MOLECULAR DESIGN OF HETEROCYCLES.
I. RECYCLIZATION GRAPHS AND STRUCTURAL HIERARCHY OF THE TRANSFORMATIONS OF HETEROCYCLES*

E. V. Babaev and N. S. Zefirov

A new method for modeling the processes involved in the transformation of the ring of heterocyclic systems by means of recyclization graphs, the vertexes and ribs of which correspond only to the atoms and skeletal bonds that enter into the starting or final heteroring, is proposed. An analysis of the structures of such graphs and the use of various types of labels for the ribs and vertexes make it possible for the first time to construct a convenient and flexible hierarchy of recyclizations from the principle of the structural similarity in their recyclization graphs. The effectiveness of the approach is illustrated by examples of the classification of known interconversions of azoles and azines and by the prediction of new recyclizations.

INTRODUCTION

The goal of the current cycle of research was to demonstrate the effectiveness of the use of non-numerical mathematical methods (the theory of graphs, combinatorial analysis, topology) in the design of heterocyclic structures and planning methods for their synthesis and reactions. Interpreting the term "design" in a broad sense, we propose to combine the development of new theoretical constructs and computer programs for the needs of the chemistry of heterocycles with their experimental realization and verification.

This communication is devoted to the problem of the structural classification of an extremely important and extensive family of reactions, viz., processes involving ring transformation (or recyclizations) of heterocyclic systems. Recyclization reactions hardly constitute the most vivid page in the chemistry of heterocycles; these reactions, which frequently are extremely elegant and discovered by chance, often lead to cyclic structures with an unexpected distribution of the heteroatoms and substituents or products, the procurement of which by other methods is difficult or impossible. Name reactions of the Yur'ev, Zincke—König, Hafner type or the Dimroth, Cornforth, Boulton—Katritzky, Kost—Sagitullin, etc. rearrangements may serve as well-known examples of recyclizations. Currently, a tendency for ring transformation has been observed for the overwhelming majority of known heterocycles, particularly for heteroaromatic heterocycles. The mechanisms of transformations of this sort are diverse, and it is often extremely difficult to propose a structure scheme of a transformation by considering only the initial and final heterorings. Despite the enormous amount of factual material in this area and the profusion of reviews that encompass large classes of recyclizations (see, for example, [1-24]), one must acknowledge that there is no general rational classification of the transformations of heterorings in accordance with a structural principle. As a consequence, it often proves to be difficult to establish, for example, the real degree of novelty of a recyclization or rearrangement declared as being "novel." It is evident that the degree and measure of the structural similarity to reactions that are already known must be initially determined in some way.

In most of the existing reviews on recyclizations the question of a unified structural classification has, for all practical purposes, not been broached; particular aspects of this problem have been touched upon only in individual review publications. In the classic monograph by van der Plas, for example, recyclizations are not so much classified as they are put in order with respect to the sizes of the starting and final rings and the number and nature of the heteroatoms [11], whereas in publications by Shvaika [2, 3] on the hydrazinolysis of azoles it was suggested that a structural feature, viz., the size of the fragment of

the starting ring used for the construction of the new heteroring, be used to ascertain the structural similarity of recyclizations. A similar idea, developed by Balaban, has found application for the classification of diverse types of transformations of pyrylium salts [4, 5], as well as 1,3-thiazinium salts [6]. In a recent review [7] L'abbe proposed a different (alternative) way for the convenient classification of types of rearrangements of azoles, viz., according to the size of the side chain in them that is involved in the formation of the new ring. (For example, in rearrangements of the Dimroth, Cornforth, and Boulton—Katritzky type this value is 1, 2, and 3.) In other reviews (see, for example, [8-24]) either particular transformations of certain classes of heterocyclic systems or correlated reaction schemes that link heterorings with different distributions of the heteroatoms or annelated rings are usually discussed.

Let us note that the most developed of the above-mentioned classification schemes use, in the final analysis, the principle of structural similarity of the heterocyclic molecules (or the different types of fragments contained in them) but not the similarity in the reactions, whereas substantial progress has been made in the last 10-15 yr in the development and creation of mathematical models, particularly the use of the language of the theory of graphs, for the description and structural classification of chemical reactions [25-32].

The key idea of the approaches associated with classifications of reactions on the basis of the theory of graphs consists in replacement of the usual chemical equation of the reaction (which contains, obviously, left-hand and right-hand parts) by a certain different object, viz., a unified structural diagram of the distribution of bonds in the course of the reaction; to obtain this diagram one must initially establish a correspondence between the atoms (and bonds) of the left-hand and right-hand parts of the chemical equation and then accomplish "superimposition" of the starting compounds and the reaction products [25, 28, 30]. The resulting diagram — a graph of the redistribution of the bonds — concentrates all of the information regarding the cleaved or newly formed (during the reaction) bonds (the remaining unchanged atoms and bonds are usually disregarded). Such "reaction graphs," which are named differently in different approaches, were subsequently used for different purposes — for example, to establish the degree of structural similarity of reactions, to codify them conveniently, and to solve problems involved in planning the synthesis of compounds or planning new reactions [27, 28, 31, 32].

We previously noted the effectiveness of the use of diagrams of the redistribution of bonds for the classification of recyclizations of heterocycles for the first time in [33]; subsequently, in analyzing the principal structural tendencies of the transformation of the ring of azoles [34], we concluded that it was necessary to create graphs of a special type, viz., recyclization graphs (or graphs of the redistribution of the ring bonds) that were oriented toward the solution of the problems of the chemistry of heterocycles. In the present research we examined a generalization of this approach that leads to a convenient [in a theoretical (and methodical) respect] hierarchical structural classification of recyclizations. This classification can be used effectively for both the systematization of the already available voluminous experimental material on the transformation of rings and for the prediction of fundamentally new — unprecedented — examples of recyclizations.

DEFINITIONS AND TERMINOLOGY

As is usually accepted, we will consider any transformation of a heteroring that includes a step involving opening of the starting ring and the formation of a final ring in any sequence as a recyclization (or ring transformation). We will limit ourselves to an examination of monocyclic heterocycles, adding to them from the number of condensed systems only those in which the annelated rings have no more than one common bond (for example, the usual type of benzannelation as in indole or acridine). We will call the ring transformation (RT) simple (SRT) if the following three conditions are fulfilled: 1) the intermediate formation of any different rings (beyond the formation of the final ring) does not occur during the recyclization; 2) intermediate migration of atoms or groups of atoms does not occur during the recyclization; 3) the starting ring is converted to no more than one ring of the final heterocyclic structure.

The overwhelming majority of known ring transformations are simple, particularly the large family of ANRORC reactions [1], "monocyclic" rearrangements of azoles [7, 14], and most intermolecular cycloaddition—elimination reactions (reactions of the Kondrat'eva type [19]). The recyclizations examined below in Schemes 1-3 give some examples of typical simple ring transformations for which conditions 1-3 are fulfilled. Less widely encountered ring transformations that are simple ring transformations (one of the conditions 1-3 is violated) include, for example, a number of photochemical rearrangements [1, 12, 21], some intramolecular cycloaddition processes (see, for example, [19, 22, 35], individual types of transformations with profound restructuring of the starting ring (of the type similar to the conversion of pyridinium salts to indoles [36]), and some other recyclizations (also see [34]). In the present communication we will restrict ourselves to an examination of only simple ring transformations.

In its most general form the central idea of our approach is based on principles that are extremely simple and clear to chemists: as soon as a recyclization mechanism is known, the skeleton of the starting ring can be readily discerned amidst the atoms and bonds of the final products, and, on the other hand, the skeleton of the final ring corresponds fully to specific atoms and bonds of the starting compounds. Having concentrated our attention only on such skeletal substructural rings (and ignoring all of the remaining details of the structures) one can markedly simplify the structural equation of the reaction. It is evident that the reaction mechanism sets up a strict (mutually single-valued) correspondence between the "ring" atoms (and bonds) of the left-hand and right-hand parts of the equation. As a result, precisely these atoms and bonds can be selected for superimposition of the starting and final structures (just as was done in the approaches examined above) in order to obtain the structural equation (or graph) of the recyclization.

Let us give a more rigorous formulation. Let us define for an arbitrary simple ring transformation (SRT) (with a previously known mechanism) two types or graphs, viz., molecular graphs (for the starting and final compounds that participate precisely in this reaction) and a recyclization graph (for any reaction of the SRT type).

As the *molecular graph* M_s of the starting reactants let us choose from the set of atoms and bonds of the left-hand part of the chemical equation only those V atoms (the vertexes of the M_s graph) and those R_s skeletal bonds (ribs of the M_s graph) that *are either present in the starting ring or are included in the final ring.* Let us similarly define the molecular graph of the final products M_f (with V vertexes and R_f ribs) as the very same V atoms and only those R_f skeletal bonds that *are present in the final ring or were present in the starting ring;* in this case we retain as labels in the M_s and M_f graphs only the symbols of the heteroatoms and discard all of the remaining symbols, viz., the hydrogen atoms, the multiple bonds, and any substituents (including annelated rings). Although the same vertexes are selected for the construction of the M_s and M_f graphs, it is clear that the number and/or distribution of the ribs in these graphs differ. An example of the isolation of the molecular M_s and M_f graphs of Ia, b for a simple ring transformation (1) [37] is presented in Scheme 1; in this example only a part of the annelated pyrrole fragment of the starting indolizine (or the resulting indole) is selected for the construction of the M_s and M_f graphs.

Scheme 1
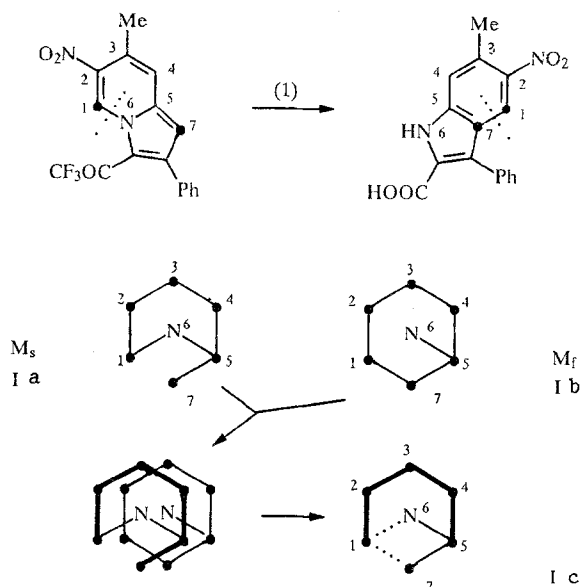
Having numbered the vertexes of the M_s graph in a certain way, let us write the corresponding numbers of the M_f graph in accordance with the reaction mechanism. Let us examine the mental superimposition of the structures of the M_s and M_f graphs in such a way that the pairs of vertexes with the same numbers are identical at the same vertex, while the corresponding pairs of ribs are identical at a certain new rib. Let us define the indicated superimposition of the M graphs as a new *recyclization graph* (RG) containing V vertexes; we will designate the ribs that are present in the M graphs and representable to the new ribs of the recyclization graph (RG) in the following way: *by a dotted line (rib) if the rib is present only in one of the M graphs but absent in another; by a bold line if the rib enters into the composition of both the starting and final rings of the M graphs; by a normal line if the rib enters into the composition of only one of the rings of the M graphs.*

An example of the superimposition of the M graphs Ia, b for SRT (1) and the resulting recyclization graph Ic are presented in Scheme 1.* From a formal point of view the RG is a certain labeled graph that contains labels at the vertexes (symbols of the heteroatoms) and ribs (the three indicated types or the "color" of the ribs, viz., dotted, bold, and normal). The inclusion of all three types of labels for the ribs in the same structure of the RG makes it possible, within the framework of a single reaction diagram, to follow the fate of each of the ring skeletal bonds of the heterocycle during the recyclization.

From the algorithm for the construction of the RG it follows that the structure of the recyclization graph does not depend on precisely which molecular graph (M_s or M_f) is taken for superimposition by the first, i.e., the recyclization graphs of the forward and reverse reactions coincide.

Let us emphasize that the selection of the atoms as the vertexes for the construction of both the RG and the M graphs is determined by their affiliation with both the starting and final rings. For example, in reactions (2a) and (2b) [15] in Scheme 2 the selection of a fragment as the molecular M_s graph is dictated by the reaction pathway, i.e., by the matter as to which of the methyl groups is involved in the construction of the new ring. As a result, the same starting structure 2a — depending on the type of transformation — is represented by two different M_s graphs 2b, c.

Scheme 2
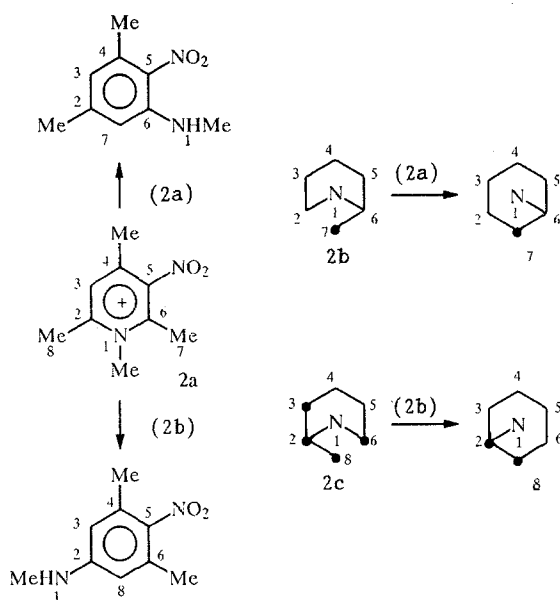
RECYCLIZATION GRAPHS AND THE STRUCTURAL SIMILARITY OF REACTIONS

Following the above-presented algorithm for the selection of M graphs and the construction of recyclization graphs, it is easy to construct RG for any known recyclizations that are simple, as, for example, for the already known to the reader transformations of pyrylium salts (3)-(8) or pyrimidine (9) in Scheme 3; the resulting RG are represented in the scheme by diagrams 3a-9a. These reactions simultaneously illustrate processes involving transformation with expansion (3), contraction (7)-(9) or retention (4)-(6) of the size of the starting ring.

Forms of Recyclizations. Let us note that the M_s and M_f graphs are linked only for rearrangements (for example M graphs 1c, 2b, c), while in all of the remaining cases (i.e., for processes involving the inclusion of an external reagent and/or the elimination of a fragment of the starting ring, which are illustrated by reactions (3)-(9), at least one of the M graphs must consist of more than one component. Let us select the number of components in the M graphs as a natural characteristic for the classification of the SRT with respect to forms. Designating the starting and final heterocycles by the symbols A and B

---

*One can proceed to obtain the RG in an even simpler manner by immediately labeling with bold ribs in the M_s graph the fragment of the ring involved in the formation of the new ring and by labeling with dotted ribs the cleaved bonds of the ring, having added the new dotted ribs that are necessary for the construction of the ring of the M_f graph.

and the external fragments incorporated into the skeleton of the ring (or eliminated from it) by the symbols X, Y..., one can immediately isolate three of the most important forms of recyclizations, which are symbolically represented by the following schemes: A → B [i.e., rearrangements such as (1) or (2a)]; A + X → B or A → B + X [reactions with the incorporation or elimination of ring fragments such as (5) and (8)]; A + X → B + Y [reactions with the incorporation and elimination of ring fragments such as (3), (4), (6), (7), and (9)]. (See Scheme 3.)

It is evident that the SRT of the same form contain an equal number of dotted ribs in the structure of the RG. Let us also note that the affiliation with one form is determined only by the connectivity of the M graphs and has nothing in common with the kinetic order of the reaction (the connected M_s graph, for example, may correspond to any bimolecular reaction such as, let us say, one of the ANRORC type).

G_0, G_1, and G_2 Graphs. A comparison of the RG in Scheme 3 shows that the recyclization graphs of individual SRT may differ from one another either with respect to the structures of the graphs themselves or with respect to the number and distribution in them of the labels of the vertexes and/or dotted labels of the ribs. Precisely this fact can be used as the main criterion for ascertaining the degree of structural similarity of different recyclizations. With this in mind, let us examine three different types of recyclization graphs: *let us call the recyclization graph defined above as the superimposition of M graphs and containing symbols of the heteroatoms and three sorts of ribs a graph of the G_2 type; let us define the G_1 graph as a G_2 graph in which the labels of the vertexes (the symbols of the heteroatoms) are omitted; let us define the G_0 graph as a G_1 graph in which the dotted labels of the ribs are omitted.*

Examples of G_1 and G_0 graphs (diagrams 3b-9b and 3c-9c, respectively) are presented for each of recyclizations (3)-(9) in Scheme 3. The significance of the introduction of additional constructions of RG is evident: by comparing various types of RG with one another one can uncover different aspects of the structural similarity of reactions that have identical G_0, G_1, or G_2 graphs.

Families, Classes, and Types of Recyclizations. Let us examine the closest type of structural similarity of recyclizations — the coincidence of the G_2 graphs corresponding to them. In this case we will state that the recyclizations belong to the same *family*. The type of family is fixed unambiguously by the skeleton and the location of the heteroatoms in the starting and final heterocycles (i.e., by the fixed pair of M graphs), and any variations in the degree of unsaturation, tautomeric structure, and the number and nature of the substituents (and/or condensed rings) therefore do not exceed the bounds of the family. Rearrangements (1), (2a), and (2b) may serve as an example of reactions of the same family.

The next type of structural similarity is identical character (or, more precisely, isomorphism) of the G_1 graphs; we will state that SRT with identical G_1 graphs form the same *class*. Structurally similar SRT of various heterocycles such as, for example, recyclizations of pyrylium salts to isoxazoles and pyrimidines to pyrazoles, viz., reactions (7) and (9) with isomorphic G_1 graphs 7c and 9c, fall into the same class.

Finally, let us assign to the same structural *type* those SRT, the G_0 graphs of which coincide. The recyclization type is the crudest type of similarity of SRT; coincidence of the size of a fragment (labeled in the G_0 graph by bold ribs) that is common to the starting and final rings serves as the determining factor of this similarity. The concept of type is conveniently used by comparing the SRT of a different form, as, for example, by comparing recyclizations with the inclusion and/or elimination of ring fragments with ordinary rearrangements. It is easy to see that rearrangements (1) and (2a, b) belong to the same type as reaction (4); the types of recyclizations (7)-(9) are also identical.
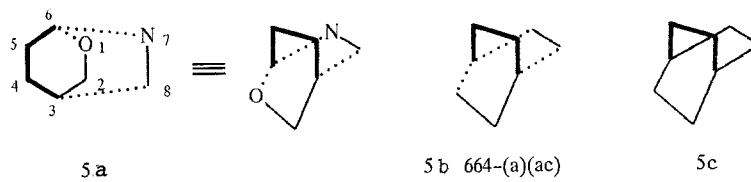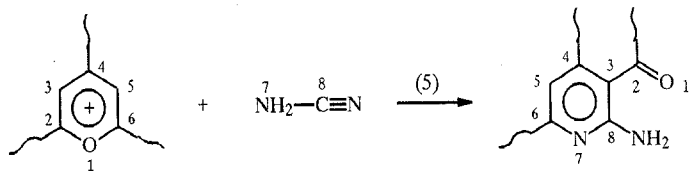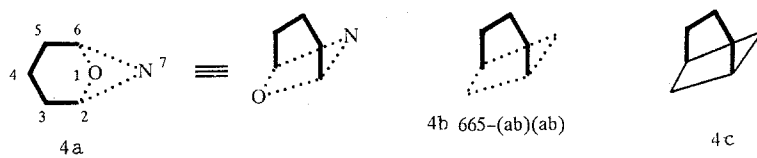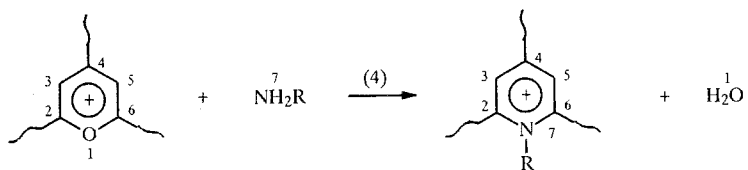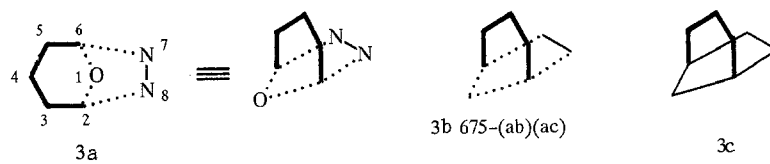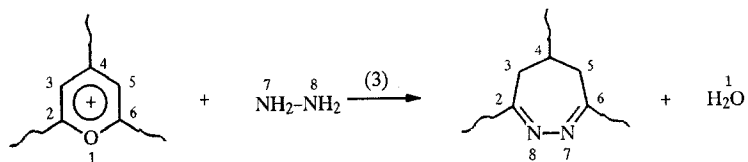
Qualitative arguments regarding the degree of similarity of SRT can be formulated in rigorous form on the basis of an analysis of the structural characteristics of different types of recyclization graphs.
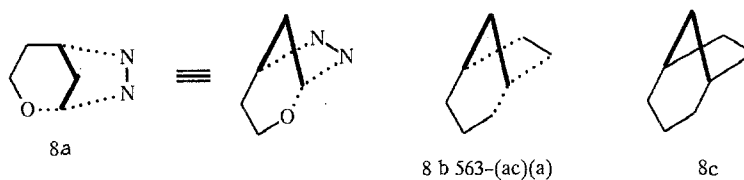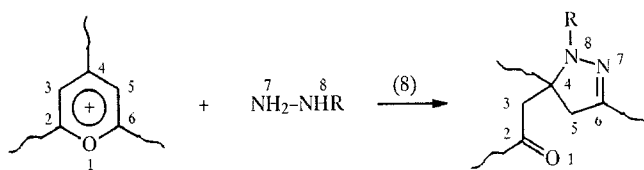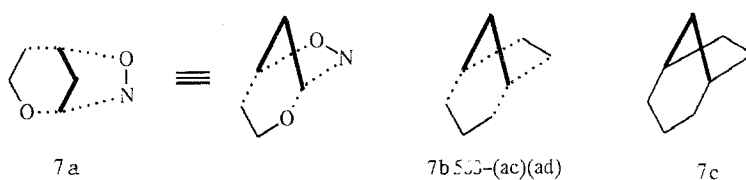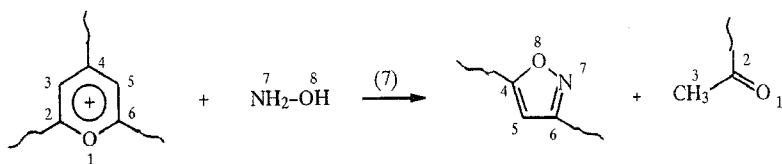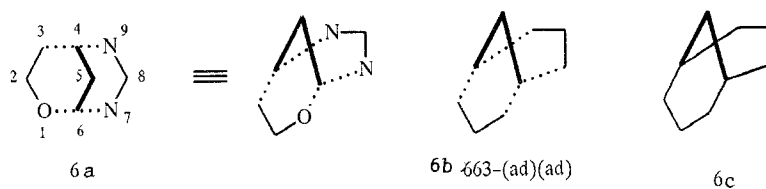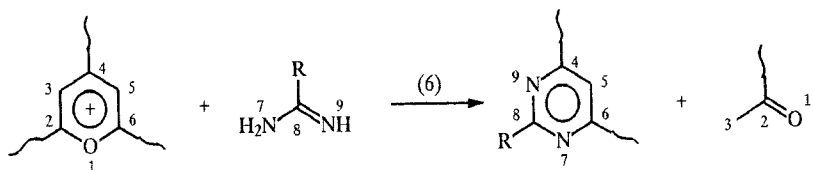
STRUCTURE, SYMMETRY, AND CODES OF RECYCLIZATION GRAPHS

Since RG reflect the different degree of structural kinship of recyclizations, let us pose the following problem: precisely how are the structures of the G_0, G_1, and G_2 graphs, i.e., the types, classes, and families of recyclizations, interrelated? The G_0 and G_1 graphs were constructed above by the removal of certain types of labels (vertex or rib) from the structure of the G_2 graph. The indicated procedure can be inverted by regarding the G_1 graph as a rib-labeled G_0 graph and the G_2 graph as a vertex-labeled G_1 graph.

Thus the relationship between various RG is purely combinatorial in character and is a variation of problems involved in the allocation of labels (rib and then vertex) in G_0 graphs. The number of nonequivalent rib labels in the G_0 graph determines the number of theoretically possible G_1 graphs, while the number of nonequivalent placements of symbols of the heteroatoms (labels of the vertexes) in the G_1 graphs determines the number of G_2 graphs. Problems of this type have been
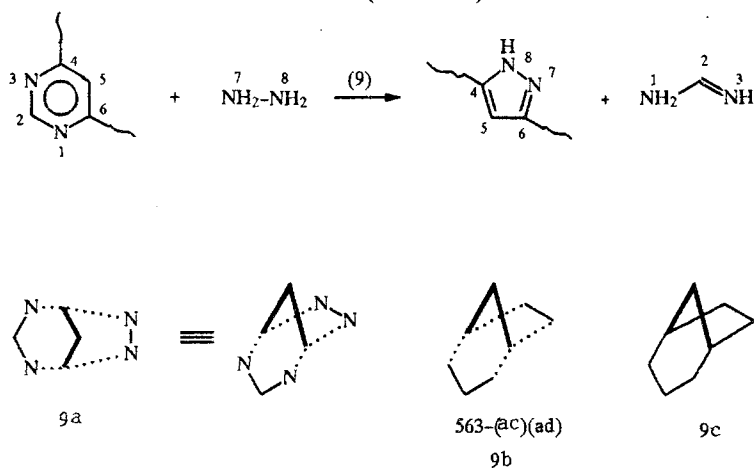
Scheme 3

3a ≡ 3b 675–(ab)(ac) 3c

4a ≡ 4b 665–(ab)(ab) 4c

5a ≡ 5b 664–(a)(ac) 5c

Scheme 3 (continued)

6a ≡ 6b 663-(ad)(ad) 6c

7a ≡ 7b 5.3-(ac)(ad) 7c

8a ≡ 8 b 563-(ac)(a) 8c

Scheme 3 (continued)

$$
\text{pyridine diagram} \quad + \quad \underset{7 \quad 8}{NH_2-NH_2} \quad \xrightarrow{(9)} \quad \text{pyrazole diagram} \quad + \quad \underset{1}{NH_2}\diagup\underset{3}{NH}
$$

9a ≡ diagrams

563-(ac)(ad)
9b

9c

investigated in detail (see the enumeration of substitution isomers in [38]) and are solved, for example, using Burnside's lemma [38-40] for cases in which it is known precisely how a fixed number of labels must be placed and what the type of symmetry of the starting graph is, i.e., a G_0 graph in our case.

Structure and Codes of G_0 Graphs. It is not difficult to demonstrate [34, 41] that G_0 graphs have a completely determined type of structure; namely, they are labeled bicyclic systems or graphs with a cyclomatic number equal to two [42]. In other words, any of them is representable in the form of a certain pair of annelated rings in which a fragment (bridge) that is common to both rings is labeled by bold lines (ribs). Diagrams 4c, 5c, and 6c, which represent three different types of fusion of six-membered rings, may serve as an example of this representation of G_0 graphs. In the general case, graphs with a cyclomatic number of two contain three rings, whereas only two of them are linearly independent [42]. (In this case any of the three rings, such as, for example, the ten-membered perimeter in the "bicyclic" molecular graph of naphthalene, can be regarded as linearly dependent.) Let us agree to regard as independent in G_0 graphs only those rings that contain a "bold" crosspiece (always only two of them) and to call them precisely *rings* of the G_0 graph. The dimensions of such annelated *rings* of the G_0 RG are evidently determined by the dimensions of the rings of the M graphs, i.e., of the starting and final heterocycles.

The relatively simple type of structure of the G_0 graphs permits the introduction of a convenient codification of recyclization types [34]. Let K and L be the sizes of the starting and final rings of the SRT and N be the number of atoms in a fragment common to both rings (i.e., the number of vertexes in the bridge of the G_0 graph labeled by bold lines). The ordered triad of numbers (or vectors) of the form KLN then unambiguously determines the structure of bicyclic system G_0 and the recyclization type corresponding to this graph. (For example, the code 564 denotes mutual transformation of the five- and six-membered heterocycles, such that the fragment made up of four atoms of the starting ring shows up in the new ring.)

Enumeration and Codes of the G_1 Graphs. In comparing the structures of the G_0 and G_1 graphs we note that the G_1 graph is obtained by a certain labeling of part of the ribs of the G_0 graph by means of dotted lines. Above, the dotted rib was identified with a cleaved or formed skeletal bond of the ring. For example, in the case of rearrangements, the G_1 graphs contained a pair of dotted ribs that were clearly located in different *rings*. If, therefore, one sets out to enumerate all of the possible G_1 graphs (classes of SRT) that correspond to rearrangements with a predesignated structure of the G_0 graph (a predesignated type of SRT), one should allocate the pair of dotted labels in the G_0 graph in all possible ways, provided that these labels turn out to be in different *rings*.*

As a result, one can easily deduce (even by hand) the number of theoretical classes of rearrangements with respect to a predesignated type. For example, if a G_0 graph with the code 554 (diagram 10 in Scheme 4) is predesignated, only two nonequivalent G_1 graphs, viz., 10a, b, which contain a dotted rib in each of the rings, i.e., only two classes of rearrangements of this type, can exist. Both of these classes are well known in the literature as rearrangements of the Dimroth type and rearrangements of furoxanes [21] and are illustrated by reactions (10a) and (10b) in Scheme 4.

---

*It is natural that only the "normal" ribs of the G_0 graph can be labeled with a dotted line, since, by definition, the ribs labeled by bold line are unchanged skeletal bonds.

The classes of recyclizations are conveniently characterized by simple *codes* of the G_1 graphs corresponding to them [34, 41]. For this it is sufficient to point out the precise location of the dotted bonds in the previous G_0 graph. In the case of rearrangements there are only two such dotted labels situated in different rings of the G_0 graph. Let us index the sequence of the normal ribs in the smaller ring of the G_0 graph: we will designate the rib closest to the bold bridge by the symbol a, the rib next to it by the symbol b, etc. Similarly, let us index by means of the same symbols a, b, c... the normal ribs in the other (larger) ring, commencing with the rib adjacent to the a rib of the first ring. The resulting code of the form KLN-ij unambiguously determines the class of the rearrangement.* The codes 554-aa and 554-ab correspond to the above-examined examples of reactions (10a) and (10b).

Scheme 4
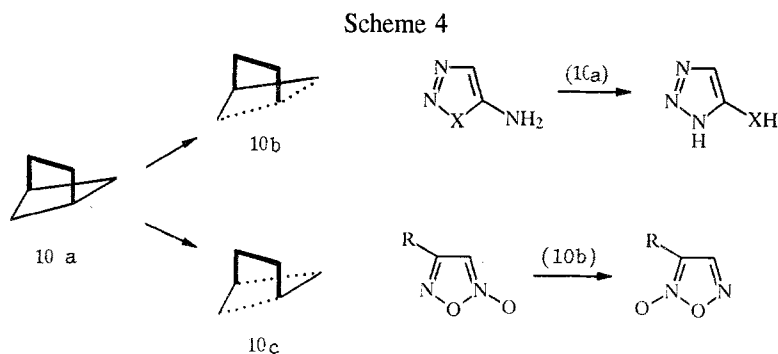
Combinatorial formulas for the number of G_1 graphs in connection with the structure of the G_0 graphs that precede them were previously deduced, and all of the possible recommendable codes of the form KLN-ij for rearrangements were tabulated [34]. Let us attempt to generalize the principle of codification of the classes of SRT for recyclization with the inclusion or elimination of fragments. In the general case the code of the recyclization class will be determined by the distribution of more than two dotted labels between the rings of the starting G_0 graph and can be written in the form KLN-$(i_K j_K k_K....)(i_L j_L k_L...)$, where the symbols in parentheses are the defined, in accordance with the same rules (and lexicographically ordered), sequence of dotted ribs in the K- and L-membered rings of the G_0 graph. In chemical practice the total number of letters in the code does not exceed four (two bonds are cleaved in the starting heterocycle, and two bonds are formed in the final heterocycle), i.e., the maximally long codes for recyclizations have the form KLN-$(i_K j_K)(i_L j_L)$. The codes of the G_1 graph for the examined recyclizations (3)-(9) are indicated in Scheme 3.

G_2 **Graphs.** The following type of RG — G_2 graphs — are vertex-labeled G_1 graphs. To calculate the number of all possible G_2 graphs (the families of SRT) with respect to a predesignated G_1 graph (the class of the SRT) one must know the symmetry of the G_1 graph (determined by its group of automorphisms) and the number of labels (heteroatoms) of each type, including carbon as a label of equal standing. In this case for the computation of the nonequivalent labels, i.e., the G_2 graphs, one may effectively use Burnside's lemma [38-40]. Without dwelling in detail on the calculations, let us note, for example, that the number of theoretically possible families of rearrangements of the Dimroth type in the azole series (if one restricts oneself to only three labels, as, for example, C, N, and S) is 486; reaction (10a) gives an example of only one such family.

Above it was noted that the G_2 graphs (obtained by superimposition of the M graphs) for the forward and reverse reactions coincide. The reverse is also valid: if recyclization graph G_2 is predesignated, a fully defined pair of M graphs, i.e., a pair of reactions, viz., forward and reverse, but not the recyclization pathway, corresponds to it.† To obtain a pair of M graphs from a given G_2 RG one can propose the following simple algorithm. Any of the rings of the G_2 graph is selected, and the dotted ribs of only this ring change to normal ribs; simultaneously with this, the dotted ribs in the second ring are removed, which gives the first of the M graphs. The repeated procedure, commencing with the second ring, gives rise to the second of the M graphs that determine the pair of reactions.

---

*For the case K = L the symbols i, j from the set a, b, c... should be lexicographically ordered.

†In a number of approaches [28, 32] the reaction graph contains special types of labels (for example, arrows or crossed-out lines) that indicate precisely which bonds are cleaved and which are newly formed; although such labels can naturally be included in the structures of the G_1 and G_2 graphs, this leads to a loss in clarity and will not be used in the present communication.

HIERARCHICAL CLASSIFICATION OF THE TRANSFORMATIONS OF HETEROCYCLES

In its most general features, the above-examined structural interrelationship of various recyclization graphs makes it possible for the first time to propose a classification of the transformations of heterocycles in accordance with a hierarchical principle.

Let us examine successively the levels of this hierarchy.

1. Only a common "archetype" is set at the top of the hierarchical tree: the sizes of the starting and final heterocycles in the form of a pair (K, L).

2. At the next level of the hierarchy one takes into account the fact that the heterocycles may have a different value (N) of a fragment corresponding to the unchanged part of the starting ring involved in the formation of the final ring. Allowing the N value to vary from unity to $K \leq L$, one can obtain all of the possible modes of annelation of the pair of K- and L-membered rings that determine the labeled G_0 graphs with KLN codes or recyclization types.

3. At the new level one takes into account the number of components into which the starting ring decomposes or from which the final ring is formed, i.e., the form of recyclization is determined. It is clear that the number of cleaved and newly formed bonds, i.e., the overall number of dotted labels that can be allocated in the structure of the G_0 RG, is thereby established.

4. The size of the fragment of the starting ring that remains unchanged and the total number of vanishing/developing skeletal bonds of the rings are known for a fixed type and form of the SRT. This makes it possible, on the basis of purely combinatorial considerations, to sort out all of the conceivable ways of cleavage and formation of bonds, i.e., the classes of recyclizations (the G_1 graphs) of a predesignated type (and form), via arrangement of a certain number of dotted labels at the ribs in each of the rings of the G_0 graph.

5. As soon as one knows in precisely which positions of the starting ring the bonds are cleaved and the formation of precisely which skeletal bonds corresponds to the formation of a new ring, it is not difficult, on the basis of symmetry considerations, to sort out the theoretically possible starting and final heterocyclic systems (the families of recyclizations) that correspond to a predesignated class. This is achieved by allocation of the heteroatomic labels at the vertexes of a predesignated G_1 graph leading to a set of nonisomorphic G_2 graphs.

6. The predesignated family of recyclizations (with an accuracy up to the reaction pathway) corresponds to strictly determined pairs of heterocyclic systems that contain only those substituents that are necessary for the construction of the final heterocycle or were previously introduced into the cyclic skeleton of the starting heterocycle (pairs of M graphs). At this level one can select a specific recyclization pathway.

7. At the lowest level the predesignated (by the pair of M graphs) symbolic writing down of the transformation of the heterocycle corresponds to a set of specific chemical transformations, the components of which differ with respect to the degree of unsaturation, tautomerism, the presence of substituents, annelated rings, etc.

Thus the hierarchical character of the proposed classification consists in the fact that the number and structure of the recyclization graphs of the lowest levels is strictly determined by the structure and symmetry of the RG of the highest levels. The indistinguishability of recyclizations at any level is eliminated on passing to the lower level.

PATHWAY OF HETEROLYSIS OF THE BONDS

The examined classification of the SRT is structural in character; nevertheless, the levels of its hierarchy can be effectively supplemented by empirical data on the details of the mechanisms of SRT, particularly chemically significant information regarding the distribution of donor-acceptor centers in the starting molecule and regarding the pathway of cleavage of the ring bonds.

Let us examine the rearrangement of the pyridine ring determined by the sequence of reactions (11a) and (11b) in Scheme 5 (see [18]). In both cases heterolysis of the C—N bond and the formation of a new six-membered ring occur. It is not difficult to see that these are reactions of the same class (with identical G_1 graph 11a with code 665-aa) that belong to different families, whereas the undoubtedly important and profound chemical difference between them consists in the fact that in the first reaction (11a) the electrophilic ring carbon atom is replaced by the electrophilic exocyclic carbon atom of the cyano group, while in the second reaction (11b) the nucleophilic ring nitrogen atom changes places with the nucleophilic exocyclic nitrogen atom. In other words, the numbers of electrophilic and nucleophilic centers in reactions (11a) and (11b) do not coincide.

Scheme 5
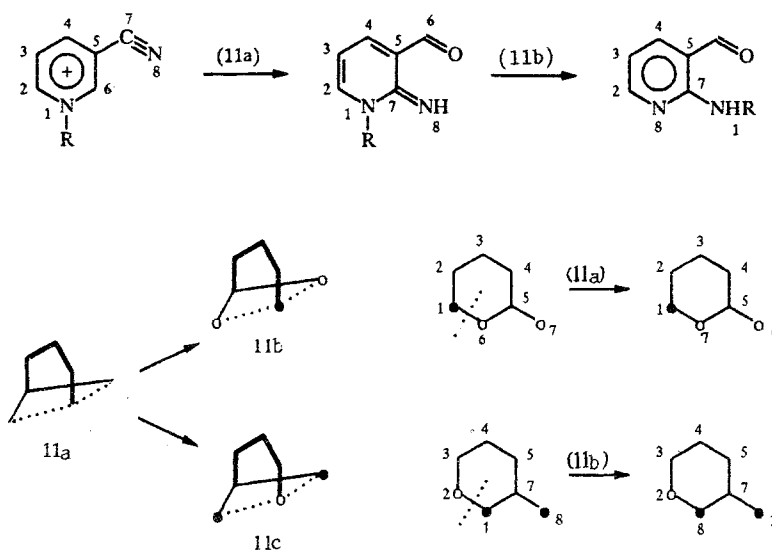
This result is readily reflected in the same structure of recyclization graph G_1 for 11a: let the electrophilic atoms (the E centers) be expressed at the vertexes of this graph by white dots and the nucleophilic atoms (the N centers) be expressed by black dots; reactions (11a) and (11b) are then representable by different (colored) diagrams 11b and 11c (see Scheme 5). We will call the resulting G_1 graphs with colored vertexes $G_1(EN)$ graphs. In such $G_1(EN)$ graphs one takes into account not only the structural aspect of the redistribution of the bonds but also the chemical factor — the donor—acceptor nature of the atoms between which bonds are cleaved or newly formed. We will state that the SRT belong to the same *sort* if the $G_1(EN)$ graphs corresponding to them coincide; thus reactions (11a) and (11b) belong to different sorts with an opposite orientation of the E,N centers.

It is not difficult to see that, from the point of view of combinatorial analysis, the $G_1(EN)$ graphs are obtained from the starting G_1 graph by nonequivalent coloring of the latter, i.e., by allocation of the labels (black and white dots) at the ends of the dotted ribs of the G_1 graph. If the dotted ribs in the G_1 graph are not adjacent, their ends must have the opposite color. We have previously solved the problem of sorting out nonequivalent ways of such coloring in general form [34]; let us note that, since the number of $G_1(EN)$ graphs changes (depending on the symmetry group of the starting G_1 graph) from one to four, it is not difficult to enumerate such families by hand.

Since the $G_1(EN)$ graphs (like the G_2 graphs) are vertex-labeled G_1 graphs, their position in the hierarchy of recyclizations examined above is somewhat special: they can be regarded as an independent sublevel of recyclization classes. In turn, from the level of $G_1(EN)$ graphs one can pass to somewhat different (than were previously determined) M graphs, which we will call M(EN) graphs, which contain only black and white labels of the vertexes; the corresponding pairs of M(EN) graphs of reactions (11a) and (11b) are presented in Scheme 5. Such M(EN) graphs with a previously designated distribution of the E,N centers contain important heuristic information, suggesting precisely which chemical functions should be present in the structures of heterocycles of any families in order that the SRT corresponding to them belong to the same sort. For example, the development of a black label in the $G_1(EN)$ graph requires the presence in the M(EN) graphs (and the corresponding heterocycles, regardless of the distribution of the heteroatoms) of a function with a strictly fixed property of electron-donor character (nucleophilicity) such as, let us say, an amino or CH-acidic group, while a "white" label of the vertex requires, correspondingly, an acceptor (electrophilic) grouping of the carbonyl, cyano, or nitroso type. It is not difficult to convince oneself that the transformations of pyrylium salts under the influence of hydroxylamine (7) and the hydrazinolysis of pyrimidines (9) are reactions of the same sort.

PROBLEM OF THE PREDICTION OF FUNDAMENTALLY NEW EXAMPLES OF RECYCLIZATIONS

The examined classification naturally leads to the problem of "empty cells" — unknown examples of recyclizations in different branches of the hierarchical tree. The prediction of unprecedented examples of rearrangements of various types,

forms, classes, etc. evidently requires reliable information regarding the recyclization graphs of already existing SRT and regarding their distribution over the levels of the hierarchy.

The GREH (Graphs of REcyclizations of Heterocycles) computer program [44], which is, on the one hand, a unique, constantly replenishable data base and educational base of knowledge with respect to recyclizations of heterocycles and, on the other, an instrument for the prediction of fundamentally new heterocyclic transformations, is being developed by us on the basis of the approach set forth above in order to solve problems of this sort. Of course, the utilization of this program for prediction purposes presupposes the use in the last stage of both expert evaluation on the part of chemists and computational methods from the arsenal of modern quantum chemistry. Let us note that the successive classification of known literature examples of SRT and the statistics of encountering various types of RG are also independent and useful criteria of the expert evaluation of predictions.

The methodology of the prediction of unknown recyclizations includes a general review of the literature, the construction for recyclizations of the RG corresponding to them, the search for various types of RG for which there are no prototypes in the form of reactions, the construction of a pair of M graphs, and expert evaluation of the starting compound that could undergo a previously unknown structural transformation.

Our previous review of the diversity of the types of rearrangements of azoles [34] was the first example of predictions of this sort. As it turned out, the most widely spread rearrangements are those, the G_1 graph of which is representable by a code of the 55N-aa type (i.e., the cleaved bond of the starting ring is adjacent to the resulting bond of the ring being formed). In particular, the well-known classes of rearrangements of azoles of the Dimroth, Cornforth, and Boulton—Katritzky type have the corresponding codes 554-aa, 553-aa, and 552-aa (see reactions I, III, V, as well as transformation VII with code 551-aa in Scheme 6). Rearrangements of other classes (with codes that differ from 55N-aa) are also encountered in the literature, although rather seldom, which leaves experimenters a broad field of activity with respect to the detection of genuinely new classes of azole—azole transformations.

Only two recyclization sorts [two G_1(EN) graphs] with opposite distributions of the donor and acceptor centers are theoretically possible for each of the classes of rearrangements with the code 55N-aa of the G_1 graph. Examples of pairs of rearrangements that differ with respect to sort are reaction pairs I and II and V and VI, while the Cornforth reaction III and SRT VII [44] do not have such antipodes. Thus the presented pairs of M graphs of transformations IV and VIII are promising prototypes for the search for azole rearrangements of previously unknown sorts. (See scheme 6).

Let us attempt to carry out at least a qualitative examination of a problem of a more global nature — the distribution along the levels of the hierarchy of mutual *rearrangements* of five- and six-membered heteroaromatic systems, the archetypes of the interconversions of which are fixed by (K,L) pairs of the form (5,6), (6,5), and (6,6). Let us note that, as before, we will restrict ourselves, first, to simple transformations of rings and, second, we will *completely reject* cases of inclusion and/or elimination of any fragments of the starting or final rings.
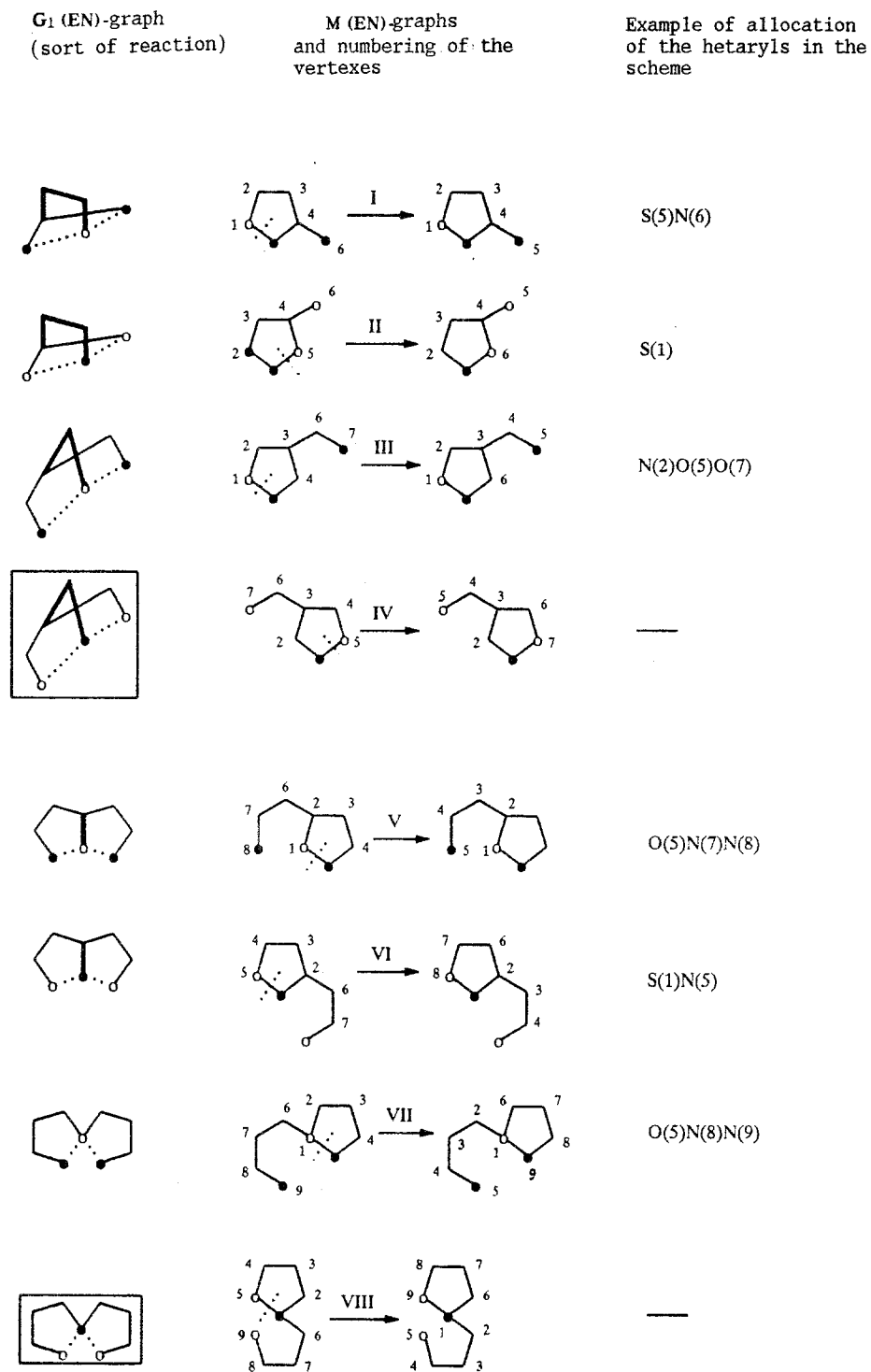
An analysis of the literature data on such rearrangements shows that (as in the case of azole—azole SRT) the most widely spread transformations are those transformations of rings, the codes of the G_1 graphs of which are representable in the KLN-aa form. The structures of all of the theoretically possible G_1 graphs with this code for the mutual SRT of five- and six-membered rings are presented in Scheme 6. Let us note that the code *aa* for the G_1 graphs is unique: the total number of reaction centers (donor and acceptor) between which the bonds are redistributed is equal to three only for this code. (For any other distribution of letters in the code of the G_1 graphs the number of such centers of necessity should increase to four.) Thus the distinctive magic rule of the "minimal number of reaction centers," the violation of which may evidently lead to competitive processes such as, for example, cross side condensations, is operative. Thus a new strategic direction in the area of recyclizations of both azoles and azines could be the purposeful search for classes of SRT with codes of the G_1 graphs that differ from KLN-aa.

Pairs of possible G_1(EN) graphs corresponding to transformations I-XXXVI are presented in Scheme 6 as an example of the prediction of new sorts of mutual transformations of five- and six-membered rings for all G_1 graphs of the *aa* type. If the corresponding reaction was found in the literature, its particular example (which does not reflect the statistics of encountering this given sort) is indicated. However, if the corresponding sort of rearrangement is still unknown to us ("empty cell"), the structural scheme of the reaction is given in the form of a pair of M(EN) graphs, and G_1(EN) is included within the framework.

Scheme 6 is preliminary in character, and in publishing it we will set several goals: 1) to draw the attention of experimental chemists to formal models of mathematical chemistry; 2) to expand the existing point of view with regard to recyclization processes as a whole and to discuss the merits and inadequacies of the idea of a structural hierarchy and the
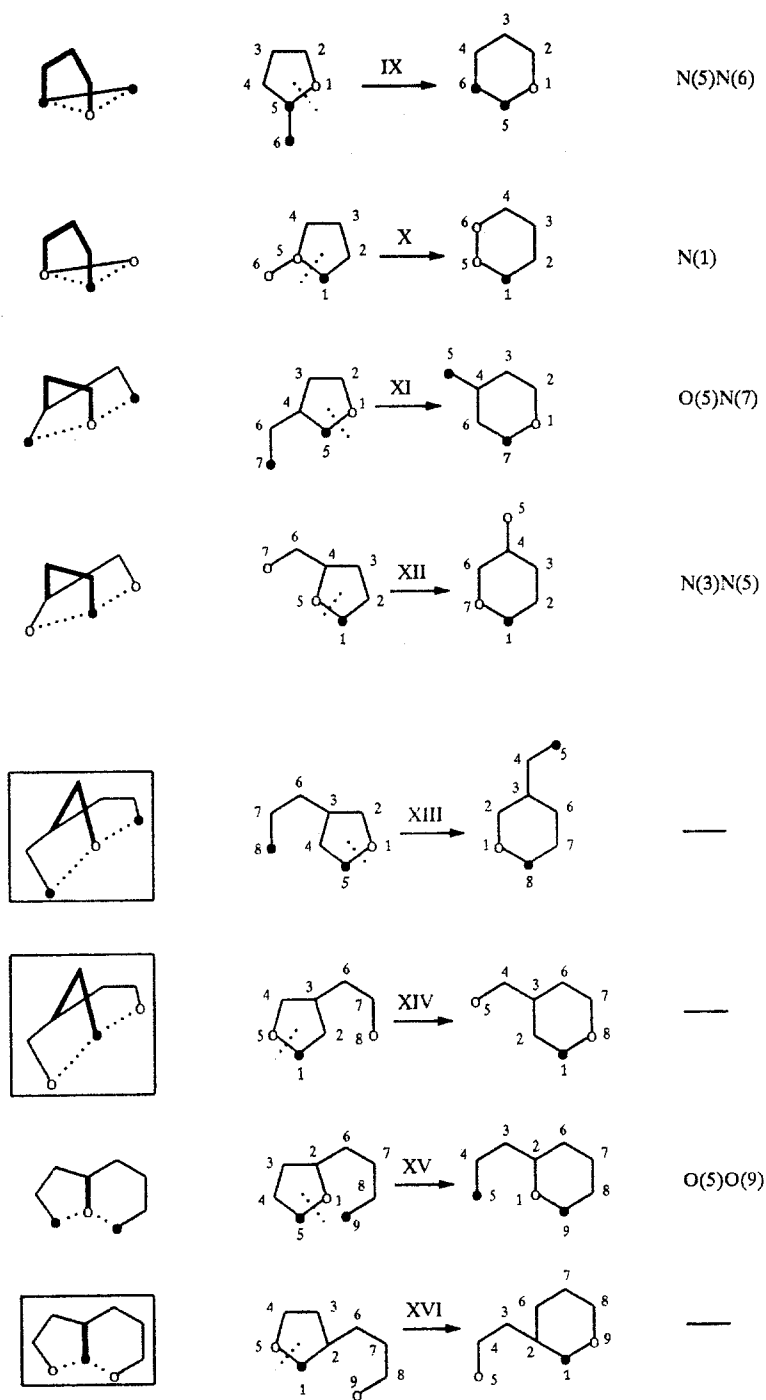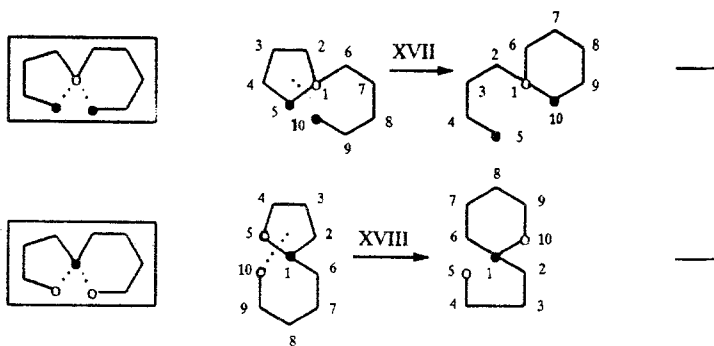
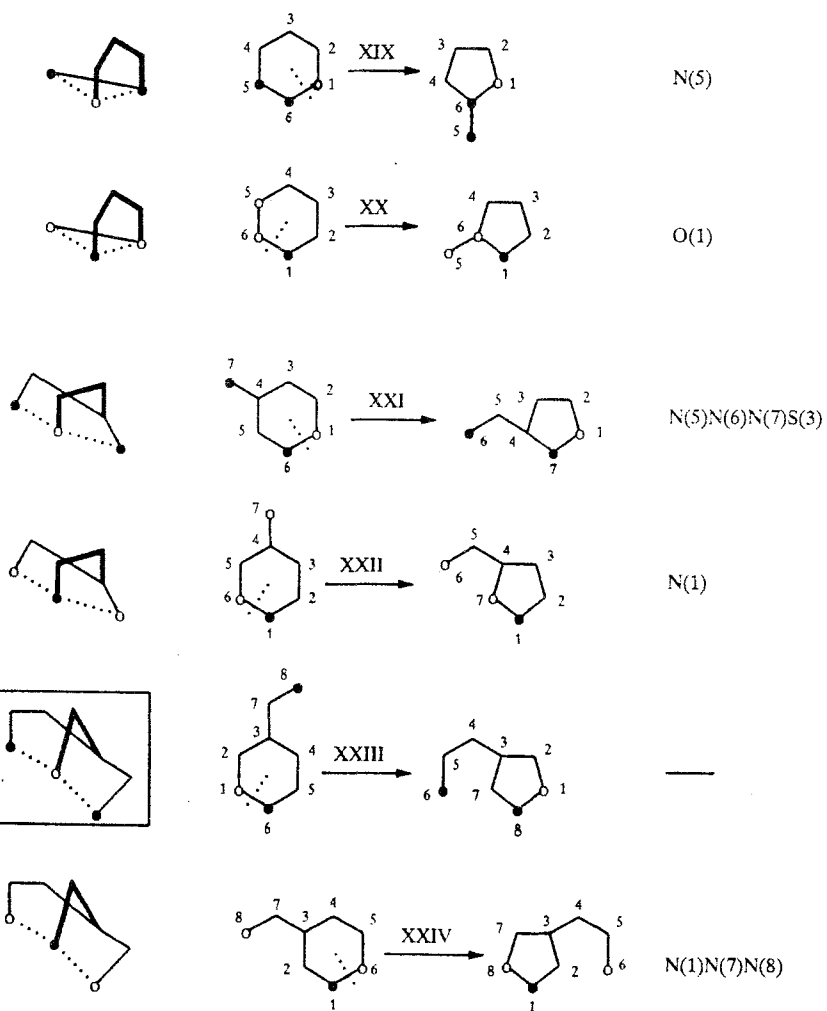# Scheme 6

(5 → 5) — Transformation of hetarenes

G₁ (EN)-graph (sort of reaction)	M (EN)-graphs and numbering of the vertexes	Example of allocation of the hetaryls in the scheme

S(5)N(6)

S(1)

N(2)O(5)O(7)

—

O(5)N(7)N(8)

S(1)N(5)

O(5)N(8)N(9)

—

Scheme 6 (continued)

(5 → 6) — Transformation of hetarenes

IX   N(5)N(6)

X   N(1)

XI   O(5)N(7)

XII   N(3)N(5)

XIII   —

XIV   —

XV   O(5)O(9)

XVI   —

684

Scheme 6 (continued)

$(6 \rightarrow 5)$ — Transformation of hetarenes

N(5)

O(1)

N(5)N(6)N(7)S(3)

N(1)

—

N(1)N(7)N(8)

Scheme 6 (continued)

N(6)N(8)O(9)

N(1)N(5)N(7)

—

—
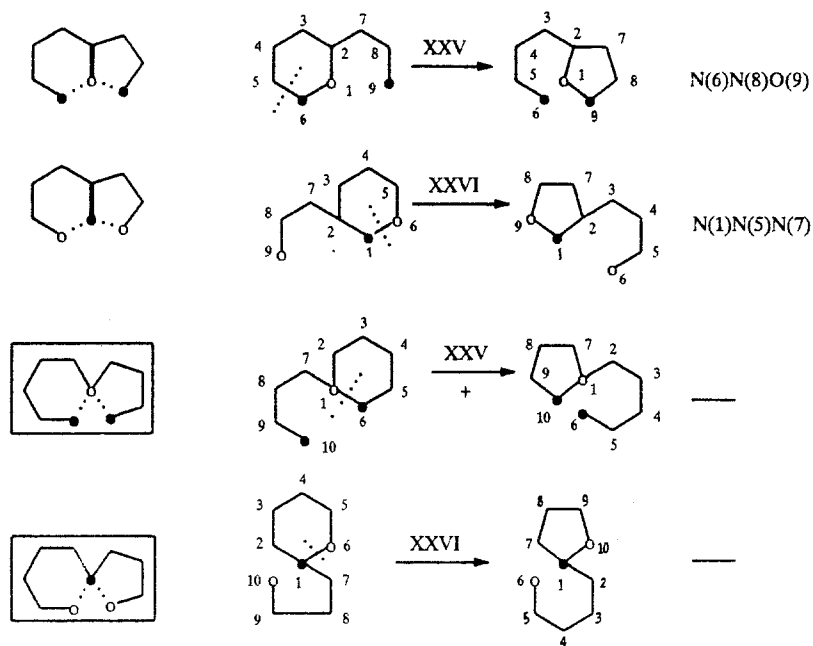
(6→6)-Transformation of hetarenes

O(6)

N(1)N(3)

N(6)

N(1)N(3)N(5)
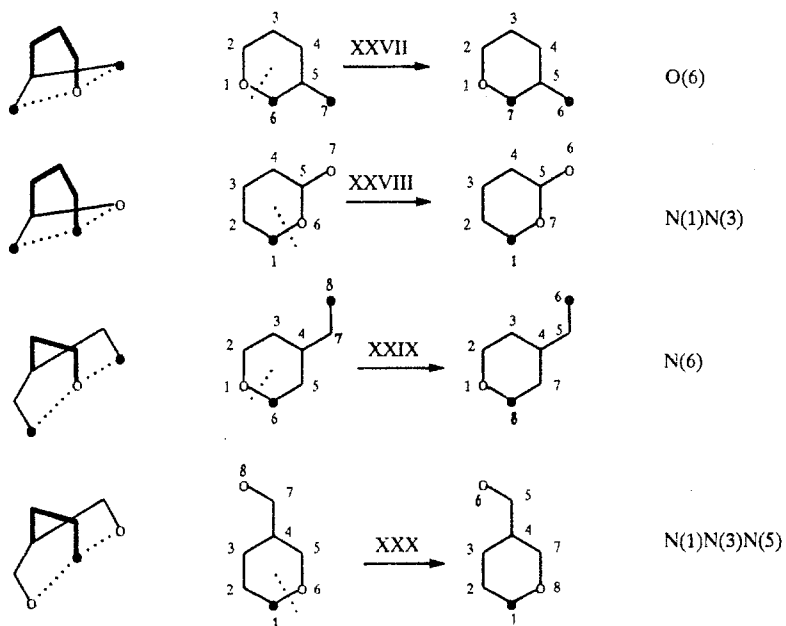
Scheme 6 (continued)
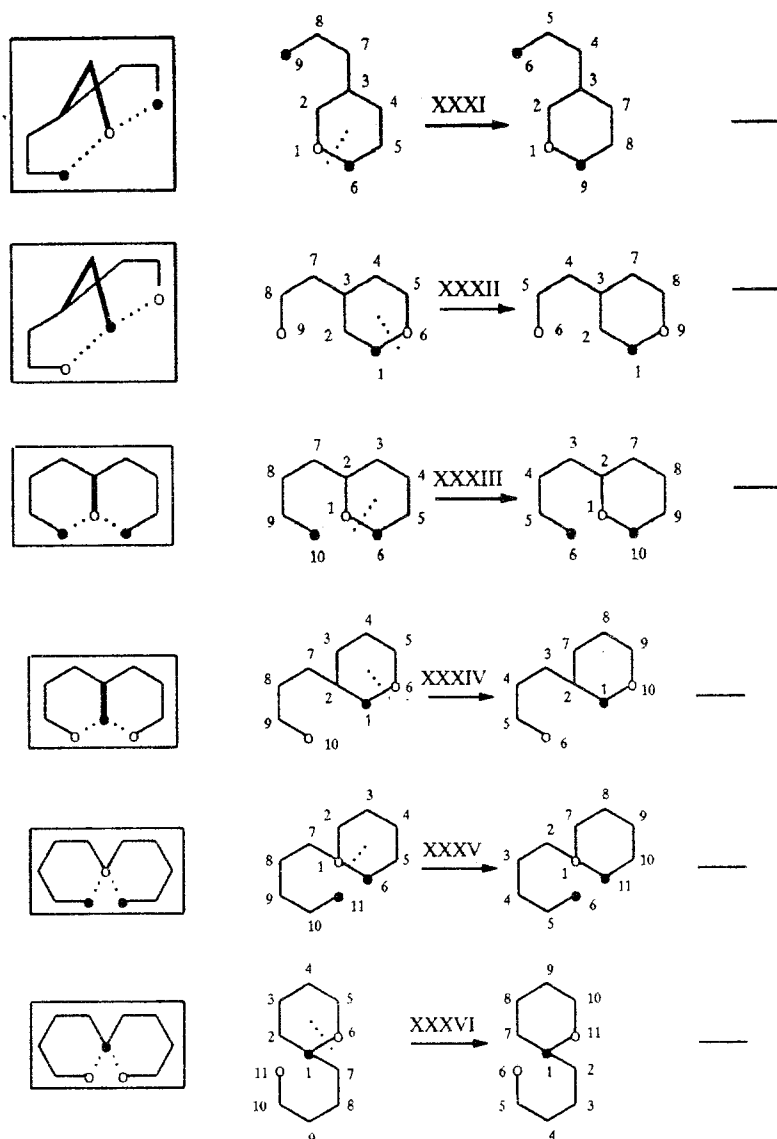
codification of recyclizations; 3) to begin the development of a unique computer data base relative to already existing recyclizations; 4) to stimulate experimental studies in the area of the search for genuinely unprecedented examples of transformations of heterocycles.

As the computer data base grows, we hope to improve the results of Scheme 6 and also to publish atlases of analogous schemes for other types, forms, classes, sorts, and families of transformations of heterocycles. Any criticism, comments, additions, and counter examples to Scheme 6 will be gratefully accepted, particularly if they have the character of experimental refutation.

With the intention in the future of attempting to limit the number of empty cells of Scheme 6 (and future analogous schemes) within the framework of our own experimental studies, the authors are open to collaboration with any other collectives.

LITERATURE CITED

1. H. C. van der Plas, *Ring Transformations of Heterocycles*, Vols. 1 and 2, Academic Press, New York—London (1973).
2. O. P. Shvaika and V. N. Artemov, *Usp. Khim.*, **41**, 1788 (1972).

3. O. P. Shvaika, "Recyclization reactions of heterocyclic compounds under the influence of hydrazine and some studies in the azole series," Author's Doctoral Dissertation Abstract, Kiev (1976).

4. A. T. Balaban et al. (eds.), *Pyrylium Salts. Syntheses, Reactions, and Physical Properties*, Academic Press, New York (1982).

5. A. T. Balaban, Plenary Paper. Papers Presented at the 6th International IUPAC Congress on Organic Synthesis, Moscow (1986).

6. W. Schroth, *The Chemistry of Heterocyclic Compounds*, J. Kovac and P. Zalupsky (eds.), Elsevier, Amsterdam (1988), p. 115.

7. G. L'abbe, *J. Heterocycl. Chem.*, **21**, 627 (1984).

8. A. J. Boulton, *Lectures in Heterocyclic Chemistry*, Vol. 2, R. N. Castle and L. B. Townsend (eds.), Heterocorporation, Orem, Utah (1974), p. S-45.

9. H. van der Plas, *Khim. Geterotsikl. Soedin.*, No. 7, 867 (1978).

10. Ya. P. Stradyn', *Khim. Geterotsikl. Soedin.*, No. 11, 1570 (1979).

11. J. Becher, *Synthesis*, No. 8, 589 (1980).

12. A. Padwa, *Rearrangements in Ground and Excited States*, Vol. 3, P. de Mayo (ed.), Academic Press, New York (1980), p. 501.

13. C. A. Ramsden, *Adv. Heterocycl. Chem.*, **26**, 1 (1980).

14. M. Ruccia, N. Vivona, and D. Spinelli, *Adv. Heterocycl. Chem.*, **29**, 141 (1981).

15. A. N. Kost, S. P. Gromov, and R. S. Sagitullin, *Tetrahedron*, **37**, 3423 (1981).

16. S. K. Robev, *Khim. Geterotsikl. Soedin.*, No. 000, 1587 (1981).

17. V. N. Charushin and O. N. Chupakhin, *Usp. Khim.*, **53**, 1648 (1984).

18. R. S. Sagitullin, *Khim. Geterotsikl. Soedin.*, No. 4, 563 (1984).

19. D. L. Boger, *Chem. Rev.*, **86**, 781 (1986).

20. A. Kato, T. Nishio, and C. Kashina, *Heterocycles*, **26**, 2223 (1987).

21. A. R. Katritzky and C. W. Rees (eds.), *Comprehensive Heterocyclic Chemistry*, Vols. 1-8, Pergamon Press, Oxford (1984).

22. A. R. Katritzky and N. Dennis, *Chem. Rev.*, **89**, 827 (1989).

23. N. V. Alekseeva and L. N. Yakhontov, *Usp. Khim.*, **59**, 888 (1990).

24. V. G. Andrianov and A. V. Eremeev, *Khim. Geterotsikl. Soedin.*, No. 11, 1443 (1990).

25. N. S. Zefirov and S. S. Trach, *Zh. Org. Khim.*, **11**, 225 (1975).

26. N. S. Zefirov and S. S. Trach (Tratch), *Chem. Scripta*, No. 15, 4 (1980).

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

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

29. I. I. Baskin, S. S. Trach, and N. S. Zefirov, Summaries of Papers Presented at the 7th All-Union Conference on the Use of Computers in the Chemical Investigation and Spectroscopy of Molecules [in Russian], Riga (1986), p. 88.

30. G. Vladutz, *Modern Approaches to Chemical Reaction Searching*, P. Willet (ed.), Andershot, Gover (1987), p. 202.

31. R. Herges, *Tetrahedron Comput. Methodology*, **1**, No. 1, 15 (1988).

32. G. A. Hopkinson, T. P. Cook, and I. P. Buchan, *Chemical Information Systems. Beyond the Structure Diagram*, D. Bawden and E. Mitchell (eds.), Ellis Horwood, New York (1990), p. 83.

33. E. V. Babaev and N. S. Zefirov, Summaries of Papers Presented at the Intercollegiate Conference on Molecular Diagrams in Chemical Research [in Russian], Kalinin (1990), p. 3.

34. E. V. Babaev and N. S. Zefirov, *Bull. Soc. Chim. Belg.*, **101**, 67 (1992).

35. M. S. Salakhov and S. A. Ismailov, *Usp. Khim.*, **55**, 2008 (1986).

36. V. I. Terenin, E. V. Babaev, M. A. Yurovskaya, and Yu. G. Bundel', *Khim. Geterotsikl. Soedin.*, No. 6 [1992, in press (this issue)].

37. E. V. Babaev, S. I. Bobrovskii, and Yu. G. Bundel', *Khim. Geterotsikl. Soedin.*, No. 11, 1570 (1988).

38. N. S. Zefirov, S. S. Trach, and O. S. Chizhov, *Carcass and Polycyclic Compounds. Molecular Design on the Basis of the Principle of Isomorphic Substitution (Advances in Science and Technology)* [in Russian], Moscow (1979).

39. Yu. G. Papulov, *Molecular Symmetry* [in Russian], Kalinin (1979).

40. M. Ch. Klin, R. Pochel, and K. Rosenbaum, *Angewandte Algebra*, VEB Deutsche Verlag der Wiss., Berlin (1988).

41. E. V. Babaev and N. S. Zefirov, Fourth International Conference on Mathematics and Computer Chemistry, 3—7.06.1991, Bled, Yugoslavia [*J. Mathematical Chem.* (1992, in press)].

42. F. Harari, *The Theory of Diagrams* [Russian translation], Mir, Moscow (1973).

43. E. V. Babaev, D. E. Lushnikov, and N. S. Zefirov, Summary of Papers Presented at the 9th All-Union Conference on Chemical Information Science [in Russian], Chernogolovka (1992), p. 93.

44. B. S. Holla and S. I. Ambekar, *J. Chem. Soc., Chem. Commun.*, No. 5, 221 (1979).

SYNTHESIS AND PROPERTIES OF FUNCTIONALLY SUBSTITUTED 1,2-AZOLIDINES

## I. A. Motorina and L. A. Sviridova

Results of investigations on the synthesis and properties of functionally substituted 1,2-azolidines are reviewed.

Compounds having azolidine fragments in their structure have a broad spectrum of biological activity. Among derivatives of substituted isoxazolidines and pyrazolidines have been found substances with fungicidal [1, 2], herbicidal [3], antiinflammatory [4, 5], and antibacterial [6, 7] activity. N-Acylisoxazolidines and benzylpyrazolidines exhibit properties of minor tranquilizers [8, 9]. Isoxazolidine derivatives that are lysergic acid antagonists [10] and pyrazolidine derivatives that are good anesthetics [11] are also known. The isoxazolidine fragment is contained in alkaloids [12, 13], terpenes [14], condensed heterocycles [15, 16], some sugar derivatives [17, 18], and antibiotics [19].

Extensive investigations in the field of the chemistry of functionally substituted azolidines began in the middle of the 1960's; in particular, there was especially rapid development of the chemistry of isoxazolidines after the appearance, in the 1960's and 1970's, of publications on the valence-molecular-orbital theory and its application (together with calculations of molecular-orbital parameters) to 1,3-dipolar cycloaddition reactions. At the same time, there is virtually no published systematization of data on the synthesis and properties of functionally substituted azolidines. (There is only a review on the chemistry of isoxazolidines [20] containing data up to May of 1974.) The present review covers the literature for mainly the last 15-20 yr. We review data on the structure, methods for synthesis, and chemical properties of isoxazolidines and pyrazolidines having heteroatomic substituents (amino, hydroxy, alkoxy, hydrazino groups, etc.) at carbon atoms.

1. STRUCTURE OF 1,2-AZOLIDINES

All possible isomers of the substituent position in the ring (3- (a), 4- (b), and 5-substituted (c) derivatives) are known for molecules of functionally substituted isoxazolidines and pyrazolidines,
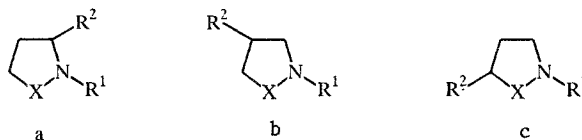
and stereoisomers are known for disubstituted azolidines.