Novel Graph-Theoretical Approach to Ring-Transformation Reactions: Hierarchic Classification and Computer Design of Heterocyclic Rearrangements[1,2]

Eugeny V. Babaev,*,† Dmitry E. Lushnikov,‡ and Nikolai S. Zefirov†,‡

Contribution from the Department of Chemistry, Moscow University, Moscow 119899, Russia, and Institute of Organic Chemistry, Moscow 117913, Russia. *Received July 27, 1992*

Abstract: A novel graph-theoretical approach based on the notion of ring-bond-redistribution graphs (RBR graphs) is suggested as a formal basis for the classification of a great number of ring transformations in heterocycles. The vertices and edges of RBR graphs exclusively correspond to those atoms and skeleton bonds which are present in both the final and initial heterocyclic rings. The simple bicyclic structure of these graphs permits us to suggest an original and effective hierarchic classification for simple heterocyclic rearrangements. Structurally similar reactions of the same level have the same RBR graphs and differ only by their labeling. This approach was realized as a computer program for the systematization of heterocyclic recyclizations. Selected predictions of previously unknown types, classes, and sorts of these reactions are briefly discussed.

Introduction

A specific feature of heterocyclic molecules is their ability to undergo ring-transformation reactions or recyclizations. Beginning with the classic work by Dimroth,[3] this type of reaction has received much attention from organic chemists, and this area has been extensively studied and reviewed (see, for example, refs 4–10). These reactions frequently lead to unexpected products, they often involve unusual mechanisms, and sometimes they prove to be the only routes to target heterocyclic structures. Recently, examples of heterocyclic ring transformations (HRTs) have been found for almost all heterocycles of any ring size and any type, number, or distribution of heteroatoms.[11] Indeed, the reactions discussed in this article (see eqs 1–30) can be regarded as good examples of such HRTs. Many of them have special names, and one can easily recall the well-known examples of the Dimroth rearrangement of azoles[3-5] (eqs 1, 3, and 22), azines[4,6] (eq 25), and fused azoloazines[6,11,12a,13] (eq 11), the Cornforth[4,5] (eq 4) and Boulton–Katritzky[5,7,10,11] (eq 5) rearrangements of 5-membered heterocycles, and the Kost–Sagitullin transformations of the pyridine nucleus to benzene[6,12a] (eqs 11 and 24). Other examples of "named" ring transformations are the well-known Hafner[6,8] and Zinke–Konig[6] reactions, as well as the processes of pyrrole, furan, and thiophene interconversions (eq 16), generally referred to in the Russian literature as the Yur'ev reactions.[12b]

However, in spite of the great experimental development in this area of heterocyclic chemistry, ring-transformation reactions are still poorly classified. Because of the great variety of recyclization types, it is sometimes difficult to evaluate the real novelty of a claimed "new ring transformation" published in the literature. There are only a few works concerned with the general classification of ring transformations.

In his first classic review on HRTs, van der Plas[4] classified reactions according to the size of the initial and final heterocyclic rings. L'abbé[5] suggested a more detailed classification of monocyclic rearrangements for 5-membered heterocycles, taking into account the number of side-chain atoms in different recyclization processes. For instance, in the case of the Dimroth reaction (eq 3) the size of the side chain corresponding to an amino group is equal to 1; for the Cornforth rearrangement (eq 4), the side chain is a carbonyl group with size 2, and for the Boulton–Katritzky rearrangement (eq 5) the size corresponding to an oxime group is equal to 3. Schwaika[7] examined a large number of azole ANRORC reactions with hydrazine and classified the resulting HRTs according to the size of unchanged ring fragments in the initial heterocycle. The proposed classification of pyrylium HRTs[8] was based on the structure of pyrylium ring fragments incorporated into new heterocyclic rings.

It should be mentioned that in all of these approaches chemists try to determine the degree of similarity for different HRTs, taking into account some structural features of the starting or final heterocycles (e.g., size of rings, side chains, etc.), i.e., comparing the molecules but not the reactions. However, the recent development of mathemetical chemistry gives reasons to compare and classify the reactions using the language of graph theory.

During the last decade, methods of graph theory became useful instruments for the analysis of various problems in theoretical and experimental chemistry.[14-22] A traditional application of graphs

† Moscow University.
‡ Institute of Organic Chemistry.

(1) Presented at meetings on the both mathematical and heterocyclic chemistry: (a) Babaev, E. V.; Zefirov, N. S. IVth International Conference on Mathematical and Computational Chemistry, Bled, Yugoslavia, 1991. (b) Babaev, E. V.; Zefirov, N. S. *Abstracts of Papers*, IInd (USSR) Conference on Molecular Graphs in Chemical Research; Kalinin University Press: Kalinin, Russia, 1990; p 3 (in Russian). (c) Babaev, E. V.; Zefirov, N. S. Presented at the Vth (USSR) Conference on the Chemistry of Nitrogen Heterocycles, Chernogolovka, Russia, October 1991; Institute of Chemical Physics: 1991. (d) Babaev, E. V.; Lushnikov, D. E.; Zefirov, N. S. *Abstracts of Papers*, IXth (USSR) Conference on Chemical Information Science, Chernogolovka, Russia, 1992; Institute of Chemical Physics: 1992; p 93 (in Russian).

(2) Preliminary communication (classification review of rearrangements for 5-membered heterocycles): Babaev, E. V.; Zefirov, N. S. *Bull. Soc. Chim. Belg.* **1992**, *101*, 67.

(3) Dimroth, O. *Ann. Chem.* **1909**, *364*, 183.

(4) van der Plas, H. C. *Ring Transformation of Heterocycles*; Academic: London, 1973; Vols. 1,2.

(5) L'abbé, G. *J. Heterocycl. Chem.* **1984**, *21*, 627.

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

(7) Shwaika, O. P.; Artyomov, V. N. *Usp. Khim.* **1972**, *41*, 1788 (in Russian).

(8) (a) *Pyrylium Salts. Syntheses, Reactions and Physical Properties*; Balaban, A. T., et al., Eds.; Academic: New York, 1982. (b) Balaban, A. T. In *New Trends in Heterocyclic Chemistry*; Mitra, R. B., et al., Eds.; Elsevier: Amsterdam, 1979; p 79. (c) Balaban, A. T. Presented at the VI International IUPAC Congress on Organic Synthesis, Moscow, 1986.

(9) Padwa, A. In *Rearrangements in Ground and Exited States*; de Mayo, P., Ed.; Academic: New York, 1980; Vol. 3, p 501.

(10) Boulton, A. J. In *Lectures in Heterocyclic Chemistry*; Castle, R. N., Townsend, L. B., Eds.; Heterocorporation: Orem, UT, 1974; Vol. 2, p S-45. Ruccia, M.; Vivona, N.; Spinelli, D. *Adv. Heterocycl. Chem.* **1981**, *29*, 141.

(11) *Comprehensive Heterocyclic Chemistry*; Katritzky, A. R., Rees, C. W., Eds.; Pergamon: Oxford, 1984; Vols. 1–8.

(12) (a) Babaev, E. V.; Bobrovskii, S. I.; Bundel, Yu. G. *Khim. Geterotsikl. Soedin.* **1988**, 1570 (in Russian). (b) Yur'ev, Yu. K.; Zefirov, N. S. In *Development of Organic Chemistry in USSR*; Nauka: Moscow, 1967 (in Russian). (c) Babaev, E. V. In *Principles of Symmetry and Systemology in Chemistry*; Stepanov, N. F., Ed.; Moscow University: Moscow, 1987; p 30; *Chem. Abstr.* **1989**, *109*, 27902b. (d) Tratch, S. S.; Zefirov, N. S. *Ibid.* p 54. (e) Bobrovskii, S. I.; Lushnikov, D. E.; Bundel, Yu. G. *Khim. Geterotsikl. Soedin.* **1989**, 1634 (in Russian). (f) Gordeeva, E. V.; Lushnikov, D. E.; Zefirov, N. S. *Tetrahedron* **1992**, *48*, 3789.

(13) Jacquier, R.; Lopez, H.; Maury, G. *J. Heterocycl. Chem.* **1973**, *10*, 755.

(14) *Chemical Applications of Topology and Graph Theory*; King, R. B., Ed.; Elsevier: Amsterdam, 1983.

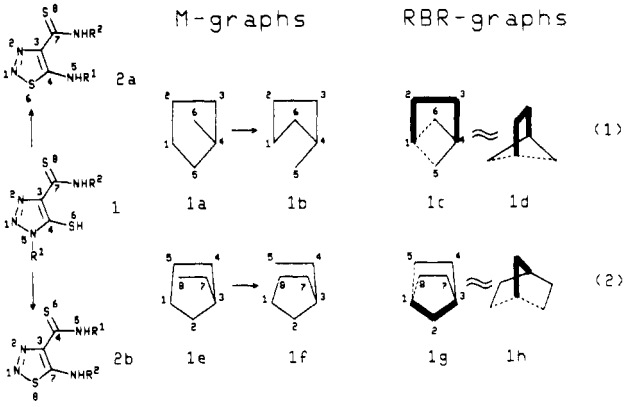
0002-7863/93/1515-2416$04.00/0

in chemistry is the description of molecular structures (indeed, a graph, as a set of vertices and edges, corresponds to the classical notion of a structural formula as a sequence of atoms and bonds). An unusual important application of graph theory is the use of a graph as a representation of a chemical reaction,[23-26] namely, for the description of bond redistribution in the course of an organic reaction. Particularly,[27-29] if the initial and final molecules in a reaction are described as labeled multigraphs, then any reaction is characterized by an edge redistribution in the initial multigraph. The redistribution is represented by a symbolic equation[29] where only the bonds whose order changes are presented. The relative positions of these bonds are uniquely characterized by the topology identifying graph; the vertices and edges of this graph correspond to reaction centers and to the bonds whose order changes during the course of the reaction, respectively. This direction (the so-called "formal-logical approach"[27,28] or FLA) has been expanded to a rigorous mathematical level.[29] The computer programs based on FLA have been elaborated;[29,30] they permit one to generate all possible symbolic equations (corresponding to a given topology identifier) and, hence, they are powerful tools in the search for new types of organic reactions.[27]

This work is stimulated by our common interest in both graph theory and in the chemistry of heterocycles.[12] Our early attempt[1,2] to describe ring-transformation reactions in heterocycles with the use of the FLA methodology showed the necessity for a certain modification of this approach. The most serious problem lies in the tautomerism of heterocycles (especially of substituted heteroaromatics), which makes it difficult to use FLA models for the general description of heterocyclic rearrangements and HRTs. Indeed, the presentation of HRTs by bond-redistribution graphs results in a great number of "waste" structures (corresponding to different tautomeric pairs), while important information on the bonds from the heterocyclic ring, which are conserved in the course of the reaction, is often lost.

In our early review on monocyclic rearrangements in azoles,[2] we came to an idea of a new type of reaction graphs—graphs of ring-bond redistribution. They seem to be a useful tool for the general classification of HRTs; the structures of these graphs do not depend on tautomerism and reflect only the topology changes of the heterocyclic nucleus in the course of the ring transformation. The objective of this article is to demonstrate applications of such graphs for the classification of numerous HRT reactions into a hierarchic system. Such a system can be used for the more strict determination of the degree of similarity (and novelty) for heterocyclic recyclizations and rearrangements and, more importantly, for computer predictions of unknown HRTs.

Scheme I. Definition of Ring-Bond-Redistribution Graph[a]



[a] Reactions 1 and 2 are symbolically represented by the pairs of molecular graphs **1a,e** and M_f graphs **1b,f**. Pairs of M graphs after superposition lead to RBR graphs **1c,g**, which may be equivalently presented by the graphs **1d,h**.

Definitions

From the qualitative point of view, our approach is based on the following chemically evident consideration: if the chemical equation of a heterocyclic ring transformation is given, the skeleton fragment of the initial ring can be easily found among the atoms and bonds of the final molecule(s) and, vice versa, the skeleton of the final ring can be found among the atoms and bonds of the starting molecule(s). Any other atoms and bonds which do not participate in the transformation of these (initial and final) rings should be omitted. The remaining atoms and skeleton bonds in the right and left parts of the chemical equation belong either to the initial or the final ring, in direct correspondence with the reaction mechanism. Thus, the left and right sides of the reaction equation can be superposed to obtain the reaction graph (as in the above-mentioned FLA approach), and one can analyze how the skeleton ring bonds are redistributed in the initial heterocycle to form the final one. Let us give a more strict definition of this qualitative idea.

We shall use the term heterocyclic ring transformation (HRT) for any reaction containing the steps of heterocyclic ring opening and ring closure in any sequence; the only requirement is that at least one (or more) atom(s) of the initial heterocyclic ring should be included as a fragment in the final ring.

We call the heterocyclic ring transformation simple (SHRT) if (a) only one ring of the initial heterocycle is transformed into only one ring of the final molecule; (b) there is no intermediate formation of rings except for the formation of the final ring; and (c) there are no intermediate permutations of the atoms which belong either to the initial or the final heterocyclic ring.

For instance, the reactions discussed below (eqs 1–27) are simple, while the HRTs (eqs 28–30) are not SHRTs because one or more of the above-mentioned conditions a, b, or c are violated. An SHRT can contain any number of starting reagents and any number of final products, but only one initial heterocyclic ring which is transformed to only one corresponding final ring should be chosen. To avoid confusion in the case of fused heterocycles, let us consider their HRTs to be SHRTs if fused rings have only one pair of common atoms (e.g., common benzannelation in indole, quinoline, etc.) and if the presence of fused rings does not violate conditions a–c.

Let us now consider any SHRT. The atoms of the final structure should be numbered in accordance with the numbers of the initial structure and with the HRT mechanism. Two types of graphs, molecular graphs of SHRTs and ring-bond-redistribution graphs, are defined.

(15) *Chemical Applications of Graph Theory*; Balaban, A. T., Ed.; Academic: London, 1976. *Chemical Graph Theory*; Trinajctič, N., Ed.; CRC: Boca Raton, FL, 1973.
(16) Merrifield, R. E.; Simmons, H. E. *Topological Methods in Chemistry*; Wiley Interscience: New York, 1989.
(17) *Advances in the Theory of Benzenoid Hydrocarbons*; Dewar, M. J. S., et al., Eds.; Topics in Current Chemistry 153; Springer: Berlin, 1990.
(18) Graovac, A.; Gutman, I.; Trinajstic, N. In *Lecture Notes in Chemistry*; Springer: Berlin, 1977; Vol. 4.
(19) Mingos, D. M. P.; Johnston, R. L. In *Structure and Bonding*; Springer: Berlin, 1987; Vol. 68, p 29.
(20) King, R. B.; Rouvray, D. H. *J. Am. Chem. Soc.* **1977**, *99*, 7834. King, R. B. *Inorg. Chim. Acta* **1986**, *116*, 126. Teo, B. K. *Inorg. Chem.* **1984**, *23*, 1251.
(21) Stankevitch, M. I.; Stankevitch, I. V.; Zefirov, N. S. *Usp. Khim.* **1977**, *57*, 337 (in Russian).
(22) *Drug Design: Fact or Fantasy*; Jolles, G., Wooldridge, K. R. H., Eds.; Academic: London, 1984.
(23) Dugundji, J.; Ugi, I. *Top. Curr. Chem.* **1973**, *39*, 19.
(24) Zefirov, N. S.; Tratch, S. S. *Zh. Org. Khim.* **1975**, *11*, 225, 1785 (in Russian).
(25) Fujita, S. *J. Chem. Inf. Comput. Sci.* **1986**, *26*, 205; **1987**, *27*, 111; **1988**, *28*, 128.
(26) Vladutz, G. In *Modern Approaches to Chemical Reaction Searching*; Willet, P., Ed.; Gover: Andershot, U.K., 1987; p 202. Herges, R. *Tetrahedron Comput. Methodol.* **1988**, *1*, 15. Hopkinson, G. A.; Cook, T. P.; Buchan, I. P. In *Chemical Information Systems. Beyond the Structure Diagram*; Bawden, D., Mitchell, E., Eds.; E. Horwood: New York, 1990; p 83.
(27) Zefirov, N. S.; Tratch, S. S. *Chem. Scr.* **1980**, *15*, 4. Zefirov, N. S. *Acc. Chem. Res.* **1987**, *20*, 237.
(28) Zefirov, N. S.; Tratch, S. S. *Zh. Org. Khim.* **1984**, *20*, 1121 (in Russian); Zefirov, N. S.; Tratch, S. S.; Gamziani, G. A. *Zh. Org. Khim.* **1986**, *22*, 1341 (in Russian).
(29) Zefirov, N. S.; Tratch, S. S. *Anal. Chim. Acta* **1990**, *235*, 115.
(30) Gordeeva, E. V. Candidate Dissertation, Moscow University, Moscow, 1986. Baskin, I. I. Candidate Dissertation, Moscow University, Moscow, 1989.

The molecular graph of SHRT (M graph) is determined for the initial reagents (symbol s) and final products (symbol f) as a set of the pairs $M_s(V,R_s)$ and $M_f(V,R_f)$, where the vertices V and the edges R of the M graph correspond only to the atoms and skeleton bonds which are present either in the starting or the final ring. The vertices of M graphs of the final and starting structures should be numbered according to the reaction mechanism; symbols of heteroatoms, as well as hydrogen atoms, multiple bonds, and substituents (including condensed rings), should be omitted in the M graph structure.

Since vertices and edges of the M_s graph are determined by the chemical equation (i.e., actually, by the resulting M_f graph), the same chemical structure can be presented by different M graphs in the case of different reactions. An excellent illustration of this thesis is shown in Scheme I: the same molecule, 5-mercapto-1,2,3-triazole-4-carbothioamide (**1**), can undergo rearrangements in two directions (reactions 1 and 2).[31a] Hence, it is presented by two different M_s graphs, **1a** and **1e**, according to the chosen reaction, i.e., corresponding to the selection of its side chains which are incorporated in the new thiadiazole ring of structures **2a** and **2b**. As a result, the M_s graph **1a** does not contain vertices, which correspond to the thioamide group, while in the M_s graph **1e** the corresponding α-amino group is omitted.

In general, the chosen M_s and M_f graphs for any SHRT contain the same number of vertices, but they differ in the number and/or distribution of edges (cf. pairs **1a,b** or **1e,f**). M graphs can be connected or disconnected, according to the chemical equation; a pair of connected M graphs corresponds to a rearrangement (as it is for the M graphs in Scheme I).

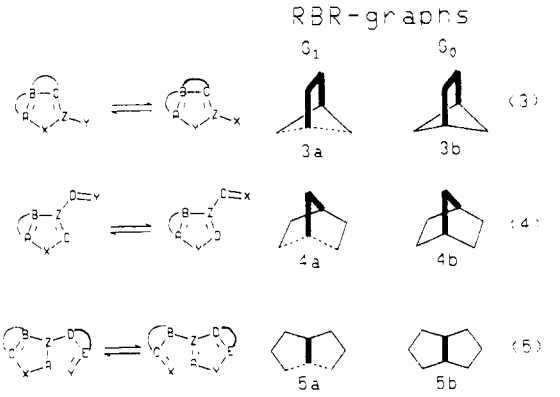
The pair of molecular M graphs (M_s, M_f) is used to construct the graph of the SHRT reaction—the ring-bond-redistribution graph (RBR graph). M_s and M_f graphs should be "superposed" according to the matching vertices with identical numbers. The resulting RBR graph of the reaction (we also call it the G_f graph) contains the same number of vertices as the M graphs and edges of different types, which are designated by the following: solid lines, if the edge is present in both the M_s and M_f graphs; dashed lines, if the edge is present only in one M graph; and bold lines, if the edge belongs to both rings of the M graph. We deliberately use all three types of labels in the G_1 graph to provide as much information as possible about the appearance (disappearance) and conservation of any ring bond in the final (initial) cycle, while the structure of the reaction graph is the same.

Examples of different pairs (M_s, M_f), **1a,1b** and **1e,1g**, and their resulting superpositions (G_1 graphs **1c** and **1g**) for reactions 1 and 2 are shown in Scheme I. Since the structures of the graphs do not depend on the manner of their drawing, we will also give an alternative presentation of the G_1 graphs (**1d** instead of **1c** and **1h** instead of **1g**). One can also conclude that the structure of the G_1 graph does not depend on the order of M graph superposition (any graph in the pair M_s, M_f can be regarded as the first one); hence, the direct and inverse reactions have one and the same RBR graph.

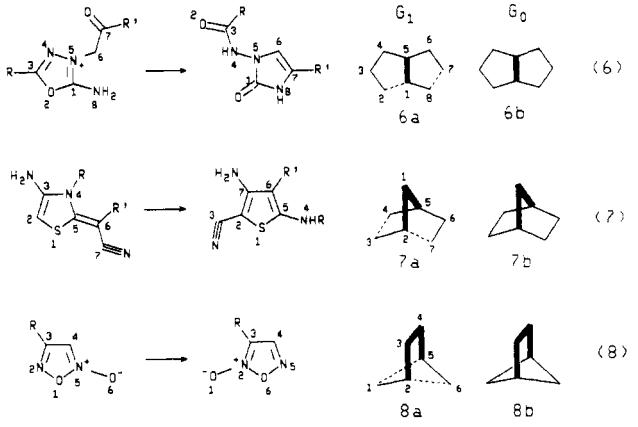
Evidently, RBR graphs such as **1d** and **1h** differ from the symbolic equations and topology identifier graphs used in the FLA approach. Indeed, redistributions of multiple bonds, as well as bonds with hydrogen atoms, are now neglected. On the other hand (also in contrast with the FLA formalism), RBR graphs contain all of the edges which correspond to strategically important skeleton cyclic bonds, even if their formal orders are conserved.

General Properties and Application of RBR Graphs

Heterocyclic Rearrangements. The suggested RBR graphs can be used to define the degree of structural similarity for different SHRT reactions as is illustrated in Scheme I. Following the above-described algorithm, one can draw the RBR graphs **3a**, **4a**, and **5a** of the known SHRTs for the examples of the "named" Boulton–Katritzky (eq 3), Cornforth (eq 4), and Dimroth (eq 5) azole rearrangements.[31b,32]



By comparing the structures of RBR graphs, one can easily establish the similarity of different HRTs. For instance, reaction 1 can be immediately regarded to be of the Dimroth class of rearrangements (eq 3), and reaction 2 should be analogous to the Cornforth rearrangement (eq 4), due to the identity of G_1 graphs **1d** and **3a** (or **1h** and **4a**). Let us say that some SHRTs belong to the same *class* if their G_1 graphs are identical. The above-mentioned classes of the Dimroth, Cornforth, and Boulton–Katritzky reactions are indeed the most widespread in the chemistry of azoles;[2] however, the existence of other classes is illustrated by reactions 6–8[31c] with different G_1 graphs **6a–8a**.



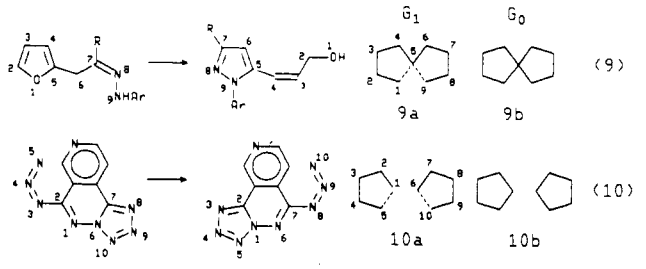
G_1 and G_0 Graphs. A more detailed comparison of G_1 graphs **3a–8a** demonstrates another possibility of HRT classification; indeed, the pairs of graphs **3a** and **8a**, **4a** and **7a**, and **5a** and **6a** have the same type of topology. To define this similarity strictly, let us introduce another RBR graph, the G_0 graph, which coincides with the G_1 graph where all of the dashed edges are changed to solid ones, i.e., it is a graph with only two sorts of edges: solid and bold (see G_0 graphs **3b–8b**). Let us say that the SHRTs are of the same *type* if their G_0 graphs are identical. Now the more global structural similarity between the reaction pairs 3 and 8, 4 and 7, and 5 and 6 is simply described by the identity of the corresponding G_0 graphs.

The difference between the types also seems clear. Thus, for each of these G_0 graphs, its structure can be interpreted as a pair of fused 5-membered rings (which are responsible for the sizes of the starting and final rings), and graphs **3b–8b** differ from each other only by the size of a boldfaced bridge between the annelated rings (i.e., the size of the fragment which is common for the initial and final heterocyclic rings). Following this interpretation, one can draw other G_0 graphs (e.g., graphs **9b** and **10b**) as pairs of 5-membered rings, containing one or no common vertices.

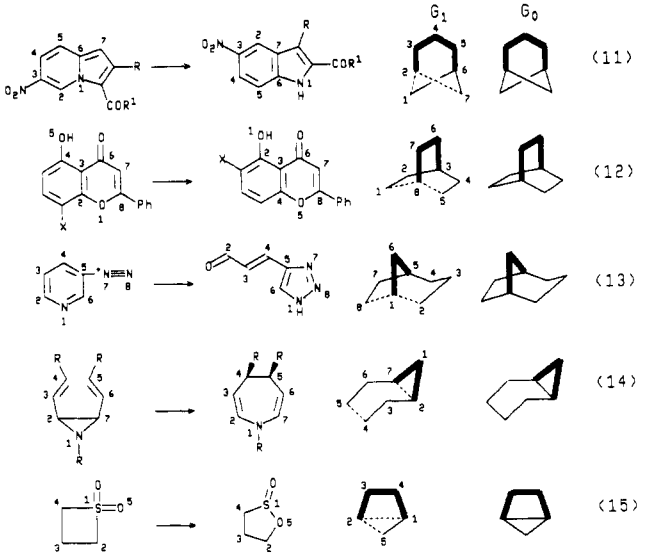
(31) (a) Reactions 1 and 2: Bakulev, V. A.; Lebedev, A. T.; Dankova, E. F.; Mokrushin, V. S.; Petrosyan, V. S. *Tetrahedron* **1989**, *45*, 7329. (b) Reactions of the general types 3–5 are taken from reviews 2, 4, 5, 10, 11. (c) Reaction 6: Hetzheim, A. *Z. Chem.* **1969**, *9*, 337. Westphal, G.; Henklein, P. *Z. Chem.* **1969**, *9*, 25. Reaction 7: Gewald, K.; Hetschel, M. *J. Prakt. Chem.* **1976**, *318*, 343. For reaction 8, see refs 10 and 11. (d) Reaction 9: Holla, B. S.; Ambekar, S. I. *J. Chem. Soc., Chem. Commun.* **1979**, 221. Reaction 10: Stanovnik, B.; Tisler, M.; Stepanov, B. *J. Org. Chem.* **1971**, *36*, 3812.

(32) A circle means the possible presence of fused rings.

However, the chemical prototypes of these G_0 graphs are comparatively rare reactions (eqs 9 and 10) (G_1 graphs **9a** and **10a**) which have also been found in the literature.[31d,33]



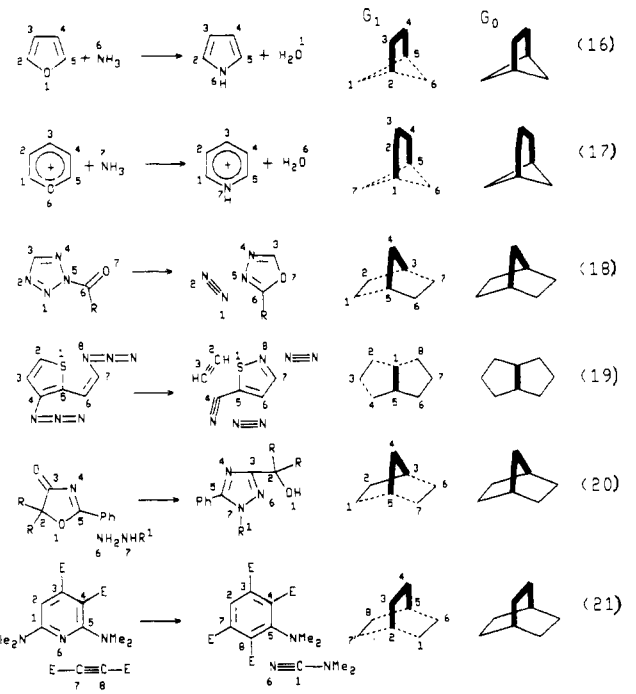
Ring Contraction/Extension Examples. One can use the idea of ring-transformation graphs G_1 and G_0 to classify not only the above-mentioned azole transformations (reactions 1–10) but also rearrangements of heterocyclic systems with any ring size. Examples 11–15 include reactions with the extension, contraction, and retention of ring size in heterocycles containing from 3 to 7 atoms.[34]



All of the RBR graphs of reactions 11–15, as in the previous examples, contain a pair of annelated rings and a bridge; according to the definition, the size of this bridge is determined by the common fragment of both heterocyclic rings. A new peculiarity is observed in the G_0 graphs of reactions 14 and 15, where the bridges coincide with the smaller cycle of the annelated rings.

Fragment Inclusion/Elimination Reactions. The developed approach is general enough to include not only rearrangements but also recyclizations in which reagents are inserted into the final ring and/or in which some fragments are eliminated from the initial ring; see, for example, eqs 16–21.[35] There are many well-known examples of this kind: for instance, reactions of heteroatom exchange, especially in pyrylium and furan chemistry (see eqs 16 and 17).

This new *kind* of SHRT should be distinguished from the previously examined rearrangements: if A and B are the labels



of the initial and final heterocyclic ring systems, and X_1, X_2, ... are the labels for any linear external fragments incorporated into the structure of a final or starting ring, we can roughly classify HRTs into different kinds by the number of independent fragments (A, B, X_i). Examples of several kinds are the following: kind $A \rightleftarrows B$, rearrangements 1–15; kind $A + X \rightarrow B$, reaction 20; kind $A \rightarrow B + X$, reaction 18; kind $A + X_1 \rightleftarrows B + X_2$, reactions 16, 17, and 21. Obviously, the number of symbols A, B, and X_i in the classification equation is not the reaction order but only the number of the fragments which are related to the starting or final ring.

The algorithm for the construction of G_1 graphs can be easily adapted for reactions of every kind (see the corresponding RBR graphs of reactions 16–21. However, in the cases of those SHRTs which are not rearrangements, the resulting molecular graphs prove to be disconnected because they must include linear fragments of reagents (or eliminated groups). In fact, the only difference between the G_1 graphs of HRTs (eqs 16–21) and the graphs of rearrangements (eqs 1–15) is the number of dashed lines corresponding to the formed or broken bonds in the heterocyclic rings. All of the G_0 graphs obtained from G_1 graphs have the usual bicyclic structure in accordance with the sizes of the initial and final rings. It is interesting that the unusual reaction 19 should be classified as $A \rightarrow B + X$, and its G_1 graph is also bicyclic. We also have to mention that reactions like 21, usually regarded as examples of cycloaddition–elimination processes, may now be considered as simple ring transformations corresponding to a G_1 graph.

Direction of Bond Heterolysis. G_2 Graphs. Following the algorithm of RBR graph construction, it is now clear how to determine the degree of structural similarity for any SHRTs. Nevertheless, there is still another chemical reason for further classification of similar SHRTs. Let us compare reactions 22 and 23[36a,b] (Scheme II).

Both reactions seem to be structurally identical, since their G_1 graphs **11a** (and obviously G_0 graphs) coincide. However, the actual difference between these SHRTs lies in the number of electrophilic and nucleophilic centers participating in the reaction mechanism. Indeed, for the Dimroth-like rearrangement (eq 22), the electrophilic carbon C-2 leaves the nucleophilic sulfur atom and forms a bond with the nucleophilic nitrogen of the *exo*-amino group.[36a] On the contrary, in reaction 23, the nucleophilic sulfur

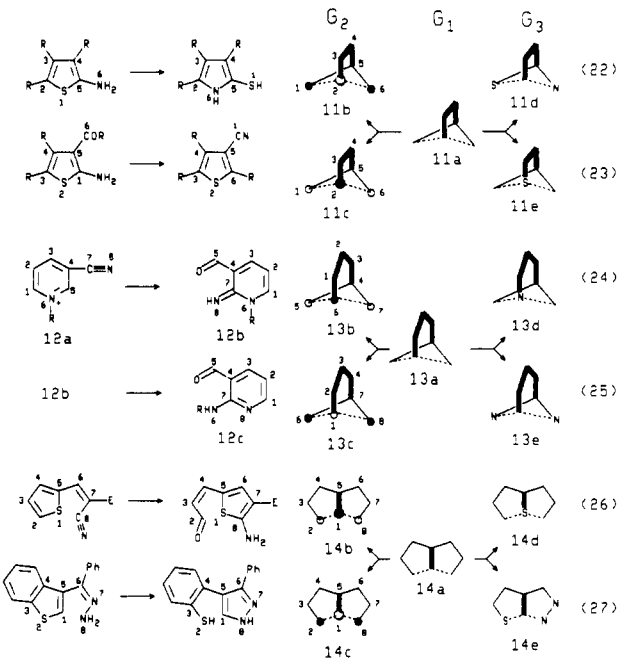
(33) Since reaction 10 violates our initial term of HRT as a recyclization with at least one atom common to both rings, we do not consider this reaction type in our approach.

(34) Reaction 11: ref 12a. Reaction 12: McCuscer, P. E.; Philbin, E. M. *J. Chem. Soc.* **1963**, *4*, 2374. Reaction 13: Sumpter, W. C.; Miller, M.; Hendrick, L. N. *J. Am. Chem. Soc.* **1945**, *67*, 1037. Reaction 14: Pomelet, J. O.; Ohuche, J. *Tetrahedron Lett.* **1974**, 3897. Reaction 15: Dorson, R. M.; Hammen, P. D.; Davis, R. A. *J. Org. Chem.* **1971**, *36*, 2693.

(35) Reaction 16: see ref 12b. Reaction 17: see refs 4 and 8. Reaction 18: see ref 4. Reaction 19: Moody, C. J.; Rees, C. W.; Tsoi, S. C. *J. Chem. Soc., Chem. Commun.* **1981**, *11*, 550. Reaction 20: see ref 7. Reaction 21: Neuenhoffer, H.; Lehmann, B. *Liebigs Ann. Chem.* **1975**, 1113. Gompper, R.; Heinemann, U. *Angew. Chem., Int. Ed. Engl.* **1980**, *19*, 216.

(36) (a) Middleton, W. G.; Enhelhardt, V. A.; Fischer, B. S. *J. Am. Chem. Soc.* **1958**, *80*, 2822. (b) Meth-Cohn, O.; Narine, B. *J. Chem. Res. S* **1977**, 294.

Scheme II. Definition of G_2 and G_3 Graphs and Their Relation to G_1 Graphs[a]



[a] In the G_2 graphs, empty circles and heavy dots correspond to electrophilic and nucleophilic centers. In the G_3 graphs the positions of the heteroatoms are shown.

atom goes away from the electrophilic C-1 atom to the electrophilic C-6 atom of the carbonyl group.[36b] Thus, we have to consider the number and distribution of potentially nucleophilic or electrophilic centers both in the ring and in the side chain.

In Scheme II, we have marked nucleophilic ("donor") atoms with heavy black dots and electrophilic ("acceptor") atoms with empty circles at the corresponding vertices in G_1 graph **11a**. The resulting new graphs **11b,c** contain useful additional information on the direction of bond heterolysis in reactions 22 and 23. We shall refer to such graphs as G_2 graphs and determine one more step in our classification: the SHRTs are of the same *sort* if their G_2 graphs coincide. Thus, reactions 22 and 23, being of the same type, kind, and class, prove to be of different sorts. In fact, the rearrangement (eq 22) is of the same sort **11b** as the Dimroth rearrangement (eq 3).

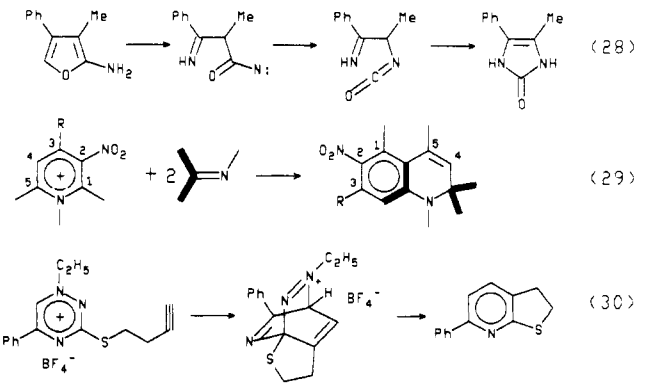
Another suitable illustration of the application of G_2 graphs is the interesting sequence of SHRTs (eqs 24 and 25)[37] in structures **12a–c** in Scheme II. Since G_1 graphs **13a** of these reactions are also identical, one can more easily grasp the resulting mechanism of this double HRT, using the structures of the corresponding G_2 graphs **13b,c**, as well as the difference in the direction of bond heterolysis in both pyridinium rings. Reaction 26 is another example of the reaction which is structurally similar to the Alberti rearrangement (eq 27)[38] (it belongs to the general Boulton–Katritzky class (eq 5)). Indeed, both reactions 26 and 27 have the same G_1 graphs **14a**, but have the opposite distribution of donor/acceptor centers (cf. G_2 graphs **14b,c**).

Evidently, the construction of G_2 graphs (which are in fact vertex-labeled graphs) is applicable only to the reactions with a heterolytic mechanism of bond redistribution; if HRT describes a process with a homolytic or synchronous mechanism, only G_0 or G_1 graphs can be used. However, the whole family of ANRORC reactions can be treated within the framework of our approach with the use of G_2 graphs.

Position of Heteroatoms. G_3 Graphs. The structures of G_0, G_1, and G_2 graphs correspond to different levels of structural similarity between SHRTs. However, these graphs do not contain the information which is most important for heterocyclic chemistry, i.e., the positions of the heteroatoms. Taking this fact into account, we can introduce G_3 graphs as G_1 graphs with heteroatom labels. The examples are G_3 graphs **11d,e**, **13d,e**, and **14d,e** of reactions 22–27 in Scheme II; just like the G_2 graphs, these new graphs also permit us to reveal the elusive similarity of these reactions.

Bicyclic Structure of RBR Graphs. The examples of RBR graphs for reactions 1–27 show that these graphs have a very simple—bicyclic—structure. In general, one can find three cycles in the structures of RBR graphs (e.g., the sequence 12345, 12346, and 1546 for G_1 graph **1c** in Scheme I); however, only two cycles are linearly independent for graphs with the cyclomatic number 2.[39] To avoid confusion, we consider the cycle to be independent only if it contains boldfaced edges, and we use the term cycle for RBR graphs only in this sense below, while the remaining (linearly dependent) cyclic fragment will be referred to as the perimeter of the RBR graph.

Now it is important to ask ourselves whether any ring-transformation reaction can be represented exclusively by bicyclic G_0 and G_1 graphs. As it is proved,[40] the bicyclic nature of the graphs should be observed only if the HRT is an SHRT. Fortunately, most of the known HRTs, independently of their kind, prove to be SHRTs (see the above-mentioned examples). Hence, the bicyclic structure of G_0 graphs should be regarded empirically as the general case for heterocyclic chemistry. Some rare counterexamples of nonsimple HRTs are ring transformations 28–30,[41] where the above-mentioned conditions a, b, and/or c are violated.



In principle, our methodology of RBR graphs can be applied not only to SHRTs; however, it can lead to a more complex RBR graph type (e.g., the appearance of more than two cycles in G_0 graphs) and will evoke more difficulties in the classification. In this article, we shall limit our analysis exclusively to SHRTs, because they are in fact the most widespread HRTs.

Restoration of M Graphs from Different RBR Graphs. Because each cycle of a bicyclic RBR graph has its ancestor, a cycle of the M_s or M_f graph, there is one-to-one correspondence between a pair of numbered molecular M graphs and the G_1 graph. The pair of M graphs can be restored from the given graph G_1 (this operation is inverse to the superposition in Scheme I) in the following manner. Let us choose any circle from the G_1 graph; the dashed line(s) of this cycle should be changed back to a solid one(s) while the dashed line(s) of another cycle should be removed. This operation restores one of the M graphs. Now the same procedure should be repeated, starting from the second cycle of

(37) Johnson, S. L.; Rumon, K. A. *Tetrahedron Lett.* **1966**, 1721.

(38) Reaction 26: Yasuda, H.; Midorikawa, H. *J. Org. Chem.* **1971**, *36*, 2196. Reaction 27: Alberti, C. *Gazz. Chim. Ital.* **1955**, *85*, 245; *Chem. Abstr.* **1956**, *50*, 9389.

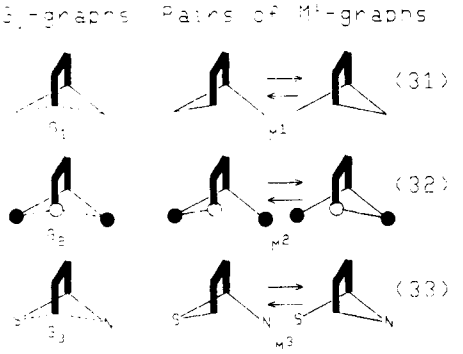
(39) Harary, F. *Graph Theory*; Addison-Wesley: Reading, MA, 1969.

(40) Really, for simple ring transformations the G_0 graph is constructed from the molecular graph of the starting system (which contains only one cycle) by addition of only one new edge into the final cycle. Independently of the place of the new edge arrangement, the resulting G_0 graph has to contain only two cycles (see ref 39).

(41) Reaction 28: Nishiwaki, T.; Saito, T.; Onomura, S.; Kondo, K. *J. Chem. Soc. C* **1971**, 2644. Reaction 29: Yurovskaja, M. A.; Afanas'ev, A. Z.; Chertkov, V.; Bundel, Yu. G. *Khim. Geterotsikl. Soedin.* **1988**, 854 (in Russian). Reaction 30: Charushin, V. N.; van Veldhuizen, B.; van der Plas, H. C. *Tetrahedron* **1989**, *45*, 6499.

the G_1 graph, to restore another M graph. The retained boldface labels should also be removed. Thus, a G_1 graph gives rise only to a pair of numbered M graphs, but not to the direction of the reaction presented by these graphs.

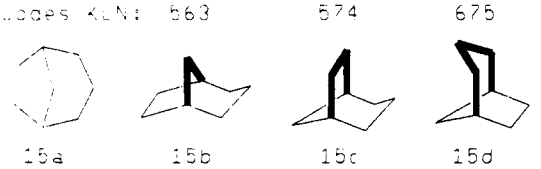
Since we introduce some other RBR graphs, G_2 and G_3 (in fact, differently labeled G_1 graphs), it also seems important to define some other types of labeled M graphs, which include either heteroatoms or donor/acceptor centers as labels. Let us refer to the M graphs which are associated with the graphs G_1, G_2, and G_3 as M^1, M^2, and M^3, respectively. As a result, each G_i graph has a one-to-one correspondence with a certain pair of M^i graphs (i = 1–3). As an example, the pairs of M^i graphs (eqs 31–33) associated with certain G_i graphs are presented.



Structure, Symmetry, Enumeration, and Codes of RBR Graphs

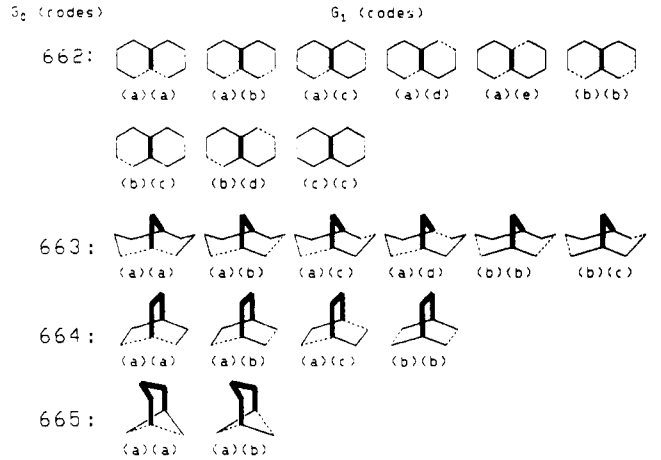
G_0 Graphs. The simple bicyclic structure of G_0 graphs permits us to suggest a simple code for any type of SHRT. The structure of a G_0 graph can be considered as a pair of annelated cycles: the size of these annelated cycles corresponds to the size of the initial and final heterocyclic rings. If the size of these heterocycles is fixed, e.g., if we consider all of the SHRTs of *K*-membered heterocycles into *L*-membered ones, the different ways of annelation between the *K*- and *L*-membered cycles (i.e., the number of vertices *N* in the boldfaced bridge) show all of the topologically nonequivalent types of ring transformations. Then, the vector *K,L,N*, where $K \leq L$ and $1 \leq N \leq K$, can be used to code the type of SHRT; the idea of such a code is somewhat analogous to the IUPAC rules for bicycloalkanes. We shall call the particular case where $N = 1$ the SHRT of spirotopology.

In this way, all of the theoretically possible types (G graphs) of ring interconversions can be exhaustively enumerated and coded. For the heterocycle sizes from 5 to 7 (covering all of the most important heteroaromatic SHRTs), the structure and the codes *KLN* of G_0 graphs were tabulated early.[2] It should be mentioned that the difference in the codes does not reflect the isomorphism of the graphs. Indeed, G_0 graphs are graphs with a labeled chain; therefore, the same graph can be labeled in different ways, leading to G_0 graphs with different codes. For instance, G_0 graphs **15b–d** with the codes 563, 574, and 675, which are responsible for HRTs with different sizes of initial and final rings (namely, 5–6, 5–7, and 6–7), without boldface labeling become isomorphic to the same graph **15a**.



G_1 Graphs from G_0 Graphs. As we defined above (cf. diagrams **3a–8a** and **3b–8b**), G_0 graphs were constructed from G_1 graphs by changing the dashed edges into solid ones; thus, SHRTs were classified into types. If we try to reverse the procedure, i.e., to subclassify or pick out the different classes in a given type of SHRT, we should, on the contrary, label some edges in a given G_0 graph as dashed ones. According to the definition of a G_1 graph, the dashed label corresponds to the formed or broken bond; for rearrangements, only two dashed edges are allowed. (Evidently, these dashed labels should be located in different annelated cycles of the G_0 graph.) Thus, all of the nonequivalent possibilities for the distribution of two dashed labels between two annelated cycles of a G_0 graph should give all of the possible G_1 graphs, i.e., the number of different classes for rearrangements of the given type.[42] As an example, let us consider G_0 graphs containing non-spiro annelated 6-membered cycles; all of the possible corresponding G_1 graphs which can be constructed by such labeling are presented in Scheme III. In the practical sense, RBR graphs in Scheme III represent exhaustively all of the possible non-spiro classes of interconversion in 6-membered heterocycles that are rearrangements.

Scheme III. Structures and Codes of G_1 Graphs Corresponding to All Possible Non-Spiro Classes of Rearrangements for 6-Membered Heterocycles



Codes for SHRT Classes. A simple convention can be used to codify any G_1 graph on the basis of the dashed label distribution. Since a G_0 graph with the code *KLN* contains two annelated cycles, let us choose any node of degree 3 in the G_0 graph and denote nonboldfaced edges in both cycles by the letters a, b, c, ... (starting from the a edges adjacent to the chosen node). As a consequence, an expression $KLN\text{-}(i_K j_K \ldots)(i_L j_L \ldots)$, showing the position of dashed labels, codifies a certain G_1 graph or a certain class of SHRT (the first is the smallest cycle; for cycles of equal size the letters should be in lexicographic order).

Earlier, the idea of such a codification was proposed to classify rearrangements of azoles;[2] now it is clear how to codify the corresponding G_1 graphs for any SHRT. For rearrangements, the possible code has the notation *KLN*-(i)(j). For SHRTs other than rearrangements, the difference lies only in the number of dashed edges used as labels of the G_0 graph and, evidently, in the corresponding number of letters in the code. Practically, the greatest number of the formed and broken bonds is not greater than 4, as it is for the SHRTs in eqs 16, 17, and 21 (two pairs of ring bonds to be formed and to be broken), where the resulting codes should have the notation $KLN\text{-}(i_K j_K)(i_L j_L)$, such as, for instance, 554-(ab)(ab), 665-(ab)(ab), and 664-(ac)(ac) for reactions 16, 17, and 21, respectively.

Enumeration of G_1 Graphs. A simple combinatorial technique can be used to enumerate the number of possible nonequivalent G_1 graphs responsible for the given G_0 graph (see the Appendix, section 1). We consider below only RBR graphs of SHRTs where the size of the heterocycle is varied from 5 to 7. Let *KLN* be the code of the G_0 graph; then the number of edges in the boldface chain of the G_0 graph is $N - 1$, and in each solid chain it is $k = K - N + 1$ and $l = L - N + 1$. The number of G_1 graphs is determined exclusively by the topology of the G_0 graph (spiro or not) and by the oddness of the values k and l. The resulting formulas for calculating the number of G_1 graphs are presented in Table I.

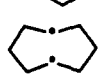
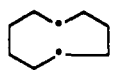
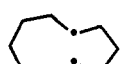
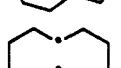
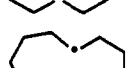
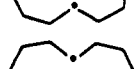
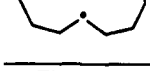
(42) It is obvious that the boldfaced edges of the bridge in the G_0 graph should not be labeled, since they are responsible for the unchanged bonds in both heterocyclic rings.

Table I. Combinatorial Equations for Calculating the Number of Possible G_1 Graphs from the Structures of Ancestral G_0 Graphs (Non-Spiro Topology)

size of cycles K and L	oddness of k, l[a]		equations for different reaction kinds		
	k	l	A ⇄ B	A + X ⇄ B	A + X_1 ⇄ B + X_2
equal	odd	odd	$(k+1)^2/4$	$(k^2+1)(k-1)/4$	$(k-1)[(k-1)(k^2+5)+4]/16$
	even	even	$k(k+2)/4$	$k^2(k-1)/4$	$k[(k-2)(k^2+6)+8]/16$
nonequal	odd	odd	$(kl+1)/2$	$(kl+1)(k+l-2)/4$	$(kl+1)(k-1)(l-1)/8$
	even	even	$kl/2$	$kl(k+l-2)/4$	$kl[(k-1)(l-1)+1]/8$
	even	odd	$kl/2$	$[kl(k+l-2)+k]/4$	$k(l-1)[l(k-1)+1]/8$

[a] The number of solid edges in every annelated cycle of G_0 graphs; for G_0 graph KLN, $k = K - N + 1$, $l = L - N + 1$.

Table II. Calculated Number of G_1 Graphs Responsible for G_0 Graphs with Different Perimeter Structures[a]

perimeter of G_0 graph	codes of G_0 graphs	number of G_1 graphs for different reaction kinds: A ⇄ B	A + X ⇄ B	A + X_1 ⇄ B + X_2
	565, 676	1	1	0
	575	2	2	0
	554, 665, 776	2	1	1
	564, 675	3	5	2
	574	4	8	4
	553, 664, 775	4	5	4
	563, 674	6	16	10
	573	8	24	16
	552, 663, 774	6	12	13
	562, 673	10	36	32
	572	12	48	48
	662, 773	9	26	31
	672	15	69	78
	772	12	45	66

[a] The perimeters of G_0 graphs are presented as polygons divided by points according to k and l values.

This result shows that the number of G_1 graphs does not depend on the size of the boldfaced bridge in the G_0 graph structure. Thus, this number can be easily calculated according to formulas from Table I, taking into account only the structure of the nonboldfaced $(k + l)$ cyclic fragment (or the perimeter) of the G_0 graph. The calculated numbers of G_1 graphs for the given perimeters are given in Table II for different kinds of SHRTs. (The points at the perimeter define the k,l values and show the position of the boldfaced bridge to be attached.)

G_2 Graphs. The enumeration of G_2 graphs corresponds to the calculation of the possible number of SHRT sorts in the given class. Because G_2 graphs are labeled G_1 graphs, it is clear that the number of possible G_2 graphs is determined by the symmetry of the initial G_1 graph. The label (a pair of black and empty circles) should be situated only at the ends of the dashed edges of the G_1 graph in all possible manners; see, for example, Scheme IV.

Scheme IV. Examples of G_2 Graph Enumeration Due to the Symmetry of the Ancestral G_1 Graph

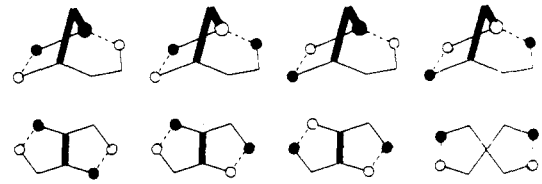


Table III. Results of Enumeration of Types, Classes, and Families for SHRTs of 5-7-Membered Heterocycles[a]

K	L	topology[b]	number of G_i graphs: G_0	G_1	G_3	non-spiro G_3/total G_3, %
5	5	spiro	1	45	5.8×10^5	24
		non-spiro	3	48	1.9×10^5	
5	6	spiro	1	108	4.8×10^6	26
		non-spiro	4	122	1.7×10^6	
5	7	spiro	1	144	1.9×10^7	27
		non-spiro	4	176	7.2×10^6	
6	6	spiro	1	78	1.0×10^7	28
		non-spiro	4	114	3.9×10^6	
6	7	spiro	1	192	8.0×10^7	29
		non-spiro	5	284	3.3×10^7	
7	7	spiro	1	136	1.6×10^8	31
		non-spiro	5	237	7.0×10^7	

[a] The number of dashed lines in G_1 and G_3 graphs varies from 2 to 4. The percent of G_3 graphs with non-spiro topology is presented. [b] For the spirotopology $N = 1$; for the non-spiro topology $N > 1$.

A special combinatorial consideration elaborated early[2] permits us to solve this combinatorial problem in the general case. The main idea is as follows: the number of G_2 graphs is determined by the structure of the automorphism group Aut(Γ) of the prototype G_1 graph, namely, it depends on the presence of special vertex permutations in this group. (Since this number varies from 1 to 4, it seems more suitable to calculate it manually.) Obviously, for G_1 graphs with codes of the (a)(a) type only two G_2 graphs are possible; this idea is illustrated in Scheme III, where the pairs of G_2 graphs **11b,c**, **13b,c**, and **14b,c** have simply the opposite manner of labeling.

G_3 Graphs. The enumeration of G_3 graphs corresponds to the following question: how many SHRT families theoretically exist for the given class? As in the case of the G_2 graph, the problem of G_3 graph enumeration is determined by the automorphism group Aut(Γ) of the G_1 graph. Due to the nature of the objects to be labeled (all of the vertices of the G_1 graph), this problem has a clear solution. In general, the number of nonequivalent labelings is described by the Polya theorem.[43] The automorphism group Aut(Γ) is easily defined for any G_1 graph. The most important labels are such heteroatoms as nitrogen (or other pnicogens) and oxygen (or other chalcogens); carbon atoms should be also present as labels. Two examples of calculations which provide the number of G_3 graphs for G_1 graphs with different Aut(Γ) are given in the Appendix, section 2.

(43) Polya, G. *Acta Math.* **1937**, *68*, 145.

Table IV. Possible (Non-Spiro) Types of SHRTs for 5–7-Membered Heterocycles and Their Homological Relationship[a]

size of cycles	oddness of k, l[b]		size of boldfaced bridge N in G_0 graph				
K and L	k	l	2	3	4	5	6
equal	odd	odd	662	773	553—664	775	
	even	even	772	552—663—774		554—665	776
nonequal	odd	odd		573		575	
	even	even	572		574		
	even	odd	672	562—673	563—674	564—675	565—676

[a] G_0 graphs are distributed according to the structure of their perimeter (K, L, k, l values) and the size of the boldfaced bridge, N. Horizontal solid lines connect codes of homologic G_0 graphs. [b] The number of solid edges in every annelated cycle of a G_0 graph; for G_0 graph KLN, $k = K - N + 1$, $l = L - N + 1$.

A special computer program has been elaborated for enumeration of all of the possible G_0, G_1, and G_3 graphs for all of the possible interconversions of 5–7-membered heterocycles of the kinds A ⇄ B, A + X ⇄ B, and A + X_1 ⇄ B + X_2. The numerical results are presented in Table III; for G_3 graphs only the approximate values are given. From Table III, the "combinatorial explosion" of the number of G_3 graphs is caused mainly by the graphs with spirotopology, over two-thirds of which are from the total set of G_3 graphs. The codes of possible G_1 graphs discussed in Table II are given in the supplementary material.

Hierarchic Classification of SHRTs

Classification Tree(s). The properties of RBR graphs discussed above permit us to suggest the first classification of simple heterocyclic ring transformations. Let us prove that this classification is a hierarchic one. The levels of such a hierarchic system could be as follows.

Level I. Only the sizes K and L of the initial and final heterocycles are known; this level corresponds to the most rough scheme which is used, for instance, in the van der Plas monograph.[4]

Level II (SHRT Types, G_0 Graphs). The pair (K,L) determines all of the possible vectors K,L,N, N = 1, 2, ..., K where $K \leq L$, and the corresponding G_0 graphs as all of the possibilities of annelation between the K- and L-membered cycles with the codes KLN, where N is the size of their common bridge; this level particularly intersects with the classification schemes by Schwaika[7] (azoles) and Balaban[8] (pyrylium salts). As an example, all structures of this level for $K = L = 5$ are exhausted by the G_0 graphs **6b–10b**.

Level III (SHRT Kinds). The number of skeleton ring bonds E ($E \leq 2$), corresponding to the number of formed and broken bonds (or dashed edges), is defined; the additional levels correspond to $E = 2$ or rearrangements.

Level IV (SHRT Classes, G_1 Graphs). The structure and symmetry of the given G_0 graph with the KLN code define the possible number of G_1 graphs (with a pair of dashed edges) with KLN-(i)(j) codes (see, for example, Scheme III and Tables I and II); this level intersects slightly with the L'abbé classification[5] (azoles).

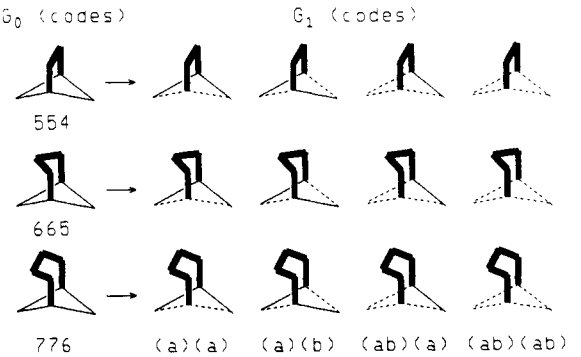
Level Va (SHRT Sorts, G_2 Graphs). The structure of the group Aut(Γ) of the given G_1 graph determines the number of possible G_2 graphs, which contain information on the direction of bond heterolysis (see Schemes II and IV). This level should be omitted (or substituted for level Vb) for pericyclic or biradical rearrangements.

Level Vb (SHRT Families, G_3 Graphs). The cycle index of the group Aut(Γ) of the given G_1 graph and the number of heteroatomic labels determine the number of possible G_3 graphs (see Scheme II, G_3 graphs **11d,e**, **13d,e**, and **14d,e**). Thus, levels Va and Vb can be derived independently from level IV.

Level VI (Pair of M^i Graphs, Skeleton Equation of SHRT). By the operation which is reversible to the definition of a G_1 graph (see eqs 31–33), a pair of M graphs is produced; the direction of the reaction (direct or inverse) is still not known. If necessary, any type of M^i graph (i = 1–3) can be chosen. Depending on the choice of i, this level can be derived directly from level IV, Va, or Vb. The corresponding sublevels can be named as VI.i (i.e., as VI.1, VI.2, or VI.3, respectively).

Level VII (Common Chemical Equation). At this level, the direction of the reaction is defined, and the necessary substituents, annelated rings, multiple bonds, etc. are added.

Scheme V. Homological Relationship between G_0 and G_1 Graphs



As one can see, the main idea of the suggested hierarchy is that the symmetry properties of the elements of each higher level strictly determine the number of elements (and their structure) at the lower level. Depending on the needs of the chemist, three alternative types of the classification tree can exist: (1) without any vertex labeling until the last level, i.e., the sequence of the levels is I, II, III, IV, VI.1, VII (G_0, G_1, M^1 graphs); (2) including the direction of bond heterolysis, i.e., levels I, II, III, IV, Va, VI.2, VII (G_0, G_1, G_2 M^2 graphs); (3) including only heteroatomic labels, i.e., levels I, II, III, IV, Vb, VI.3, VII (G_0, G_1, G_3, M^3 graphs).

Homological Relationship between SHRTs. Let us call G_0 graphs (SHRT types) homologic if they have the same k and l values of the perimeter and differ by their N values. For instance, the G_0 graphs in Scheme V are homologic. The homologic G_0 graphs have an interesting combinatorial peculiarity: the possible number and the codes for their G_1 graphs are the same. (It simply follows from their symmetry properties; see Table I and Appendix section 1.) For instance, the homologic G_0 graphs of Scheme V give rise to an analogous series of G_1 graphs (rows) with the same codes (columns).

Let SHRT classes (G_1 graphs) be also homologic if they have the same letter code and belong to the same homologic type (like the columns of G_1 graphs in Scheme V). The homological relationship between SHRT classes reflects the well-known similarity between SHRTs of heterocycles with different ring sizes. For instance, the Dimroth classes of rearrangements either for azoles (eqs 1, 3, and 22) or for azines (eq 25) are homologic, as is the Kost–Sagitullin SHRT (eq 11) of azoloazine. On the other hand, the reactions of heteroatom exchange in pyrylium and furan chemistry are also homologic (cf. eqs 16 and 17).

The homological relationship brings to our hierarchic classification the elements of a "periodic system", since the structure of some branches and the number of sublevels in the classification tree are periodically repeated. This thesis is illustrated in Table IV, where the possible (non-spiro) types of SHRTs for 5–7-membered heterocycles are distributed according to the structure of the perimeter (rows), the size of the boldfaced bridge (columns), and the homological relationship (horizontal solid lines between the codes of homologic structures) of G_0 graphs.

When such a hierarchic (and periodic) classification exists, it is easy to locate any known SHRT at the fixed branch of such a classification tree(s) and to attribute the proper code to its type,

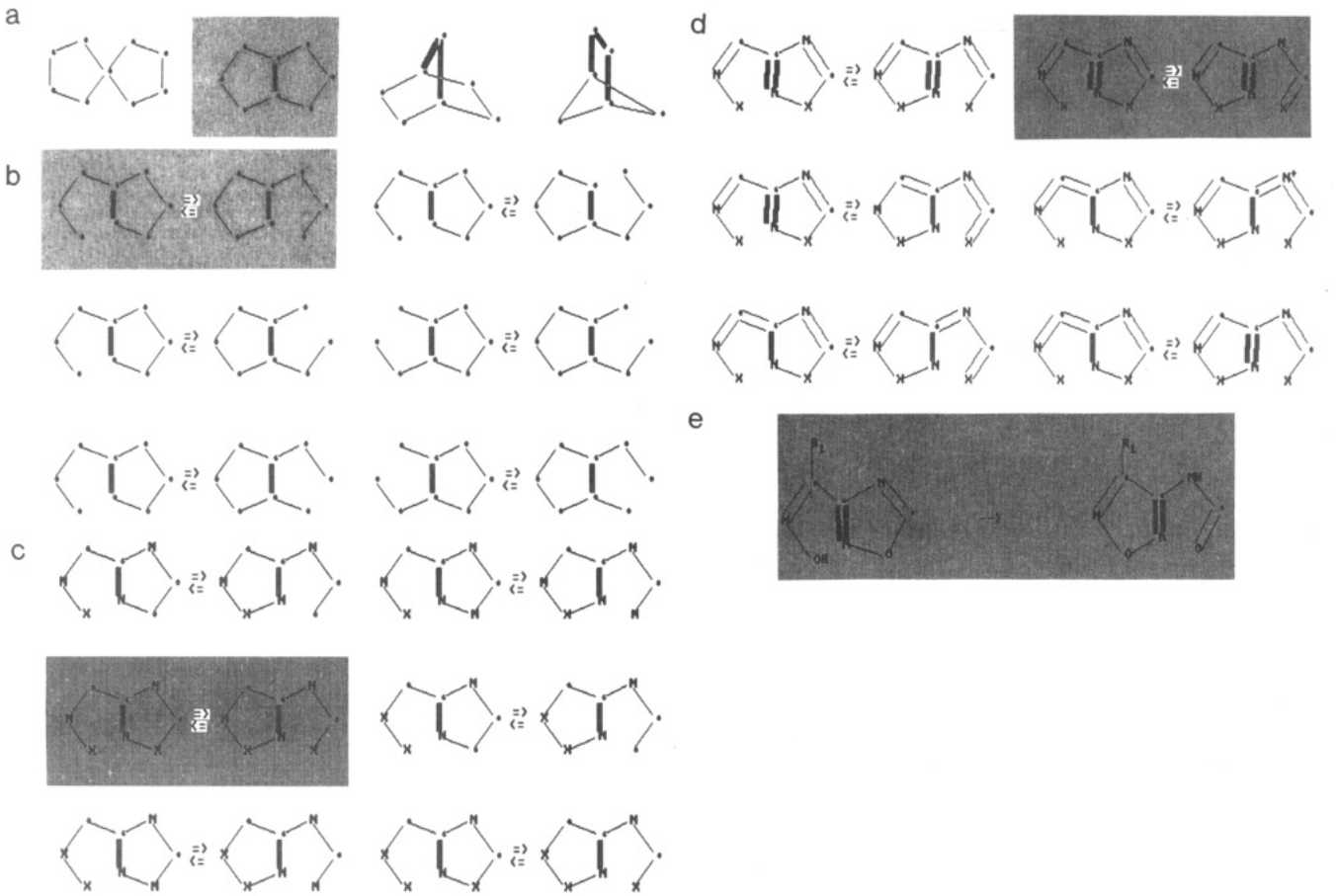


Figure 1. Computer program presentation of RBR graphs: (a) G_0 graphs; (b) G_1 graphs as pairs of M^1 graphs; (c) G_3 graphs as pairs of M^3 graphs; (d) presentation of different tautomers; (e) simple entry in the database of known ring transformations.

kind, and sort. (Since the general classification of reactions is an urgent problem in chemical information science,[26] we hope the recommended codes will be accepted for the nomenclature of SHRTs.) On the other hand, it is easy to find which types (classes, sorts, etc.) are widespread in heterocyclic chemistry and which ones are rare (or even exotic) examples, i.e., one can establish the real degree of novelty for any HRT. Moreover, this scheme permits us to make probable predictions about the existence of previously unknown ring transformations. It is obvious, however, that this kind of problem would be better examined with the use of computers. The next section describes our experience with the elaboration of such a computer program.

Computer Generation of SHRTs

The GREH Program. The program GREH (generation of rearrangements in heterocycles) is designed on the IBM PC in order to generate, visualize, and investigate the SHRT hierarchic tree. Each new branch of the tree is generated when the user attempts to display it. All of the branches are stored in separate files on the hard disk for the subsequent viewing and search. For the user's convenience, RBR graphs are shown only at the top level (G_0 graphs); all of the lower levels present information in the form of differently labeled pairs of M graphs.

Levels I and II. The program starts from the root of the tree. First, the user should select the sizes of the initial and final heterocyclic rings (level I) from the menus. The current choice (it may vary from 3 to 8) is shown at the top of the screen. After this, the program displays (or generates, if necessary) all of the possible ways of annelation for these cycles as bicyclic G_0 graphs, giving the possible types of recyclizations, i.e., the process passes to level II (Figure 1a). The highlighted rectangle in Figure 1a denotes the existence of lower levels in the tree for the given G_0 graph (for example, from the existing database). By cursor movement, one can choose the current structure, which becomes the ancestor for the lower levels.

M^1 Graph Presentation of G_1 Graphs. At the next level, the program does not show the structure of G_1 graphs for the chosen ancestor and the reaction kind (i.e., levels III and IV); instead, the pair of unlabeled M^1 graphs from level VI.1 presents the structure of the corresponding G_1 graph (Figure 1b). (Due to the established one-to-one correspondence between the G_1 graph and the pair of M^1 graphs, such a representation is equivalent to the definition of the G_1 graph structure or the class of SHRTs). In the current version, the possible number of broken or formed bonds (which defines the reaction kind) varies from 1 to 2. The user can scroll the complete list of M graph pairs and jump to the top or bottom of the list.

Level of Families. The next level of the tree corresponds to all of the possible families for the given SHRT class; i.e., it represents all of the possible arrangements of heteroatom labels (G_3 graphs) among the vertices of the G_1 graph. As in the previous case, the resulting G_3 graphs are shown on the screen as a pair of labeled M graphs from level VI.3 (Figure 1c). As the most important labels, the symbols N (nitrogen) and X (element of main group VI, e.g., oxygen or sulfur) are used.

Level of Tautomers. After a certain distribution of heteroatoms is selected, the novel classification level, tautomeric structures, is generated (Figure 1d). Before such generation, the user is requested for the minimum and maximum numbers of allowed double bonds; it is postulated here that the degree of saturation is conserved during a SHRT. The resulting set of M graphs at this level represents all of the reasonable arrangements of double bonds in the left and right M graphs containing heteroatomic labels. The completely independent arrangement of double bonds in both M graphs leads to a lot of "waste"; therefore, double bonds are arranged exhaustively only at the left side of the equation. Due to the direct correspondence between the matching edges of the left and right M graphs, each double bond from the left M graph is transferred to another (right) M graph according to the following empirical rules: (1) the double bond may remain at the

same position; (2) it may appear at one of the adjacent edges; (3) it may be converted into a single bond if it is adjacent with a dashed edge of the G_1 graph (i.e., the bond of the M graph which is to be formed). Thus, the new double bond appears in one of the adjacent places with another dashed edge (or a single bond to be broken). Another essential rule of rejection of unreal tautomers of heteroaromatics is to discard molecules with only one sp^3-hybridized carbon atom in a 5-membered heterocycle.

The last (bottom) level of the tree is the database to be filled by the recyclizations which are described in the literature. Each reaction has its index according to our classification; thus, each branch of the tree has access only to relevant reactions and references. A simple example of the Boulton–Katritzky rearrangement is shown in Figure 1e.

Selection Options. The GREH program can generate a vast number of rearrangements; therefore, it requires an effective search and selection of a small relevant part out of the whole rearrangement array. For instance, the program permits one to carry out a quick search for the next (after the current one) equation which has descendants at the lower level (highlighted equations). A special option for the user is the possibility to choose the skeleton of the required heterocyclic nucleus, as well as the sequence of heteroatoms in the side chain (both for the right and left structures) at the level of families. In the same way, the required tautomeric structure can be easily found. Hence, at each level up to 4 consequent searches may be done. After each search session, only the structures found are visible, and the following search will be conducted with the currently selected subset. For "symmetrical" rearrangements where the sizes of both rings are equal, the first search on the ring is carried out both at the left and right side of the equation up to the first matching, and the second ring search will be done exactly at the opposite side of the equation. During ring selection, one can fix the position of a side chain(s) as it is drawn on the screen or permit "rotation" of the ring, which means the search for all nonequivalent arrangements of the ring in relation to the side chain(s). As a result, one can carry out easily adaptable searches for rearrangement subsets of particular interest. For instance, the user would ask unusual questions, such as to find all of the SHRTs of pyrylium salts into pyrazoles, heterocyclic rodanides into thiazoles, or heterocyclic hydrazones into heterocyclic azides.

Predictions of New SHRTs

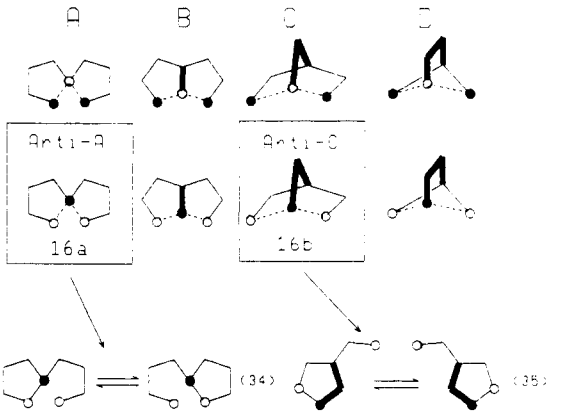
As we have mentioned, chemists usually make predictions of new SHRTs intuitively and only at the level of molecular graphs (e.g., by the variation of substituents, annelated rings, degree of saturation, etc. in heterocyclic rings), i.e. at the last two levels of our classification tree, but not by use of reaction RBR graphs from the higher levels.

We present here our experience in the theoretical prediction of two previously unknown sorts of heterocyclic rearrangements in 5-membered rings (e.g., azoles) on the basis of the suggested RBR graph concept. Probable experimental confirmation of our prediction would be a good starting point for tens of previously unknown families of rearrangements or hundreds of concrete reactions.

Actual Distribution of SHRTs into Sorts. We have analyzed the distribution of the available azole and azoline rearrangements between different theoretically possible types, classes, and sorts. Concrete numeric values of such a distribution[2] show that all of the possible types of such SHRTs (see Scheme VI) have already been discovered; the most important (97% of all rearrangements) are the G_0 graphs with the codes 552, 553, and 554, i.e., graphs of non-spiro topology. The most interesting is the distribution into classes of each type: the (a)(a) class turns out to be the most widespread class (more than 90% in every type) among the azole (azoline) rearrangements. In particular, the well-known examples of the Dimroth (D), Cornforth (C), Boulton–Katritzky (B), and Holla–Ambekar (A) rearrangements belong to the (a)(a) classes (in parentheses are the mnemonical abbreviations which are used below for these named reactions).

As we have already mentioned (cf. Scheme II), there are only two possible sorts of G_2 graphs corresponding to G_1 graphs of

Scheme VI. Known Examples of Reaction Sorts Shown by Their G_2 Graph Structure[a]



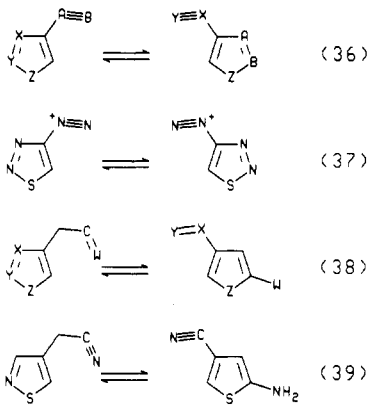
[a] In the rectangles are given the unknown sorts with the corresponding pairs of molecular M^2 graphs for new ring-transformation predictions.

(a)(a) classes: the possible structures should contain either two nucleophilic (N) and one electrophilic (E) center (NNE sort) or vice versa (EEN sort). The rearrangements of the first NNE sort turn out to be more widespread in azole chemistry, and they are also better studied. The sorts which include the named rearrangements A, B, C, and D have the corresponding codes 551-(a)(a)-NNE, 552-(a)(a)-NNE, 553-(a)(a)-NNE, and 554-(a)-(a)-NNE. The contrary EEN sort, nevertheless, has also been found, and its contribution varies from 20% in the 554-(a)(a) class to 10% in the 552-(a)(a) class. Due to the opposition in the distribution of donor/acceptor centers with the known named reactions A, B, C, and D, these sorts of rearrangements could be conventionally named as the "anti-D" (G_2 graph **11c**, eq 23) and "anti-B" sorts (G_2 graph **14b**, eq 26) in Scheme II, respectively. As far as we are aware, the corresponding examples for the rather interesting "anti-C" and "anti-A" rearrangements of the sorts 553-(a)(a)-EEN and 551-(a)(a)-EEN are unknown; the structures of G_2 graphs **16a,b** and the pairs of the labeled M^2 graphs which correspond to these really new predicted sorts of SHRTs (eqs 34 and 35) are shown in Scheme VI.

Selected Examples of Computer Predictions. By use of the GREH program, we have also enumerated the possible number of pairs of M^3 graphs which are responsible for eq 35. The total number of families, i.e., pairs of M^3 graphs, was about 1200, and after the exclusion of unreasonable chemical structures (e.g., with 5 heteroatoms or 3–4 atoms from main group VI) it was near 1000. This result was conjugated with the chemical sense of the black and empty labels of eq 35. Similar to the known types of ring-opening reactions of azoles, the black label in the rings of M^2 graphs in eq 35 was postulated to be a pyrrole-like heteroatom Z (Z = NR, O, or S in the structures of Scheme VII); that is why the most appropriate heterocycles were heteroanalogs of pyrrole, pyrazole, imidazole, and 1,2,3-triazole (Z = NR) skeletons. Other reasonable chemical arguments gave the optimum sequence of atoms in the side chain of M graphs from eq 35 as CC, NC, and NN. Because the right (terminal) atom should be electrophilic, these atoms either must be connected by a triple bond or, in the case of the fragment CC, should contain an *exo*-carbonyl or cyano group (all other combinations of heteroatoms seem to be unreasonable). With the use of the search option, the necessary combinations of heterocycles and side chains were generated. The addition of a maximum number of multiple bonds (three) at the tautomer level resulted in the possible heteroaromatic structures. Some of the selected examples are shown in Scheme VII; one of the most attractive examples of our predictions seems to be the degenerate rearrangement (eq 37) of 1,2,3-thiadiazole-4-diazonium.

Let us now compare the results of our prediction with the literature data on the behavior of azole 4-diazonium salts (Scheme VIII). Indeed, according to our predictions, in the isoxazole-4-

Scheme VII. Predicted Rearrangements of the Previously Unknown Sorts (General Equations and Concrete Examples Are Given)



(36)

(37)

(38)

(39)

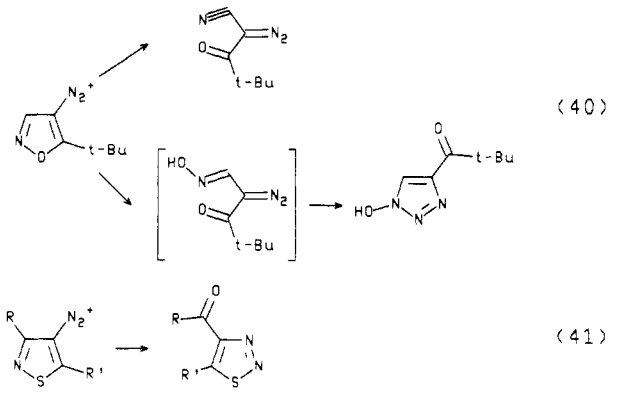
. Z: NR, O, S

X, Y: CH, N

A≡B: C≡CH, $^{+}$N≡CR, $^{+}$N≡N

C≡W: COR, CN

Scheme VIII. Known Chemical Behavior of Azole 4-Diazonium Salts



(40)

(41)

diazonium ion (generated in situ),[44] the desired heterolytic cleavage of the N–O ring bond occurs (reaction 40). However, further reactivity of the open intermediate is determined by the nucleophilic media, which is used by the authors: it easily undergoes either a proton loss or the addition of a hydroxyl group to electrophilic nitrogen. As a result, only side products instead of the expected 1,2,3-oxadiazole have been observed. On the other hand, isothiazole-4-diazonium ion,[45] after the expected cleavage of the S–N ring bond (reaction 41), does indeed undergo the desired recyclization into 1,2,3-thiadiazole. However, the presence of a substituent in the 3-position of isothiazole excludes the possibility of proton elimination (which would lead to a cyano group). As a result, the nitrogen atom of the initial ring is eliminated in the course of the reaction, and the resulting SHRT no longer belongs to the (a)(a) rearrangement type.

The above-discussed examples prove our prediction of the new 553-(a)(a)-EEN sort of rearrangements to be reasonable from the chemical viewpoint: both predicted directions of ring opening and ring closure for this sort are true. The most serious problem, which explains why this SHRT sort has been unknown for such a long time, seems to lie in the probable unstability of opened bis-electrophilic intermediates in the presence of nucleophiles. (Indeed, the usual reaction media for HRTs is water or alkanoles.) Hence, an investigation (and reinvestigation) of the behavior of these models in nonnucleophilic media is recommended.

(44) L'abbé, G.; Godts, F.; Toppet, S. *Tetrahedron Lett.* **1983**, 3149.
(45) Lee, F. T.; Volpp. G. P. *J. Heterocycl. Chem.* **1970**, *7*, 415.

Conclusion

Among the different computer applications related to organic reactions, such as computer-assisted synthesis, construction, and handling of reaction databases, or molecular graphics, one of the most serious problems is how to make computer programs really useful for experimental chemistry. In our approach we demonstrate the solution of such a problem from the beginning in a particular branch of organic chemistry—heterocyclic ring transformations: (1) An appropriate mathematical model for the description and classification of a great amount of empirical literature facts on SHRTs is suggested. (2) The computer program,[46] which is based on this model, is elaborated with the aim of hierarchic SHRT classification and predictions of unknown examples. (3) Reasonable predictions of previously unknown reaction sorts which are based on our model and computer program are given; they can be easily confirmed by members of the wide scientific community.

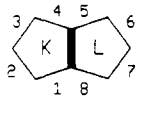
Acknowledgment. We thank S. Tratch and Dr. M. Klin for the idea on the enumeration of G_1 and G_3 graphs.

Appendix

1. Counting the Number of G_1 Graphs. The number of nonequivalent G_1 graphs is calculated by the Cauchy–Frobenius–Burnside lemma:[47]

$$t = \frac{1}{|\Gamma|}\sum_{\gamma\in\Gamma}\chi(\gamma)$$

where $|\Gamma|$ is the order of the permutation group acting on the set Φ of all possible edge labelings and $\chi(\gamma)$ is the number of labelings fixed by permutation γ from this group.



552

For example, the number of G_1 graphs for the G_0 graph 552 can be counted as follows: The order of edge automorphism group **H** for graph 552 is 4.

$$\mathbf{H} = \{h_1,h_2,h_3,h_4\}$$
$$h_1 = (1)(2)(3)(4)(5)(6)(7)(8)$$
$$h_2 = (1,4)(2,3)(5,8)(6,7)$$
$$h_3 = (1,5)(2,6)(3,7)(4,8)$$
$$h_4 = (1,8)(2,7)(3,6)(4,5)$$

The group **H** of a G_0 graph induces the group $\Gamma = \{\gamma_1,\gamma_2,\gamma_3,\gamma_4\}$ acting on the set of all possible labelings Φ. Let k be the number of edges of the cycle K which do not belong to a bridge and let a be the allowed number of dashed edges in this cycle; only the values $a = 1$ and $a = 2$ will be considered here. Similarly, l is the number of edges of the cycle L which do not belong to a bridge, b is the allowed number of dashed edges in this cycle, and b is equal to either 1 or 2. The cardinality of the set Φ consisting of all allowed labelings is counted as

$$|\Phi| = \begin{cases} \binom{k}{a}\binom{l}{b} & \text{if } a = b \\ 2\binom{k}{a}\binom{l}{b} & \text{if } a \neq b \end{cases}$$

The identity permutation γ_1 fixes all labelings: $\chi(\gamma_1) = |\Phi|$. Thus, for graphs similar to 552 ($k = l$) we obtain $\chi(\gamma_1) = \binom{k}{a}^2$ if $a =$

(46) A more detailed description of computer implementation and algorithms will be discussed elsewhere.
(47) Kerber, A. *Algebraic Combinatorics Via Finite Group Actions*; BI-Wiss.: Mannheim, 1991; p 11. Klin, M. Ch.; Pochel, R.; Rosenbaum, K. *Angewandte Algebra*; WEB Deutscher Verlag der Wissenschaften: Berlin, 1988.

b and $\chi(\gamma_1) = 2\binom{k}{a}\binom{k}{b}$ if $a \neq b$. The numbers $\chi(\gamma_2)$, $\chi(\gamma_3)$, and $\chi(\gamma_4)$ differ for each reaction kind and may be counted by simple combinatorial considerations.

For the reaction kind A $\rightleftarrows$ B ($a = 1$, $b = 1$):

$$\binom{k}{1} = k \qquad \chi(\gamma_2) = 0 \qquad \chi(\gamma_3) = \chi(\gamma_4) = k$$

$$t = (k^2 + k + k)/4 = k(k + 2)/4$$

For the reaction kind A + X $\rightleftarrows$ B ($a = 1$, $b = 2$):

$$\binom{k}{1} = k \qquad \binom{k}{2} = k(k-1)/2 \qquad \chi(\gamma_2) = \chi(\gamma_3) = \chi(\gamma_4) = 0$$

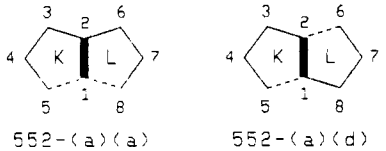
$$t = (2k^2(k-1)/2)/4 = k^2(k-1)/4$$

For the reaction kind A + X_1 $\rightleftarrows$ B + X_2 ($a = 2$, $b = 2$):

$$\binom{k}{2} = k(k-1)/2 \qquad \chi(\gamma_2) = k^2/4 \quad \chi(\gamma_3) = \chi(\gamma_4) = k(k-1)/2$$

$$t = ([k(k-1)/2][k(k-1)/2] + k^2/4 + k(k-1)/2 + k(k-1)/2)/4 = k((k-2)(k^2+6) + 8)/16$$

2. Counting the Number of G_3 Graphs. At present we use three vertex labels: C = carbon, N = nitrogen, X = element of main group VI (oxygen, sulfur, or their analogs). The total number of labelings (i.e., G_3 graphs) for three labels and G_1 graphs without symmetry is 3^N, where N is the number of vertices of the G_1 graph. In general, the number of nonequivalent labelings can also be counted by the Cauchy–Frobenius–Burnside lemma, but calculations based on simple sequences from the Polya theorem[43] are easier. For example, for the G_1 graph 552-(a)(a) the number of labelings can be counted in the following way.



(1) The automorphism group Aut(G_1) of this graph consists of two permutations:

$$\Pi = \mathrm{Aug}(G_1) = \{\pi_1, \pi_2\}$$

where π_1 = (1)(2)(3)(4)(5)(6)(7)(8) and π_2 = (1)(2)(3,6)(4,7)(5,8). (2) Calculate the cycle index for each permutation:

$$Z(\pi_1) = x_1^8 \qquad Z(\pi_2) = x_1^2 x_2^3$$

(3) Calculate the cycle index for the group Π:

$$Z(\Pi) = 1/2(x_1^8 + x_1^2 x_2^3)$$

(4) Substitution of x_i, i = 1,2, by the number of labels (3) gives the result:

$$t = 1/2(3^8 + 3^5) = 3402$$

For the G_1 graph 552-(a)(d) the number of labelings can be counted similarly:

$$\Pi = \mathrm{Aut}(G_1) = \{\pi_1, \pi_2\} \qquad (1)$$

where π_1 = (1)(2)(3)(4)(5)(6)(7)(8) and π_2 = (1,2)(3,8)(4,7)(5,6).

$$Z(\pi_1) = x_1^8 \qquad Z(\pi_2) = x_2^4 \qquad (2)$$

$$Z(\Pi) = 1/2(x_1^8 + x_2^4) \qquad (3)$$

$$t = 1/2(3^8 + 3^4) = 3321 \qquad (4)$$

Supplementary Material Available: Table of the codes of all possible G_1 graphs which correspond to interconversions of 5–7-membered heterocycles (5 pages). Ordering information is given on any current masthead page.

A diskette with the GREH program for the IBM PC and compatible personal computers is available from Prof. Steve R. Heller, Model and Database Coordination Lab., Agriculture Research Service, BARC-W, Beltsville, MD 20705.

Dynamic Models for the Thermal Deazetization of 2,3-Diazabicyclo[2.2.1]hept-2-ene

Barbara A. Lyons, Jörg Pfeifer, Thomas H. Peterson, and Barry K. Carpenter*

Contribution from the Department of Chemistry, Baker Laboratory, Cornell University, Ithaca, New York 14853-1301. *Received September 25, 1992*

Abstract: Kinetic studies on the thermal nitrogen extrusion from 2,3-diazabicyclo[2.2.1]hept-2-ene-*exo,exo-5,6-d*$_2$ are reported. The ratio of rate constants for formation of the label-isomeric products (bicyclo[2.1.0]pentane-*exo,exo-2,3-d*$_2$ and *-endo,endo-2,3-d*$_2$) is found to exhibit no statistically significant temperature dependence. A comparison of gas-phase and solution-phase results is presented. The results are interpreted in terms of two complementary dynamic models. In the first, classical trajectory calculations are run on a three-dimensional projection (two geometric coordinates) of the potential energy hypersurface. These calculations correctly identify the major product, and reproduce the near temperature-independence of the rate-constant ratio, but do not match the ratio quantitatively. Modification of the trajectory calculations to simulate the effect of collisions with solvent molecules also qualitatively matches the observed difference between gas-phase and solution-phase behavior. In the second model, the vector of atomic displacements corresponding to the reaction coordinate at the transition state for nitrogen loss is identified by both semiempirical and ab initio calculations. The components of this vector pointing along the paths to the post-transition-state minima are computed and are shown to lead to a prediction of the product ratio that is in good (if partly fortuitous) agreement with the experimental result. The vector model is used to predict isotope effects on the product ratio, which are then investigated experimentally with 2,3-diazabicyclo[2.2.1]hept-2-ene-*endo,endo-1,4,5,6,7,7-d*$_6$ and *-endo-7-d*.

Introduction

The thermal deazetization of 2,3-diazabicyclo[2.2.1]hept-2-enes has been a source of mechanistic fascination ever since the report by Roth and Martin that the deuterium-labeled parent structure (compound **1** in Scheme I) underwent the reaction in the gas phase with a preference for inversion of configuration, i.e., giving the exo-labeled bicyclo[2.1.0]pentane **2x** in preference to the endo-labeled isomer **2n**.[1]

0002-7863/93/1515-2427$04.00/0