MOLECULAR DESIGN OF HETEROCYCLES.

4.* UTILIZATION OF COMPUTERS IN THE CHEMISTRY OF

HETEROCYCLES (REVIEW)

D. E. Lushnikov and E. V. Babaev

Fundamental principals of the utilization of computers in the design of heterocyclic structures and reactions are examined and systematized in this review. The authors' original GREN program, which lists the recyclizations of heterocycles, and Heterocycland program, which is based on the previously proposed [2] "structure-synthesis" concept and is used for the retrosynthesis of six-membered heterocycles, are examined concisely.

The intent of this review is to achieve several goals: first, to familiarize a wide audience of heterocyclic chemists with the methodology and the most recent advances in the utilization of computers in organic chemistry in general and the chemistry of heterocycles in particular, and, secondly, to place emphasis upon the attention of the specialists themselves in the field of computer chemistry on the specific characteristics of the problems that exist in the chemistry of heterocycles and on the known methods for their solution. Finally, the review can be addressed simply to thinking organic chemists who are not indifferent to research at the boundaries of the sciences. For specialists who have a vague idea regarding the subject of this review we have striven to set forth a maximally approachable exegesis of the idea of computer chemistry.

In examining the utilization of computers in chemistry, one cannot help being amazed at the diversity and wide gamut approaches. Quantum-chemical calculations of polyatomic molecules [3] or optimization of the conformations of polycycles by the methods of molecular mechanics [4] can be realized only by means of computers. Independent reviews [5, 6] have been devoted to the key role of computers in the processing of chemical information, and computer programs for establishing the structures of new compounds on the basis of the spectra and other information available regarding them have been created (and are already being used extensively) [7, 8]. Finally, the entire trend of computer chemistry oriented toward the quest for regularities of the "structure-property" and "structure-activity" type (English abbreviations QSPR and QSAR) has been formulated [9-11].

It goes without saying that in the cited approaches, in which the chief goal is the analysis of the static (physicochemical or structural, for example) properties of substances, heterocyclic substances play a far from minor role (simply by virtue of the significance of heterocycles, let us say, as medicinal preparations, dyes, etc.). Moreover, understanding the chief task and discipline of chemistry as being the study of the utilization of substances in chemical processes, we will strive to concentrate our principal attention on the use of computers in the analysis of precisely chemical transformations, i.e., we will examine the computer aspects of the synthesis and reactions of organic substances, namely heterocyclic compounds. Despite the existence of reviews in this field [12-16], the specific characteristics of the computer chemistry of heterocycles, have evidently not yet been the subject of independent study.

COMPUTER DESIGN OF STRUCTURES AND REACTIONS: CLASSIFICATION
OF GOALS AND CONCEPTS

In discussing the use of computers for the solution of synthetic problems one may speak of three major "components" of synthesis — *the starting reactants, the reactions, and their products*. Usually, one or several of these components are unknown and are the subject of investigation in the use of a computer, while others are known. We will assume that these

---

*For Communication 3, see [1].

components are independent and, consequently, that they can be combined with one another with a "plus" sign (if starting information is available) or with a "minus" sign (if no information is available or it constitutes the task of the investigation). Simple sorting gives us eight possible combinations, which are presented in Table 1 ("Goals and Concepts of Computer Chemistry"), which is, evidently, presented for the first time. The fundamental difference between the information on the substances and the information on the reactions should be emphasized; the former is subject to quite simple processing, while the latter is not.* We therefore arbitrarily divided the table into four possible types of tasks that reflect the existence or absence of *data* regarding the reactants and products and the different nuances of *knowledge* regarding the way in which the reactions occur.

Let us note that the mathematics- and logic-oriented abstract conception of the reactions in the literature is not infrequently typified by the words "formal," "logical," or a combination of these words [12]. The opposite approach is designated as "empirical" or "informational-oriented."

Cases I-III (plus signs in the "Products" column) correspond to retrosynthetic planning of the synthesis, i.e., to finding ways to synthesize the *desired structure* (subsequently abbreviated DS), regardless of which method and regardless of the substances from which the synthesis is effected. In actual practice, it is necessary to introduce restrictions on the methods (strategy) of synthesis. In case I, moreover, information regarding the available reactants is available.

Case IV (the DS and reactants are available) corresponds to a "lack of freedom" in the selection of the reactants for the DS (suitable reactants should be available in the laboratory or in a catalog). In addition, a certain degree of knowledge regarding the reactions is not explicitly implied; this makes cases IV and I similar.

Cases V and VI (a set of reactants is predesignated). The task of the direct (or synthetic) planning of the synthesis consists in prediction of the products of the reaction between the reactants under predesignated conditions. Elementary knowledge (case VI) regarding the rules under which the reactions proceed is possible.

Case VII (knowledge regarding the reactions is available) corresponds to the classification and quest for new reactions and the study of their mechanisms. The selection of the optimum substances for the realization of abstract ideas is relevant here.

There are idiosyncrasies for extreme cases I and VIII. The first case (the goals are set, the reactions are known, and the reactants are given) correlates with experimental practice. The second case (the lack of reactants, knowledge of the reactions, and the goals of the investigation) also implies the lack of computers.

Below we will discuss all of the cases using heterocyclic structures as the reactants and/or products and placing emphasis upon attention to the specific characteristics of knowledge of heterocyclic reactions (for example, recyclizations or "magic" rules for the cyclizations of heterocycles). The examples presented in the schemes will play the chief role: all of them are predicted using computers (in place of the reactants, above the arrows we will place the name of the program in which the reaction is proposed). The experimental correlation will be specified individually for each case.


RETROSYNTHETIC ANALYSIS (APPROACHES I AND II)

In giving a unique "retrospective of the ideas of retrosynthesis" let us note that the idea to use computers to search for pathways for the synthesis of organic compounds was first proposed back in 1963 by Vleduts, a former coworker in the All-Union Institute of Scientific and Technical Information (VINITI, Moscow). In 1967, Corey, whose name is usually associated with the birth of computer retrosynthesis, proposed the concept of *retrosynthetic analysis* of a desired molecule and developed the general principles of synthesis planning [18].

In the Corey concept the result of analysis is a set of synthetic precursors, and the application of the retrosynthesis procedure to it leads to the formation of a *synthesis tree*. Thus two features are fundamental in the development of a plan of synthesis: 1) both to generate synthetic precursors starting from predesignated desired structures (DS) (tactics) and 2) to control the construction of the synthesis tree in order to avoid "combinatorial explosion" and to find the optimum pathways in this tree (strategy).

---

*Information regarding the substances — the reactants and products — is usually *data* (either structures or a list of structures such as a catalog of commercially available reagents): one either has them, or one does not. On the other hand, the information entered into the computer regarding reactions is a certain degree of *knowledge* as to how the transformations of the compounds occur.

Empirical computer synthesis planning, or retrosynthetic analysis, imitates, on the whole, the way in which a chemist thinks. It is based on the singling out in the DS *retrons* — certain structural features that give the chemist a hint regarding a specific reaction (or sequence of reactions) leading to the creation of the DS. Transformations of the structure that are the reverse of the reactions have been called *retroreactions*. When the structural fragment is revealed, the computer evaluates whether the environment of the fragment is favorable for accomplishing the contemplated transformation. If the answer is yes, the computer replaces the fragment with its precursor. The decision as to which retroreactions can be used is made on the basis of empirical rules that constitute the strategy of the synthesis. The classic set of Corey rules (strategies) realized in the LHASA program is widely known [19].

The following example illustrates how the LHASA program uses the strategy of a change in the functional groups for the "disassembling" of a pyrrole obtained by the Knorr method [20] (unfortunately, there is a misprint in the original paper — an ethyl group is presented instead of an ethoxycarbonyl group):
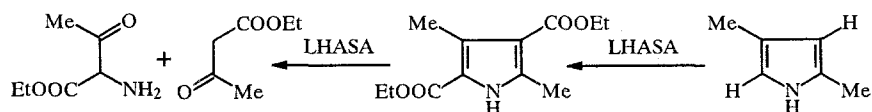

Let us note that the greatest effectiveness of the Corey approach was demonstrated thoroughly in an analysis of complex alicyclic or saturated heteropolycyclic systems [19]. As regards heterocycles with conjugated bonds, they have been used as protective [21] or auxiliary [22] groups:
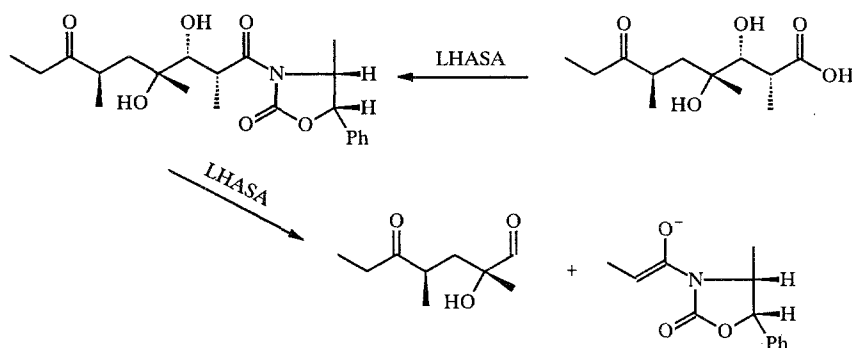

A pyridine ring with suitable functionality was considered as a precursor of saturated bicyclic systems in only one of the studies:
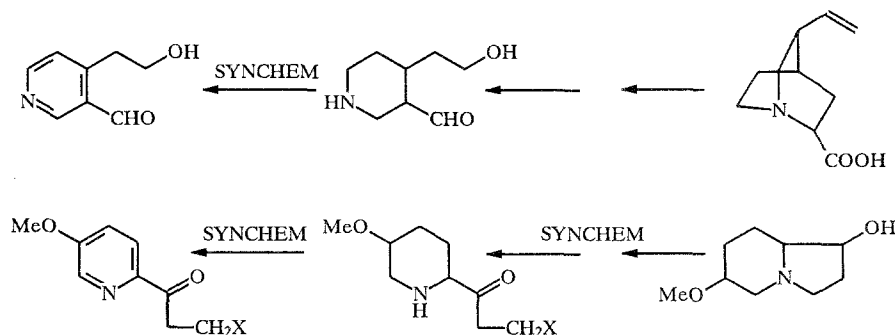
while the furan ring was considered as a precursor of a chain with the required allocation of functions [23]:
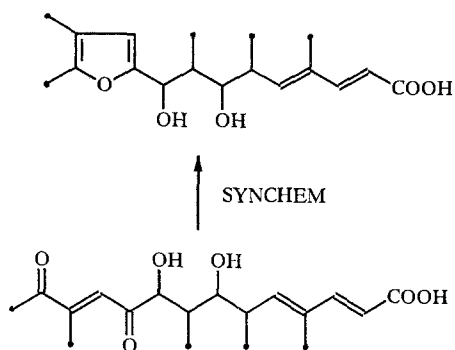
In studies and programs devoted to the further development of the retrosynthetic approach heterocyclic structures have also begun to emerge as desired structures (DS), whereas retroreactions that have been proposed for the assembly of heterocycles have not gone beyond the bounds of the simplest two-component syntheses, chiefly of the 5+1-cyclization type as, for example, in the retrosynthesis of chlorpromazine [24]:
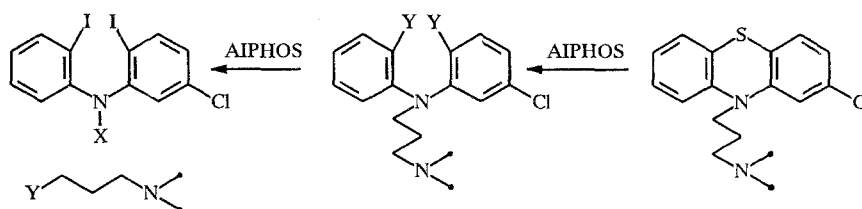

A similar 5+1 cyclization (with replacement of a 1,1-binucleophile by a 1,1-bielectrophile) was used in the retrosynthesis of the alkaloid yohimbine [25]:
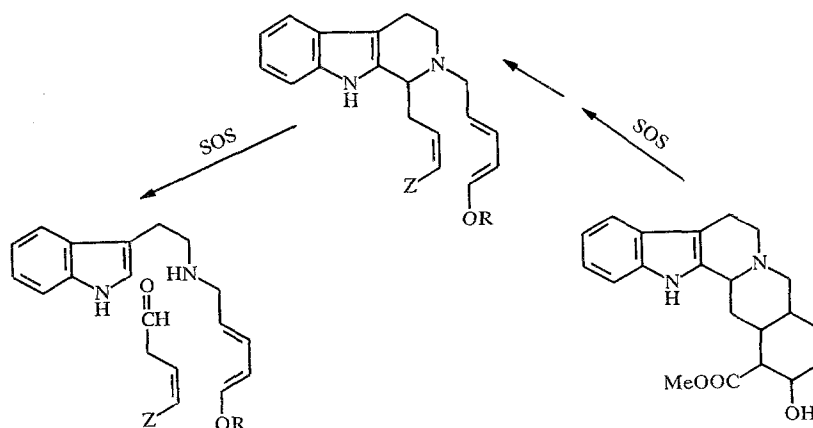

It should be emphasized that the methodology of programs of the LHASA type cannot go beyond the scope of the library of retroreactions. As a result, in the retrosynthesis of, for example, a heterocyclic ring, the method for the production of which has not been introduced into the library of reactions, the program cannot suggest an adequate synthesis pathway.

The greatest number of heterocyclizations (> 300) are evidently contained in the library of transformations of the RDSS program [26]. A part of the tree for the synthesis of 4-hydroxyindole is presented as an example of its operation. Of the 11 best precursors, only two pathways lead to the desired indole:
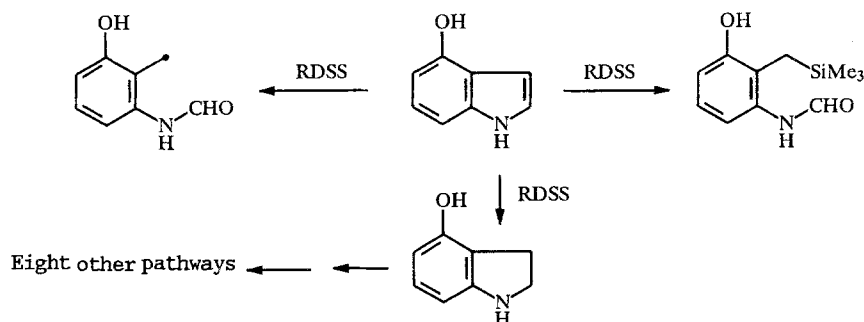

The remaining pathways are based on the development of the hydrogenated 4-hydroxy-2,3-dihydroindole structure as a precursor and were constructed in the usual transformations of aliphatic fragments and aromatic substitution reactions that lead to it.

In a number of approaches one can construct a model for the evaluation of the selectivity of the introduction of functions into aromatic and heteroaromatic systems. For example, in the SECS program [27] the preferred reaction pathway of this type is determined on the basis of a comparison of the electron energies of localization for various positions calculated by the Hückel method. The example presented illustrates the idea of this sort of planning of the synthesis of a disubstituted pyrrole:
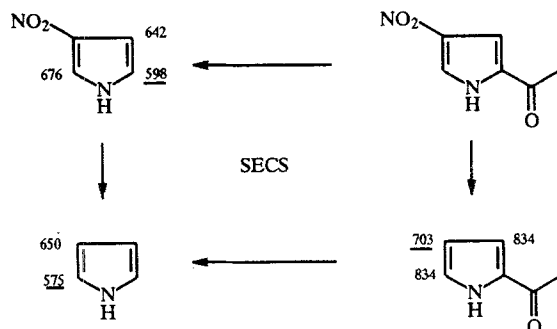
The development of programs for searching for synthesis pathways that operate noninteractively (i.e., without the participation of the user) impose quite a number of additional restrictions on the retroreactions used and the synthesis tree (for example, the number of steps, the possible yield, the cost of the reactants, the possibility of complication of the precursors during the synthesis, etc.). The complexity of the problem can be demonstrated in the case of the operation of the Sumimoto Chemical Company program [28], which is oriented toward the maximally lowest prices of the reactants used. In the case of the maximally allowed number of steps (seven) the program did not find even one "cheap" pathway for the "thiophene-saccharine" synthesis, while a method described in the literature was reproduced for an eight-step synthesis:
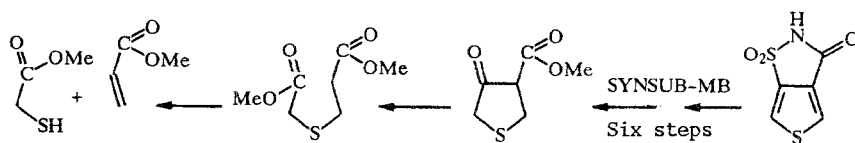
The examples presented above show that, although heteroaromatic systems have been encountered in a number of subjects used for retrosynthetic analysis, no decisive correlations whatsoever regarding methods for the retrosynthesis of heterocycles have been made.

## SEARCH FOR STARTING REACTANTS FOR THE SYNTHESIS OF DESIRED STRUCTURES (APPROACH IV)

One of the principles used by chemists in practical work is the interchangeability or synthetic equivalence of the reactants, whereas retrosynthesis leads to certain specific structures, and thus if one instructs the computer to recognize chemically equivalent or similar structures, each precursor generated by the computer can be compared with its analogs such as those contained in commercial catalogs.

A key feature is the definition of what precisely one should consider to be a measure of similarity. Evidence for the complexity of this problem is provided by the fact that in one of the recent programs [29] 48 (!) definitions of the structural similarity of molecules were used to search for starting reactants. One such criterion introduced into the computer made the following aromatic and heteroaromatic systems from the Jannsen catalog identical:
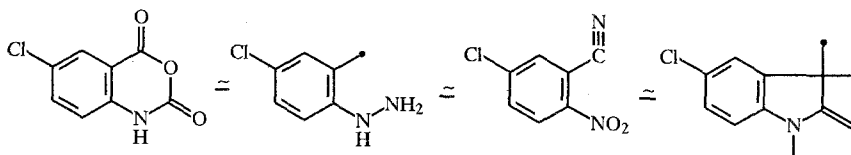
(The structures presented above were declared to be similar, since they all have a benzene fragment and atoms of identical nature adjacent the ring.)

The second variation of the optimum selection of reactants for the synthesis of desired structures (DS) can be compared with an "intuitive leap" [30], which makes it possible for the chemist to immediately see in the starting reactants the prototype of the final structure without planning the steps of the synthesis. In the SST program [30] Wipke and Rogers attempted to imitate this approach by means of a computer. In particular, the realization of the idea of structural similarity in the program made it possible to find in the catalog heterocyclic structures that are suitable as precursors for the synthesis of the carbocyclic compound agarospirol:
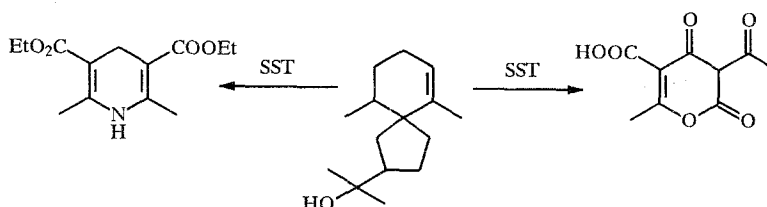
Nevertheless, we do not know of other examples of the use of heterocycles as specific precursors of open forms in computer synthesis.

The CHIRON program was developed specially for the search for chiral synthetic precursors. It contains a library of more than 1000 available optically pure compounds and ≈ 300 racemic compounds and makes it possible to find chiral tem-

plates, including latent types, for the synthesis of optically pure natural compounds [31]. For example, the following template was found for the synthesis of the chiral heterocycle thienamycin by means of a computer:
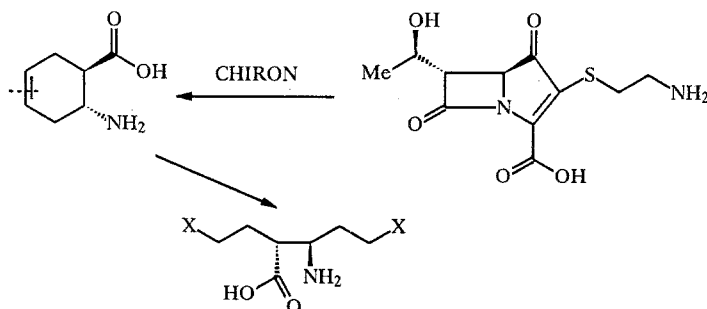

ANALYSIS OF REACTION MECHANISMS (APPROACH V)

The computer analysis of complex mechanisms requires the introduction of "elementary data on elementary steps." For example, the AHMOS program examines eight types of elementary reactions (addition, substitution, dissociation, protonation, sextet rearrangement, polarization, electrophilic substitution, and elimination) and six types of reaction centers (hard and soft electrophiles and nucleophiles, nucleofuges, and electrofuges) [32]. A small data base contains reactivity indexes that quantitatively reflect the nucleophilic and electrophilic properties of various functional groups. On its basis the program determines the highest priority modes of interaction of electrophilic and nucleophilic centers. This relatively simple approach has made it possible to accurately predict the results of the occurrence of uncomplicated ionic reactions [33]:
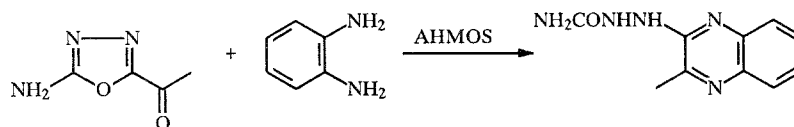

Modeling of reaction mechanisms was realized most consistently in research by the Jorgensen group [34], which attempted to thoroughly analyze fundamental types of reaction mechanisms and the competition among them. The CAMEO program that was created as a result can predict nucleophilic, electrophilic, pericyclic, redox, free-radical, and carbonium reactions.

Thus the CAMEO program makes it possible to solve problems of a rather high level. If one introduces the structures of the starting reactants and predesignates the reaction conditions, the program itself determines which classes of mechanisms can be used. The examples of predicted heterocyclizations that are in agreement with the experimental results are impressive [35]. For example, the following base-catalyzed nucleophilic addition-substitution was predicted:
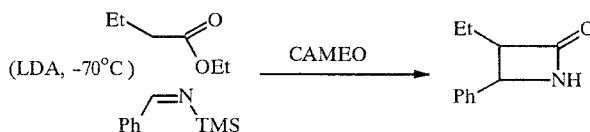
Another example is heterocyclization via intramolecular, electrophilic, aromatic substitution:

(HCl, 180°C)
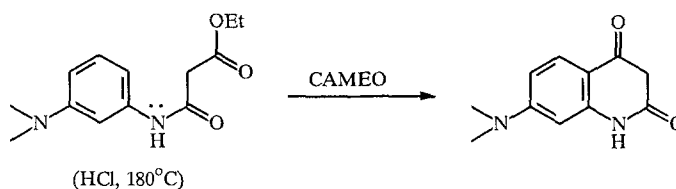
Predictions of yet another type are represented by condensation reactions:

(PPA)
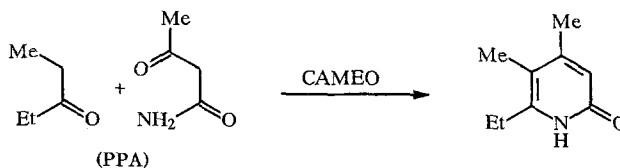
A distinctive feature of the CAMEO program is the possibility of investigating complex syntheses in which several types of mechanisms are realized. Thus the program accurately predicts a sequence of reactions such as the sequence corresponding to the Fischer indole synthesis:
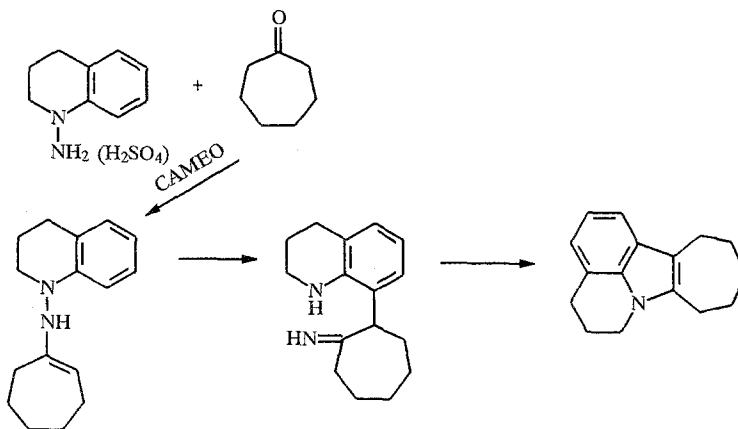
MODELING OF THE REACTIVITIES OF COMPOUNDS (APPROACH VI)

Another approach to the prediction of the results of the occurrence of organic reactions is being developed by the Gasteiger group in the EROS program [36, 37]. Reactions in the EROS program are generated formally via shifting of the electrons and removal and allocation of the bonds. In other words, the subject of the research is not the mechanism but rather the overall result of the reaction. The chief task of the program consists in the automatic singling out of chemically realizable reactions from the multitude of those that are formally possible. With this end in mind, numerical models of the calculation of the enthalpies of reactions and the ability of the bonds to undergo homolytic and heterolytic cleavage were developed. Taking into account both the enthalpy and the reactivity of a bond in evaluating the formal intermediates, the program accurately predicts the results of extremely complex rearrangements [36]:
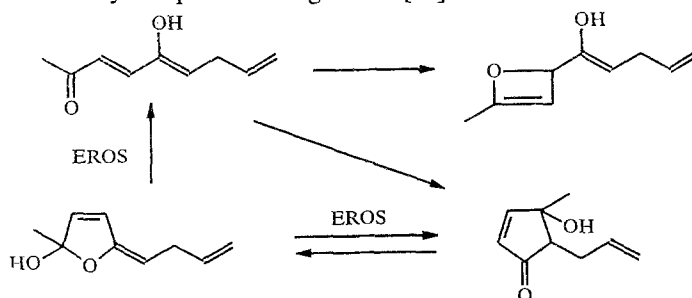
The examined approaches encompass rather broad divisions of chemistry and, on the whole, have extremely high predictive ability. An evaluation of alternative pathways through which the reactions proceed makes it possible to predict the formation of side products.

METHODS OF CLASSIFICATION OF ORGANIC REACTIONS (APPROACH VII)

A universal systematic classification of organic reactions that would be suitable in all computer applications — from the systematic search for reactions data bases to efficient synthesis planning and the search for new reactions — has not yet been developed in organic chemistry [5]. The first mathematical approaches to the description of structural changes curing a chemical reaction were developed in the 70's by Balaban [38, 39], Ugi and Dugundji [40, 41], Hendrickson [42], and Trach and Zefirov [43, 44]. Later publications in this direction are represented by the research of Fujita [45, 46], Roberts [47], Arens [48], Kvasnicka [49, 50], Wilcox [51], and others.

The central idea of virtually all of the approaches consists in the establishment of a correlation between the atoms of the reactants and the products by means of their "superimposition" upon one another. After this confluence or superimposition, one can unambiguously describe the *changes in the bonds* during a chemical reaction. The problem of expressing this redistribution of bonds in any computer-accessible form in the language of mathematics then arises.

Let us note that ways of representing the information on reactions were different for different research groups; however, two sides of the same idea dominated: either a matrix representation of the reactants and reactions or the utilization of graphical diagrams, which is closer to the language of structural chemistry. Without plunging into a discussion of the peculiarities of each approach, we have presented in Scheme 1 a conceivable heterocyclization reaction and have attempted to demonstrate how the diagrams of the same reaction look in the languages of the various approaches. The more cumbersome method of matrix description of the same reaction is presented in the lower part. Various aspects of the representation of the reactions were examined in greater detail in previous reviews [13, 52].
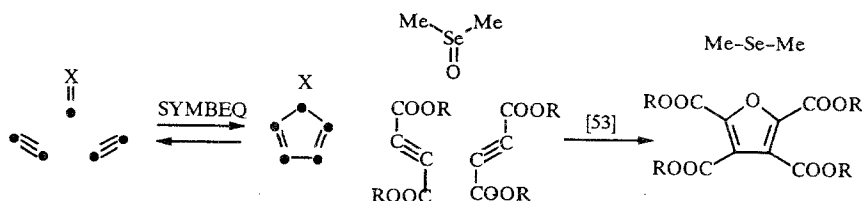
Let us note that all of the approaches presented describe the overall results of the reactions and do not take into account their mechanisms at all.

NONEMPIRICAL DESIGN OF NEW REACTIONS (APPROACH VII)

The various diagrams for the description of the reactions presented in Scheme 1 are similar in that they substantially simplify the usual chemical equations. The graphical aspect ("skeleton") of the reaction, which, as a rule, contains atoms that change the oxidation state (or any environment) and the bonds that change their multiplicity during the reaction, remains in its usual form. Thus chemical reactions can be reduced to a finite number of bond-redistribution diagrams. A comparison of such diagrams for various reactions makes it possible, first and foremost, to reveal the similarity or degree of novelty of the reactions themselves [40, 44].

Another way to use bond-redistribution diagrams is the prediction of new reactions. For this, one is required to build on real substituents, chains, rings, etc. to these diagrams. In a number of cases this sort of prediction was heuristic.

The indicated principle was the cornerstone of the SYMBEQ [53, 54] and IGOR [55] programs. Generation of the following symbolic equation was accomplished by means of the SYMBEQ program, the synthesis of the furan ring corresponding to it was experimentally realized [53]:

1119

Scheme 1

Various Languages of the Formal Description of the Reactions

| Herges–Ugi (1985) [83]	Trach–Zefirov (1975) [43]	Fujita (1986) [45]
Heteroreaction	Reaction equation	
Base reaction	Symbolic reaction	Wilcox (1985) [51]
Reaction category	Topology	Hendrickson (1974) [42]
	$C_{A_4}\ C_{A_2} \longrightarrow R_{A_4}R_{A_2}$ Classification equation	

Ugi-Dugundju (1971) [41]

```
      C C N C N C F O H H H H            C C N C N C F O H H H H            C C N C N C F O H H H H
  C [/ 2 0 0 0 0 0 0 0 1 1 0 0]      [/-1 0 0 0 1|0 0 0 0 0 0 0]      [/ 1 0 0 0 1 0 0 0 1 1 0 0] C
  C |2 / 1 0 0 0 0 0 1 0 0 0 0|      |-1 / 1 0 0 0|0 0 0 0 0 0 0|      |1 / 2 0 0 0 0 0 1 0 0 0 0| C
  N |0 1 / 2 0 0 0 0 0 0 0 0 0|      | 0 1 /-1 0 0|0 0 0 0 0 0 0|      |0 2 / 1 0 0 0 0 0 0 0 0 0| N
  C |0 0 2 / 0 0 0 2 0 0 0 0 0|      | 0 0-1 / 1 0|0 0 0 0 0 0 0|      |0 0 1 / 1 0 0 2 0 0 0 0 0| C
  N |0 0 0 0 / 2 0 0 0 0 0 1 0|      | 0 0 0 1 /-1|0 0 0 0 0 0 0|      |0 0 0 1 / 1 0 0 0 0 0 1 0| N
  C |0 0 0 0 2 / 1 0 0 0 0 0 1| +    |1 0 0 0-1 /|0 0 0 0 0 0 0| =    |1 0 0 0 1 / 1 0 0 0 0 0 1| C
  F |0 0 0 0 0 1 / 0 0 0 0 0 0|      |0 0 0 0 0 0 / 0 0 0 0 0 0|      |0 0 0 0 0 1 / 0 0 0 0 0 0| F
  O |0 0 0 2 0 0 0 / 0 0 0 0 0|      |0 0 0 0 0 0 0 / 0 0 0 0 0|      |0 0 0 2 0 0 0 / 0 0 0 0 0| O
  H |0 1 0 0 0 0 0 0 / 0 0 0 0|      |0 0 0 0 0 0 0 0 / 0 0 0 0|      |0 1 0 0 0 0 0 0 / 0 0 0 0| H
  H |1 0 0 0 0 0 0 0 0 / 0 0 0|      |0 0 0 0 0 0 0 0 0 / 0 0 0|      |1 0 0 0 0 0 0 0 0 / 0 0 0| H
  H |1 0 0 0 0 0 0 0 0 0 / 0 0|      |0 0 0 0 0 0 0 0 0 0 / 0 0|      |1 0 0 0 0 0 0 0 0 0 / 0 0| H
  H |0 0 0 0 1 0 0 0 0 0 0 / 0|      |0 0 0 0 0 0 0 0 0 0 0 / 0|      |0 0 0 0 1 0 0 0 0 0 0 / 0| H
  H [0 0 0 0 0 1 0 0 0 0 0 0 /]      [0 0 0 0 0 0 0 0 0 0 0 0 /]      [0 0 0 0 0 1 0 0 0 0 0 0 /] H

        B matrix of the                  R matrix of the                  E matrix of the
        starting reactants               reaction                         products
```

Note: a description of a hypothetical Diels—Alder reaction (top) by means of various formal approaches of mathematical chemistry (see the text) is presented just as the authors themselves of the approaches would describe it.

For example, an experimentally new method for the generation of 1,4-dienes based on a heterocyclic system was found and realized by means of the IGOR program [56]:

Y = NN⁻-Tos, N₂, S
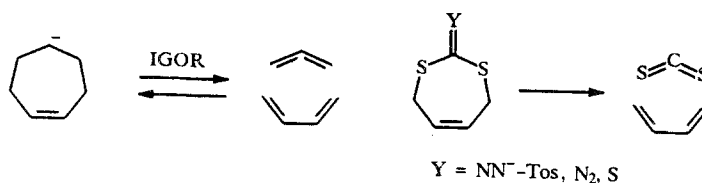
It should be noted that both the SYMBEQ and IGOR programs only generate comprehensively examples of reactions, whereas the evaluation of their reliability depends entirely on the user and his chemical experience and intuition.

## INVESTIGATION OF THE SCOPE OF APPLICABILITY OF REACTIONS (APPROACH VII)

The SCORE program [57] attempts to answer the question: what forms of cyclic structures can be obtained by means of a designated reaction? For this, the program adds chains consisting of from two to four carbon atoms successively to all pairs of atoms in the designated abstract reaction diagram. One of the solutions suggested by this program for the De Mayo reaction is presented below:

General scheme of the De Mayo reaction

One of the schemes after the addition of new chains
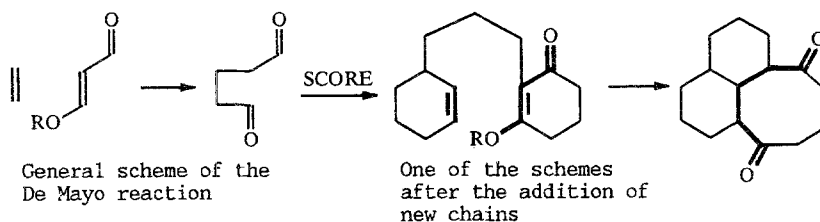
In this scheme, to the left we show the scheme of the reaction designated by the user; the rest of the bonds were added successively by the program. Replacement of the carbon atoms in the resulting skeletons by heteroatoms and allocation of the multiple bonds are carried out by the chemist independently. Thus the program can be used for the design of new heterocyclic systems. (Let us note that the SCORE and IGOR programs were written for an IBM-compatible computer and were submitted to the journals on diskette [55, 57].)


CLASSIFICATION AND DESIGN OF THE RECYCLIZATIONS OF HETEROCYCLES (APPROACH VII)

Recyclization reactions are known for many heterocycles with rings of any size and almost any type, number, and distribution of heteroatoms [58]. In communication 1 of this series [59] we proposed a new methodology for the classification of heterocyclic rearrangements with subsequent development in [60]. Its principal idea consists in the above-discussed superimposition of the starting and final heterocyclic structures. The specific characteristics of the approach (and its chief difference from the abstract diagrams in Scheme 1) consist in taking into account only those bonds that enter into both the starting ring and into the ring that is being formed. Special types of markings show precisely which ring bonds are cleaved, formed, or retained and those that belong to both rings.

M-graphs RBR-graphs

G_1 G_0
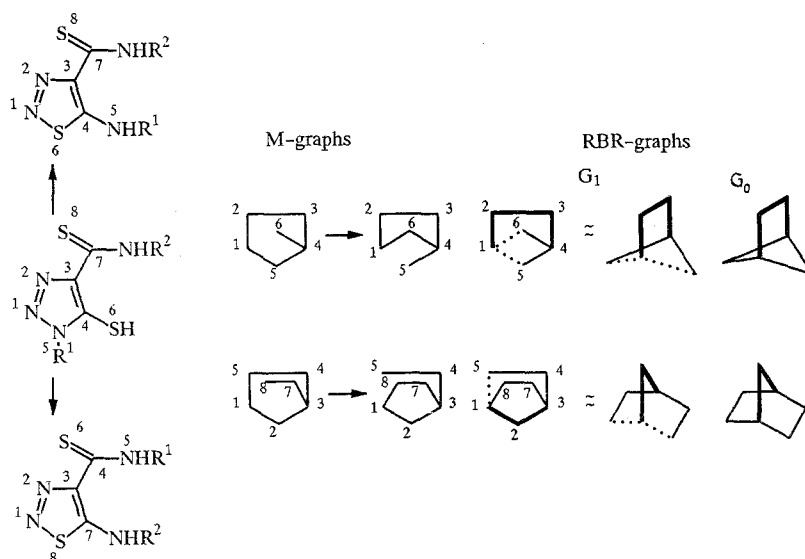
The reaction diagram obtained has a natural chemical significance: in a single scheme the chemist sees simultaneously the former and future heterocycles, as well as the locations of the cleaved and newly formed bonds.

The resulting reaction diagrams always have a simple two-ring structure; this substantially simplifies the analysis of complex rearrangements. A successful hierarchical classification of the transformations of heterocycles that makes it possible to also take into account a purely chemical factor — the distribution of the electrophilic and nucleophilic centers in the bonds that are being cleaved and the bonds that are being formed — was developed on the basis of these diagrams. Algorithms of exhaustive enumeration of all of the rearrangements at each level of the hierarchy were realized in the GREN computer program [60]. The simple possibility of introducing into the program of schemes of reactions that are already known made it possible to use this approach to predict unknown structural types of rearrangements. Several new types of recyclizations were proposed by means of this program:

Z: NR, O, S; X, Y: CH, N; A≡B: C≡CH, N^+≡CR, N^+≡N ; C≡W: COR, CN
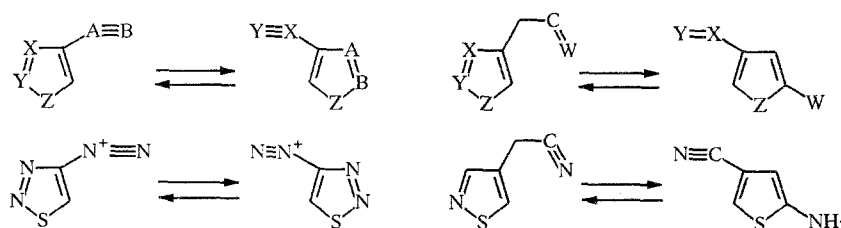
LOGICAL METHODS FOR THE SEARCH FOR SYNTHESIS PATHWAYS (APPROACH III)

The simplest approach to retrosynthesis that does not require any chemical knowledge from the program at all is to cleave one several bonds in the desired structure (DS) and to hand over the structures to the chemist for further analysis. The first version of the SOS microcomputer program worked in precisely this way [61]. (This approach has been called "the *ab initio* method of computer synthesis," although the authors themselves justifiably place it at the same level with the Hückel method [13].) Approaches of this type that make it possible to rivet the attention of the chemist on less obvious possibilities for the assembly of DS was found to be useful in discussing the pathways for the synthesis of simple heterocyclic molecules. For example, in [62] a computer generated 253 ways to cleave two bonds in the skeleton of ellipticine (A), from which the researchers selected 41 combinations for careful consideration. For azaadamantane (B) Barone and coworkers [61] generated 2510 ways of cleavage, from which (manually) they selected 25 schemes for analysis.
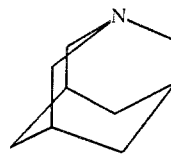
TABLE 1. Goals and Concepts of Computer Chemistry

Type	Reactants	Reactions	Products
I	+	+	+
II	-	+	+
III	-	-	+
IV	+	-	+
V	+	+	-
VI	+	-	-
VII	-	+	-
VIII	-	-	-

Notes. +) Indicates that empirical knowledge of the subject is used; −) indicates that empirical knowledge of the subject is not used.
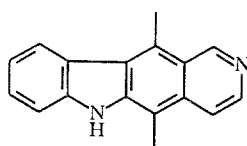

A   B

Despite the fact that this nonempirical approach to computer synthesis is, of course, capable of suggesting new synthetic pathways, the laborious process of evaluating a large number of precursors and the complexity of recognizing known synthesis pathways and discerning new synthesis pathways markedly limit the range of its application.

Nevertheless, the introduction of even minimal information regarding elementary reactions significantly facilitates the understanding and evaluation of the results of the operation of the program with retention of the potential novelty of the solutions. For example, a small set of schemes that describe the principal types of reactions, viz., nucleophilic substitution and addition, elimination, aromatic electrophilic and nucleophilic substitution, tautomeric shifts, and oxidation and reduction, was used in the same SOS program [63]. The pathways of the synthesis of indazoles were analyzed by means of this program. The predicted [64] new pathway for the synthesis of the indazole ring was realized experimentally by the same research group:


R = H, Ph
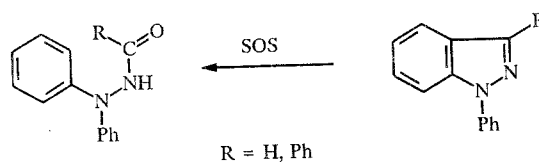
In the development of new pathways for the synthesis of thiazole [65] the computer suggested 800 intermediates, of which 140 were selected for analysis. The four reactions presented below were judged to be the most promising:
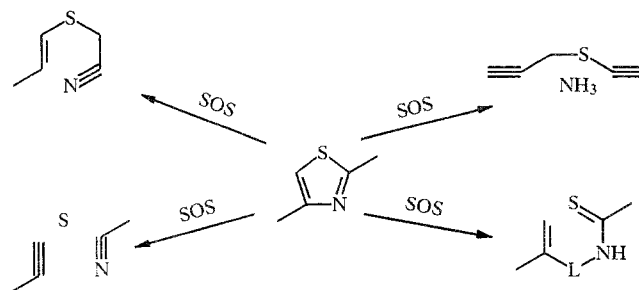
One of the methods suggested in [66] was later confirmed experimentally [67].

In a publication by the same authors the computer suggested 250 intermediates for the synthesis of 6-azauracil, of which they present and carefully consider only 12 synthesis pathways [68]. Let us note that recyclizations were also among the synthesis pathways suggested by the computer:
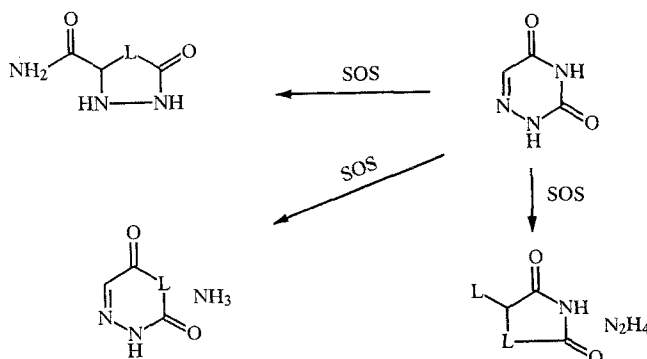
The EROS and TOSCA programs use sets of "reaction generators" that describe in the most general form the redistribution of the bonds during chemical reactions [36, 69, 70]. The synthetic precursors in the EROS program are selected primarily on the basis of evaluation of the reaction enthalpies (see the "Evaluation of the Reactivities" section). This approach may be useful, in principle, to establish pathways for the biosynthesis of simple natural compounds. For example, ammonia, glyceraldehyde, and pentose were found among the possible precursors for pyridoxal [69]:
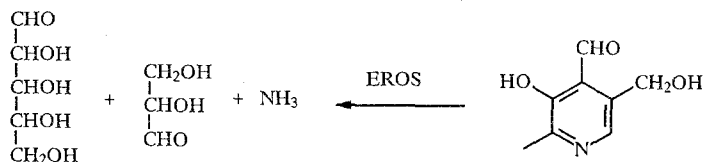

The TOSCA program (Hoechst) uses the idea of the so-called "consonant" and "dissonant" structures and reactions for planning the synthesis of desired structures (DS) [70]. The donor or acceptor character of the atoms entering into the structure is determined, on the basis of which the molecule is classified as dissonant (adjacency of two donor or two acceptor centers is possible) or consonant if the opposite is true. The use of "consonant syntheses for consonant structures" is postulated a priori [70]. The synthesis pathways presented below, which were found by the TOSCA program, constitute an example of the use of this strategy:
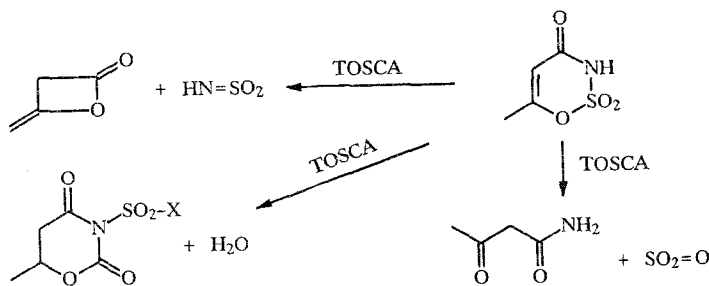
The most serious problem that arises in the practical use of the program consists in the development of a large number of noise structures in the step involving the generation of the reactants for the synthesis of a heterocycle. Thus the structures presented above were obtained by selection from more than 700 suggested structures. Let us note that the idea of consonant

Scheme 2

Example of the Generation of Syntheses by the Heterocycland Program
(Printout from the Computer Screen)
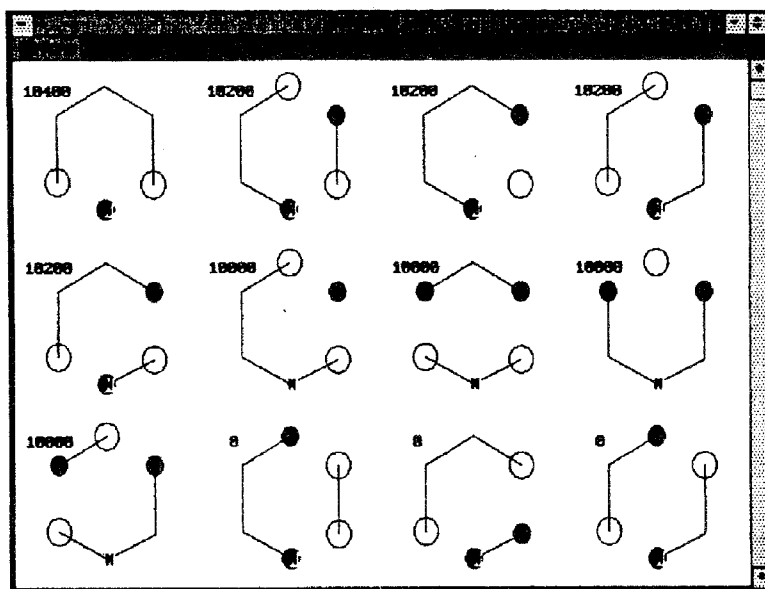
Note: polar types of reactants (the black marks pertain to nucleophilic centers, while the blank marks pertain to electrophilic centers) suggested by the program for "permitted" two-component syntheses of the pyridine ring are seen. The numbers next to the synthesis schemes correspond to the preferableness of one or another method ("zero" corresponds to rare or experimentally unknown synthesis schemes).

molecules and reactions was purely speculative in nature and, nevertheless, is closest to the problematic character of heterocyclic synthesis.

In a previous communication of this series, on the basis of correlation of the experimental data on the synthesis of six-membered heterocycles [2], we proposed a set of magic "structure-synthesis" rules that is analogous in some respects to the idea of consonancy. The central difference consists in the fact that the "structure-synthesis" rules suggest a class of synthones, while the TOSCA program generates specific (including improbable) reactants, leading to "combinatorial explosion."

We have developed the Heterocycland computer program [71], which realizes the previously proposed "structure-synthesis" rules [2]. The program exhaustively generates all possible ways to accomplish the heterolytic formation of bonds in desired structures (DS).

As an example, let us examine the generation and selection of the most promising types of heterolytic reactions that lead to a pyridine ring. The first step in the operation of the program is the exhaustive nonrepetitive generation of heterolytic dissections of the bonds of the DS. Thus 33 such divisions that also include, in particular, improbable combinations of polar centers and reactants correspond to possible two-component syntheses of the pyridine ring.

The user is given two possibilities: 1) to sever the combinations of synthones that are forbidden by the heteroalternation rule; 2) to rank the syntheses on the basis of the preferred formation of some type of bond (for example, a carbon—heteroatom bond). An example of the realization of this sort of selection is presented in Scheme 2. It is apparent that the number of types of syntheses was decreased to nine (high rating), eight of which correspond to the known syntheses of the pyridine ring, while the ninth was observed experimentally in a previous communication [1]. The remaining synthesis schemes with a low rating correspond to unknown or extremely rare syntheses of the pyridine ring [2].

One cannot fail to note the enormous contribution to the development of the idea of synthesis planning introduced by the research of Hendrickson [42, 72-77]. In his approach the purely chemical idea of the heterolytic formation of a C—C bond was successfully linked with the numerical expression of the electrophilicity and nucleophilicity of the carbon atom [42, 74].

The immediate environment of the carbon atom (for example, the number of protons, multiple bonds, and leaving groups) plays a key role in the methodology that he developed for the use of half-reactions [73, 75], which include most of the known reaction schemes [76]. Unfortunately, heterocyclic structures have not yet been the subjects of analysis in this approach.

The application of the Hendrickson idea to the formation of carbon—heteroatom bonds was realized by Moreau in his MASSO program [78]. This promising direction [79], however, considers only one published example:
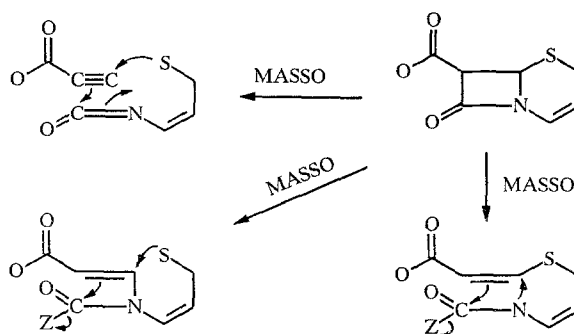
TWO-WAY COMPUTER SYNTHESIS (APPROACH IV)

Ugi's group is developing an original two-way method of synthesis planning [80, 81]. Its idea consists in the simultaneous construction of a synthesis tree from two sides — from the DS side and from the reactant side — which decreases extremely effectively the overall number of intermediates as compared with a unidirectional search. Two-way synthesis was realized in the RAIN program [81], which was used to search for obscure reaction mechanisms and an explanation of the pathways of the formation of side products. For example, for an obscure transformation such as the Straight reaction presented in the scheme, the RAIN program suggested several completely plausible mechanisms [82]:
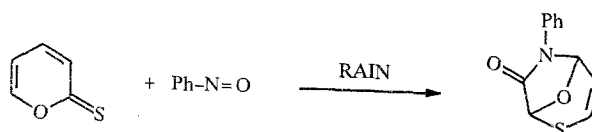
Thus, in principle, most of the examined approaches of computer chemistry as applied to heterocycles do not go beyond the bounds of the usual reactions involving the formation of C—C or C—X bonds, whereas, as we have attempted to demonstrate, individual approaches to the design of heterocyclic structures and reactions that use the specific characteristics of heterocyclic synthesis are only in their formative stage.

REFERENCES

1. E. V. Babaev, Khim. Geterotsikl. Soedin., No. 7, 962 (1993).
2. E. V. Babaev, Khim. Geterotsikl. Soedin., No. 7, 937 (1993).
3. T. Clark, Handbook of Computational Chemistry, Wiley, New York (1985) [Russian translation, Mir, Moscow (1990)].
4. U. Burkert and N. L. Allinger, Molecular Mechanics, Washington, American Chemical Society (1982) [Russian translation, Mir, Moscow (1990)].
5. D. Bawden and E. Mitchell (Eds.), Chemical Information Systems: Beyond the Structure Diagram, E. Horwood, Chichester (1990).
6. G. Vernin and M. Chanon (Eds.), Computer Aids to Chemistry, E. Horwood, Chichester (1986) [Russian translation, Khimiya, Leningrad (1990)].
7. T. H. Pierce and B. A. Hohne (Eds.), Artificial Intelligence Applications in Chemistry, American Chemical Society, Washington (1986) [Russian translation, Mir, Moscow (1988)].

8. R. K. Lindsay, B. G. Buchanan, E. A. Feigenbaum, and J. Lederberg, Applications of the Artificial Intelligence for Organic Chemistry. The DENDRAL Project, McGraw Hill, New York (1980).

9. A. V. Rozenblit and V. E. Glender, Logical-Combinatorial Methods in the Design of Medicinals [in Russian], Zinatne, Riga (1983).

10 A. B. Stankevich, I. V. Stankevich, and N. S. Zefirov, Usp. Khim. **57**, 337 (1988).

11. O. A. Raevskii and A. M. Sapegin, Usp. Khim., **57**, 1565 (1988).

12. N. S. Zefirov and E. V. Gordeeva, Usp. Khim., **56**, 1753 (1987).

13. R. Barone and M. Chanon, Computer Aids to Chemistry, Edited by G. Vernin and M. Chanon, E. Horwood, Chichester (1986) [Russian translation, Khimiya, Leningrad (1990), p. 11].

14. J. H. Winter, Chemische Syntheseplanung, Springer-Verlag, Berlin (1982).

15. I. B. Repinskaya, Retrosynthetic Approach to Planning the Synthesis of Organic Compounds [in Russian], Novosibirsk (1989).

16. F. Serratosa, Organic Chemistry in Action. The Design of Organic Synthesis, Elsevier, Amsterdam (1990).

17. G. Vleduts, Inf. Storage Retr., **1**, 117 (1963).

18. E. J. Corey, Pure Appl. Chem., **14**, Pure Appl. Chem., **14**, 19 (1967).

19. E. J. Corey, Chem. Soc. Rev., **17**, 111 (1988).

20. A. P. Johnson, Chem. Brit., **21**, 59 (1985).

21. E. J. Corey, A. K. Long, T. W. Greene, and J. W. Miller, J. Org. Chem., **50**, 1920 (1985).

22. E. J. Corey, A. K. Long, and S. D. Rubenstein, Science, **228**, 408 (1985).

23. H. L. Gelernter, A. F. Sanders, D. L. Larsen, K. K. Agarwal, R. H. Boivie, G. A. Spritzer, and J. E. Searleman, Science, **197**, 1041 (1977).

24. K. Funatsu and S. I. Sasaki, Tetrahedron Comput. Methodol., **1**, 27 (1988).

25. P. Azario, R. Barone, and M. Chanon, J. Org. Chem., **53**, 720 (1988).

26. F. Haasea and K. Biedka, Tetrahedron Comput. Methodol., **3**, No. 6B, 461 (1990).

27. W. T. Wipke, H. Braun, G. Smith, H. Choplin, and W. Sieber, Computer-Assisted Organic Synthesis. ACS Symposium, Series No. 61 (1977), p. 97.

28. M. Takahashi, I. Dogane, M. Yoshida, H. Yamachika, T. Takabatake, and M. Bersohn, J. Chem. Inf. Comput. Sci., **30**, 436 (1990).

29. J. Gasteiger, W. D. Ihlenfeldt, and P. Rose, Rec. Trav. Chim., **111**, 270 (1992).

30. W. T. Wipke and D. Rogers, J. Chem. Inf. Comput. Sci., **24**, 71 (1984).

31. S. Hanessian, J. Franco, and B. Larouche, Pure Appl. Chem., **62**, 1887 (1990).

32. A. Weise, Z. Chem., **13**, 155 (1973).

33. A. Weise, G. Westphal, and H. Rabe, Z. Chem., **21**, 218 (1981).

34. W. L. Jorgensen, E. R. Laird, A. J. Gushurst, J. M. Fleischer, S. A. Gothe, H. E. Helson, G. D. Paderes, and S. Sinclair, Pure Appl. Chem., **62**, 1921 (1990).

35. M. G. Bures and W. L. Jorgensen, J. Org. Chem., **53**, 2504 (1988).

36. J. Gasteiger, M. G. Hutchings, B. Christoph, L. Gann, C. Hiller, P. Low, M. Marsili, H. Saller, and K. Yuki, Topics Curr. Chem., **137**, 19 (1987).

37. J. Gasteiger, M. Marsili, M. G. Hutchings, H. Saller, P. Low, P. Rose, and K. Rafeiner, J. Chem. Inf. Comput. Sci., **30**, 467 (1990).

38. A. T. Balaban, Rev. Chim., **12**, 875 (1967).

39. A. Barabas and A. T. Balaban, Rev. Chim., **19**, 1927 (1974).

40. J. Dugundji and I. Ugi, Topics Curr. Chem., **39**, 19 (1973).

41. I. Ugi and P. Gillespie, Angew. Chem., Int. Ed. Engl., **10**, 914 (1971).

42. J. B. Hendrickson, Angew. Chem., Int. Ed. Engl., **13**, 47 (1974).

43. N. S. Zefirov and S. S. Trach, Zh. Org. Khim., **11**, 1785 (1975).

44. N. S. Zefirov, Acc. Chem. Res., **20**, 237 (1987).

45. S. Fujita, J. Chem. Comput. Sci., **26**, 205 (1986).

46. S. Fujita, J. Chem. Soc., Perkin Trans. II, 597 (1988).

47. D. C. Roberts, J. Org. Chem., **43**, 1473 (1978).

48. J. F. Arens, Rec. Trav. Chim., **98**, 155 (1979).

49. V. Kvasnicka, Coll. Czech. Commun., **49**, 1090 (1984).

50. V. Kvasnicka, M. Kratochvil, and J. Koca, Coll. Czech. Commun., **48**, 2284 (1983).

51. C. S. Wilcox and R. A. Levinson, Artificial Intelligence Applications in Chemistry, Edited by T. H. Pierce and B. A. Hohne, American Chemical Society, Washington (1986), p. 209 [Russian translation, Mir, Moscow (1988), p. 238].

52. J. B. Hendrickson, Rec. Trav. Chem., **111**, 323 (1992).

53. N. S. Zefirov and S. S. Trach (Tratch), Anal. Chim. Acta, **235**, 115 (1990).

54. I. I. Baskin, V. A. Palyulin, and N. S. Zefirov, J. Chem. Inf. Comput. Sci. (in press).

55. J. Bauer, Tetrahedron Comput. Methodol., **2**, 269 (1989).

56. R. Herges and C. Hoock, Science, **255**, 711 (1992).

57. R. Barone, M. Arbelot, and M. Chanon, Tetrahedron Comput. Methodol., **1**, 3 (1988).

58. A. R. Katritzky and C. W. Rees (Eds.), Vols. 1-8, Pergamon, Oxford (1984).

59. E. V. Babaev and N. S. Zefirov, Khim. Geterotsikl. Soedin., No. 6, 808 (1992).

60. E. V. Babaev, D. E. Lushnikov, and N. S. Zefirov, J. Am. Chem. Soc., **115**, 2416 (1993).

61. R. Barone, A. Boch, M. Chanon, and J. Metzger, Comput. Chem., **3**, 83 (1979).

62. R. Barone and M. Chanon, Heterocycles, **16**, 1357 (1981).

63. R. Barone, M. Chanon, M. Cadiot, and J. M. Cense, Bull. Soc. Chim. Belge, **91**, 333 (1982).

64. R. Barone, P. Camps, and L. Elguero, Ann. Quim., **75**, 736 (1979).

65. R. Barone, M. Chanon, and J. Metzger, Chimia, **32**, No. 6, 216 (1978).

66. R. Barone, M. Chanon, and J. Metzger, Tetrahedron Lett., 2761 (1974).

67. P. Dubs and R. Stuessi, Synthesis, No. 10, 696 (1976).

68. R. Barone and M. Chanon, Nouv. J. Chim., **2**, 659 (1978).

69. J. Gasteiger and C. Jochum, Topics Curr. Chem., **74**, 93 (1978).

70. R. Doenges, B. T. Groebel, H. Nickelsen, and J. Sander, J. Chem. Inf. Comput. Sci., **25**, 425 (1985).

71. D. E. Lushnikov, E. V. Babaev, and S. V. Tsitovskii (unpublished data).

72. J. B. Hendrickson, J. Am. Chem. Soc., **99**, 5439 (1977).

73. J. B. Hendrickson, D. L. Grier, and A. G. Toczko, J. Am. Chem. Soc., **107**, 5228 (1985).

74. J. B. Hendrickson, Acc. Chem. Res., **19**, 274 (1986).

75. J. B. Hendrickson, J. Chem. Inf. Comput. Sci., **29**, 137 (1989).

76. J. B. Hendrickson and T. M. Miller, J. Am. Chem. Soc., **113**, 902 (1991).

77. J. B. Hendrickson and C. A. Parks, J. Chem. Inf. Comput. Sci., **32**, 209 (1992).

78. G. Moreau, Nouv. J. Chim., **2**, 187 (1978).

79. P. Pollet, New J. Chem., **14**, 957 (1990).

80. I. Ugi, J. Bauer, A. Dengler, E. Fontain, M. Knauer, and S. Lohberger, J. Mol. Struct. (THEOCHEM), **230**, 73 (1991).

81. A. Dengler, E. Fontain, M. Knauer, N. Stein, and I. Ugi, Rec. Trav. Chim., **111**, 262 (1992).

82. E. Fontain, J. Bauer, and I. Ugi, Z. Naturforsch, **42b**, 889 (1987).

83. J. Bauer, R. Herges, E. Fontain, and I. Ugi, Chimia, **39**, 43 (1985).