

**THE ARGENT PROGRAM SYSTEM: A
SECOND-GENERATION TOOL AIMED AT
COMBINATORIAL SEARCH FOR NEW TYPES OF
ORGANIC REACTIONS. 1. MAIN CONCEPTS AND
POTENTIALITIES**

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

[†] Department of Chemistry, Moscow State University;
Leninskie Gory, Moscow 119899, Russia

[‡] Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences;
Leninsky Prospekt 47, Moscow 117913, Russia

Abstract

The main principles and prospects of ARGENT, a new software package for multistage reaction design on the basis of our Formal-Logical Approach to Organic Reactions, are presented. The distinctive features that make ARGENT a potentially versatile and efficient tool for finding new types of chemical interconversions are discussed. A general outline of the reaction hierarchy within the present-state Formal-Logical Approach is given, and a representation of this hierarchy by a multistage graph labeling technique is described. The basic principles of reaction design and their implementation in the program ARGENT-1 (the first executable module of the ARGENT system) are illustrated by several examples.

1. INTRODUCTION

In the preceding publications,¹ the current state of the Formal-Logical Approach to Organic Reactions^{2,3} was described in detail, and the theoretical principles that can form the basis of an enhanced program system for solution of reaction design problems were considered. We are currently planning to implement these principles in a new multipurpose reaction design system called ARGENT (Automatic Reaction Generation and Evaluation of New Types). In this paper, we discuss the general foundations, main features, and future prospects of this system; a description of the mathematical models that

make it possible to successively solve all the generation problems appearing in the first program of the ARGENT 1 series (as well as some related problems) are outlined in the next paper in this issue.⁴

Of course, our program system is not the first computerized tool aimed at systematic search for chemical interconversions. Actually, creation of such programs has been evoking the attention of many specialists in computer chemistry over the last thirty years. However, the overwhelming majority of existing software products in this field deal with search for new synthetic pathways (computer-assisted *synthesis design*) rather than with computer-assisted *reaction design* in the narrow sense of the term, that is, with search for new reaction types. The difference between these two kinds of programs was first mentioned by Herges,⁵ who considered reaction planning (and reaction design) exactly as prediction of unprecedented types of chemical interconversions.

In fact, programs of the first aforementioned type, i.e., of synthesis design, may concern retrosynthesis (search for possible synthetic routes to a given substance) or prosynthesis (search for possible reactions of a given compound),⁶⁻⁸ but, in any case, they must start from some preselected structure. On the other hand, programs of the second type are oriented to creation of *novel types of chemical interconversions* and prediction of their educts and products. Following the notation of Dugundji-Ugi's well-known model,⁹ let us characterize an organic reaction by the chemical systems at its beginning (*B*) and end (*E*) and by the corresponding transformation matrix (*R*). Then, computer-assisted synthesis design means listing of possible (*R*, *B*)-pairs for a given *E* (retrosynthesis) or (*R*, *E*)-pairs for a given *B* (prosynthesis), whereas reaction design means generation of various acceptable *R* and listing of possible (*B*, *E*)-pairs for each preselected *R*. (See ref 10 for a more detailed discussion.)

Elaboration and improvement of programs for computer-assisted synthesis design was and still is very popular,⁷ and this is natural because of their evident importance for practice. However, problems of reaction design — in the strict sense of the term — have received comparatively little attention. Several papers (e.g., see refs 11–14) dealing with generation and classification of specific kinds of reactions have been published, but ready-to-use universal software for reaction design seems to be very scarce. To our knowledge, the only full-scale programs are IGOR and IGOR2,^{5a,10a,15} elaborated by the Munich group, and SYMBEQ,^{3c,d,16} developed within the framework of the Formal–Logical Approach at Moscow State University.

In contrast to SYMBEQ, other software products of the Moscow University group are oriented to solution of less general problems. In particular, the program ELSE^{3d} can generate ionic, radicalic, or redox processes with linear electron transfer;¹⁷ these processes typically describe elementary steps of complex reactions. The second program, GREH,¹⁸ enables one to construct bicyclic graphs corresponding to interconversions of heterocyclic systems. This program makes allowance for the cyclic structure and tautomerism of heterocyclic compounds but completely neglects all acyclic appendages and substituents.

All the above programs have been successfully used for predicting new types of reactions, mostly pericyclic ones (see refs 3c,d, 5, 15a, 16b); several other generic classes of reaction processes have been systematically treated in separate works.^{19,20} However, new reaction types really predicted by a computer are not numerous, and future progress in reaction design must hence be associated with development of highly improved and versatile, “second-generation” program tools aimed at computer-assisted search for new chemical interconversions.

Actually, analysis of existing reaction design programs reveals many limitations and drawbacks, such as inadequate generation capabilities, insufficient development of selection criteria, inconvenient implementation of the man-machine dialog, etc. For example, neither IGOR nor IGOR2 can directly produce complete lists of reactions corresponding to a given (e.g., cyclic) topology of bond redistribution, because these programs do not contain modules²¹ generating all relevant transformations *R*. Similarly, the results of SYMBEQ cannot be considered as exhaustive, because they lack interconversions of organic ions, radicals, and ion-radicals.

2. SPECIFIC FEATURES OF THE NEW PROGRAM SYSTEM

In our work on the ARGENT program system, we are trying to overcome the difficulties mentioned in the preceding section. Our aim is to combine the advantages of earlier programs with new means available in the present-state Formal-Logical Approach¹ in order to create an efficient and chemically versatile multipurpose program system with the following distinctive features:

1. ARGENT is aimed to be a universal software package for design of reactions with any topology of bond redistribution (i.e., disposition of bonds that change their multiplicity during the reaction); the topology identifying graphs can, in principle, be introduced by the user or generated by some external program. Also, in contrast to SYMBEQ, the new program system treats not only interconversions of neutral molecules but also those of charged and radicalic species. Therefore, it can produce nearly all reactions of practical interest: neutral, ionic, radicalic, and even monoredox or diredox processes. Charges and unpaired free electrons at atoms are always considered explicitly, in contrast to IGOR/IGOR2.
2. As in other existing reaction generators, the generation process in ARGENT is based on a hierarchical classification of reactions. The difference between the detailed hierarchy of the Formal-Logical Approach (which is the theoretical basis of ARGENT) and other previously suggested multilevel classification systems²² is responsible for a difference in the degree of the program adaptation to solution of some actual problems. For example, special hierarchy levels (in future versions of ARGENT) are planned to rely on explicit consideration of additional ring closures existing between atoms that participate in the reaction. This possibility has never been taken into account in other reaction-generating programs; it seems to be an especially useful tool in prediction of heterocyclic rearrangements or, for example, in search for new synthetic routes to polycyclic and caged molecules.
3. A new system of hierarchical coding^{1b} has been developed for purposes of the ARGENT program system in order to provide the basis for saving the resultant reactions in a very compact new notation. In contrast to other notation systems,²³ our "hierarchical organization" of any reaction code significantly facilitates the search for its closest analogs in relevant databases and hence enables one to estimate the degree of novelty for the corresponding reaction.
4. Rigorous mathematical models⁴ elaborated for the purposes of the ARGENT program system enable us to suggest powerful analytical enumeration (counting) tech-

niques^{24a} for a manual or computer-aided solution of many reaction design problems. The corresponding combinatorial algorithms^{24b,25} always ensure exhaustive, irredundant, and efficient constructive enumeration (i.e., generation) of organic reactions. In addition, as is demonstrated in the next paper,⁴ the same models can be used to solve certain structural design problems.

5. An enhanced system of formal selection criteria makes it possible to perform efficient automatic screening of the results generated by ARGENT, leaving mainly those of interest for the problem currently considered. The system contains a number of new criteria in comparison with SYMBEQ and in some respects makes more extensive use of formal chemical knowledge in comparison with IGOR/IGOR2. The selection criteria used within the ARGENT-1 program will be described in future papers of this series.²⁶
6. Last but not least: in contrast to IGORs and SYMBEQ, which may be mastered only through considerable time and effort, ARGENT is planned to be an easy-to-use tool for any chemist, not necessarily a specialist in computer chemistry. For this purpose, a well-developed graphic interface for MS Windows and a help system based mostly on the language of chemistry (rather than on the language of combinatorics or graph theory) are being created. The user is additionally provided with the possibility to work either with a simple, specially reduced set of selection criteria or with a more complicated one, corresponding to the *novice* and *expert* work levels, respectively. All programs within ARGENT are meant to be very versatile; in particular, generation may be started and terminated at any level of the reaction hierarchy (see below), thus providing easy access to reaction types of the desired degree of novelty.

We believe that the above combination of features must make ARGENT a universal program system in the field of reaction design, and, since it will represent a new level of adaptation to solution of actual chemical problems, we claim it to be the first second-generation tool aimed at systematic search for new types of organic reactions.

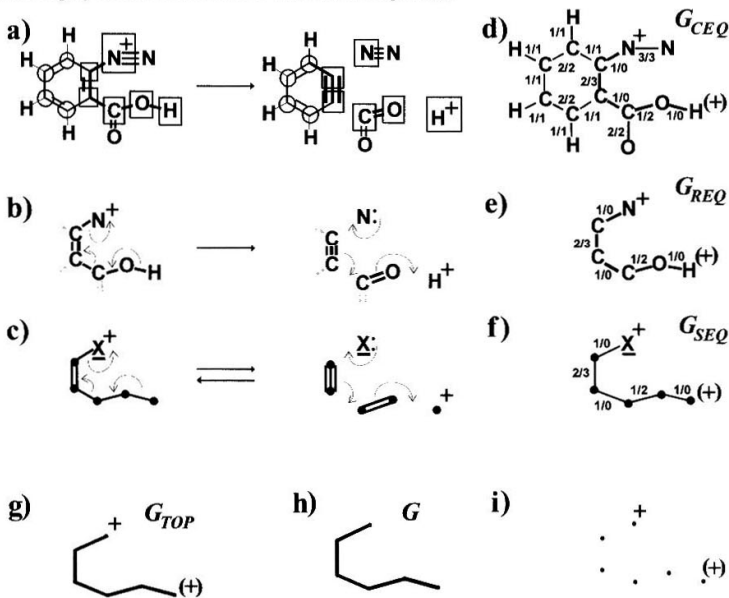
In this communication, we outline the present-day potentialities and future prospects of the ARGENT system with special attention to the first program implementation of this system, hereafter referred to as ARGENT-1. To clearly understand the representation of organic reactions and the main generation stages in ARGENT-1, one must be familiar with several ideas and terms that have already been introduced in our earlier publications and with the present development of these ideas. That is why some main notions of the Formal-Logical Approach are briefly explained and illustrated in the next two sections.

3. FUNDAMENTALS OF THE FORMAL-LOGICAL APPROACH TO ORGANIC REACTIONS

The Formal-Logical Approach (see refs 1,3 and references therein) provides a hierarchical graph-theoretical description of chemical interconversions, both known and unprecedented. In accordance with this hierarchy, the design of new reactions, which is the main aim of the approach, may be reduced to solution of graph-theoretical problems — more

precisely, to generation of some graphs that can represent novel types of chemical transformations. In the following paragraphs, the main notions of the Formal-Logical Approach are introduced in a concise form.

Figure 1: An organic reaction as described within the Formal-Logical Approach: (a) a complete chemical equation (with reaction centers denoted by rectangles and structural centers denoted by circles); (b) reaction and (c) symbolic equations corresponding to the given complete equation; (d)–(f) graphs G_{CEQ} , G_{REQ} , and G_{SEQ} ; (g) signed and (h) unsigned topology identifiers; (i) an empty signed graph corresponding to the c/c -subcategory of ionic reactions for a six-centered process.



1. The initial and final systems contain one or more components; each of these components is a connected structure and represents a definite molecule, or ion, or radical, etc. In the present-day state of the Formal-Logical Approach, only bonds of integer multiplicities — single, double, and triple — are considered. (In principle, the approach can be further expanded to fractional bond orders, which have been shown²⁷ to satisfactorily describe chemical interconversions involving nonclassical structures.) Tautomeric forms²⁸ and even nonidentical resonance structures of the same compound are regarded as different chemical species.

2. Any chemical interconversion — that is, conversion of some initial (educt) system into some final (product) one — is represented by a change in the multiplicity of some *bonds*; this change may be accompanied by appearance or disappearance of *signs* (designating charges or free electrons) at some atoms. Note that complete formation or breakup of a bond is also regarded as a change in its multiplicity from zero or to zero, respectively. Thus, any reaction is considered as a *bond-and-sign redistribution*, that is, as a process which involves a change in the multiplicity of some bonds and maybe also a change in the sign of some atoms. Direct and inverse processes (conversions from the initial into the final system and vice versa) are not differentiated. The mechanism of the reaction,^{29–32} the actual possibility of its implementation, and also its reversibility (or irreversibility) and stereochemical aspects are not explicitly taken into account.
3. All atoms incident to any bonds that change their multiplicity during the reaction are referred to as *reaction centers*, or RCs. Examples of RCs (marked with rectangles) may be found in Fig. 1a, which represents the well-known decomposition process resulting in dehydrobenzene. Some RCs can bear charges or free electrons in the educt or product system; they are supplied with the respective plus (“+”), minus (“−”), or radical (“·”) symbols and referred to as *signed reaction centers* (SRCs). Lone electron pairs or vacant orbitals at reaction centers are not explicitly taken into account. Further analysis is confined to systems where each signed reaction center bears only one definite sign “+”, “−”, or “·”, either in the educt or in the product system but not in both.³³ It can be easily proved^{1a} that the total number of signed reaction centers in both systems (initial and final) is always even; within ARGENT-1, it may be equal to 0 or 2.
4. One can distinguish two types of bonds that do not change their multiplicity during the reaction: those which form additional rings with participation of reaction centers (in the initial or final system) and those which do not; for a more precise distinction, see ref 1b. Atoms that belong to the aforementioned rings but are not reaction centers themselves are referred to as *structural centers*. For example, a chain consisting of four structural centers (circled carbon atoms) may be found in the chemical equation in Fig. 1a. Obviously, account for structural centers and specification of ring structures in educt and product molecules enable one to more precisely characterize the degree of novelty for a particular organic interconversion.
5. Additional notions in the Formal-Logical Approach are associated with *reaction fragments*,^{1b} i.e., with structural units (consisting of RCs and modified bonds between them) that “behave as a whole” in the process under consideration. In Fig. 1a, two monoatomic (N and H) and two diatomic (CC and CO) reaction fragments can be observed. Note that substitution of one or several reaction fragments by similar fragments from a preselected list can result in prediction of new analogs of well-known reaction processes; a nontrivial example of this specific approach to reaction design problems was briefly considered in ref 1b.

In this brief survey, we do not explicitly consider various structural and numerical characteristics of bond and bond-and-sign redistributions or characteristics of individual reaction centers, structural centers, and reaction fragments. This information (see refs 1a,b)

is more suitable for discussions of classification problems rather than problems of reaction design. At the same time, the main features and generation stages in our current design program, ARGENT-1, depend on hierarchically organized specific representations of bond-and-sign redistributions in the form of graphs with labeled vertices and/or edges. Consideration of these graphs and more or less generalized equations corresponding to them is the subject of the next section.

4. LEVELS OF THE REACTION HIERARCHY

For the purposes of reaction design, the most important feature of any formal approach to chemical interconversions is the hierarchy of their description, that is, the choice of representations that describe the corresponding bond-and-sign redistributions at different levels of generalization. (For a detailed comparison of the hierarchy used in the Formal-Logical Approach with those developed by other authors, see the review in ref 1b and references therein.)

Let us describe different levels of generalization with respect to a particular organic reaction — in our case, to the dehydrobenzene formation process. Naturally, the least generalized level of the hierarchy is the level of a *complete chemical equation* (Fig. 1a), which explicitly represents all information on the structures of components (molecules, ions, radicals, etc.) in the educt and product systems, including not only all reaction and structural centers and bonds between them but also all substituents and all hydrogen atoms.

However, some atoms and groups in the educt and product systems do not actually participate in the reaction and therefore may be neglected. So, to reflect all and only information on structural changes, one should consider only RCs and bonds between them, and a complete equation is thus transformed into a *reaction equation* (Fig. 1b).

At the next level of generalization, a bond-and-sign redistribution may be described independently of the actual chemical nature of reaction centers (H, O, N, C, P, etc.). For such a description, one should regard them as some "abstract" reaction centers, denote them by special symbols (preserving only the sign, if any), and thus obtain a *symbolic equation*; see Fig. 1c for an example.³⁴ This equation plays a very important role in the Formal-Logical Approach: it provides a clear and evident representation of the most essential information on the overall structural changes but completely ignores less important evidence such as unchanged bonds or the nature of atoms involved.

It is important that equations at all generalization levels are intended to depict not only the structures of educts and products but also the pattern of valence reorganization. To unambiguously represent this information, organic chemists traditionally use curved arrows similar to those in Figs. 1b,c (in the complete equation of Fig. 1a, these arrows are omitted due to lack of space).

Up to this moment, each kind of equation was represented by two graphs corresponding to the initial and final systems. However, a more compact representation would be convenient for discussion of generation problems. To obtain it, let us consider a complete, reaction, or symbolic equation and superimpose its left-hand and right-hand parts. As a result, one gets a single graph whose vertices (corresponding to reaction centers, structural centers, or "non-centers") are connected by an edge if and only if a bond between these centers exists in the educt or product system. Edges of this graph are marked

with “double” labels of the form a/b , with a and b ($0 \leq a, b \leq 3$; $a + b > 0$) being the multiplicities of bonds between the corresponding atoms in the initial and final systems, respectively. If some RC is signed in the initial system, it retains the same sign in the superimposed graph; if it is signed in the final system (note that it cannot be signed in both systems, see above), its sign is parenthesized: “(+)”, “(-)”, or “(.)”.

Evidently, a graph thus constructed unambiguously characterizes the equation under consideration. Three kinds of such graphs, i.e., G_{CEQ} , G_{REQ} , and G_{SEQ} , are exemplified in Figs. 1d–f; these graphs are very similar to Fujita’s^{14a} “imaginary transition structures”, “reaction center graphs”, and “reaction graphs”, respectively. Note that various kinds of labeled superimposed graphs have also been used by many other authors; only selected references³⁵ are given here.

Further, a representation more generalized than G_{SEQ} can be obtained if one considers only the topology of the bond-and-sign redistribution rather than the magnitudes of the actual changes in bond multiplicities. In this way, one can construct the *topology identifying graph* G_{TOP} , which is similar to G_{SEQ} but contains only unlabeled ordinary edges (the graph in Fig. 1g can serve as example).

Note that G_{TOP} is still a signed graph; that is, it contains information on the positions and types of signed reaction centers, if any. Disregarding this information results in the *unsigned topology identifying graph* G (see Fig. 1h); this graph may be regarded as the “parent” for all graphs successively produced by labeling of its vertices and/or edges. On the other hand, if one disregards the connectivity information in G_{TOP} , then an empty graph with 0 or 2 signed vertices (such as the graph in Fig. 1i) will be obtained. This graph makes it possible to recognize the number and nature of non-neutral atoms in the initial and final systems (and the total number of added or removed electrons in the case of redox processes). In the Formal-Logical Approach, this information is associated with the *category* of the process (e.g., neutral, ionic, radicalic, etc.) and with its *subcategory*, which is determined by the actual number of SRCs bearing each of the possible signs (“+”, “-”, or “.”) in the initial and final systems.³⁶

Finally, note one additional important detail. We may also choose a slightly different strategy of generalization and consider equations and graphs that include not only reaction centers and bonds between them but also structural centers and additional ring systems containing them.^{1b} The corresponding levels of the hierarchy are represented by *skeletal equations*, *structural equations*, and (*signed and unsigned*) *expanded topology identifiers*—i.e., by analogs of reaction equations, symbolic equations, and graphs G_{TOP} and G . The skeletal and structural equations and the labeled superimposed graphs corresponding to them have not been considered in other general approaches developed up to now. These equations and graphs also play an important role in solution of reaction design problems; their generation will be implemented in future versions of the ARGENT program system.

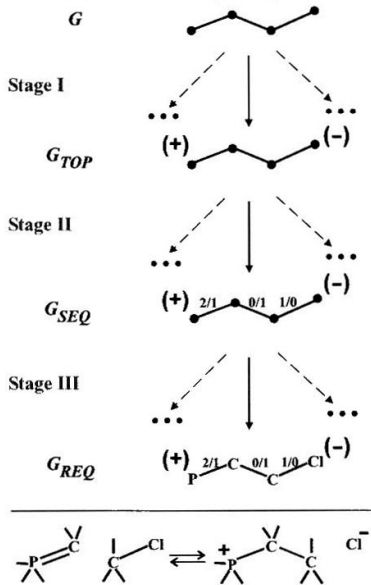
5. MAIN STAGES OF THE GENERATION PROCESS

Let us discuss the generation process starting from a preselected parent graph G that represents some topology of bond redistribution. In ARGENT-1, the stages of the generation process are in one-to-one correspondence with the above-described levels of the reaction hierarchy. Thus, the first generation stage consists in obtaining all possible graphs G_{TOP} from G (if the category of the process is not neutral, because otherwise there are no signed

reaction centers and $G_{TOP} \equiv G$). At the second stage, all symbolic equations represented by graphs G_{SEQ} are to be constructed for each of the graphs G_{TOP} . Finally, at the third stage, the program must produce all reaction equations represented by graphs G_{REQ} for each graph G_{SEQ} . Generation of graphs G_{CEQ} , which identify complete chemical equations, lies outside the scope of this work.³⁷

That is, generation may be regarded as a multistage process $G \Rightarrow G_{TOP} \Rightarrow G_{SEQ} \Rightarrow G_{REQ}$, where each stage means derivation of more detailed reaction representations from less detailed ones. The three successive stages — **I** ($G \Rightarrow G_{TOP}$), **II** ($G_{TOP} \Rightarrow G_{SEQ}$), and **III** ($G_{SEQ} \Rightarrow G_{REQ}$) — are exemplified in Fig. 2 for a well-known nucleophilic substitution reaction belonging to the **ca/n** subcategory. It is important that generation processes at all these stages have much in common: they consist in assignment of some labels to vertices or edges of a graph. As a result, the overall problem is reduced to a *multistage graph labeling problem*.

Figure 2: Three labeling stages ($G \Rightarrow G_{TOP} \Rightarrow G_{SEQ} \Rightarrow G_{REQ}$) illustrated for the reaction equation that depicts a phosphonium ylide alkylation process.



At the first stage, generation of signed topology identifiers G_{TOP} from an unsigned graph G consists in assignment of sign labels — “+”, “-”, “.”, “(+)”, “(-)”, “(.)”, and the blank label corresponding to unsigned RCs — to all graph vertices; in ARGENT-1, the actual choice of sign labels to be used depends on the preselected subcategory (or

subcategories). For example, to obtain equations representing dipolar (**ca/n**) processes, the sign labels “+” and “-” (or “(+)” and “(-)”, see stage I in Fig. 2) must be used. In general, each graph G_{TOP} within the Formal-Logical Approach contains two signed reaction centers for any non-neutral subcategory. Therefore, two vertices in each graph G_{TOP} must be assigned non-blank labels, while all other vertices must bear the blank label. Note that six out of seven labels used for generation of G_{TOP} are actually paired labels, that is, have counterparts that are converted into them by reversal of the reaction direction: “+” and “(+)”, “-” and “(-)”, “.” and “(·)”. The only unpaired label is the blank one, indicating that the atom bears no charge or free electron both in the initial system and in the final one.

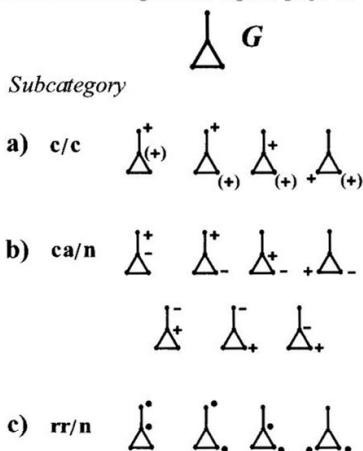
At the next stage, generation of graphs G_{SEQ} — and hence symbolic equations — from an unsigned or signed topology identifier G_{TOP} consists in assignment of “double” labels of the form a/b ($a, b = 0, 1, 2, 3; a \neq b$) to all edges of G_{TOP} . All labels used for this purpose are evidently paired: labels a/b and b/a are considered as counterparts. Indeed, each label a/b is converted into b/a if the direction of a reaction process is reversed. Assignment of the bond labels 2/1, 0/1, and 1/0 to three edges of graph G_{TOP} is illustrated by stage II in Fig. 2.

At the last stage, generation of graphs G_{REQ} — and hence reaction equations — starting from some graph G_{SEQ} is reduced to assignment of atom labels from a preselected set (H, O, N, C, P, S, I, ...) to all graph vertices, i.e., reaction centers. (Note that atom labels must unambiguously represent the valence states of corresponding atoms; for this reason, different labels may be attributed in ARGENT-1 to, e.g., divalent, tetravalent, and hexavalent sulfur atoms.) Evidently, reversal of the reaction direction does not affect the atom name and its actual valence state(s); that is why all corresponding labels are regarded as unpaired. In the example of Fig. 2, the graph G_{REQ} constructed at stage III via vertex labeling of graph G_{SEQ} uniquely corresponds to the reaction equation under consideration (see the bottom of this figure).

It should be stressed that, at any of the labeling stages, the *lists* of resultant graphs rather than a single “best” labeled graph are typically produced. In Figs. 3a–c, the complete lists of signed topology identifiers for the “methylocyclopropane-like” graph G are represented; in the case of the selected subcategories (**c/c**, **ca/n**, and **rr/n**), these lists consist only of 4, 7, and 4 graphs, respectively. In the general case, the number of results obtained at stage I is relatively small.

On the contrary, complete lists of graphs G_{SEQ} generated at stage II from a single graph G_{TOP} can hardly be examined. In accordance with theoretical calculations (based on the general methodology of ref 24a), the total number of symbolic equations for an unsigned graph $G_{TOP} \equiv G$ in Fig. 3 equals 656; surely, only a small part of these equations is of interest to an organic chemist. For illustrative purposes, in Figs. 4a–h we present all edge-labeled graphs G_{SEQ} that correspond to symbolic equations^{19b} with (1) only one RC changing its valence (by two units) and (2) without RCs whose formal valences in the left-hand and right-hand parts of the symbolic equations exceed three. The symbolic equations are exemplified by (a) really observed^{39a} and (b) predicted^{19b} processes of phosphonium ylide formation from dibenzoyl ethylene and (substituted) cyclopropene, respectively; (c) deoxygenation of a peroxide^{39b} and (d) desulfurization of an episulfide^{39c} by tertiary phosphines and phosphites, respectively; (e) hypothetical formation of sulfonium ylides from sulfides and azirines as methylenating agents; decomposition processes of (f) substituted aminonitrene,^{39d} (g) cyclopropylnitrene,^{39e} and (h) cyclopropenylni-

Figure 3: Complete lists of signed topology identifiers G_{TOP} corresponding to (a) cationic, (b) dipolar, and (c) diradicalic subcategories for a given graph G .



trene.^{39f} Several additional examples can be found in refs 20a and 25b.

The numbers of vertex-labeled graphs G_{REQ} (and reaction equations) produced at stage III for a preselected G_{SEQ} can be very large; their multiplicity has been discussed in literature (see, e.g., refs 11c and 40) for some specific cases. In this paper, we present only four graphs G_{REQ} (see Figs. 5) constructed for the edge-labeled graph G_{SEQ} of Fig. 4c; the corresponding reaction equations represent the oxidative cleavage of (a) carbon–boron^{41a} and (b) silicon–silicon^{41b} bonds, (c) desulfurization of disulfides,^{41c} and (d) imination of organic sulfides by the sodium salt of N-chloro-p-toluenesulfonamide.^{41d} Note that the example equations of Figs. 4 and 5 need not necessarily evoke the thoughts about concerted processes; many reaction equations actually represent elementary steps or overall results of multistage organic reactions.

It is important that a novel, unprecedented type of an organic reaction may be observed at any stage of the multistage labeling process: a reaction or symbolic equation generated by the ARGENT-1 program (and sometimes even a topology identifier) may be of the type never encountered before. It is also evident, however, that the degree of novelty depends on the actual stage at which this “novelty” was observed. For example, a symbolic equation that represents a new bond-and-sign redistribution pattern possesses a higher degree of novelty than a reaction equation, because in the latter case the corresponding reaction process differs from previously known ones only by the nature of some atoms.

Finally, note one important detail. Of course, not all graphs generated in the above-discussed multistage labeling process must represent realistic chemical interconversions: many of them can be chemically infeasible or even formally forbidden. Here are three

Figure 4: The restricted list of graphs G_{SEQ} , the corresponding symbolic equations, and examples of reaction equations for a four-centered unsigned identifier of a linear-cyclic topology. The reaction equations (b) and (e) correspond to hypothetical processes; all other examples describe already known organic reactions.

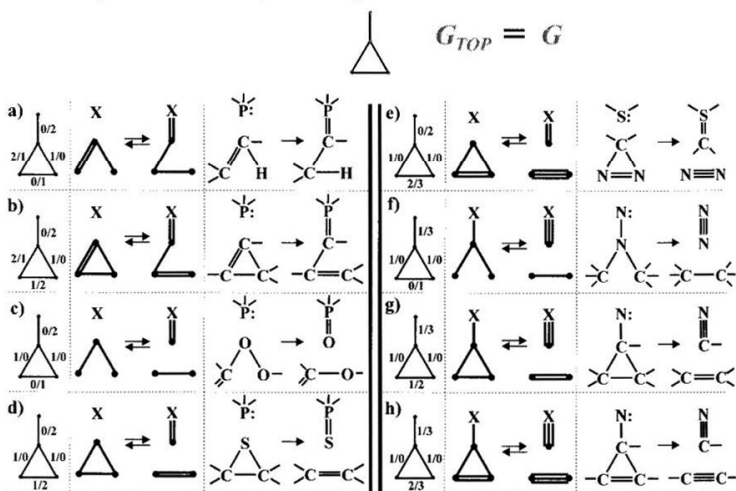
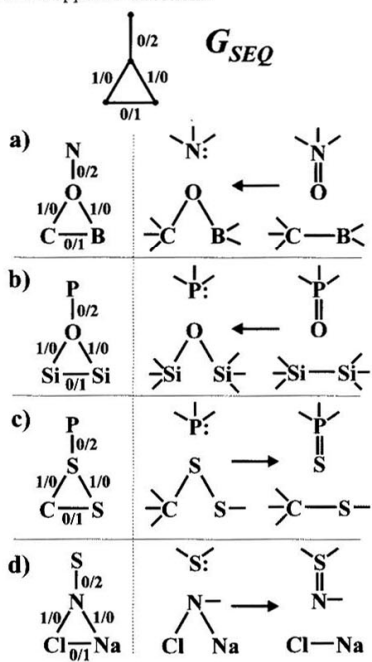


Figure 5: Selected examples of graphs G_{REQ} corresponding to the edge-labeled graph G_{SEQ} in Fig. 4c. The known processes represented by reaction equations (a)–(b) and (c)–(d) actually proceed in opposite directions.



simple rules related to stages I, II, and III of the generation process, respectively; the corresponding selection criteria are incorporated into any version of the ARGENT system and therefore called built-in criteria.

- All signed topology identifiers G_{TOP} where the number of signed reaction centers is not equal to 0 or 2 are automatically disregarded because of the general limitations of the present-state Formal-Logical Approach (see above).
- As was proved in refs 3a,d, the so-called *rank* of a reaction center — i.e., the sum of the absolute values of all changes in the multiplicities of its adjacent bonds — must always be even for an unsigned RC and odd for a signed one. (In the example of Fig. 1f, the rank equals 1 for both SRCs and 2 for all unsigned RCs.) As a result, all graphs G_{SEQ} that do not satisfy this “parity rule” are automatically disregarded.
- If the maximal possible valence of some element is exceeded in the educt or product system, this reaction is *a priori* infeasible. For this reason, all graphs G_{REQ} that contain one or more “unrealistically multivalent” atoms (such as hexavalent carbon or trivalent hydrogen) are automatically rejected.

Furthermore, even if some reaction found by the program is correct from the formal standpoint, this fact does not necessarily mean that this reaction should be considered as one of the desired results: it may be infeasible because of some chemical reasons or just be out of the user’s personal scope of interest. Therefore, ARGENT 1 enables the chemist to specify a set of diverse criteria that can be used for selection of promising interconversions among those generated by the program. The formulation of these “user-defined” criteria and the most important features of their algorithmic implementation will be discussed in future communications of this series.²⁶

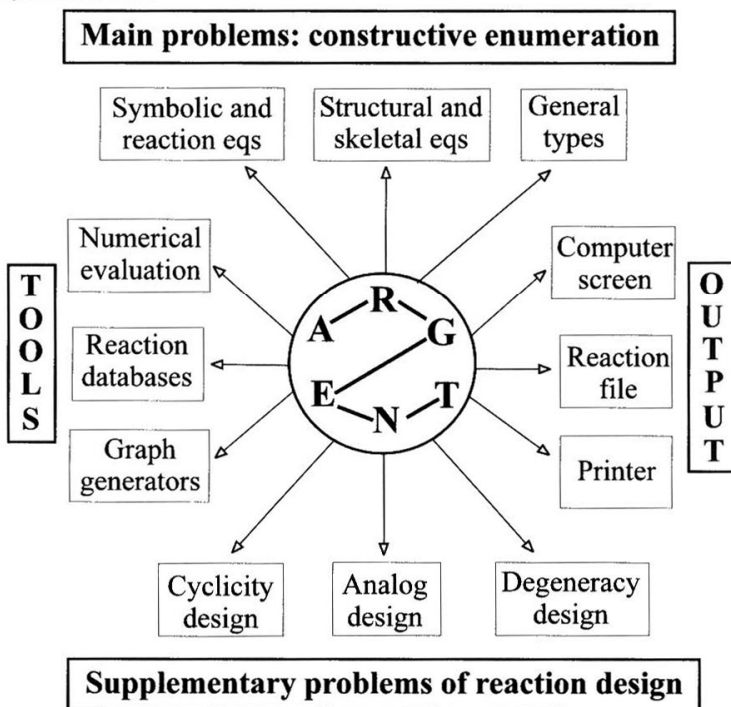
6. POTENTIALITIES AND FUTURE PROSPECTS

The above-considered main principle of ARGENT-1 — i.e., multistage construction of labeled graphs representing chemical interconversions at different levels of the reaction hierarchy — remains valid for other programs of the ARGENT system. However, it is planned that future programs of this system will provide a wider range of generation modes and supplementary means, which are shown in a condensed form in Fig. 6.

Thus, in addition to the built-in and user-defined selection criteria already implemented, future versions of ARGENT can be supplied with new evaluation tools based on some easily calculated numerical (e.g., thermodynamic or other physicochemical) characteristics of output reactions. Another possible tool (see the left-hand part of Fig. 6) is associated with the use of databases containing evidence on already known elementary or multistep organic reactions. In order to organize such databases, one must solve separate sophisticated problems (such as completion of traditionally truncated chemical equations^{42a} or canonization of reaction representations^{23,42b}), but the use of databases offers attractive new possibilities. For example, the degree of novelty could be directly estimated for any reaction devised by a computer via its comparison with similar chemical (or reaction, or symbolic, etc.) equations from a database.

In contrast to the first implementation, next versions of the ARGENT program system will not require explicit input of the parent graph by the user; appropriate lists of

Figure 6: The purposes, techniques, and future possibilities of the ARGENT program system as a whole.



such graphs may be supplied by a separate program module. For this purpose, one can use the known generation technique for "mathematical" graphs⁴³ or utilize any of the structure generators aimed at construction of "chemical" multigraphs (which can satisfy numerous supplementary constraints, see refs 44 for description of some generation techniques). It is important that one of the natural output modes ("to the reaction file", see the right-hand part of Fig. 6) makes it possible to accumulate the resultant equations into virtual "reaction libraries". We anticipate that relevant manipulations with both reaction databases and libraries will reveal new methodologies that can notably enhance the reliability of the predictions obtained.

Now let us turn to general consideration of design problems that we are planning to solve using the present-state and future programs of the ARGENT system. As it is apparent from Fig. 6 (see its top part), generation of symbolic and reaction equations that conform to given constraints is one of these problems; the three stages of the generation process and the basic principles of the corresponding program, ARGENT-1, were outlined in preceding sections. A related problem consists in construction of structural and then skeletal equations (see section 4 and ref 1a) starting from an expanded topology identifier. The principles and stages of the generation process are quite similar in this case, but supplementary "intact" ring structures in one or both parts of the resultant equations must be taken into account.

Another closely related problem can be formulated as search for reactions that belong to some generalized type, such as rearrangements, additions/eliminations, substitutions, etc. Actually, this task may be viewed as construction of appropriate symbolic and reaction (or structural and skeletal) equations combined with subsequent screening for appropriate "general types". Note that all the aforementioned generation problems can also be formulated for a whole family of topology identifying graphs (e.g., produced by some graph generator, see above) rather than for a single parent graph introduced by the user. Surely, more severe restrictions must be used in this case in order to prevent a combinatorial explosion, i.e., a drastic increase in the number of equations generated.

In addition to main generation problems of reaction design, there exist several other problems whose solution must be based on similar techniques; the corresponding supplementary directions of design are listed in the bottom part of Fig. 6. Firstly, we should mention a very interesting task of "cyclicity design", which consists in the use of some well-investigated organic reaction for prediction of synthetic routes to (either already known or new) polycyclic systems. To find novel synthetic possibilities, one should construct and then analyze all ways of supplementary ring attachment to a preselected graph G_{SEQ} or G_{REQ} ; upon our knowledge, this intriguing problem has never been formulated for the general case. The second promising task is based on the idea of "analog design", i.e., of producing structurally analogous symbolic or reaction equations starting from the equation that corresponds to a known chemical interconversion. For this purpose, one or several reaction fragments (unbreakable units in the preselected equation, see section 3 and ref 1) should be successively replaced by other fragments taken from some user-defined list that reflects the similarity relations between fragments under discussion.

Another very attractive problem is "degeneracy design", which is associated with search for new types of degenerate isomerizations. Although theoretical investigations as well as reviews (e.g., see refs 45) on symmetry properties of reaction graphs describing highly degenerate isomerizations have been published, only few attempts⁴⁶ have been made to find new interesting degenerate processes and to clarify the mathematical nature

of degeneracy. Maybe the only example of a constructive approach described in literature is the search^{46a} for bullvalene-like rearrangements in annulene structures; even in this specially chosen case, the exhaustiveness of results was not rigorously proved. However, although the general problems associated with the search for degenerate (and especially highly degenerate) processes are far from being solved,^{46c} some techniques for revealing new possible types of degeneracy have been incorporated into ARGENT-1.

In our final comment, let us note that the results of generation, i.e., graphs G_{TOP} , G_{SEQ} , and G_{REQ} , evidently depend on the symmetry of the starting graph (G , G_{TOP} , or G_{SEQ} , respectively) and on the nature of labels used. (For example, out of 12 formally possible "methylcyclopropane-like" graphs with 2 vertices bearing the "+" and "(+)" signs, only 4 are actually nonequivalent and hence drawn in Fig. 3a). This trivial observation shows that one must thoroughly analyze *the symmetry of the starting graph and the interconvertibility of paired labels* in order to formulate rigorous mathematical models for all stages of the generation process. In turn, such models are necessary not only for devising efficient combinatorial algorithms within the ARGENT system but also for estimating the number of resultant graphs prior to their actual production at any generation stage. These fairly complicated mathematical models are briefly described in the next paper in this issue, and their similarity to more traditional formal models of structural design problems is discussed therein.

ACKNOWLEDGMENT

Partial financial support of this work by the Russian Foundation for Basic Research (Grant No. 00-03-32778a) is gratefully acknowledged.

REFERENCES AND NOTES

1. (a) Tratch, S.S.; Zefirov, N.S. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 331-348. (b) Tratch, S.S.; Zefirov, N.S. *Ibid.*; pp 349-366.
2. The Formal-Logical Approach was suggested and is still used as a tool for (1) unambiguous description of bond (or bond-and-sign) redistributions in the course of real chemical interconversions and (2) systematic search for new, unprecedented types of organic reactions. The main notions as well as results of early applications of this approach have been published in Russian^{3a,b} and in English;^{3c} new concepts and other innovations within this approach were summarized in the doctoral dissertation by one of the authors of this paper^{3d} and in ref 1.
3. (a) Tratch, S.S.; Zefirov, N.S. *Zh. Org. Khimii* **1982**, *18*, 1561-1583. (b) Zefirov, N.S.; Tratch, S.S. *Zh. Org. Khimii* **1984**, *20*, 1121-1142. (c) Zefirov, N.S.; Tratch, S.S. *Anal. Chim. Acta* **1990**, *235*, 115-134. (d) Tratch, S.S. Doctoral Dissertation; Moscow, 1993; Vol. 1, pp 10-153; Vol. 2, pp 27-55, 145-263.
4. Tratch, S.S.; Molchanova, M.S.; Zefirov, N.S. *Next paper of this issue.*

5. (a) Herges, R. *Tetrahedron Comput. Methodol.* **1988**, *1*, 15–25. (b) Herges, R. In *Chemical Structures: The International Language of Chemistry*; Springer: Berlin, 1988; pp 385–398.
6. Several dozens of retro- and prosynthetic design programs have been devised up to date.⁷ In most cases, the search for synthetic routes is empirical in essence and based on previously investigated reactions stored in a database. However, several nonempirical and semi-empirical approaches are known (e.g., see ref 8); the formal rules used in such approaches are typically similar to those implemented in reaction design programs.
7. (a) Barone, R.; Chanon, M. In *Computer Aids to Chemistry*; Ellis-Horwood: Chichester, 1986. (b) Zefirov, N.S.; Gordeeva, E.V. *Russian Chem. Revs.* **1987**, *56*, 1002–1014. (c) Dengler, A.; Fontain, E.; et al. *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 262–269.
8. (a) Hendrickson, J.B.; Grier, D.L.; Toczko, A.G. *J. Am. Chem. Soc.* **1985**, *107*, 5228–5238. (b) Zefirov, N.S.; Gordeeva, E.V.; Tratch, S.S. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 188–193. (c) Gasteiger, J.; Hutchings, M.G.; et al. In *Chemical Structures: The International Language of Chemistry*; Springer: Berlin, 1988; pp 343–359. (d) Hladká, E.; Koča, J.; et al. *Top. Curr. Chem.* **1993**, *166*, 121–197.
9. (a) Dugundji, J.; Ugi, I. *Top. Curr. Chem.* **1973**, *39*, 19–64. (b) Ugi, I.; Bauer, J.; et al. *MATCH* **1979**, *6*, 159–176.
10. (a) Bauer, J.; Herges, R.; et al. *Chimia* **1985**, *39*, 43–53. (b) Fontain, E.; Reitsam, K. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 96–101.
11. (a) Balaban, A.T. *Rev. Roum. Chim.* **1967**, *12*, 875–898. (b) Brownscumbe, T.F. *Diss. Abstr. Int.* **1973**, *B34*, 1035. (c) Hendrickson, J.B. *Angew. Chem., Int. Ed. Engl.* **1974**, *13*, 47–76.
12. (a) Zefirov, N.S.; Tratch, S.S. *Zh. Org. Khimii* **1975**, *11*, 225–231. (b) Zefirov, N.S.; Tratch, S.S. *Ibid.*; pp 1785–1800. (c) Zefirov, N.S.; Tratch, S.S. *Zh. Org. Khimii* **1976**, *12*, 7–18. (d) Zefirov, N.S. *Acc. Chem. Res.* **1987**, *20*, 237–243.
13. (a) Arens, J.F. *Recl. Trav. Chim. Pays-Bas* **1979**, *98*, 155–161. (b) Arens, J.F. *Ibid.*; pp 395–399. (c) Arens, J.F. *Ibid.*; pp 471–483.
14. (a) Fujita, S. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 212–223. (b) Fujita, S. *Ibid.*; pp 224–230. (c) Fujita, S. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 99–104. (d) Fujita, S. *Ibid.*; pp 104–110.
15. (a) Bauer, J.; Ugi, I. *J. Chem. Res.* **1982**, *S11*, 298. (b) Bauer, J. *Tetrahedron Comput. Methodol.* **1989**, *2*, 269–293.
16. (a) Tratch, S.S.; Podimova, E.V.; Zefirov, N.S. In *Application of Computers in Molecular Spectroscopy and Chemical Investigations* (Proc. of the 6th All-Union Conf.); Novosibirsk, 1983; pp 205–206. (b) Zefirov, N.S.; Baskin, I.I.; Palyulin, V.A. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 994–999.
17. (a) Zefirov, N.S.; Tratch, S.S.; Gamziani, G.A. *Zh. Org. Khimii* **1986**, *22*, 1341–1359. (b) Tratch, S.S.; Gamziani, G.A.; Zefirov, N.S. *Zh. Org. Khimii* **1987**, *23*, 2488–2507. (c) Zefirov, N.S.; Tratch, S.S.; Gamziani, G.A. *Description of Organic Reactions in Frame of the Formal-Logical Approach*; Metsniereba Press: Tbilisi, 1993; 208 pp.

18. (a) Babaev, E.V.; Zefirov, N.S. *Bull. Soc. Chim. Belg.* **1992**, *101*, 67–84.
(b) Babaev, E.V.; Lushnikov, D.E.; Zefirov, N.S. *J. Am. Chem. Soc.* **1993**, *115*, 2416–2427.
19. (a) Tratch, S.S.; Baskin, I.I.; Zefirov, N.S. *Zh. Org. Khimii* **1988**, *24*, 1121–1133.
(b) Tratch, S.S.; Baskin, I.I.; Zefirov, N.S. *Zh. Org. Khimii* **1989**, *25*, 1585–1606.
20. (a) Herges, R. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 91–102. (b) Herges, R. *Angew. Chem., Int. Ed. Engl.* **1994**, *33*, 255–276.
21. (a) In fact, there exists a program that generates *R*-matrices on the basis of a given set of constraints. However, this program (briefly mentioned in note 179 of ref 21b) is not a built-in part of IGOR/IGOR2; to our knowledge, no information about the algorithm of this program or the results of its work has been published in journals readily available. (b) Ugi, I.; Bauer, J.; et al. *Angew. Chem., Int. Ed. Engl.* **1993**, *32*, 201–227.
22. (a) Roberts, D.C. *J. Org. Chem.* **1978**, *43*, 1473–1480. (b) Brandt, J.; Bauer, J.; et al. *Chem. Scripta* **1981**, *18*, 53–60. (c) Kvasnička, V. *Coll. Czech. Chem. Commun.* **1984**, *49*, 1090–1097. (d) Fujita, S. *J. Chem. Soc., Perkin Trans. II* **1988**, 597–616. (e) Hendrickson, J.B. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 852–860.
23. (a) Fujita, S. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 128–137. (b) Fujita, S. *Ibid.*; pp 137–142.
24. (a) Tratch, S.S.; Molchanova, M.S.; Zefirov, N.S. Contribution 3 of this series. To be submitted. (b) Tratch, S.S.; Molchanova, M.S.; Zefirov, N.S. Contribution 4 of this series. To be submitted.
25. (a) The implementation and efficiency of procedures that produce “reaction codes” (with simultaneous or subsequent rejection of unsatisfactory codes) are but scarcely discussed in literature; ref 25b seems to be an exception. In one of the forthcoming publications,^{24b} we are trying to partially fill these gaps by describing the back-track generation procedure and the very effective canonicity testing subprocedures. (b) Herges, R. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 377–383.
26. Molchanova, M.S.; Tratch, S.S.; Zefirov, N.S. Contributions 5 and 6. In preparation.
27. (a) Ugi, I.; Stein, N.; et al. *Top. Curr. Chem.* **1993**, *166*, 199–233. (b) Stein, N. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 305–309.
28. (a) Zefirov, N.S.; Tratch, S.S. *Zh. Org. Khimii* **1976**, *12*, 697–718. (b) Zefirov, N.S.; Tratch, S.S. *Chem. Scripta* **1980**, *15*, 4–12.
29. The elucidation of the mechanism for a given complex chemical interconversion is a separate important problem; however, its discussion lies outside the scope of this paper. We can mention only selected references: descriptions of two working computer programs (RAIN^{10b} and PEGAS^{8d}) and several papers by German,³⁰ Czechoslovak,³¹ and other investigators.³²
30. (a) Jochum, C.; Gasteiger, J.; et al. *Z. Naturforsch.* **1982**, *37b*, 1205–1215. (b) Ugi, I.; Fontain, E.; Bauer, J. *Anal. Chim. Acta* **1990**, *235*, 155–161.
31. (a) Koča, J.; Kratochvil, M.; Kvasnička, V. *Coll. Czech. Chem. Commun.* **1985**, *50*, 1433–1449. (b) Pospichal, J.; Kvasnička, V. *Theor. Chim. Acta* **1990**, *76*, 423–435.

32. (a) Zefirov, N.S.; Baskin, I.I. *Zh. Org. Khimii* **1993**, *29*, 449–460. (b) Fujita, S. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 238–242.
33. Actually, “two-sign” (i.e., dicationic, diradicalic, dianionic, and also ion–radicalic) reaction centers might be considered within the same approach. However, this possibility would require a somewhat more complicated model and significantly increase the number of equations generated. At the same time, many processes with such centers can still be easily considered in ARGENT-1: for example, all carbenes may be treated as divalent carbon atoms rather than as diradicals.
34. Designations of RCs in symbolic equations depend on the actual changes in the valences of the corresponding atoms; see section 2 in ref 1a for details. In particular, the designations \bullet and \bar{X} denote ordinary and specific *unsigned* centers whose valence remains unchanged or increases (decreases) by two units, respectively. In the case of *signed* centers, \bullet and \bar{X} denote two possible kinds of atoms changing their valence by one unit; the RC \bar{X} (whose valence in the signed form exceeds its valence in the unsigned form) is referred to as a *pseudospecific* center.^{1a}
35. (a) Kiho, Yu.K. *Reactivity of Organic Compounds (Tartu)* **1971**, *8*, 429–444. (b) Tratch, S.S.; Zhidkov, N.P.; et al. In *Papers Contributed to the 4th Intern. Conf. on Computers in Chemistry*; Novosibirsk, 1979; Vol. 2, pp 479–490. (c) Kvasnička, V.; Kratochvíl, M.; Koča, J. *Coll. Czech. Chem. Commun.* **1983**, *48*, 2284–2304. (d) Fujita, S. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 205–212. (e) Jauffret, P.; Hanser, T.; et al. *Tetrahedron Comput. Methodol.* **1990**, *3*, 323–333.
36. Subcategories are usually identified by a “double” notation of the form “**A/B**”, where **A** and **B** are character strings denoting the presence of cationic (**c**), anionic (**a**), and/or radicalic (**r**) SRCs in the educt (**A**) and product (**B**) systems. The absence of charges and free electrons is denoted by the letter **n**, which stands for “neutral”. For example, the notation **c/c** corresponds to a conversion of one cationic system into another (two signed centers bear the “+” and “(+)” labels). The notation **ca/n** identifies an interconversion between a neutral system and a system that contains a cationic center and an anionic one (the labels are “+” and “–”). The two labels for the biradicalic reaction **rr/n** are identical: “.” and “.”. Note that the designation **B/A** corresponds to the same subcategory as **A/B**, because, as was noted before, the direct and inverse reactions are not explicitly differentiated in the Formal–Logical Approach.
37. Although generation of complete chemical equations may seem to be the natural last stage of the whole generation process, there are two reasons for ignoring this problem in ARGENT-1. Firstly, such generation is of a doubtful practical value, because it would produce a very large number of equations that represent virtually the same process (only substituents at reaction centers can be varied). Secondly, construction of graphs G_{CEQ} (starting from G_{REQ}) necessitates consideration of more complicated mathematical models and elaboration of new algorithms; in contrast to models and algorithms used in ARGENT-1, generalized wreath products of permutation groups must be explicitly taken into account in this case. Wreath product groups and their generalizations are typically applied to description of symmetry associated with non-rigid molecules and constitutional formulas of organic compounds; see refs 38a and 38b, respectively.

38. (a) Balasubramanian, K. *J. Chem. Phys.* **1980**, *72*, 665–677. (b) Zefirov, N.S.; Kaluzhnin, L.A.; Tratch, S.S. In *Investigations on Algebraic Theory of Combinatorial Objects*; VNIISI Press: Moscow, 1985; pp 175–186.
39. (a) Ramirez F.; Madan, O.P.; Smith, C.P. *Tetrahedron Lett.* **1965**, 201–205. (b) Greenbaum, M.A.; Dennie, D.B.; Hoffmann, A.K. *J. Am. Chem. Soc.* **1956**, *78*, 2563–2565. (c) Neureiter, N.P.; Bordwell, F.G. *J. Am. Chem. Soc.* **1959**, *81*, 578–580. (d) Hinman, R.L.; Hamm, K.L. *J. Am. Chem. Soc.* **1959**, *81*, 3294–3297. (e) Baldwin, J.E.; Jackson, D.A.; et al. *Chem. Commun.* **1985**, 206–207. (f) Closs, G.L.; Harrison, A.M. *J. Org. Chem.* **1972**, *37*, 1051–1052.
40. (a) Fujita, S. *Bull. Chem. Soc. Jpn.* **1988**, *61*, 4189–4206. (b) Fujita, S. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 22–30. (c) Fujita, S. *Bull. Chem. Soc. Jpn.* **1989**, *62*, 662–667.
41. (a) Davies, A.G.; Roberts, B.P. *J. Chem. Soc., C* **1968**, 1474–1478. (b) Naumann, K.; Zon, G.; Mislou, K. *J. Am. Chem. Soc.* **1969**, *91*, 7012–7023. (c) Harpp, D.N.; Gleason, J.G. *J. Am. Chem. Soc.* **1971**, *93*, 2437–2445. (d) Mann, F.G.; Pope, W.J. *J. Chem. Soc.* **1922**, *121*, 1052–1055.
42. (a) Ugi, I.; Dengler, A. *J. Math. Chem.* **1992**, *9*, 1–10. (b) Brandt, J.; von Scholley, A. *Comput. & Chem.* **1983**, *7*, 51–59.
43. (a) Faradjev, I.A. In *Algorithmic Investigations in Combinatorics*; Nauka Press: Moscow, 1978; pp 3–11. (b) Faradjev, I.A. *Ibid.*; pp 11–19. (c) Zaitchenko, V.A.; Ivanov, A.V.; et al. *Ibid.*; pp 19–25.
44. (a) Masinter, L.M.; Sridharan, N.S.; et al. *J. Am. Chem. Soc.* **1974**, *96*, 7702–7714. (b) Grund, R.; Kerber, A.; Laue, R. *MATCH* **1992**, *27*, 87–131. (c) Kvasnicka, V.; Pospichal, J. *Chemom. Intell. Lab. Syst.* **1993**, *18*, 173–181. (d) Molchanova, M.S.; Shcherbukhin, V.V.; Zefirov, N.S. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 888–899.
45. (a) Jones, G.A.; Lloyd, E.K. *Stud. Phys. Theor. Chem.* **1983**, *28*, 252–267. (b) Klin, M.H.; Tratch, S.S.; Zefirov, N.S. *J. Math. Chem.* **1991**, *7*, 135–151. (c) Klin, M.H.; Zefirov, N.S. *MATCH* **1991**, *26*, 171–190. (d) Balaban, A.T. *Understanding Chem. React.* **1994**, *9*, 137–180.
46. (a) Lushnikov, D.E.; Gordeeva, E.V.; Zefirov, N.S. *Tetrahedron Comput. Methodol.* **1990**, *3*, 417–427. (b) Nourse, J.G. *J. Am. Chem. Soc.* **1980**, *102*, 4883–4889. (c) Tratch, S.S.; Molchanova, M.S.; Zefirov, N.S. In *Molecular Modeling* (Proc. 2nd All-Russian Conf.); Moscow, 2001; p 18.