated in just the same manner. Moreover, the applications of the classification procedure are in no case limited to processes which can be simulated by appropriate computer-assisted generation techniques. (For example, ionic reactions with four or more charged atoms can easily be classified, although the generation of corresponding equations is not covered by combinatorial models discussed in section 6 of ref 1.) As a result, our classification scheme needs only two important requirements to be necessarily fulfilled: (1) all educts and products of the reaction (including those related to inorganic substances) must be represented by definite resonance structures with localized bonds, charges, and free electrons and (2) the redistribution of all chemical bonds (characterized by multiplicity changes in the course of the reaction process) must be exactly specified. These limitations directly follow from the nature of the Formal-Logical Approach, cf. ref 1.

The next important feature of the classification scheme is that almost all hierarchy levels (with the reaction category

being an exception) can be, in fact, characterized by some combinatorial objects, and the complete sets of these objects may be more or less directly constructed starting from the object which corresponds to the preceding level or sublevel. This fact results in the "tree-like" pictures (cf. Chart 8) in which almost all classification attributes, i.e., contours, graphs, and equations, contain the complete information produced by all higher hierarchy levels. This principle is slightly violated for interconversion type and subtype because from classification equations, some part of structural information (e.g., the topologies of bond redistribution, the "internal" structures of RFs, and even the total numbers of RCs in the case of standard equations) cannot be extracted. This fact shows that the interconversion level and sublevel may be regarded as specific tools which are very useful for methodology purposes but seem to be hardly applicable to the solution of problems associated with storage and retrieval of individual organic reactions.

An attentive reader can note that there are two real "sorts" of hierarchy levels in the suggested classification scheme. The levels of the first sort—i.e., reaction category, class, interconversion type and subtype—are not well-adapted for computer-assisted manipulations but make it easily possible to recognize some similarity in organic reactions with the same or different numbers of reaction centers. Thus, the processes belonging to the same reaction category are characterized by the same nature of educts and products (only neutral molecules, neutral molecules plus ions, etc.). The processes belonging to the same reaction class are also characterized by very similar topologies of bond redistribution (e.g., linear, cyclic, linear-cyclic, bilinear-cyclic, etc.). The reactions belonging to the same interconversion type or subtype possess an even higher degree of structural similarity which is typically regarded by organic chemists as "the same type of reaction". Two- and three-component cycloaddition-cyclofragmentations of Chart 6 (the processes with cyclic electron transfer) and electrophilic addition-fragmentation reactions represented in Chart 8 (the processes with linear electron transfer) can serve as convincing examples.

In contrast, the attributes associated with all other levels of our classification scheme—i.e., with reaction subcategory and subclass, with topological type and subtype, and with reaction type and subtype—are applicable not only to methodology purposes but also to computer-aided manipulations. For the latter purpose, the special Hierarchical Coding System has recently been elaborated; this system makes it possible to sequentially recognize in computer memory all information related to subcategory designations, topology identifying graphs, symbolic, structural, skeletal, and chemical equations. The example which illustrates the main principles of the Hierarchical Coding System is briefly considered in the last section of this paper.

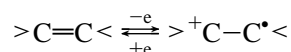
4. ADDITIONAL COMMENTS AND EXAMPLES

In this section, all hierarchy levels of the suggested classification scheme are discussed with more details. The additional examples and explanations are expected to clarify some specific features and applications of our classification principle.

4.1. Reaction Categories and Subcategories. The notion of *reaction category* is introduced in order to take

into account the nature of chemical species involved in the process and also the number of "extra" electrons which can be added to or removed from initial systems.

In the case of oxidation–reduction reactions, the categories—monoredox, diredox, triredox, etc.—are completely defined by numbers of "extra" electrons; note that in corresponding equations, the signed RCs (i.e., either only charged or charged and radical reaction centers) are necessarily involved. The formation of ethylene cation–radical and the reverse process



can serve as typical examples of oxidative and reductive reactions.

The processes in which no atoms change their oxidation states are sometimes called *isohypsic*; see, e.g., refs 4. In the Formal–Logical Approach, this notion is used in a somewhat broader sense; the bond-and-sign redistribution is named isohypsic if no "extra" electrons are involved in corresponding reaction process. In isohypsic case, the initial and final systems (with only reaction centers being taken into account) can evidently be formed out of (a) neutral (i.e., uncharged and nonradical) molecular substructures, (b) organic or inorganic ions (together with or without neutral substructures), (c) organic or inorganic radicals (together with or without neutral substructures), and (d) ions, radicals, and/or ion–radicals;³⁰ the components corresponding to neutral molecules can also be present or absent in reaction systems. The above consideration shows that only four reaction categories—i.e., neutral, ionic, radicalic, and ion–radicalic—are formally possible in the case of isohypsic reactions.

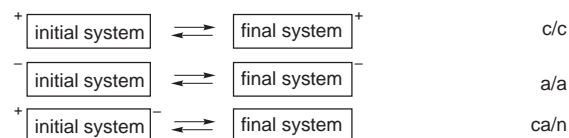
The *reaction subcategory* represents even more detailed information on the number and nature of signs "+", "•", and "−". The corresponding designations consist of two parts separated by a slash and related to initial and final systems, respectively. In the left part, the symbols "c", "r", and "a" ("c" = cation, "r" = radical, "a" = anion) must be arranged in nondescending order ("c" < "r" < "a"); the symbol "n" ("n" = neutral) is conventionally associated with the final system. It should be noted that in any subcategory designation, the total number of charge and radical symbols is always even; the brief explanation may be found in note 13 of ref 1.

In contrast to reaction categories, the subcategories can be associated with some combinatorial objects,³¹ and this fact allows the subcategory designations to be systematically generated and analytically enumerated. It is evident that a single ("n/n") subcategory relates to interconversions of neutral systems. The combinatorial objects characterizing subcategories with exactly two charge and radical symbols were enumerated (cf. note 31), constructed, and then classified in accordance with possible reaction categories. The results (represented by schematic pictures and subcategory designations) are shown in Scheme 1.

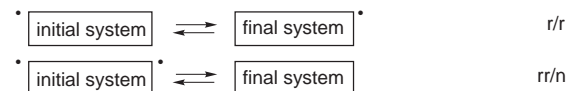
The subcategories associated with ion–radical (as well as with "multiredox") processes cannot be found in the above list. This fact shows that these processes are possible if and only if four or more charge and radical symbols are present in the subcategory designation.

Scheme 1

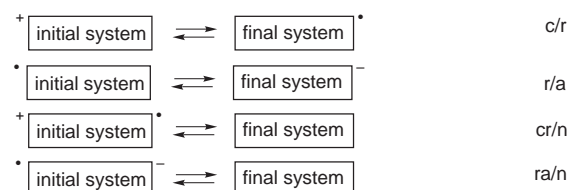
The subcategories corresponding to ionic reactions:



The subcategories corresponding to radicalic reactions:



The subcategories corresponding to monoredox reactions:



The subcategories corresponding to diredox reactions:

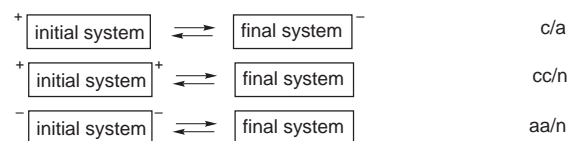
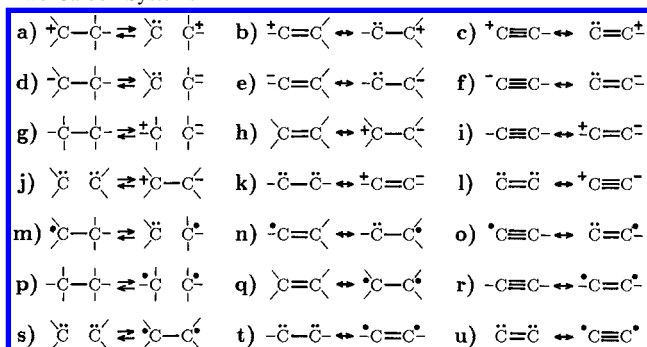


Chart 9. The Complete Lists of Two-Centered c/c– (a–c), a/a– (d–f), ca/n– (g–i) Ionic as well as r/r– (m–o) and rr/n– (p–u) Radicalic Processes Are Represented by Interconversions of Two-Carbon Systems^a

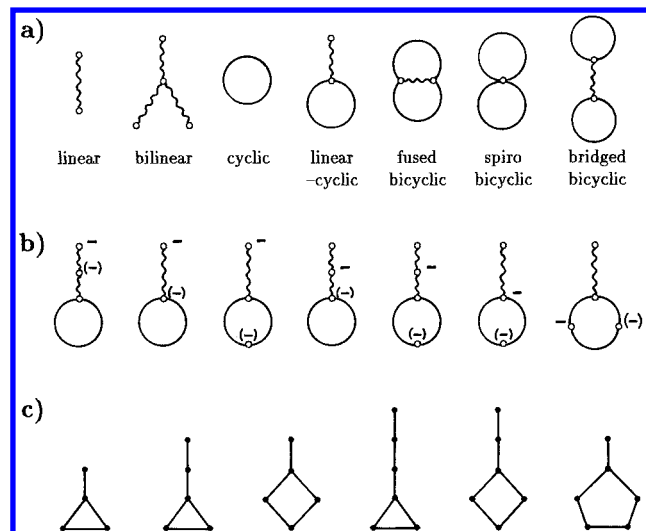


^a The dissociation–recombination reactions (cf. the first column of this chart) and transformations of resonance structures (cf. the second and third columns) are the only isohypsic single mode two-centered processes with di-, tri-, and tetravalent carbon atoms being involved.

In order to exemplify the application of categories and subcategories to classification purposes, let us consider the complete list of symbolic equations³² which correspond to the simplest two-centered processes with linear electron transfer^{10c,d} (in other words, to two-centered, noncanonical, linear processes with the multiplicity of the bond between centers being changed by one unit). Of 63 possible equations (cf. note 32), 12, 9, 24, and 18 symbolic equations relate to ionic, radicalic, monoredox, and diredox categories, respectively. The equations which describe 21 isohypsic, i.e., ionic and radical, reactions are represented in Chart 9 by interconversions of two-carbon systems; the valencies of carbon atoms are allowed to be equal to 4, 3, and 2 in all reaction equations of this chart.

The dissociation–recombination processes as well as resonance transformations of Chart 9 can evidently be classified in accordance with five possible subcategories (cf.

Chart 10. Several Reaction Classes Are Represented by Unsigned (a) and Signed (b) Topology Contours; the Four-, Five-, and Six-Centered Graphs (c) Characterize Subclasses Corresponding to the Fourth Contour of Chart 10a



the above list). It is easy to see that “c/c”, “a/a”, “ca/n”, “r/r”, and “rr/n” subcategories are represented by reaction equations of Chart 9a–c, d–f, g–l, m–o, and p–u, respectively.

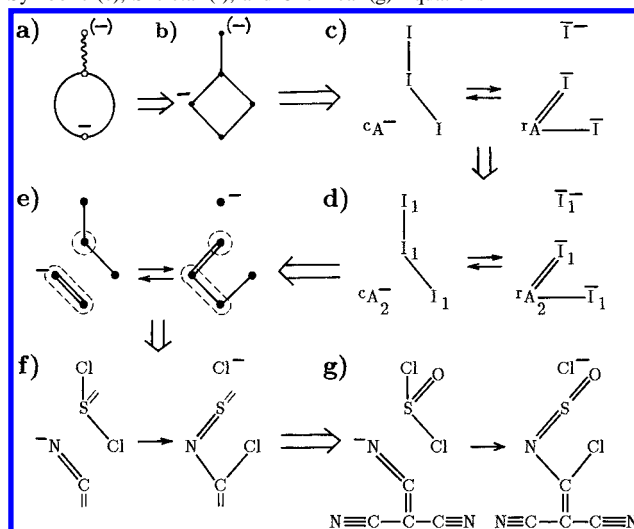
4.2. Reaction Classes and Subclasses. These hierarchy levels are associated with topology of electron transfer (or with topology of bond-and-sign redistribution, in the more exact terminology). The difference between classes and subclasses follows from the fact that *reaction classes* are characterized by more or less complex contours formed by unspecified (e.g., curvilinear or wavy) lines, while corresponding *reaction subclasses*—by topology identifying graphs G_{TOP} which edges uniquely correspond to all bonds changing their multiplicities in the course of the reaction.

In topology contours related to the neutral category (n/n subcategory), the nodes are associated with those RCs which degree values are not equal to 2; in schematic representations of such contours (cf. Chart 10a), all nodes are visualized by small circles being incident with 1, 3, 4, ...curvilinear or wavy lines. The additional nodes can, however, appear in signed topology contours which relate to non-neutral processes. The signed nodes of any degree must be present in this case. In the example of Chart 10b, all possible linear–cyclic contours related to a/a subcategory are explicitly shown; two nodes signed by “–” and “(–)” denote reaction centers which are negatively charged in initial and final systems, respectively.

In contrast to topology contours,³³ the unsigned and signed graphs G_{TOP} are typical combinatorial objects; the complete lists of unsigned topology identifiers can be successfully constructed by standard graph generation techniques (e.g., by the highly effective Faradjev’s technique, cf. the key ref 23b in the preceding paper). As an example, all possible unsigned “linear–cyclic” (i.e., corresponding to the fourth contour of Chart 10a) graphs with 4, 5, and 6 vertices are explicitly represented in Chart 10c.

In order to illustrate how the notions of reaction class and subclass can be applied to a real organic transformation, let us consider the α -chloro- β , β -dicyanovinylsulfanylamine formation process^{34a} (cf. Chart 11g). All bonds changing their multiplicities in the course of this reaction evidently form

Chart 11. Classification Attributes Corresponding to α -Chloro- β , β -dicyanovinylsulfanylamine Formation Process: Signed Linear-Cyclic Topology Contour (a); Signed Topology Identifying Graph (b); Standard (c) and Particular (d) Classification Equations; Symbolic (e), Skeletal (f), and Chemical (g) Equations^a



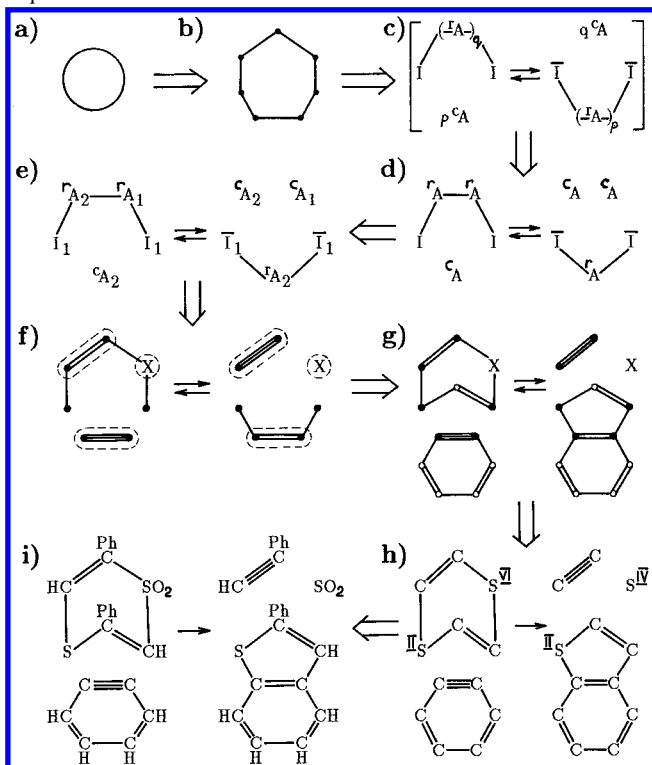
^a Note that in this example, the structural equation is identical to symbolic equation due to the absence of any cycles in initial and final systems. For the same reason, reaction and skeletal equations also coincide.

unsigned linear–cyclic contour (i.e., the fourth contour of Chart 10a) and unsigned linear–cyclic graph (i.e., the third graph of Chart 10c). The process under discussion is, however, ionic and belongs to a/a subcategory, and, for this reason, both the topology identifying contour and graph must be signed by two symbols of negative charge. The signed topology contour and graph G_{TOP} are visualized in Chart 11a and 11b, respectively; just these contour and graph represent the class and subclass to which the reaction process of Chart 11g corresponds.

It should be stressed here that symbols “–” and “(–)” are in some sense “interchangeable” (cf. section 4 in ref 1). This fact leads to a conclusion that the signed contour of Chart 11a and the third signed contour of Chart 10b (and also graph G_{TOP} of Chart 11b and its counterpart with symbols “–” and “(–)” being interconverted) must be considered as isomorphic. The rigorous combinatorial model which allows for generation of all nonisomorphic signed topology identifiers (starting from “parent” graphs similar to those of Chart 10c) is discussed in subsection 6.1 of the preceding paper.

4.3. Interconversion Types and Subtypes. The attributes which characterize *interconversion types* and *subtypes*, i.e., standard and particular classification equations, were discussed in section 2 with all necessary details. The specific role of these equations in the hierarchical system follows from the fact that in classification equations, only the most significant, i.e., “newly formed” and “completely destroyed” bonds, are taken into account. Disregard for all “inner” bonds of RFs undoubtedly leads to partial loss of structural information in classification equations which are therefore poorly adapted to solution of storage and retrieval problems for organic reactions. But on the other hand, just classification equations make it possible to recognize similarity in the very unlike processes (e.g., in those with different numbers of RCs), and this fact makes it possible to regard

Chart 12. Classification Attributes Corresponding to 2-Phenylbenzo[*b*]thiophene Formation Process: Cyclic Topology Contour (a); Seven-Centered Topology Identifying Graph (b); Generalized (c), Standard (d), and Particular (e) Classification Equations; Symbolic (f), Structural (g), Skeletal (h), and Chemical (i) Equations^a



^a The equation in square brackets (c) is associated with the optional hierarchy level of the classification scheme.

interconversion types and subtypes as “methodologically” central hierarchy levels in our classification system.

The position of interconversion types and subtypes in the suggested classification scheme can be observed in the (fifth and sixth rows of Chart 8) and also in Chart 11c,d; in both examples, charged reaction fragments are incorporated into the initial and final fragment systems. The same hierarchy levels but represented by interconversions of unsigned fragment systems can be found in Chart 12. The difference between the standard equation of Chart 12d (representing interconversion type) and the particular equation of Chart 12e (representing interconversion subtype) is that in the second equation, the numbers of reaction centers in all RFs are explicitly indicated. Both classification equations of Chart 12d,e correspond to the neutral (*n/n*), cyclic, seven-centered 2-phenylbenzo[*b*]thiophene formation^{34b} process, cf. Chart 12i.

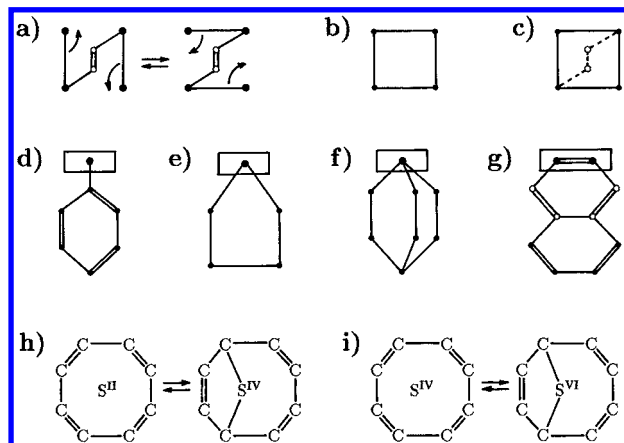
It should be mentioned here that generalized classification equations are typically not associated with some hierarchy levels of our classification scheme. In special situations, these equations can, however, be optionally regarded as attributes corresponding to an additional hierarchy level, i.e., “interconversion supertype”. The example is depicted (in square brackets) in Chart 12c. The generalized equation of Chart 12c evidently accumulates all cyclodismutation processes with cyclic electron transfer in which one sequence of (say *p*, *p* ≥ 1) addition RFs simultaneously substitutes the other sequence of (e.g., *q*, *q* ≥ 1) addition RFs at two inversion fragments.

4.4. Topological Types and Subtypes. *Topological type* is the highest hierarchy level whose attributes, i.e., symbolic equations, contain the complete information about redistribution of bonds (as well as of charge and radical symbols) in the course of any organic reaction. In contrast to the characteristics of lower hierarchy levels (structural, skeletal, and chemical equations, see below), these equations represent no evidence on the chemical nature of reaction centers and other atoms and also of all bonds which multiplicities remain unchanged. Numerous examples of symbolic equations can be found in this and the preceding paper (and in our early publications^{9d,15c} in English). For this reason, only two additional examples (cf. Charts 11 and 12) are demonstrated here; the visual examination of Charts 11e and 12f clearly shows that both symbolic equations contain all information related to higher hierarchy levels. For example, from the equation in Chart 12f, one can see that the process is neutral, cyclic (cf. Chart 12a), and seven-centered (cf. Chart 12b); the redistribution of “outer” bonds between RFs (being reflected by classification equations of Chart 12c–e) can also be easily recognized in the symbolic equation under discussion. The combinatorial models which make it possible to generate all symbolic equations (such as those of Charts 11e and 12f) starting from topology identifying graphs (e.g., those of Charts 11b and 12b) were discussed in detail in subsection 6.2 of the preceding paper.¹

In contrast to symbolic equations, the attributes related to *topological subtypes*, i.e., structural equations, were not completely characterized in the preceding sections (and also in ref 1). It was only mentioned that structural equations typically reflect cyclic structures of educts and products and, additionally, contain unchanged bonds and/or chains which connect pairs of RCs. From the above-considered examples, one could see that in structural equations, either slightly modified cycles (cf. Chart 2d and the left equation in the eighth row of Chart 8) or newly formed and disclosed cycles (cf. the right equation in the eighth row of Chart 8) are taken into account. All three possibilities can be observed in an additional example represented in Chart 12; it is evident that the structural equation of Chart 12g simultaneously describes modification of one (“benzene-like”) six-membered cycle, rupture of another (“cyclohexadiene-like”) six-membered cycle, and formation of a new (“cyclopentadiene-like”) five-membered cycle.

The first important comment on structural equations is that some of these equations reflect cycles in *transition structures* rather than in initial or final systems. The example is represented in Chart 13a by the structural equation which relates to the so-called dyotropic rearrangement of type II. (Two types of these specific rearrangements were considered in detail by Reetz.^{34c,d}) It is easy to see that in a structural equation under consideration, both the initial and final systems contain no cycles, while the intermediate transition structure (whose connectivity corresponds to superposition of left and right parts of the equation) should be, in fact, bicyclic. It is important to clearly understand why the dyotropic process of Chart 13a and the “parent” four-centered [(1,1)+(1,1)] process (see the left symbolic equation in Chart 4f) are considered as belonging to different topological subtypes. The reason is that (due to the presence of additional chain) the former reaction is an *intramolecular rearrangement*, while the latter is a typical *intermolecular*

Chart 13. The Dyotropic Rearrangement Is Characterized by Structural Equation (a) and Corresponding Graphs G_{TOP} (b) and G_{ETOP} (c); the Typical Situations in Which Some Atoms Form Cycles But Are Not Considered as Structural Centers (d–g)^a the Skeletal Equations (h and i) Differing Only by Valence States of Sulfur Atoms Are Demonstrated^b



^a In all examples d–g, the reaction systems are schematically represented by one or two RCs (heavy points) in a box; the structural centers and “noncenters” are designated by circles and small points, respectively. ^b The corresponding [1+4] cycloadditions and reverse processes (with SO and SO_2 molecules, respectively) are very similar but belong to different reaction types.

(cycloaddition) reaction. The more formal argument is that cyclic structures of graphs G_{TOP} (cf. Chart 13b) and G_{ETOP} (cf. Chart 13c) are very different for the dyotropic process under discussion. The first of graphs is evidently monocyclic, and the second one contains three cycles (of which two are linearly independent).

The main difficulty in the exact definition of structural equation is associated with the fact that *from a chemist's viewpoint*, many cycles are not significant enough to be registered in structural equations. Any cyclic fragment attached to a single reaction center (by ordinary, double or triple bond) can serve as an example; all such fragments are undoubtedly considered as simple substituents and hence must not be used for differentiation of individual topological subtypes. The most typical example is the phenyl group, cf. Chart 13d. (Note that six-membered cycles corresponding to two phenyl groups cannot be found in the structural equation of Chart 12g.) The monocyclic or polycyclic systems containing a *single reaction center* can also be regarded as more complex (divalent or polyvalent) spiro substituents; the corresponding “insignificant” cycles are therefore ignored in structural equations. Two examples are schematically depicted in Chart 13e,f. It should be mentioned that the first example represents, e.g., ethylene glycol ($-OCH_2CH_2O-$) protecting group in which oxygen and carbon atoms—small points in Chart 13e—are considered as “noncenters” but not structural centers.

The last situation which needs some atoms to be regarded as noncenters can be observed in condensed systems which contain “attached” cycles, i.e., cycles with no bonds changing their multiplicities. In the example of Chart 13g, the upper six-membered cycle (with two RCs designated by heavy points) is somewhat modified in the course of any reaction and must therefore be considered as chemically significant. This is not the case for the lower six-membered cycle which remains unchanged; only two centers of this cycle—i.e., those represented by circles in Chart 13g—are structural centers.

Note that the schematic picture of Chart 13g is closely related to the reaction process of Chart 12i but with the dehydrobenzene component being substituted by the dehydronaphthalene one. Both processes are characterized by the same structural equation of Chart 12g and hence belong to the same topological subtype.

An attentive reader can see that in the above-mentioned cyclic systems, all atoms designated by small points (i.e., all “noncenters”) do not belong to the shortest chains which connect pairs of reaction centers. That is why we consider the presence of such chains as an important criterion which makes it often possible to differentiate between “significant” and “insignificant” cycles and hence to recognize structural centers in real situations. The more general problem, i.e., generation problem for structural equations, can be considered as consisting of two subproblems: (1) “How to construct the proper graphs G_{ETOP} (i.e., graphs which ensure all corresponding structural equations to contain only significant cycles) starting from a given graph G_{TOP} ?” and (2) “How to generate the complete lists of structural equations starting from a given graph G_{ETOP} ?” The first of these subproblems needs more formal heuristic rules and the special generation algorithm to be additionally elaborated. The second subproblem is very similar to the generation problem for symbolic equations; its graph-theoretical formalization is briefly outlined in section 4 of ref 1.

Our last comment on topological subtypes and structural equations is associated with the fact that all processes with no unchanged bonds (and structural centers) must also belong to some topological subtype. It seems reasonable to represent this specific subtype by the “parent” structural equation which contains no additional bonds and centers and hence formally coincides with the symbolic equation. An example can be found in Chart 11; the equation of Chart 11e uniquely characterizes both the topological type and subtype to which the α -chloro- β , β -dicyanovinylsufinylamine formation process of Chart 11g evidently belongs.

4.5. Reaction Types and Subtypes. In the modern state classification scheme, the *reaction types* are represented by skeletal equations; these equations are very similar to reaction equations but reflect cyclic structures in just the same way as was described above for structural equations. In special cases (with no unchanged bonds and additional centers), the skeletal and reaction equations can also coincide, cf. Chart 11f for an example.

In contrast to other hierarchy attributes, the skeletal equations afford the very detailed description of corresponding reaction processes; only the atoms and bonds which form substituents and “insignificant” cycles are not covered by these equations. The high level of detail leads, in turn, to a conclusion that it is reasonable to specify valence states of centers in skeletal equations. In practice, only unusual valence numbers of heteroatoms and specific RCs must be explicitly indicated, cf. Chart 12h for an example. It is important that specification of valence states makes it sometimes possible to differentiate between reaction types to which quite similar processes correspond. For example, two skeletal equations of Chart 13h,i are nonidentical, and this fact shows that cheletropic cycloaddition reactions of SO^{3+} and SO_2^{3+} to cyclooctatetraene belong to different reaction types. It should be additionally noted that the account of valence states allows easy calculation of the

numbers of "free" valencies for all atoms; these valencies were explicitly depicted in many of the above examples (cf. Charts 2e,f, 5a–c, 7a, 9a–u, and 11f) in order to facilitate the visual perception of skeletal and reaction equations.

The complete lists of skeletal equations can evidently be produced from structural equations by systematical substitution of abstract symbols (related to reaction and structural centers) for symbols of preselected chemical elements. Different valence states of some atoms (e.g., C^{IV} and C^{II}) must be necessarily taken into account in the appropriate generation procedures, cf. subsection 6.3 in ref 1.

The lowest hierarchy level, that is *reaction subtype*, needs few additional words for its characterization; the corresponding attributes—traditional chemical equations (cf., e.g., Charts 11g and 12i)—accumulate all structural information which is conventionally associated with real organic reactions. The stereochemistry of these reactions, the necessary conditions, catalysts, etc. are not covered, however, by the classification scheme or by the modern state Formal–Logical Approach.

5. FURTHER APPLICATIONS

Throughout this paper, we have concentrated our attention mainly to methodological aspects of the suggested classification scheme. Other important applications can, however, be associated with some hierarchy levels in computer representations of organic reactions. The clearly defined levels and additional specific rules can result in construction of *hierarchically organized codes* which parts might afford all structural information corresponding to preselected hierarchy (sub)levels. This possibility seems to be very useful for information searching in reaction libraries and databases: not only individual reactions but also more or less broad collections of "structurally similar" reactions are expected to be easily recognized in hierarchically organized databases. The elaboration of such databases seems to be important for further developments of computer-assisted search for new types of organic interconversions. The degree of novelty for any process corresponding to just generated symbolic or reaction, structural or skeletal equation could be immediately evaluated by means of hierarchically organized codes.

Although many reaction storage and retrieval systems are available,^{35a–c} most of them are oriented to the search for molecular fragments (related to substrate, reagent, or product substructures) rather than to the search for specific types of bond or bond-and-sign redistribution. Some approaches making it possible for the formal structural changes to be taken into account, are, however, reported; the Grethe and Moock's similarity searching technique (illustrated by the search for 1,3-dipolar cycloaddition reactions^{35d}) can serve as an example.

It may be that a single highly developed information system which was elaborated for manipulations with reaction types (rather than with educt and product structures) is the CASTOR system, see ref 36a and section 7 in the extensive review^{36b} on Munich group projects. In this system, the main levels of Dugundji-Ugi's approach (i.e., R-category, RB-class, and RA-class, cf. Introduction to this paper) are characterized by hierarchically organized codes; the original canonization and coding rules are also described^{5b,36c} in detail. The other, conceptually different notation systems are known to provide the direct access to essentially the same hierarchy

levels (i.e., to Roberts's^{7d} three levels of mechanistic information and to basic reaction graph, reaction graph, and reaction center graph in Fujita's approach^{7c}). Unfortunately, in all three cases, the hierarchical coding schemes cannot be extended to complete structures of educts and products and hence are inapplicable to individual organic reactions. In contrast, the single known coding system, i.e., Fujita's CANITS system^{36d} for individual reactions, is nonhierarchical in principle. No information related to reaction center graph of any level can be directly extracted from the CANITS code of a particular reaction.

At the end of this paper, we demonstrate the main features of our recently developed Hierarchical Coding System for organic reactions. The principal idea is that the complete code of any particular reaction process can be represented by an ordered sequence of several *partial codes*, and the definite subsequences formed out of these partial codes can uniquely characterize (all or the selected) levels of the above classification scheme.

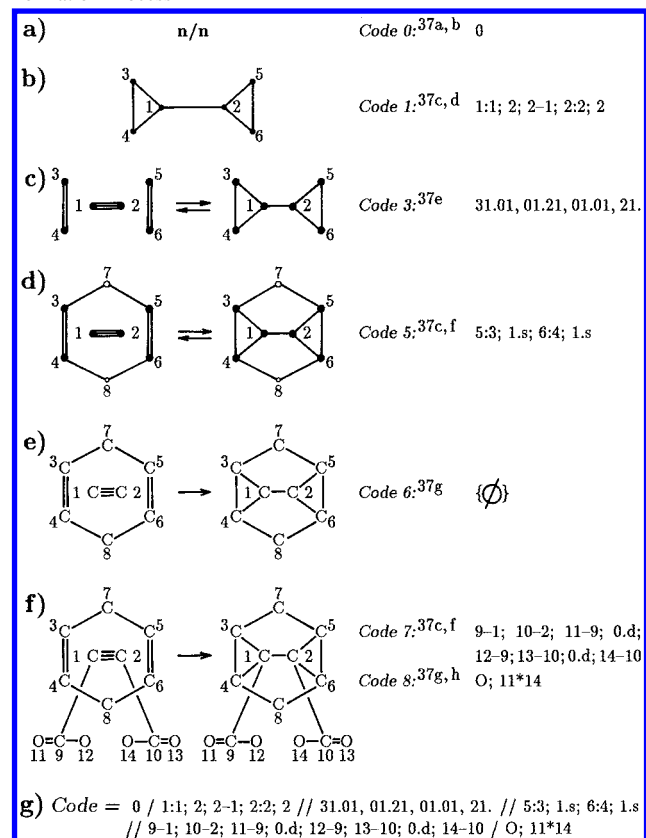
It is important that, in principle, any of hierarchical coding systems must not necessarily be an exact copy of the "methodology oriented" classification scheme. Some hierarchy levels may be neglected if corresponding classification attributes are not well-adapted to computer-aided manipulations or if the relevant information can be easily extracted from other, more informative partial codes. On the other hand, the properly organized codes are thought to provide direct access to additional specific attributes of organic reactions *if these attributes are associated with typical, practically useful queries to reaction databases or information systems*.

At the present stage of our investigations, we prefer six hierarchy levels—i.e., reaction subcategory and subclass, topological type and subtype, and reaction type and subtype—to be reflected in machine-readable codes and corresponding human-readable notations. The application of the Hierarchical Coding System is illustrated for the "double" cycloaddition process of acetylenedicarboxylic acid to cyclohexadiene-1,4.^{34g} The manually calculated partial codes (related to the above-mentioned six hierarchy levels) are represented in the right side parts of Chart 14a–f in the "human-readable" form.

The starting partial code of Chart 14a, i.e., *Code 0*, specifies **n/n** subcategory (and also neutral category) for the process under discussion. The other possible categories (cf. subsection 4.1) can also be recognized from *Code 0* notations of corresponding subcategories, and this is the reason why we do not assign specific codes to reaction categories.

In the general case, the subsequence of three partial codes, i.e., *Code 0*, *Code 1*, and *Code 2*, specifies the topology identifying graph G_{TOP} and corresponding reaction subclass; the last of these codes, *Code 2*, indicates what vertices of the "parent" graph (uniquely characterized by *Code 1*) are signed by symbols "+", "(+)" etc. For all neutral processes, and hence for the process under consideration, *Code 2* is necessarily empty, and for this reason, only *Code 1* is shown in Chart 14b. It should be noted here that topology contours (and corresponding reaction classes) are not explicitly reflected in the coding system; the information related to these contours can, however, be easily extracted from *Codes 0*, *1*, and *2* in all possible cases.

Chart 14. An Application of the Hierarchical Coding System Is Illustrated for Tetracyclo[3.3.0.0^{2,8}.0^{4,6}]octane-1,5-dicarboxylic Acid Formation Process^a



^a The selected hierarchy attributes characterizing reaction subcategory (a) and subclass (b), topological type (c) and subtype (d), and reaction type (e) and subtype (f) are explicitly shown together with corresponding partial codes. The actual direction of the process is specified in *Code* 3 and also in the complete notation of Chart 14g.

The sequence of four partial codes, i.e., *Code* 0 to *Code* 3, represents all structural information associated with the topological type of any reaction process. In the example of Chart 14c, *Code* 3 designates bond multiplicity changes in the symbolic equation, cf. note 37e for details. The chemical nature of all reaction centers involved can also be taken into account by means of additional *Code* 4 which is empty in the example under discussion. It is important that the sequence *Code* 0–*Code* 4 (or its subsequence *Code* 1, *Code* 3, *Code* 4 in the case of neutral processes) specifies reaction equations; these equations *do not correspond* to any hierarchy level of the classification scheme but are very important objects in the computer-assisted search for new types of organic interconversions.

The concatenation of the sequence representing symbolic equation with *Code* 5 (or with *Code* 5 and *Code* 6) affords the direct access to structural (or skeletal) equations; the resulting codes completely characterize topological subtype and reaction type of the classification scheme. In the example under consideration, all reaction and structural centers are carbon atoms with “normal” valence number 4 (cf. the skeletal equation of Chart 14e). The partial *Code* 6 (which specifies only heteroatoms and carbon atoms with unusual valencies) is empty in such cases.

In the general situation, the ordered sequence *Code* 0–*Code* 3, *Code* 5, *Code* 7, and *Code* 8 (note that *Code* 4 and *Code* 6 are not taken into account) represents the reaction

subtype of any particular organic reaction. The first of two terminal codes, i.e., *Code* 7, specifies the multiplicities of those bonds which are absent in the skeletal equation. The second of the partial codes (*Code* 8) describes the relative disposition of heteroatoms and also of signed or divalent carbon atoms. The examples of *Code* 7 and *Code* 8 can be found in Chart 14f. Note that in the complete chemical equation of this chart, “noncenter” hydrogen atoms are erased; these atoms are not explicitly taken into account in the Hierarchical Coding System.

In addition to the above-mentioned possibilities, the Hierarchical Coding System allows for easy recognition of some “prereactions” which represent all information on redistribution of bonds (and signs) in a reaction process but completely ignore the chemical nature of atoms involved. For that purpose, *Code* 8 should be deleted from the sequence related to a chemical equation. The other additional possibility is associated with isotope labelings; in order to specify the distribution of isotope labels, supplementary *Code* 9 (which is evidently empty in the example of Chart 14) must be added to the ordered sequence consisting of *Codes* 0–3, 5, 7, and 8.

Thus, in the most general case, the complete set of partial codes consists of 10 “subunits”, i.e., *Code* 0 to *Code* 9. All partial codes can be assembled into the unique *complete code*; for the example under consideration, the resulting code is demonstrated in Chart 14g (with slashes being used as separators and with four empty codes—*Codes* 2, 4, 6, and 9—omitted). The similar complete codes can, in fact, be constructed for all organic interconversions in which the educt and product structures are represented by classical structural formulas.

It should be noted that the partial codes of Chart 14a–f are actually “human-readable” notations; for clarity purposes, the components of these notations are separated by semicolons (or by commas if similar components are listed, cf., e.g., *Code* 3). In the corresponding machine-readable codes, each component is typically represented by one byte; the separators “;”, “:”, and “–” are used in *Codes* 1, 3, 5, and 7 in order to visualize different information units related to just the same byte. The other significant feature of machine-readable codes consists in the absence of any separators; instead, the special “preamble” part of the code is used in order to specify the number of bytes related to complete as well as to partial codes. The rules which make it possible to construct both the machine codes and the man-readable notations are quite complicated; the very brief explanations for particular notations of Chart 14a–f can be found in note 37a–h.

It is important that the machine-readable codes (in contrast to man-readable notations) are typically very short even for complex organic interconversions. Thus, the complete machine code (with preamble bytes^{37b} being included) for the reaction process of Chart 14 consists of only 36 bytes; compare this number with the length of a systematic name for tetracyclic reaction product (“tetracyclo[3.3.0.0^{2,8}.0^{4,6}]octane-1,5-dicarboxylic acid”) which needs 55 characters to be written.

Our final comment on Hierarchical Coding System is that the very similar system can be applied to molecular structures if some “significant parts” of these structures are clearly predefined. Fortunately, the unequivocal formulation of

hierarchy levels related to structural formulas has been recently reported. The notions of core, framework, skeleton, hydrogen-depleted, and complete molecular graphs were exactly defined for homo- as well as heteroatomic organic structures (see sections 7 and 9 in our paper³⁸ dedicated to application of algebraic criteria to classification of chiral molecular objects). It is important that the construction of both coding systems, i.e., for reactions and structures, needs some specific canonization rules and corresponding algorithms to be additionally elaborated. These rules are, in turn, based on the concept of *hierarchically organized graphs*, and just this concept we are planning to discuss in the next paper of this series.

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 Prof. T. Sello for his valuable remarks and suggestions which helped us to significantly improve the initial version of our manuscript.

REFERENCES AND NOTES

- Tratch, S. S.; Zefirov, N. S. Systematical Search for New Types of Chemical Interconversions: Mathematical Models and Some Applications. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 331–348.
- (a) Satchell, D. P. N. The Classification of Chemical Reactions. *Naturwissenschaften* **1977**, *64*, 113–121. (b) Bawden, D. Classification of Chemical Reactions: Potential, Possibilities, and Continuing Relevance. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 212–216.
- In some other classification systems, only the structural changes at carbon atoms and adjacent heteroatoms are taken into account. These systems are based on J. B. Hendrickson's skeletal and functional characteristics of carbon atoms (see, e.g., refs 4a,b) and also on well-defined subgraphs of S. Fujita's Imaginary Transition Structures. In the latter approach,^{4c,d} the significant parts of even-nodal and odd-nodal subgraphs, i.e., "reaction kernels" and "terminal descriptors", afford information on changes of carbon skeleton and on changes of oxidation states, respectively.
- (a) Hendrickson, J. B. A Systematic Characterization of Structures and Reactions for Use in Organic Synthesis. *J. Am. Chem. Soc.* **1971**, *93*, 6847–6854. (b) Hendrickson, J. B. A Logical Characterization of Organic Chemistry. *J. Chem. Educ.* **1978**, *55*, 216–220. (c) Fujita, S. Description of Organic Reactions Based on Imaginary Transition Structures. 4. Three-Nodal and Four-Nodal Subgraphs for a Systematic Characterization of Reactions. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 231–237. (d) Fujita, S. A Novel Approach to Systematic Classification of Organic Reactions. Hierarchical Subgraphs of Imaginary Transition Structures. *J. Chem. Soc., Perkin Trans. II* **1988**, 597–616.
- (a) Dugundji, J.; Ugi, I. An Algebraic Model of Constitutional Chemistry as a Basis for Chemical Computer Programs. *Topics Curr. Chem.* **1973**, *39*, 19–64. (b) Brandt, J.; Bauer, J.; Frank, R. M.; von Sholley, A. Classification of Reactions by Electron Shift Pattern. *Chem. Scripta* **1981**, *18*, 53–60. (c) Bauer, J.; Herges, R.; Fontain, E.; Ugi, I. IGOR and Computer Assisted Innovation in Chemistry. *Chimia* **1985**, *39*, 43–53. (d) Herges, R. Reaction Planning: Computer-Aided Reaction Design. *Tetrahedron Comput. Methodol.* **1988**, *1*, 15–25. (e) Ugi, I.; Bauer, J.; Bley, K.; Dengler, A.; Dietz, A.; Fontain, E.; Gruber, B.; Herges, R.; Knauer, M.; Reitsam, K.; Stein, N. Computer-Assisted Solution of Chemical Problems—The Historical Development and the Present State of the Art of a New Discipline of Chemistry. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 201–227.
- (a) Ainsworth, C.; Kuo, Y.-N.; Chen, F. Non-Rigid Carbon–Carbon Double Bonds: Conformation Assignments of Ketene Methyl Trimethylsilyl Acetals. *Chem. Commun.* **1971**, 137–139. (b) Goldstein, M. J.; Judson, H. A. Degenerate [3,3] Sigmatropic Shift of Acetyl Peroxide. *J. Am. Chem. Soc.* **1970**, *92*, 4119–4120.
- (a) Kvasnička, V. Classification Scheme of Chemical Reactions. *Coll. Czech. Chem. Commun.* **1984**, *49*, 1090–1097. (b) Fujita, S. Formulation and Enumeration of Five-Center Organic Reactions. *Bull. Chem. Soc. Jpn.* **1989**, *62*, 662–667. (c) Fujita, S. Canonical Numbering and Coding of Reaction Center Graphs and Reduced Reaction Center Graphs Abstracted from Imaginary Transition Structures. A Novel Approach to the Linear Coding of Reaction Types. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 137–142. (d) Roberts, D. C. A Systematic Approach to the Classification and Nomenclature of Reaction Mechanisms. *J. Org. Chem.* **1978**, *43*, 1473–1480.
- (a) Balaban, A. T. Chemical Graphs. 3. Reactions with Cyclic Six-Membered Transition States. *Rev. Roum. Chim.* **1967**, *12*, 875–898. (b) Hendrickson, J. B. The Variety of Thermal Pericyclic Reactions. *Angew. Chem., Int. Ed. Engl.* **1974**, *13*, 47–76. (c) Hendrickson, J. B. Comprehensive System for Classification and Nomenclature of Organic Reactions. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 852–860. (d) Vladutz, G. E.; Geivandov, E. A. *Automatized Information Systems for Chemistry*; Nauka Press: Moscow, 1974; Chapter 12, pp 189–204 [in Russian]. (e) Arens, J. F. A Formalism for the Classification and Design of Organic Reactions. 1. The Class of $(-+)_n$ Reactions. *Recl. Trav. Chim. Pays-Bas* **1979**, *98*, 155–161. (f) Sicouri, G.; Sobel, Y.; Picchiottino, R.; Dubois, J.-É. *Localizing Variations onto the Reaction Invariant: the Transformation Structure Concept*. *C. R. Acad. Sci. Paris* **1984**, *299*, 523–528. (g) Babaev, E. V.; Lushnikov, D. E.; Zefirov, N. S. Novel Graph-Theoretical Approach to Ring-Transformation Reactions: Hierarchic Classification and Computer Design of Heterocyclic Rearrangements. *J. Am. Chem. Soc.* **1993**, *115*, 2416–2427.
- (a) 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]. (b) 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]. (c) 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]. (d) Zefirov, N. S.; Tratch, S. S. Systematization of Tautomeric Processes and Formal–Logical Approach to the Search for New Topological and Reaction Types of Tautomerism. *Chem. Scr.* **1980**, *15*, 4–12.
- (a) 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]. (b) 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]. (c) 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]. (d) Zefirov, N. S.; Tratch, S. S.; Gamziani, G. A. *Description of Organic Reactions in Frame of the Formal–Logical Approach*; Metznierbeba Press: Tbilisi, 1993; 208 pp [in Russian].
- (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) Zefirov, N. S. An Approach to Systematization and Design of Organic Reactions. *Acc. Chem. Res.* **1987**, *20*, 237–243. (d) Tratch, S. S. *Logical–Combinatorial Approaches to Design Problems for Organic Structures, Reactions, and Configurations* (Doctoral Dissertation); Moscow, 1993; Vol. 1, pp 10–60, 81–153 [in Russian].
- (a) Boersma, M. A. M.; de Haan, J. W.; Kloosterziel, H.; van de Ven, L. J. M. Stereospecificity and Stereochemistry of a Thermal Sigmatropic [1,5]-Shift of an Alkyl Group. *Chem. Commun.* **1970**, 1168–1169. (b) Reeves, P.; Devon, T.; Pettit, R. On the Possible Rectangular Nature of Cyclobutadiene. *J. Am. Chem. Soc.* **1969**, *91*, 5890–5891.
- The designations $[3,3\alpha\gamma]$, $[3\alpha+3\alpha]$, and $[3\beta+3\beta]$ are very similar to those traditionally used by organic chemists but uniquely correspond to symbolic equations (other examples of "identifying" designations can be found in the footnote to Chart 4). The Greek letters are used in order to unambiguously indicate the position(s) of one or two specific RCs X in reaction fragments. The unified notation systems were, in fact, elaborated for all possible types of *single mode* processes with cyclic^{9a,b} and linear^{10d,11d} topologies of electron transfer. The term "single mode" means that all bonds change their multiplicities by one unit; other structural characteristics of the Formal–Logical Approach are systematically considered in section 3 of ref 1.
- Oligoadditions and oligofragmentations correspond to the case $q > 1$ in the schematic representation of Chart 4e. The processes under discussion can be additionally partitioned^{9a,b} into "normal" and

- sigmatropic ones. In sigmatropic addition–fragmentation reactions, one or two migrating σ -bonds are observed. The well-known ene and retroene reactions (cf. Chart 5a) can serve as typical examples; in these reactions, the addition (or fragmentation) process is accompanied by a migration of σ -bond from one to the other end of the allylic system.
- (15) (a) Fujita, S. Description of Organic Reactions Based on Imaginary Transition Structures. 2. Classifications of One-String Reactions Having an Even-Membered Cyclic Reaction Graph. *J. Chem. Inf. Comput. Sci.* **1986**, 26, 212–223. (b) 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]. (c) Zefirov, N. S.; Tratch, S. S. Symbolic Equations and Their Applications to Reaction Design. *Anal. Chim. Acta* **1990**, 235, 115–134.
- (16) In this paper, we do not differentiate between reaction fragments formed out of atom symbols (C, H, N, O, etc.) and out of abstract RCs (\bullet , X, X', etc.). In the more rigorous terminology, the fragments obtained from several reaction equations (corresponding to the same symbolic equation) may be regarded as different realizations of the same parent “abstract” RF.
- (17) Simple *I*-fragments (such as allylic group, cf. Chart 5a) are inverted in the sense that their single external valency migrates from one to the other end of the fragment, while the internal structure of RF remains unchanged. In the general situation, the word “inversion” is in no case associated with degeneracy, it only shows that nonzero values of RN_{mi}^* and RN_{in}^* (and hence the total multiplicities of migrating external valencies) are necessarily equal for the both forms of all *I*-fragments.
- (18) In simple cases, i.e., in reactions with cyclic electron transfer,^{9a} A-fragments really participate in cycloaddition–cyclofragmentation and addition–fragmentation processes; see the symbolic equations of Chart 4c,e for examples. The names of particular forms, i.e., “chain” and “ring”, were also initially suggested^{9a} for addition fragments (of degree 2) being involved in the processes with cyclic electron transfer. In the left example of Chart 4c, two ethylene-like moieties (or two-atomic chains) form four-membered “cyclobutane” ring.
- (19) For all fragments with $\rho^* \geq 1$ outer bonds, the codes consist of ρ^* negative or positive integers which are arranged in nondescending order. Each negative (or positive) integer is equal to multiplicity of broken (or formed) outer bond in a particular form of the fragment. Two forms of *I*-fragments are often characterized by identical coding designations; the codes of *c*- and *r*-forms of A, A', etc. fragments are never identical. The codes (–1, +1), (–1, +1), (+1, +1), and (–1, –1) for *I*₃, *I*₃, 'A₂, and 'A₂ fragments (cf. third and sixth rows in the Table below Chart 5) can serve as typical examples. Note that all RFs with no outer bonds, i.e., O-fragments, can be characterized by the same zero code.
- (20) Three possibilities mentioned in the text directly follow from main restrictions¹ of the Formal–Logical Approach: any reaction center can be signed by a single symbol “+”, “•”, “–” in an initial or final system, and the total number of signed RCs is an even integer not exceeding 2. The simple examples of other, “multiply signed” RFs (as well as corresponding designations) can be easily constructed but these fragments—may be, with very rare exceptions—seem to be inapplicable to description of real organic transformations.
- (21) Let the signed fragment (with fixed values of RN_{mi}^* and RN_{in}^*) contain *a* and *b* symbols “+”, “•”, or “–” in initial and final systems, respectively. The symbols of sign can be mentally substituted by new, “phantom” centers being connected by additional ordinary bonds with real centers of the original fragment. The resulting RF is unsigned in all cases (because any “phantom” center is connected by an outer bond and hence forms a separate fragment). The unsigned fragment is evidently characterized by reaction numbers equal to $RN_{mi}^* + a$ and $RN_{in}^* + b$ and by even specificity value $|RN_{mi}^* + a - RN_{in}^* - b| = 2k$. The last expression makes it possible to conclude that the specificity value $r^* = |RN_{mi}^* - RN_{in}^*|$ of original signed fragment and the value of $|a - b|$ must be of the same parity. Note that in cases 1, 2, and 3, the values of $|a - b|$ are equal to 1, 2, and 0, and this fact proves that specificity values of signed fragments are necessarily odd (in case 1) and even (in cases 2 and 3).
- (22) In fragment systems (and in classification equations), one loop or several loops in *r*-form of some addition fragment cannot be neglected without loss of information; this is a case, e.g., for an 'A₃ fragment representing a three-membered ring in the left side parts of equations in Chart 5b. For the sake of brevity, the loop(s) are omitted only in the case of the simplest fragment systems, i.e., those consisting of a single addition fragment. The right part of the classification equation describing electrocyclic reactions (cf. Chart 4b) can serve as an example.
- (23) The examples of real [2+3], [1+4], and [1+2+2] cycloaddition–cyclofragmentation reactions (together with numerous examples of [2+2], [2+4], [2+2+2], and other cycloadditions–cyclofragmentations) can be found in classical monographs^{24a,b} on conservation of orbital symmetry. In the literature on orbital symmetry principles, the integers *i*, *j*, ... in designations [i+j+ ...] denote, however, the numbers of electrons involved rather than numbers of centers in corresponding A-fragments.
- (24) (a) Woodward, R. B.; Hoffmann, R. *The Conservation of Orbital Symmetry*; Academic Press: New York, 1970; 178 pp. (b) Gilchrist, T. L.; Storr, R. C. *Organic Reactions and Orbital Symmetry*; Cambridge University Press: London, 1971; 271 pp. (c) Wiley, R. H.; England, D. C.; Behr, L. C. The Preparation of Thiazoles. In *Organic Reactions*, v. 6; Adams, R., Ed.; John Wiley & Sons, Inc.: New York, 1951; pp 367–409.
- (25) In the symbolic equation of Chart 6d, the four-centered addition fragment relates to (probably not investigated) 1,4-dipolar species similar to $R_2C^+ - NR - CR = N^-$. The nonpolar resonance structures of such dipoles (“azomethynylidene-nitrenes”, $R_2C = NR = CR - N:$) contain two specific centers (N) in α and γ positions of the fragment. That is why any real or hypothetical reaction corresponding to a symbolic equation of Chart 6d must be considered as [1+4 $\alpha\gamma$] cycloaddition (cyclofragmentation); note that “1” in this and similar notations is a conventional shortening of “1 α ” and relates to one-centered addition fragment (X).
- (26) The most important difference is that in contrast to symbolic and reaction equations, the classification equations can contain “parallel” bonds which connect two RFs in an initial and/or final fragment system; the “parallel” loops (corresponding to bonds which connect different pairs of RCs in the same fragment) can also appear in classification equations. The parallelism of two or more valence lines should not be mixed with their multiplicity; any of the parallel lines relates to a single outer bond and hence is characterized by its own multiplicity. The arcs representing parallel lines can be observed in three two-fragment classification equations (see second and third columns of Chart 6) and in the seven-fragment classification equation (see Chart 7c).
- (27) It should be noted that all bond redistributions in fragment systems are trivial and principal; it directly follows from the fact that any outer bond (and also any loop) in a classification equation can change its multiplicity either from zero to 1, 2, 3 or from 1, 2, 3 to zero.
- (28) The general combinatorial models described in sections 5 and 6 of ref 1 are inapplicable to classification equations just due to possible presence of parallel edges and loops in corresponding graphs *G*_{TOP}; the specific permutation groups (similar to J. G. Nourse's²⁹ Edge Symmetry Groups) must be used in order to rigorously formulate generation problems for classification equations. In the most general case, special group-theoretical constructions are additionally needed for correct description of symmetries associated with individual reaction fragments.
- (29) Nourse J. G. The Configuration Symmetry Group and Its Application to Stereoisomer Generation, Specification, and Enumeration. *J. Am. Chem. Soc.* **1979**, 101, 1210–1223.
- (30) In ion–radicals, the relative disposition of charge and radical symbols (i.e., on the same or different reaction centers) is not taken into account. The “monocentric” dications, dianions, and similar molecular species (which are not covered by combinatorial models of ref 1) are also not distinguished from other ions in our classification system.
- (31) In accordance with the general combinatorial model (cf. section 5 in ref 1), the finite set *W* consists of $p = 2k > 0$ sites being related to an even number of charge and radical symbols; the symmetric group $H = S_p$, $|H| = p!$, acts on the set *W* and interconverts completely indistinguishable sites of this set. The label set *M* consists of $q = 2r = 6$ labels $m_1 = “+”, m_2 = “•”, m_3 = “–”, m_4 = “(-)”, m_5 = “(•)”, m_6 = “(+)”$; the opposite labels “+” and “(+)”, “•” and “(•)”, “–” and “(-)” are converted one into the other by nonidentical permutation s_1 of a group $S_2 = \{s_0, s_1\}$ acting on *M*. As a result, the induced group $\Gamma = S_2^p$ of order $2 \cdot p!$ acts on the set $F = M^W$ of q^p functions *f* from *W* to *M*, and subcategory designations can be associated with the orbits of this action. In order to calculate total numbers of orbits, the general enumeration formula (1) with $q = 2r$ (cf. section 7 in ref 1) may be used; the application of this formula to the groups $H \equiv S_2$ and $H \equiv S_4$ shows that there exist 12 and 66 subcategories with two and four symbols “c”, “r”, and “a” in corresponding designations.
- (32) In order to construct such lists, the transversals of orbits for induced groups Γ'' and Γ' on the complete sets *X* of functions from the site sets *W*₂ into the label sets *M*₂ must be generated, cf. subsection 6.2 in ref 1. In the example under consideration, the sets *W*₂, *M*₂, and *X* consist of 1 bond, 6 labels (“0/1”, “1/2”, “2/3”, “3/2”, “2/1”, “1/0”), and 6 functions χ , respectively. The group Γ'' (related to “c/c”, “a/a”, and “r/r” subcategories) and the group Γ' (related to nine other subcategories) are isomorphic to symmetric group *S*₂ and to identity group *E*, and this fact shows that numbers of orbits are equal to 3 and 6, respectively. The total number of two-centered symbolic equations is therefore equal to $3 \cdot 3 + 9 \cdot 6 = 63$.

- (33) Strictly speaking, topology contours may be regarded as *pseudographs*, i.e., graphs which can contain parallel edges and loops. In the exceptional case of cyclic topology (cf. the third contour in Chart 10a), a supplementary node is needed in order to construct the simplest pseudograph with one vertex and one loop. Although topology pseudographs play the specific role in some applications, they are not specially considered in this paper.
- (34) (a) Zefirov, N. S.; Chapovskaya, N. K.; Dyachkova, L. Ya.; Tratch, S. S. Reaction of Potassium Tricyanomethide with Sulfuryl Chloride and Thionyl Chloride. *Zh. Org. Khimii* **1975**, *11*, 1981 [in Russian]. (b) Kobayashi, K.; Mutai, K. Novel Cycloaddition–Elimination Reactions of 2,5-Diphenyl-1,4-dithiin-1,1-dioxide with Electron–Deficient Acetylenic Compounds. *Tetrahedron Lett.* **1978**, 905–906. (c) Reetz, M. T. Dyotropic Rearrangements, a New Class of Orbital–Symmetry Controlled Reactions. Type I, II. *Angew. Chem.* **1972**, *84*, 161–163. (d) Reetz, M. T. Primary and Secondary Orbital Effects in Dyotropic Rearrangements. *Tetrahedron* **1973**, *29*, 2189–2194. (e) Anastassiou, A. G.; Chao, B. Y.-H. 9-Thiabicyclo[6.1.0]nona-2,4,6-triene. *Chem. Commun.* **1971**, 979–980. (f) Gasteiger, J.; Huisgen, R. 9-Thiabicyclo[4.2.1]nona-2,4,7-triene 9,9-Dioxide. *J. Am. Chem. Soc.* **1972**, *94*, 6541–6543. (g) Takahashi, M.; Itahara, Y.; Murata, I.; Nitta, T.; Woods, M. C. Tetracyclo[3.3.0.0^{2,8}.0^{4,6}]octane-1,5-dicarboxylic Acid. *Tetrahedron Lett.* **1968**, 3387–3390.
- (35) (a) Willett, P., Ed. *Modern Approaches to Chemical Reaction Searching*; Gower: Aldershot, 1986; 266 pp. (b) Zass, E. A User's View of Chemical Reaction Information Sources. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 360–372. (c) Barth, A. Status and Future Developments of Reaction Databases and Online Retrieval Systems. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 384–393. (d) Grethe, G.; Moock, T. E. Similarity Searching in REACCS. A New Tool for the Synthetic Chemist. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 511–520.
- (36) (a) Brandt, J.; v. Scholley, A.; Wochner, M. Making Chemical Reaction Data Accessible to the Non-Chemist. *Comput. Phys. Commun.* **1984**, *33*, 197–203. (b) Ugi, I.; Bauer, J.; Blomberger, C.; Brandt, J.; Dietz, A.; Fontain, E.; Gruber, B.; v. Scholley-Pfab, A.; Senff, A.; Stein, N. Models, Concepts, Theories, and Formal Languages in Chemistry and Their Use as a Basis for Computer Assistance in Chemistry. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 3–16. (c) Brandt, J.; v. Scholley, A. An Efficient Algorithm for the Computation of the Canonical Numbering of Reaction Matrices. *Comput. Chem.* **1983**, *7*, 51–59. (d) Fujita, S. Canonical Numbering and Coding of Imaginary Transition Structures. A Novel Approach to the Linear Coding of Individual Organic Reactions. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 128–137.
- (37) (a) In a machine-readable code, any subcategory is represented by terminal byte(s) of the *preamble*. In the case of a neutral subcategory, “*n/n*” is substituted by a single zero byte. (b) The nonterminal bytes of the preamble designate total lengths (in bytes) of the complete and partial codes. The separators—points, hyphens, colons, semicolons, commas, and slashes—are absent in machine-readable codes. (c) In *Codes 1, 5, and 7*, two integers *i* and *j*, *i* > *j*, are separated by a colon if corresponding graph vertices or atoms are not adjacent but connected by one or several unbranched chains; the terms “*i:i*” relate to closed chains (or *loops*). The colon symbol is substituted by a hyphen if *i* and *j* are adjacent; the additional unbranched chains are also possible in this case. (d) The numbers of vertices in chains and loops are specified in the next component(s) of *Code 1* by one or more one-byte integers. The “isolated” edges between vertices *i* and *j*, *i* > *j*, are not explicitly indicated in *Code 1*. (e) Every component of *Code 3* is a byte which identifies multiplicity changes corresponding to two edges of the graph *G_{top}*; the point is a separator between two nibbles (half-bytes). For example, “31” and “01” relate to bonds 2–1 and 3–1 and designate labels “3/1” and “0/1”; cf. section 4 in ref 1. (f) In *Code 5* and *Code 7*, the one-byte terms “*k.s*”, “*k.d*”, and “*k.t*”, *k* > 0, correspond to sequences of *k* + 1 single, double, or triple bonds. The chains (subchains) formed of conjugated bonds can be similarly specified by means of symbols “*sd*”, “*ds*”, “*st*”, and “*ts*”, respectively. Note that isolated double and triple bonds—but not single *i*–*j* bonds—must be explicitly represented (“*0.d*”, “*0.t*”) in *Code 5* and *Code 7*. (g) Only heteroatoms, charged or radicalic carbon atoms, and all atoms with unusual valence numbers are specified in *Codes 4, 6, and 8*. That is why *Code 6* (and also *Code 4* which corresponds to reaction equation) are empty in the example of Chart 14. (h) The asterisk is not a separator; the corresponding byte in a machine code indicates that all four atoms (i.e., those with numbers 11–14) are oxygens.
- (38) Tratch, S. S.; Zefirov, N. S. Algebraic Chirality Criteria and Their Application to Chirality Classification in Rigid Molecular Systems. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 448–464.

CI960098U