Computer-aided design of new organic transformations: exposition of the ARGENT-1 program[†]

Marina S. Molchanova,[1] Serge S. Tratch[2] and Nikolai S. Zefirov[2]*

[1]Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, Leninskii pr. 47, Moscow 117913, Russia
[2]Department of Chemistry, Moscow State University, Leninskie Gory, Moscow 119899, Russia

ABSTRACT: The principles and distinctive features of ARGENT, a new software series aimed at a systematic search for novel types of organic interconversions, are described; its advantages in comparison with other reaction design programs are mentioned. The selection criteria used at different stages of the generation process are considered. Several interesting reactions proposed using ARGENT-1, the first program of the ARGENT series, are discussed. Prospects for the further development of ARGENT are outlined. Copyright © 2003 John Wiley & Sons, Ltd.

KEYWORDS: reaction design; organic reaction types; ARGENT program series; Formal–Logical Approach; selection criteria; novel proposed reactions

INTRODUCTION

In the present-day literature on mathematical chemistry, there are many dozen of computer programs that systematically generate certain 'objects' of organic chemistry, e.g., structural formulas of chemical compounds (structural design), synthetic pathways leading from or to a certain compound (computer-assisted organic synthesis, CAOS) and schemes of possible interconversions of organic compounds (reaction design). In all these cases, a computer program plays the role of an assistant to a chemist's imagination. Indeed, such a program should exhaustively produce all the relevant results (e.g. all isomers with a given molecular formula or, say, all possible skeletal transformations of a given compound) on the basis of some formal rules; this means that it will not miss even a single result, however exotic or chemically infeasible it might appear. Hence the chemist's task is just to select the solutions that are of interest for this or another specific problem. There is a well-known example in the field of structural design: there theoretically exist 217 constitutional isomers of the benzene molecule C_6H_6, and only ca 50 of them seem more or less probable from the chemical point of view. However, without a suitable computer program at hand, a chemist will hardly draw either all these 217 isomers or

even only all feasible isomers. Most probably, his or her imagination will not go beyond some 15–20 structures.

The same situation is observed for reaction design. Indeed, a manual systematic listing of all organic reactions belonging to a certain class is actually possible only if this class is very simple and narrow. For example, a well-known class of pericyclic reactions are represented by six-centered bond redistributions, such as the Diels–Alder process, ene/retroene reaction, Cope rearrangement, etc. If we (a) specify that all bonds change their multiplicity during the reaction from 0 to 1 or from 1 to 2 (or vice versa), (b) require that the valences of all atoms involved in the reaction remain unchanged and do not exceed 3 and (c) do not distinguish between direct and inverse reactions (this is very important for our further discussion), then it is easy to construct manually the 13 theoretically possible redistributions of bonds belonging to this extremely narrow class. This was first accomplished by Balaban;[1] the results were also obtained independently by other researchers via manual calculations[2–5] and then supported using computer programs.[6–8] The 12 bond redistributions are reproduced in Fig. 1 (the 13th one represents not a reaction but a resonance in the benzene ring and hence is omitted). For example, the Diels–Alder cycloaddition/cyclofragmentation process, the ene/retroene reaction and the Cope rearrangement are shown in Figs. 1(b), (d) and (j), respectively. Of course, this number of results is moderate, and examples of all the corresponding reactions are actually known.[1,2,5,9] However, now let us set some less strict conditions, i.e. allow atom valences to change (by two units, e.g., from 2 to 4 or vice versa) during the reactions. The number of bond redistributions immediately increases[3,8] to 226.

Moreover, consideration of all formally possible changes in the bond multiplicities yields more than 11 000 results. Hence manual listing becomes impossible even in slightly more complicated situations, and a computer is needed to generate all bond redistributions of interest.

Before describing our new program series for the generation and evaluation of unprecedented organic reactions, we should note another important aspect. Some people may confuse reaction design programs with programs for computer-assisted organic synthesis, because programs of both types are aimed at generating organic interconversions. However, these two classes of programs are actually different from each other.[10] Strictly, computer-assisted organic synthesis means the construction of reaction pathways leading either to (retrosynthesis) or from (prosynthesis) a *given compound*,[11–13] and bilateral generation of reaction networks (as in the well-known RAIN program[14]) combines these two possibilities. In contrast, reaction design in the strict sense of the phrase means that a program searches for some *novel ways* in which *any* chemical compounds (not specified beforehand) can react. That is, the emphasis is on the essence of the transformation type itself rather than on how it can be implemented for this or another definite compound. From this standpoint, reaction design is probably less useful than CAOS for synthetic purposes but more important for theoretical purposes, because it is

the diversity of chemical reactions that forms the basis and beauty of organic chemistry.

Two general software products for the purpose of reaction design have been known so far: IGOR/IGOR2, developed by the Munich group in the 1980s,[7,15,16] and SYMBEQ, elaborated by the Moscow University group several years later.[8,17,18] Both programs have been extensively used (see Refs 7, 10, 19 and 20 and Refs 8 and 21–23, respectively) to construct complete lists of certain bond redistributions (similar to those in Fig. 1) and to search systematically for new reactions; some interesting predictions made using these programs have been experimentally verified. During the last decade, new attractive possibilities (associated with the algorithmic efficiency, chemical versatility and user-friendly interface of reaction design programs) have arisen, and hence both IGOR/IGOR2 and SYMBEQ now seem somewhat obsolete. That is why we started to develop new, 'second-generation' software products forming the ARGENT (Automatic Reaction Generation and Evaluation of New Types) program series. In this paper, we outline the main advantages of the first of our programs, ARGENT-1, and illustrate its practical applications. A detailed account of the theoretical principles, models, algorithms and user-oriented features of this program will be given in subsequent papers.
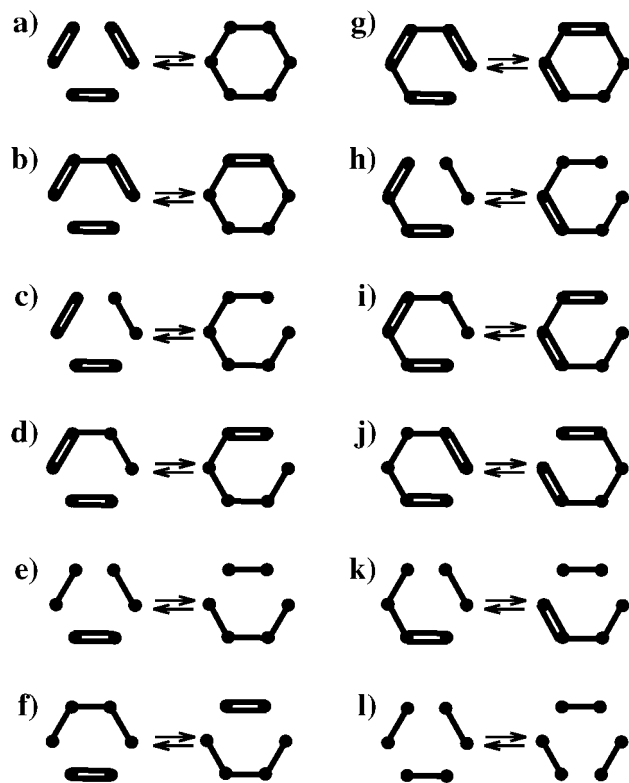
FORMAL–LOGICAL APPROACH: REACTION HIERARCHY

For the systematic generation of chemical reactions, one needs to have their unambiguous formal description and rigorous classification. Unfortunately, traditional organic chemistry does not provide these possibilities in the general case;[24] reactions are mostly characterized by trivial names or systematized using general words such as 'addition,' 'substitution,' 'recyclization,' etc. Such classification attributes are hardly applicable to reaction design problems. On the other hand, the structures of the reaction educts and products cannot form the basis of a systematic search for new reactions or estimation of their degree of novelty, because it is the inverconversion itself (rather than the set of structures involved) that is essential in formulating the reaction type.

Realizing this, many researchers invented their own formal tools to describe and classify any simple (elementary) or complex (multi-stage) organic reactions. Although several non-hierarchical approaches are known,[25–27] most scientists admitted that the solution of design problems must be based on hierarchically organized representations of organic reactions, suitable for describing any interconversion at different levels of generality. It is interesting that hierarchies used by many researchers[28–32] are closely similar; the differences between them and our hierarchical classification scheme were briefly outlined in a previous paper.[33]

Figure 1. The 12 traditionally considered bond redistributions corresponding to six-centered pericyclic reactions

Similarly to several other hierarchically organized representations of organic reactions, the Formal–Logical Approach was initially elaborated[34,35] for a narrow class of chemical transformations (pericyclic reactions), then applied[36,37] to any inverconversions of neutral educt and product systems, and finally extended[33,38] to cover ionic, radical and redox processes; the simplest (linear) multi-centered reactions with the participation of ions and/or radicals were treated in separate papers.[39,40]

The Formal–Logical Approach to organic reactions, in accordance with its name, considers any organic inter-conversion only from the formal standpoint, i.e. just as a conversion of the initial (educt) system of reagents into the final (product) system without any regard for the reaction mechanism, catalysts, conditions, etc. From this standpoint, the transition from one system to another is reduced just to breaking, forming or changing the multiplicity of bonds between some atoms (and possibly also to a redistribution of formal charges and signs of free radical, if any). Within such an approach, any inter-conversion (for example, the one representing the actually implemented[41] synthetic route to α-chloro-β, β-dicyanovinylsulfinylamine, see Fig. 2) can be described at the following levels of generality:

1. (Unsigned) topology identifier G. This is a graph whose edges correspond to all bonds actually participat-

ing in the reaction (i.e. forming, breaking or changing their multiplicity) and whose vertices represent atoms (*reaction centers*) adjacent to these bonds. For example, all six-centered pericyclic reactions in Fig. 1 are characterized by the same cyclic graph G with six vertices; the example reaction in Fig. 2 is characterized by the graph in Fig. 2(a), etc. Simple, one-stage processes usually have linear or cyclic topology identifiers, whereas multi-stage reactions are typically represented by more complex graphs G.

2. Signed topology identifier G_{TOP}. If none of the reaction centers are charged or radicalic, this graph is identical with G. However, if a reaction proceeds with the participation of charged or radicalic species, then G_{TOP} is obtained from G by assigning the '+,' '–,' and/or '·' labels to the relevant reaction centers in G. [If an atom is signed in the product rather than educt system of the process, its 'sign label' in G_{TOP} is parenthesized; cf. the '–' and '(–)' labels in the graph in Fig. 2(b).] This level in the hierarchy of the Formal–Logical Approach is unique compared with other known hierarchies, and its consideration makes ARGENT-1 especially well adapted to generation of ionic, radical and redox processes.

3. Symbolic equation. The actual bond redistribution during any reaction is formally represented by the changes in the multiplicities of all participating bonds (a complete formation or breaking of a bond may also be regarded as a change in its multiplicity—from zero or to zero, respectively). That is, each edge in graph G_{TOP} must be assigned a 'bond change' label *a/b* (0/1, 1/2, 2/3, 0/2, …, or vice versa, 1/0, 2/1, 3/2, 2/0, …), where a denotes the multiplicity of the bond between the corresponding atoms (zero if there is no bond) in the initial system and b denotes its multiplicity in the final system. After such labels have been assigned to all edges of G_{TOP} to form an edge-labeled graph G_{SEQ} [Fig. 2(c), left-hand part], one can easily extract the structures of the initial and final systems corresponding to the relevant bond (and maybe sign) redistribution. The equation thus obtained [see the right-hand part of Fig. 2(c)] evidently resembles a traditional scheme of an organic interconversion, but with reaction centers denoted by some special symbols. Therefore, such an equation is referred to as a *symbolic equation* in our papers.

4. Reaction equation. A more detailed representation of an organic reaction must make allowance for the real chemical nature of reaction centers. Therefore, at the next level of the hierarchy, definite 'atom labels' are assigned to all vertices of graph G_{SEQ}, producing a graph hereafter referred to as G_{REQ} [Fig. 2(d), left-hand part]. Restoring the initial and final systems of the chemical transforma-tion from this graph in the same way as before, one arrives at the *reaction equation* [see the right-hand part of Fig. 2(d) for an example]. Note that some information of chemical significance [e.g. the unaffected cyano groups and the oxygen atom of the S=O group in Fig. 2(d)] is still lost in these equations.
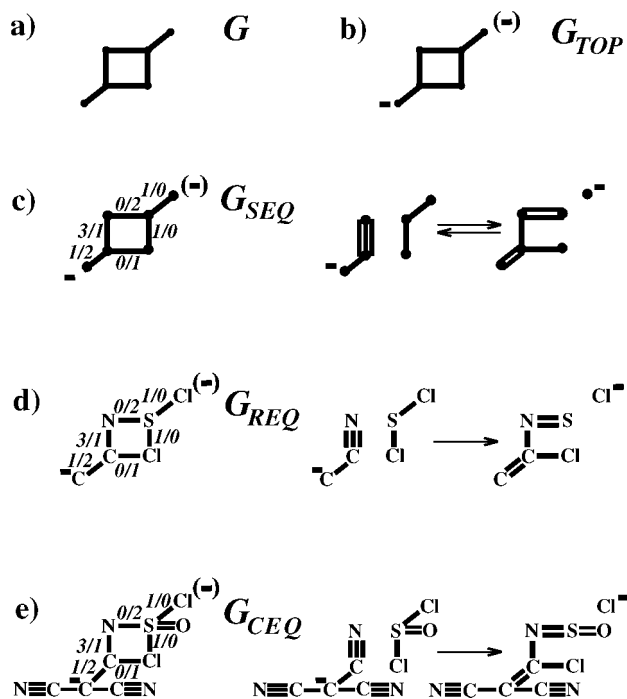

Figure 2. The hierarchy levels exemplified for the reaction of the tricyanomethide anion with thionyl chloride: (a) unsigned topology identifying graph G; (b) signed topology identifying graph G_{TOP}; (c) graph G_{SEQ} and symbolic equation; (d) graph G_{REQ} and reaction equation; and (e) graph G_{CEQ} and complete chemical equation

5. Complete chemical equation. A reaction equation contains the complete information associated with the atoms and bonds directly participating in a reaction. However, it does not include the individual atoms, functional groups, chains or rings that remain unaffected in the course of the reaction although, naturally, they may play an important role in the process (e.g. substituents can reduce or increase the electron density at reaction centers). Considering the complete structures of all components in the educt and product systems of a reaction, one obtains its *complete chemical equation* and the corresponding graph G_{CEQ} [see the right-hand and left-hand parts of Fig. 2(e), respectively].

Now let us confine ourselves to this five-level hierarchy and ask ourselves; which level or levels of this hierarchy must play the key role in reaction design? And a similar question: when should a predicted reaction be regarded as novel? Maybe when it is represented by an unprecedented complete chemical equation? Or a reaction equation never encountered before? Or a novel symbolic equation? Or a unique signed/unsigned topology identifier?

Of course, it would be appropriate to say that an organic reaction will be novel in all these cases, but its 'degree of novelty,' which actually characterizes its value for design purposes, will be different. Indeed, if only the complete chemical equation (but not reaction equation, etc.) of a certain predicted inverconversion belongs to a type never encountered before, this only means that such a new reaction may be obtained from some previously known one just by changing some substituent atoms or atom groups not directly participating in this reaction—a substitution of a methyl at the reaction periphery by an ethyl or cyano group may serve as an example. Evidently, the degree of novelty for such a reaction is typically very low; it is not a novel reaction in the strict sense of the word but probably just a variant of a formerly known process. So, partly for this reason, and partly to avoid an enormous number of possible results, complete chemical equations are not generated by the ARGENT-1 program.

If a proposed interconversion is represented by an unprecedented reaction equation (but its symbolic equation corresponds to one or several known processes), it means that this reaction equation may be obtained from another reaction equation by substituting the symbols of some participating atoms by symbols of other chemical elements. Sometimes this substitution is trivial (e.g. Br for Cl or Se for S), but sometimes it can produce interesting results, as is shown later.

An even higher degree of novelty is encountered if a new reaction suggested by a program is represented by an unprecedented symbolic equation. Indeed, a symbolic equation reflects the essence of the bond redistribution in the process, that is, the way bonds break, form and change their multiplicity. In addition, it is easily seen that this equation actually provides information on other important characteristics of the reaction, such as the changes in the valence numbers of all reaction centers and in the numbers of components and rings in the initial and final systems. As a result, a symbolic equation is the *central* level in the above hierarchy from the standpoint of reaction design.[8] This is why most examples in this paper (see later) are associated just with symbolic equations.

Of course, an even higher degree of novelty is achieved if the topology identifying graph (signed or unsigned) constructed by the program has never been encountered before for known reactions. However, such situations are comparatively rare and, further, topology identifying graphs are of little value by themselves, because their look usually gives no idea of the actual processes that they can represent. Therefore, hereafter we will center our attention only on new symbolic and reaction equations.

STAGES OF THE GENERATION PROCESS

Being based on the reaction hierarchy specified in the preceding section, the currently developed ARGENT-1 program allows one to produce sequentially the symbolic and/or reaction equations for a given unsigned topology identifier G. That is, the generation procedure can be described as follows.

1. In the first stage, all **signed topology identifiers** G_{TOP} are constructed for the current graph G by assignment of 'sign labels' to some of its vertices. The types of signed (positively/negatively charged or radicalic) centers in the initial and final systems, if any, must be specified prior to generation. In accordance with the Formal–Logical Approach, it is required that exactly two reaction centers in G_{TOP} must be signed (this requirement is introduced to reduce the number of results to a reasonable level) and that any vertex may bear a sign either in the initial or in the final system but not in both. Of course, if only equations without any ions or radicals are to be considered, then $G_{TOP} = G$.

2. In the second generation stage, the possible **symbolic equations**—actually, graphs G_{SEQ} representing them—are produced from each graph G_{TOP} generated in the first stage. To produce graphs G_{SEQ}, all edges of G_{TOP} are assigned various 'bond change' labels, 0/1, 1/0, 1/2, 2/1, etc., where the numerator and denominator in each label denote the multiplicities of the corresponding bond in the initial and final systems, respectively (or 0 if this bond is absent in the system in question).

Since the number of symbolic equations can be fairly large (see earlier), it is very important to reduce it by means of certain selection criteria. The optional criteria, the actual choice of which is left to the user, are associated with the chemist's specific interests; they will be briefly discussed in the next section. The only 'built-in' (independent of the user's preferences) criterion is that the difference between the valence numbers of any

reaction center in the initial and final systems must always be odd for a reaction center that bears a sign in either of these systems and be even for an unsigned center. It was theoretically proved[36] that an equation for which this criterion is violated just cannot be regarded as symbolic and must be rejected.

Several examples illustrating the generation of symbolic equations from a given graph G_{TOP} will be considered in detail later.

3. One may either generate only symbolic equations using ARGENT-1 or proceed to the next stage, i.e. generation of **reaction equations**. These equations (actually, graphs G_{REQ} representing them) are produced by assignment of various 'atom labels' (which include not only the symbol of the atom but also its valence number) to all vertices in each graph G_{SEQ}. Note that ARGENT-1 is the first reaction design program that explicitly takes valence numbers of atoms into account: for example, divalent sulfur (S^{II}) is regarded as a label different from tetravalent sulfur (S^{IV}) and hexavalent sulfur (S^{VI}), etc. This distinction makes it possible to use more sensitive criteria for the selection of chemically feasible results.

The number of reaction equations generated from a single symbolic equation is often very large. Therefore, the use of a versatile set of selection criteria is essential. The only built-in criterion here reflects the evident valence limitation: if an atom label is assigned to a vertex, the allowed valence number of this atom cannot be smaller than the summarized multiplicity of all its bonds in the initial or final system. (However, the valence number can evidently be greater than these sums, because it also makes allowance for the bonds unaffected by the reaction.)

Two example sets of reaction equations generated by ARGENT-1 from a preselected symbolic equation are considered later.

The three-stage generation process described here may be represented in the form of a so-called generation tree (Fig. 3)—a rooted tree where the root node is the starting graph, i.e. the unsigned topology identifier G (level 0); the terminal nodes (level 3) are reaction equations, and the nodes of the intermediate levels (1 and 2) correspond to signed topology identifiers and symbolic equations, respectively. ARGENT-1 traverses this tree in the depth-first manner (see dashed arrows in Fig. 3):[42] the program passes to the next symbolic equation only after all reaction equations have been generated for the previous one and passes to the next graph G_{TOP} only after all symbolic (and reaction) equations have been generated for the previous one. However, ARGENT-1 also enables the user to modify this strictly defined order and pass directly to any signed topology identifier or symbolic equation and then resume the generation from this point, either in the forward (from left to right at any level of Fig. 3) or in the backward (from right to left in Fig. 3) direction. Moreover, the generation process may actually

be started at any level of the tree—not only from G but possibly from a given graph G_{TOP} or symbolic equation. Such a convenient navigation about the generation tree makes ARGENT-1 unique among other reaction-generating programs and significantly enhances the efficiency of the search for desired reactions.

The actual generation procedure at each of the three stages, as was repeatedly noted above, consists in listing all possible ways of the assignment of certain labels to vertices or edges of a certain graph. The problem of finding all and only essentially nonidentical labelings necessitates a consideration of two factors: (a) equivalence of labelings due to the symmetry of the starting graph and (b) equivalence associated with some labels (such as '+' and '(+)' or 1/0 and 0/1) being interconverted as the direction of the reaction is reversed. Generation of only non-equivalent (with respect to both factors) labelings is needed in order to avoid the appearance of duplicate results, which can otherwise be intolerably numerous (a combinatorial explosion).

A highly effective algorithm for the assignment of 'non-interconvertible' labels has been known since 1974;[43] its modifications have been successfully applied to generation of molecular graphs[44,45] and to the solution of some design problems.[46] More enhanced algorithms capable of additionally treating interconversions of labels are only mentioned in many of the aforecited publications; a detailed description of such an algorithm implemented in the ARGENT-1 program will be presented in a subsequent paper.

All the models and algorithms mentioned above are mathematically rigorous, and their program implementation in ARGENT-1 is highly efficient: for a moderate-size graph G (up to 6–12 vertices), the program can produce many thousands of symbolic or reaction equations per second on a Pentium PC.
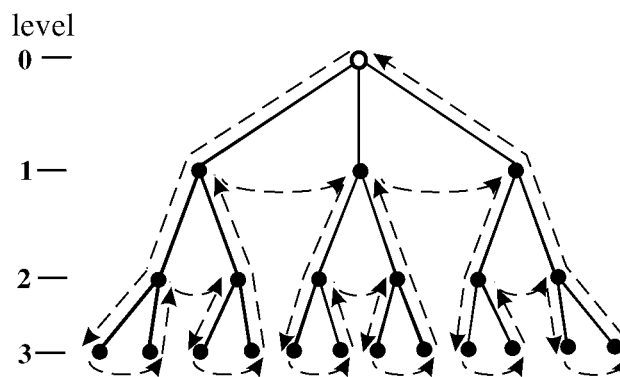

Figure 3. A representation of the multi-stage generation process in ARGENT-1 in the form of a rooted tree. Levels 0, 1, 2 and 3 correspond to the unsigned topology identifier, signed topology identifiers, symbolic equations and reaction equations, respectively

USER-DEFINED SELECTION CRITERIA

As was noted above, the total numbers of symbolic and especially reaction equations that can be generated by ARGENT-1 often amount to many thousands. Hence the crucial problem is to develop a suitable set of optional selection criteria for excluding the symbolic and reaction equations that are chemically infeasible or of little interest for the problem in question. Note that these criteria have to be purely formal and heuristic, because direct calculations of the geometric, thermodynamic, etc. characteristics of several (often incompletely specified) structures forming the initial and final systems are not only time consuming but also are hardly applicable for evaluating the resultant reactions. The criteria currently available in the ARGENT-1 program are briefly considered in this section; they are more powerful and versatile but sometimes more complicated than those used in earlier reaction-generating software.

Criteria for the selection of symbolic equations may be partitioned into two groups: local (referring to separate bonds and reaction centers) and global (referring to the symbolic equation as a whole).

The most natural local criterion is a specification of the *forbidden, allowed* and *required* types of changes in the bond multiplicity during reactions. The six possible types of changes are $0 \leftrightarrow 1$, $1 \leftrightarrow 2$, $2 \leftrightarrow 3$, $0 \leftrightarrow 2$, $1 \leftrightarrow 3$ and $0 \leftrightarrow 3$. For example, bond changes of type $0 \leftrightarrow 3$ (complete formation/breaking of a triple bond) are rare in known organic interconversions and may usually be forbidden. Another example: if one specifies that at least one of the bond changes $0 \leftrightarrow 1$, $0 \leftrightarrow 2$ or $0 \leftrightarrow 3$ is required to be present, then the resultant reactions will always represent 'non-resonance' processes, that is, processes involving a breaking or formation of at least one σ-bond.

Another important local criterion enables the user to limit the valence numbers and their changes for atoms participating in the desired reactions. For a reaction center in the initial or final system of a symbolic equation, the sum of the multiplicities of all its bonds is termed its *reaction number*; this notion of the Formal–Logical Approach is analogous to the usual notion of an atom valence number but calculated with allowance only for the bonds that change their multiplicity during a reaction. The absolute value of the difference between the reaction numbers (i.e. between the atom valences) in the initial and final systems is called the *specificity* of a reaction center; as pointed above, it must always be even for unsigned and odd for signed reaction centers.[38] In order to ensure the generation of only desirable results, the user may input the minimum and maximum values of reaction numbers for reaction centers in the reactions, indicate if they must be actually reached for any center, input the maximum specificity of an individual center in reactions and the maximum (or exact) summarized specificity of all reaction centers. The use of these

criteria makes it possible to exclude many unsatisfactory results; indeed, reaction numbers of individual centers in the majority of real reactions do not exceed 4, and their specificities seldom exceed 2.

The main global criteria in the generation of symbolic equations are associated with the *numbers of components and rings* in the educt and product systems of resultant equations. Apart from the possibility of avoiding chemically unlikely processes (e.g. those with the participation of many components in both systems or with the simultaneous formation of many rings), these criteria also enable a user to produce reactions that may be easily characterized by means of conventional terminology of organic chemistry. For example, to generate rearrangements, a user should specify that the desired reactions are to be one-to-one-component, elimination/addition reactions are usually one-to-two-component, cycle openings/formations are typically one-to-zero-ring reactions, recyclizations are one-to-one-ring reactions, etc.

Other kinds of global criteria make allowance for the topological *symmetry* (associated with the number of automorphisms of graph G that are preserved in graph G_{SEQ}; see Refs 38 and 40 for details) and the *degeneracy* of the desired processes. A degenerate organic interconversion is a reaction where the educt and product systems are identical. The notion of degeneracy can evidently be applied to symbolic equations as well; for example, the equations in Fig. 1(f) and (i)–(l) represent degenerate bond redistributions. ARGENT-1 distinguishes several types of degeneracy.[47] The characterization of the new types lies beyond the scope of this paper, but one illustrative example is given in the next section (Task 5).

Criteria for the selection of reaction equations, compared with those available for symbolic equations, are more evidently chemistry oriented and hence clearer for a user. Hence we will discuss them even more briefly.

The most important local criterion for reaction equations is, of course, the selection of actual *chemical elements* to be used in generation. For each chemical element, its *valence states* to be used in generation should be specified: valence numbers, possible signs, and the allowed and forbidden *distributions* of bonds (single, double and triple) formed by each of these elements in the initial and final systems.

Another type of a local criterion is the specification of *allowed, required and forbidden changes* in the bond multiplicities—similarly to the criterion used in the generation of symbolic equations, but this time referring to bonds between definite kinds of chemical elements. For example, if the user wants to generate reactions involving any kind of addition to the double bond of a carbonyl group, the required type of bond change should probably be $C{=}O \rightleftharpoons C{-}O$. This possibility is very important for the 'chemical orientation' of the program to certain classes of reactions.

The next local criterion is the specification of fragments that must be present in the initial or final system of resultant reactions or, on the contrary, must be absent in both systems. Such sets of fragments are termed GOODLIST and BADLIST, respectively, by analogy with the use of these terms in some structural generators.[48] For example, if the user wishes to consider only reactions of hydroxylamines, it would be appropriate to add the structural fragment N=O—H to GOODLIST; in contrast, to exclude peroxides, the O—O fragment should be added to BADLIST.

The global criteria for reaction equations include the minimum and maximum *contents* of all kinds of atoms participating in desired reactions (e.g. the minimum required number of carbon atoms in organic reactions is usually set to 1 or greater) and, just as in the case of symbolic equations, the degeneracy and symmetry of the equations in question.

EXAMPLES OF PROPOSED REACTIONS

In this section, we present and discuss the results produced by ARGENT-1 during the solution of five tentative problems, formulated as Tasks 1–5. The feasibility and novelty of some proposed reactions are considered, and the diversity of the constraints used illustrates the research potentialities of ARGENT-1.

Task 1

This task is aimed at generating symbolic equations pertaining to a preselected interconversion type. Similarly to the example in Fig. 1, the cyclic topology identifying graph $G = G_{TOP}$ has six vertices; the allowed changes in the bond multiplicities are $0 \leftrightarrow 1$ and $1 \leftrightarrow 2$, and the reaction numbers are from 1 to 3. However, up to two atoms with non-constant valence numbers (specificity equal to 2) may now also be present. Centering our attention only on *addition–elimination processes* without any rings, we arrive at the 12 symbolic equations shown in Fig. 4.

Most of these symbolic equations contain specific centers (X) and hence are absent in Fig. 1, and the equations of Fig. 4(k) and (l) [cf. Fig. 1(d) and (h)] without specific centers actually describe most known and theoretically predicted[7,10] non-cheletropic addition/elimination reactions. Numerous examples of ene/retro-ene reactions [see Fig. 4(k)] and 1,4-addition/elimination processes [see Fig. 4(l)] may be found in literature.[1,5,8,9,49]

With regard to processes with specific centers in Fig. 4(a)–(j), we did not manage to find them among those studied in the literature [possibly with one exception: the bond redistribution in Fig. 4(a) may be said to represent the ene/retroene reaction with the participation of

acetylenes[50] if the dicarbene resonance form $\ddot{C}–\ddot{C}$ is used for the acetylene group instead of C≡C].

On the other hand, many symbolic equations in Fig. 4 provide attractive possibilities for finding new reactions. For example, unusual elimination and addition processes with the participation of *P*-chloro-substituted phosphonium ylides [Fig. 4(e) and (g), respectively] seem interesting, but the corresponding ylide reagents are probably not yet available. The unsaturated bisphosphine in Fig. 4(f) seems to be much more feasible; note that [in contrast to the example in Fig. 4(e)] the valence numbers of its two P atoms simultaneously increase in the hypothetical chlorination process. Naturally, numerous other possibilities can be found from the symbolic equations in Fig. 4 with the use of other heteroatoms, different from P, O and Cl.
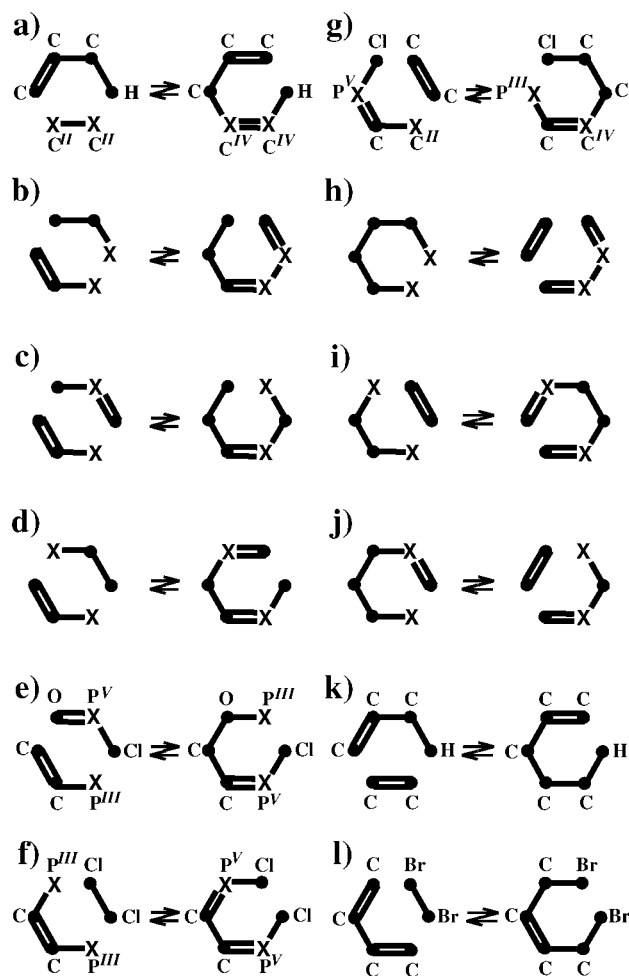

Figure 4. The 12 symbolic equations generated for Task 1 (cyclic topology of bond redistribution, non-cheletropic addition/elimination reactions). Hereafter, the symbols X denote specific reaction centers; the atom labels (with valence numbers if needed for clarity) correspond to examples discussed in text

Task 2

This is an example of searching for *degenerate* bond redistributions corresponding to six-centered pericyclic reactions. The main constraints are similar to those in the previous task: the allowed multiplicity changes are $0 \leftrightarrow 1$ and $1 \leftrightarrow 2$ and no more than two specific centers X are permitted. However, the allowed range of reaction numbers is now from 0 to 4 and the (acyclic or cyclic) initial and final systems may contain one or two components each. Using the supplementary requirement of degeneracy, one finally arrives at 12 symbolic equations (Fig. 5).

Note that only the five equations in Fig. 5(b), (c), (g), (h) and (j) represent degenerate *rearrangements*, that is, actual conversions of a one-component system into itself; five other equations [see Fig. 5(a), (d)–(f) and (i)] correspond to degenerate transformations of two-component systems, and the remaining two equations [Fig. 5(k) and (l)] describe transitions between resonance forms rather than actual chemical reactions.

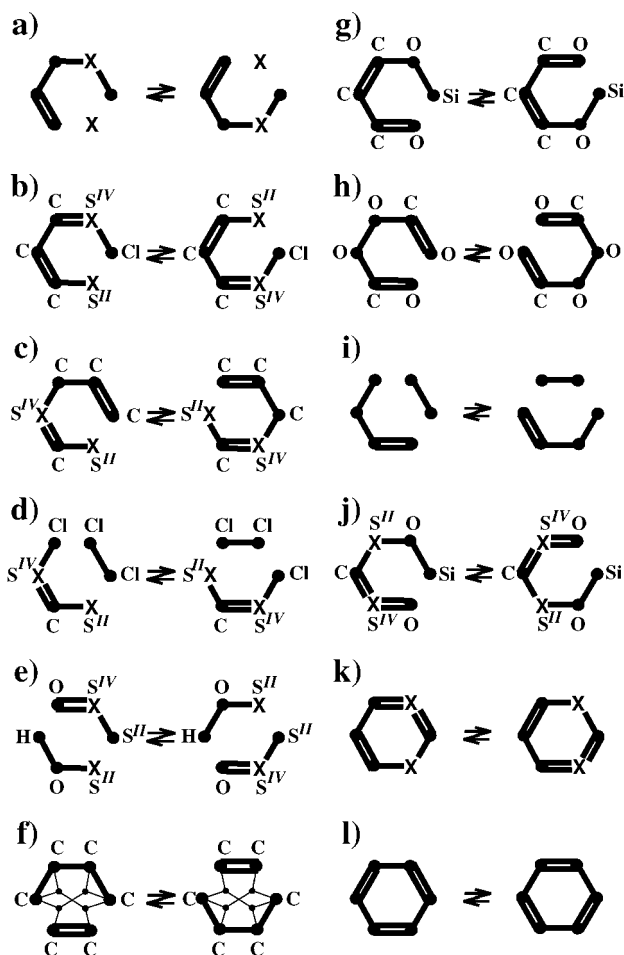The equations in Fig. 5(f)–(i) contain no specific centers and correspond to most of the known[1,5,9,49] degenerate pericyclic processes. For example, Fig. 5(g) and (h) represent the well-known [1,5]- and [3,3]-sigmatropic shifts by two of the many[51] examples: a silylotropy in the enolic form of a β-diketone[52] and a hetero variant of the Cope rearrangement,[53] respectively. Note that the assignment of the chemical element symbols to reaction centers must preserve the degeneracy of the bond redistribution.

An interesting type of degeneracy is exemplified in Fig. 5(f); the known[54] rearrangement of snoutene is formally described [because of the presence of supplementary bonds that do not change their multiplicities, see the thin lines in Fig. 5(f)] by an interconversion of two-component systems. Even more unusual is the actually observed[55] reaction of alkyl thiosulfinates with alkylsulfenic acids [see Fig. 5(e)]; this is an actual example of an intermolecular degenerate process with two specific centers.

Possibly the most intriguing one-component sigmatropic six-centered rearrangement is shown in Fig. 5(c); although this reaction was mentioned in our earlier papers,[8,21] it seems that nobody has yet tried to implement it. It is interesting that this [3,3]-rearrangement is a close analog of the Cope reaction but contains two specific centers, i.e. sulfur atoms.

The symbolic equations in Fig. 5(b) and (j) describe [1,5]-sigmatropic shifts with two specific centers X in the α- and β-positions with respect to the migrating group. The hypothetical process in Fig. 5(j) with a migration of a trialkylsilyl group seems more likely, owing to its similarity to the smoothly proceeding reaction in Fig. 5(e).

Symbolic and reaction equations corresponding to two- and multi-component degenerate interconversions can also describe the total results of two or several (not necessarily degenerate) reactions. For example, the equation in Fig. 5(i) corresponds to allylic substitution; this reaction is known[56] to proceed according to the bimolecular S_N2' mechanism. The bond redistributions in Fig. 5(a) and (d) can also represent the overall results of multi-stage processes, e.g. in systems such as $CR_2=CR'-CR_2''-SCl + S$ and $RS-CR'=S(R'')Cl + Cl_2$; the second of these hypothetical processes should involve the formation of an intermediate symmetrical bis-S-chlorosulfonium cation.

Figure 5. The 12 symbolic equations generated for Task 2 (cyclic topology of bond redistribution, degenerate processes)

Task 3

The symbolic equations considered in this task pertain to interconversions of anionic systems, namely those with a single negatively charged reaction center. In this case, the generation procedure starts from a linear–cyclic graph $G_{TOP} \neq G$ (see top of Fig. 6) and uses the following constraints: the allowed multiplicity changes are $0 \leftrightarrow 1$, $0 \leftrightarrow 2$, $1 \leftrightarrow 2$, and $1 \leftrightarrow 3$; the reaction numbers are from 0
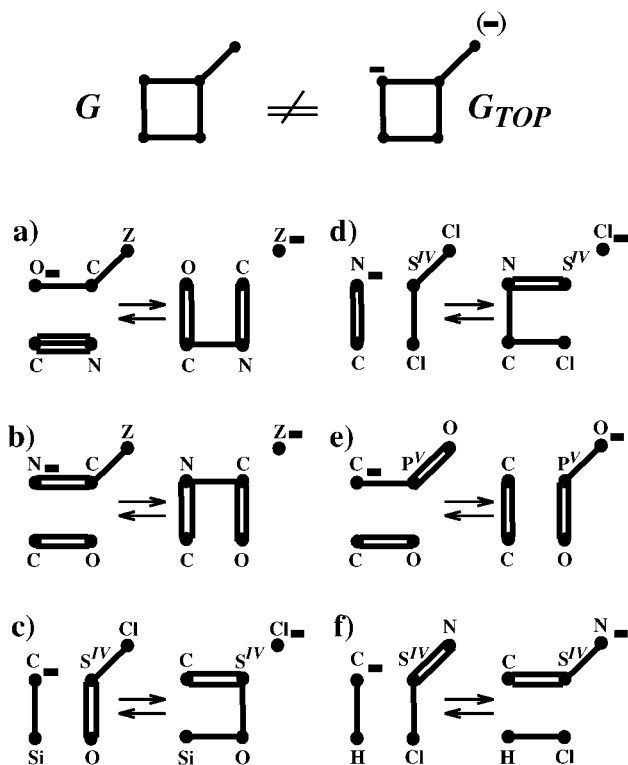
Figure 6. The six symbolic equations generated for Task 3 (linear–cyclic topology of bond redistribution, transformations of anionic systems)

to 3; and the specificity values for unsigned and signed centers are 0 and 1, respectively. Selecting only the results that represent *transformations of acyclic two-component systems*, we arrive at the six equations shown in Fig. 6.

Many actually known organic reactions may be associated with these symbolic equations. For examples, the symbolic equations of Fig. 6(c) and (f) can represent processes known to any organic chemist such as harsh alkaline hydrolysis of vinyl chlorides [HO$^-$ + CR$_2$=CR′Cl → CHR$_2$—C(R′)=O + Cl$^-$] and acyl chlorides [HO$^-$ + RC(=O)Cl → HCl + RC(O)O$^-$]. However, more interesting synthetic possibilities may be proposed for the same two equations; that corresponding to Fig. 6(f) (the formation of *N*-alkylaminosulfonium methylides in the reaction of carbanions with *N*-alkyliminosulfinic acid chlorides) was experimentally confirmed 25 years ago.[57]

The symbolic equation in Fig. 6(e) describes a very popular synthesis of olefins from phosphonate carbanions and carbonyl compounds (the Wittig–Horner reaction),[58] and the example in Fig. 6(d) is associated with the known[41] reaction considered in the second section. Note that the same process is represented in different ways in Fig. 2(c) and (d) and in Fig. 6(d), because different resonance structures are used for the cyanomethide anion; this ambiguity is an inherent feature of the Formal–Logical Approach.

Concerning attractive new possibilities of Fig. 6, one

of them is associated with the aforementioned equation in Fig. 6(c): the (probably not yet studied) reaction of sulfinyl chlorides with silyl-substituted carbanions can result in *S*-silyloxysulfinyl ylides. Two other hypothetical multi-stage reactions [see Fig. 6(a) and (b)] can possibly proceed in the systems consisting of an aldehyde (or ketone), a nitrile and some nucleophilic species Z$^-$. The addition of Z$^-$ to the C=O or C≡N bond and subsequent transformations [in accordance with Fig. 6(a) and (b), respectively] can result in the same final product, *N*-acylazomethine.

Task 4

This task demonstrates the construction of reaction equations rather than symbolic equations. In this case, the generation process starts from a preselected graph G_{SEQ} corresponding to the symbolic equation in Fig. 6(d) and is aimed at searching for close analogs of the aforementioned reaction between KC(CN)$_3$ and SOCl$_2$.

The set of constraints is as follows: the maximum numbers of C, N and O atoms are 3, 2 and 1, respectively; the allowed number of Cl atoms is 1 or 2; and all atoms can be in their standard valence states (C^{IV}, N^{III}, O^{II}, Cl^{I}) or, except for C, in negatively charged states with the valence number reduced by one. The presence of one fragment (C=N$^-$) in GOODLIST and another fragment (O—Cl) in BADLIST is required to search for reactions of imine and keteneimine anions without the participation of undesirable O—Cl bonds.

As a result, the program produces 10 reaction equations (Fig. 7). We have found no literature examples corresponding to them. However, at least four of these equations, i.e. those in Fig. 7(a), (f), (g) and (j) seem to be promising targets for experimental verification. Other hypothetical transformations in Fig. 7 seem to be less plausible, e.g. because NR$_2$ and OR are typically neither good migrating groups nor easy leaving groups in nucleophilic substitution processes.

The closest analogy to the process in Fig. 6(d) is represented by the reaction equation in Fig. 7(f) with sulfur (of the S=O group) simply substituted by carbon (of the C=O group). Hence, the interaction between KC(CN)$_3$ and COCl$_2$ will probably result in α-chloro-β,β-dicyanovinyl isocyanate; thiophosgene CSCl$_2$ and dichloroazomethines RN=CCl$_2$ can also react in a similar manner. An analogous possibility is associated with a migration of an appropriate non-chlorine substituent, such as the benzyl [see Fig. 7(a)] or silyl group.

An interesting group of proposed reactions involves *N*-chloro- and *N,N*-dichloroamines (or amides), as in the reaction equations in Fig. 7(g) and (j), respectively. These reagents are known to be highly active in ionic and radical processes; the readily proceeding hydrazone–azo interconversion can also facilitate the formation of anticipated substituted azo compounds. A reaction of
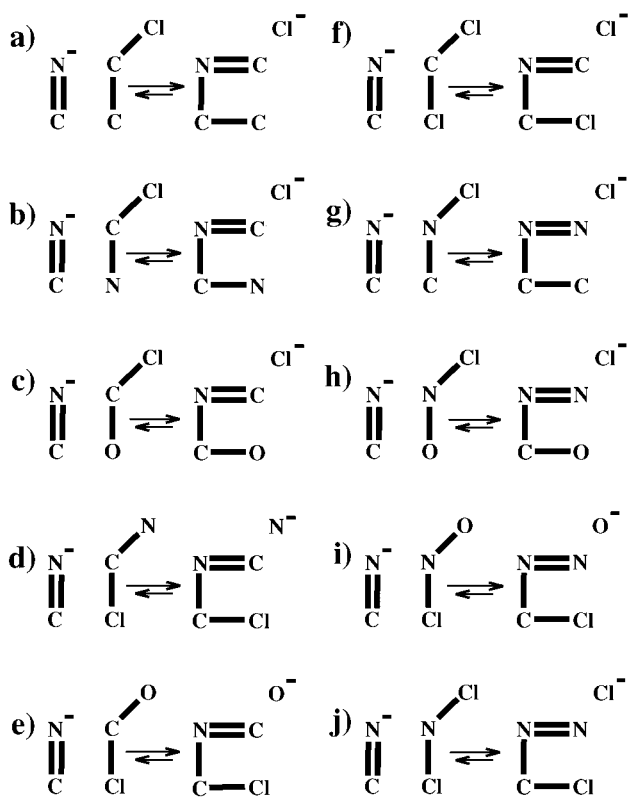
Figure 7. The 10 reaction equations generated for Task 4 [list of equations corresponding to the preselected symbolic equation in Fig. 6(a) and satisfying certain specific constraints]

$KC(CN)_3$ with the readily available *N,N*-dichlorosulfonamide $PhSO_2NCl_2$ seems especially attractive for experimental investigation.

Task 5

In this last task, we do not consider the complete lists of symbolic and reaction equations for a certain set of starting constraints but present only some symbolic equations selected manually among those generated by ARGENT-1 for a certain graph G_{TOP} [see Fig. 8(a)–(c)] and some of the reaction equations produced for one of these symbolic equations [Fig. 8(d)–(k)].

The topology identifier $G = G_{TOP}$ in this case is a comparatively complicated tricyclic graph (see top of Fig. 8), suggesting that all resultant interconversions must be actually multi-stage ones. The generation process using simple local constraints (allowed multiplicity changes $0 \leftrightarrow 1$, $1 \leftrightarrow 2$, $0 \leftrightarrow 2$; reaction numbers from 1 to 3; no specific centers) and global limitations (initial and final systems contain only one component and no more than one cycle each) produced seven symbolic equations, and three of them [see Fig. 8(a)–(c)] were chosen for more detailed discussion.

The symbolic equation in Fig. 8(a) represents the

synthetically important Beckmann rearrangement.[59] Its commonly accepted mechanism is depicted in Fig. 9(a) in a somewhat simplified form; other schemes in Fig. 9 also depict reasonable mechanisms for certain hypothetical processes in Fig. 8. Note that ARGENT-1, in contrast to some other reaction generators (such as that in Ref.14), does not suggest reaction mechanisms, and so the user can do it only according to his or her intuition.

The symbolic (and also reaction) equation in Fig. 8(b) presents a degenerate rearrangement; its possible mechanism is shown in Fig. 9(b). Note that this process differs from all known degenerate processes (e.g. those in Fig. 5), because the isomorphism between the initial and final systems in this case is not associated with the symmetry of graph G_{TOP}. This unusual type of degeneracy, first described in our studies,[47] is called *irregular degeneracy*, and the discussed rearrangement of *N*-(β-chloroethyl) azomethine is one of its most attractive examples.

Several reaction processes corresponding to the symbolic equation in Fig. 8(c) were mentioned in our previous publications.[8,38] In this study, ARGENT-1 generated all reaction equations containing three C^{IV} atoms, no more than two N^{III} and N^{V} atoms and no more than one O^{II} and one P^{V} atoms. Out of the 116 resultant equations, we selected eight [see Fig. 8(d)–(k)] contain-
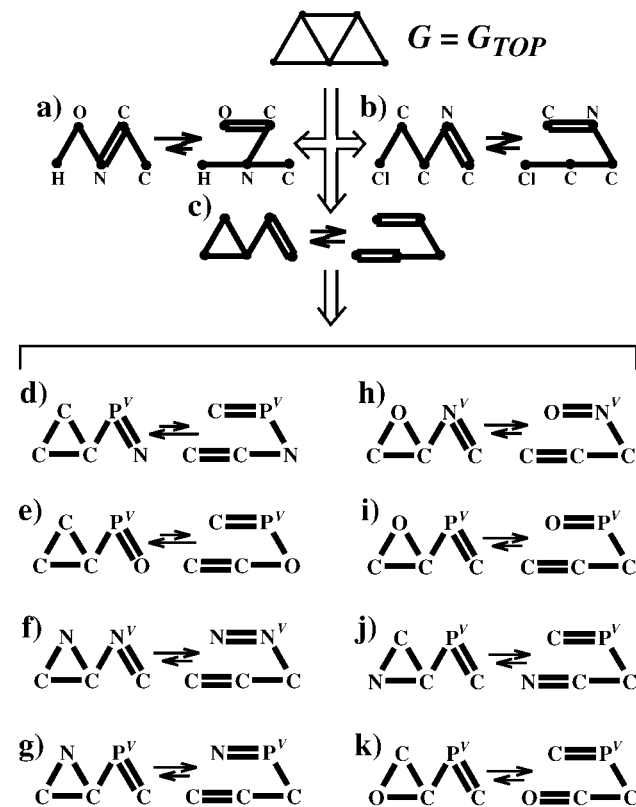

Figure 8. The three symbolic and eight reaction equations generated for Task 5 (tricyclic topology of bond redistribution, manual selection of interesting synthetic possibilities). The expected directions of the reactions are indicated by the longer arrows
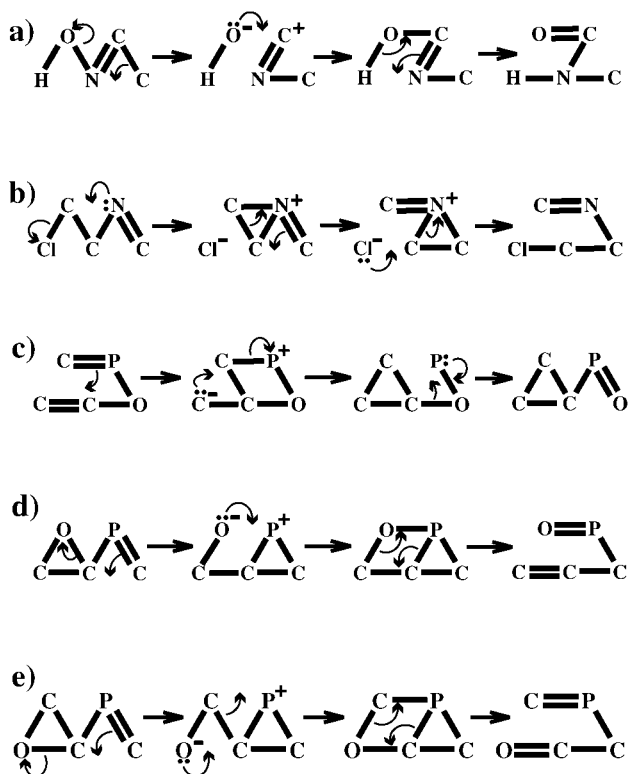
Figure 9. The possible mechanisms of rearrangements represented by some reaction equations in Fig. 8: (a)-(e) correspond to Fig. 8(a), (b), (e), (i) and (k) respectively

ing cyclopropane, aziridine or oxirane rings; it seems that none of them have been experimentally implemented.

The transformations in Fig. 8(d) and (e) will probably proceed in the opposite direction, towards more stable phosphinimines and phosphine oxides. The three-stage mechanism of Fig. 9(c) seems to be reasonable for such reactions; its implementation can probably provide an original synthetic route to cyclopropyl-substituted phosphine oxides (or corresponding phosphinimines).

The reaction equations in Fig. 8(f)–(i) describe the hypothetical formation of allyl-substituted azoxy compounds, phosphinimines, nitro compounds and phosphine oxides from appropriate compounds containing aziridine or oxirane rings [note that the unaffected N → O bond is not shown in the reaction equations in Fig. 8(f) and (h)]. The opening of the unstable three-membered rings and the formation of a P=N or P=O bond seem to be the driving forces for these unusual interconversions. The mechanism proposed for one of such reactions [Fig. 9(d) for Fig. 8(i)] also seems reasonable. The last two reactions [Fig. 8(j) and (k)] represent possible interconversions of the same phosphonium ylides as in Fig. 8(g) and (i) but the products (ylides) and the mechanistic scheme [Fig. 9(e)] are different.

Note that, in all the above examples, a chemical process is said to be 'not yet implemented' only if we found no information on it during a brief survey of the literature data and Internet sites available to us. To be more certain that a suggested process is novel, one should perform a comprehensive study of literature sources and databases on known organic reactions.

DEVELOPMENT PROSPECTS

The above examples demonstrate that ARGENT-1, the first program of the ARGENT series, makes it possible to propose new and interesting reactions. However, there exist some promising directions for its future development.

The most natural way to enhance the efficiency of working with ARGENT-1 would be to introduce new selection criteria. For example, one type of a promising future criterion seems to be associated with 'small components:' the user may specify that a reaction must have no more than one product with a complex organic skeleton, and all other products must actually be 'small' by-products such as H_2O, HCl, NH_3, CO_2, etc.; the prediction of such reactions may be especially useful for finding new synthetic routes to polycyclic or caged compounds. However, a final development of an extended set of criteria is possible only in collaboration with chemists who will use ARGENT-1 in their everyday work and inform us which selection criteria are, in their opinion, needed most.

We are also planning to develop several new programs within the ARGENT series. For example, one of them should be similar to ARGENT-1 but based on slightly different hierarchical levels—the levels that include the consideration of *additional rings* [formed not only of bonds whose multiplicity changes in the reaction but also of unchanged bonds; e.g. see Fig. 5(f)] in the initial and final systems. For this purpose, the present-state Formal–Logical Approach envisages the hierarchy levels[33] that are somewhat different from those shown in Fig. 2. The elaboration of algorithms and selection criteria for computer-aided construction of equations corresponding to these levels is expected to provide an even higher degree of adaptation to the solution of actual reaction design problems.

Yet another planned direction in the development of ARGENT is the use of automated generation of starting graphs *G* instead of their manual input. For this purpose, the inclusion of a special graph generator in ARGENT is necessary; it may be especially useful for a systematic listing of topology identifiers in the case of complex topologies. Of course, to obtain a reasonable number of results in this case, the appropriate selection criteria should be especially strict.

Finally, a very promising direction in the development of the ARGENT series consists in its linkage to actual *reaction databases*. Indeed, it is not enough just to generate some interesting hypothetical reaction using a program; the next step is to compare it with known

transformations in some reaction database(s) available at present and to estimate its degree of novelty. For such a comparison, the existing databases must be efficiently and unambiguously converted into a specific format that is consistent with the hierarchy levels used in the Formal–Logical Approach;[33] the recoded reactions must then be hierarchically arranged and indexed to facilitate the search for a given predicted reaction or its analogs in the database. Such a recoding necessitates finding a solution to a number of non-trivial problems, including the restoration of the exact stoichiometry for incompletely represented reactions, a reasonable choice of resonance and tautomeric forms for educts and products and the determination of the exact atom-to-atom mapping between the initial and final systems of any known reaction. Although almost all these problems have actually been discussed in the literature,[13,60] they are far from being completely solved. Finally, known transformations forming an existing database should be recoded into a special hierarchically organized format[33] that would allow one to compare the 'significant parts' of reactions and hence to estimate the 'structural similarity' for any pair of transformations, e.g. a known one and a transformation found by ARGENT.

We expect that the future development of ARGENT in all these directions must make it a reaction design software with essentially new possibilities and lead to the discovery of many new and interesting organic reactions.

Acknowledgements

REFERENCES

1. Balaban AT. *Rev. Roum. Chim.* 1967; **12**: 875–898.
2. Hendrickson JB. *Angew. Chem., Int. Ed. Engl.* 1974; **13**: 47–76.
3. Zefirov NS, Tratch SS. *Zh. Org. Khim.* 1976; **12**: 7–18.
4. Arens JF. *Recl. Trav. Chim. Pays-Bas* 1979; **98**: 155–161.
5. Fujita S. *J. Chem. Inf. Comput. Sci.* 1986; **26**: 212–223.
6. Brownscombe TF. *Diss. Abstr. Int.* 1973; **B34**: 1035.
7. Bauer J, Herges R, Fontain E, Ugi I. *Chimia* 1985; **39**: 43–53.
8. Zefirov NS, Tratch SS. *Anal. Chim. Acta* 1990; **235**: 115–134.
9. Mathieu J, Valls J. *Bull. Soc. Chim. Fr.* 1957; **11–12**: 1509–1541.
10. Herges R. In *Chemical Structures: the International Language of Chemistry*, Warr WA (ed). Springer: Berlin, 1988; 385–398.
11. Bersohn M, Esack A. *Chem. Rev.* 1976; **76**: 269–282.
12. Zefirov NS, Gordeeva EV. *Russ. Chem. Rev.* 1987; **56**: 1002–1014.
13. Dengler A, Fontain E, Knauer M, Stein N, Ugi I. *Recl. Trav. Chim. Pays-Bas* 1992; **111**: 262–269.
14. Fontain E, Reitsam K. *J. Chem. Inf. Comput. Sci.* 1991; **31**: 96–101.
15. Bauer J, Ugi I. *J. Chem. Res.* 1982; **S11**: 298.
16. Bauer J. *Tetrahedron Comput. Methodol.* 1989; **2**: 269–293.
17. Zefirov NS. *Acc. Chem. Res.* 1987; **20**: 237–243.
18. Zefirov NS, Baskin II, Palyulin VA. *J. Chem. Inf. Comput. Sci.* 1994; **34**: 994–999.
19. Ugi I, Bauer J, Bley K, Dengler A, Dietz A, Fontain E, Gruber B, Herges R, Knauer M, Reitsam K, Stein N. *Angew. Chem., Int. Ed. Engl.* 1993; **32**: 201–227.
20. Herges R. *J. Chem. Inf. Comput. Sci.* 1994; **34**: 91–102.
21. Zefirov NS, Tratch SS. *Chem. Scr.* 1980; **15**: 4–12.
22. Tratch SS, Baskin II, Zefirov NS. *Zh. Org. Khim.* 1988; **24**: 1121–1133.
23. Tratch SS, Baskin II, Zefirov NS. *Zh. Org. Khim.* 1989; **25**: 1585–1606.
24. Bawden D. *J. Chem. Inf. Comput. Sci.* 1991; **31**: 212–216.
25. Hendrickson JB. *J. Am. Chem. Soc.* 1971; **93**: 6847–6854.
26. Barabás A, Balaban AT. *Rev. Roum. Chim.* 1974; **19**: 1927–1940.
27. Fujita S. *Pure Appl. Chem.* 1989; **61**: 605–608.
28. Roberts DC. *J. Org. Chem.* 1978; **43**: 1473–1480.
29. Brandt J, Bauer J, Frank RM, von Scholley A. *Chem. Scr.* 1981; **18**: 53–60.
30. Kvasnička V. *Collect. Czech. Chem. Commun.* 1984; **49**: 1090–1097.
31. Fujita S. *J. Chem. Inf. Comput. Sci.* 1986; **26**: 205–212.
32. Hendrickson JB. *J. Chem. Inf. Comput. Sci.* 1997; **37**: 852–860.
33. Tratch SS, Zefirov NS. *J. Chem. Inf. Comput. Sci.* 1998; **38**: 349–366.
34. Zefirov NS, Tratch SS. *Zh. Org. Khim.* 1975; **11**: 225–231.
35. Zefirov NS, Tratch SS. *Zh. Org. Khim.* 1975; **11**: 1785–1800.
36. Tratch SS, Zefirov NS. *Zh. Org. Khim.* 1982; **18**: 1561–1583.
37. Zefirov NS, Tratch SS. *Zh. Org. Khim.* 1984; **20**: 1121–1142.
38. Tratch SS, Zefirov NS. *J. Chem. Inf. Comput. Sci.* 1998; **38**: 331–348.
39. Zefirov NS, Tratch SS, Gamziani GA. *Zh. Org. Khim.* 1986; **22**: 1341–1359.
40. Tratch SS, Gamziani GA, Zefirov NS. *Zh. Org. Khim.* 1987; **23**: 2488–2507.
41. Zefirov NS, Chapovskaya NK, Dyachkova LYa, Tratch SS. *Zh. Org. Khim.* 1975; **11**: 1981.
42. Reingold EM, Nievergelt J, Deo N. *Combinatorial Algorithms: Theory and Practice*. Prentice-Hall: Englewood Cliffs, NJ, 1977; chapt. 4.
43. Masinter LM, Sridharan NS, Carhart RE, Smith DH. *J. Am. Chem. Soc.* 1974; **96**: 7714–7723.
44. Masinter LM, Sridharan NS, Lederberg J, Smith DH. *J. Am. Chem. Soc.* 1974; **96**: 7702–7714.
45. Grund R, Kerber A, Laue R. *Discr. Appl. Math.* 1996; **67**: 115–126.
46. Tratch SS, Zhidkov NP, Shtchedrin BM, Popov AI, Zefirov NS. In *Papers Contributed to the IVth International Conference on Computers in Chemistry*, Inst. Org. Chem. Novosibirsk, 1979; Part II, 479–490.
47. Tratch SS, Molchanova MS, Zefirov NS. In *Molecular Modeling (Proceedings of 2nd All-Russian Conference)*, Moscow, 2001; 18.
48. Lederberg J, Sutherland GL, Buchanan BG, Feigenbaum EA, Robertson AV, Duffield AM, Djerassi C. *J. Am. Chem. Soc.* 1969; **91**: 2973–2976.
49. Gilchrist TL, Storr RC. *Organic Reactions and Orbital Symmetry*. Cambridge University Press: Cambridge, 1971.
50. Hoffmann HMR. *Angew. Chem., Int. Ed. Engl.* 1969; **8**: 556–577.
51. Zefirov NS, Tratch SS. *Zh. Org. Khim.* 1976; **12**: 697–718.
52. Kuo Y-N, Chen F, Ainsworth C. *Chem. Commun.* 1971; 137–139.
53. Goldstein MJ, Judson HQ. *J. Am. Chem. Soc.* 1970; **92**: 4119–4120.
54. Paquette LA, Stowell JC. *J. Am. Chem. Soc.* 1971; **93**: 2459–2463.
55. Block E. *J. Am. Chem. Soc.* 1972; **94**: 642–644.
56. Ingold CK. *Structure and Mechanism in Organic Chemistry*. Cornell University Press: Ithaca, NY, 1969.
57. Markovskii LN, Shermolovich YuG, Vasil'ev VV, Boldeskul IE, Tratch SS, Zefirov NS. *Zh. Org. Khim.* 1978; **14**, 1659–1664.
58. Boutagy J, Thomas R. *Chem. Rev.* 1974; **74**: 87–99.
59. Donaruma LG, Heldt WZ. *Org. React.* 1960; **11**: 1–156.
60. Gasteiger J, Hutchings MG, Saller H, Löw P. In *Chemical Structures: the International Language of Chemistry*, Warr WA (ed). Springer: Berlin, 1988; 343–359.