

## REACTIONS GRAPHS\*

Vladimír KVASNIČKA<sup>a</sup>, Milan KRATOCHVÍL<sup>b</sup> and Jaroslav KOČA<sup>c</sup><sup>a</sup> Department of Mathematic,

Slovak Institute of Technology, 812 37 Bratislava,

<sup>b</sup> Research Institute of Pure Chemicals, Lachema, 621 33 Brno and<sup>c</sup> Department of Chemistry, Faculty of Education,

J. E. Purkyně University, 662 80 Brno

Received October 25th, 1982

The reaction graphs for systematic and unambiguous representation of classes of reactions are suggested. This concept offers very simple and transparent method to analyze the complicated reaction schemes from the standpoint of their mechanistic realization. It is demonstrated that the reaction graphs can be used also for the canonisation of chemical reactions.

A mathematical model of organic chemistry should cover the following two aspects: First, the static aspect of organic chemistry, *i.e.* the description of molecular topology. In our graph-theory approach<sup>1-2</sup>, the molecules are represented by multi-graphs with loops, where the vertices are evaluated by atomic symbols. Second, the dynamic aspect corresponds to those phenomena that are called the chemical reactions (or transformations). A molecular system (composed of one or more molecules) is transformed onto another molecular system, both these systems are represented by isomeric molecular graphs<sup>1</sup>. Initially, we have introduced the so-called reaction operator<sup>1</sup>, its action on a substrate molecular graph produces a product molecular graph. Its matrix representation is closely related to the reaction matrix of Dugundji and Ugi model of constitutional chemistry<sup>3</sup>.

The purpose of the present communication is to suggest the reaction graphs (RGs). In fact, this concept is nothing else than a graph-theory representation of the above-mentioned reaction matrices. In contrast to the original reaction-matrix method, the RGs are of great heuristic and classification importance. This is obvious, the clumsy considerations with matrices are effectively substituted by the considerations with the corresponding graphs with very simple and transparent "topology". Moreover, the RGs make possible to canonize the chemical reactions and their directions. Recently, the canonisation problem of reaction matrices was studied by Brandt and coworkers<sup>4</sup>. Their canonisation process is relatively very time-con-

\* Part III in the series Mathematical Model of Organic Chemistry; Part II: This Journal 48, 2118 (1983).

suming, it needs to check all the permutations whether or not a permutation produces a string classified as greatest. Our approach is fully based on the graph-theory concepts and notions. The string notation of RGs simply follows from the existence of the so-called Euler alternating walks, their total permissible number is much smaller than the number of all column-row permutations in the reaction matrix.

The graph theory has been used in the enumeration problem of chemical reactions by Balaban and coworkers<sup>5</sup>, Sinanoglu<sup>7</sup>, and Nemes and coworkers<sup>6</sup>.

### REACTION GRAPHS

Let us have a pair of isomeric<sup>1</sup> molecular graphs (MGs)  $G_1 = (V, E_1, L_1, \varphi, \mathfrak{B})$  and  $G_2 = (V, E_2, L_2, \varphi, \mathfrak{B})$ . Formally, postulate that the MG  $G_1$  ( $G_2$ ) corresponds to an educt (product) molecular system, then we would like to treat within the pure graph-theory formalism the following "chemical" transformation process.



Applying the concept of the sum of graphs (see Appendix), these two MGs may be mutually related as follows

$$G_1 + \tilde{G} = G_2, \quad (2)$$

where the graph  $\tilde{G}$  is determined by (A4) and (A5). The edges and loops from  $\tilde{G}$  are evaluated with the respect to the transformation (1) by the surjective mapping  $\psi$ ,

$$\psi : E \rightarrow \{-1, +1\}, \quad (3a)$$

$$\psi : L \rightarrow \{-1, +1\}, \quad (3b)$$

*i.e.* each edge/loop for  $\tilde{G}$  is evaluated by the integer +1 and -1, respectively. In order to specify the mapping  $\psi$  we use the following defining relations

$$\psi(e) = \begin{cases} -1 & (\text{if } e \in E_1), \\ +1 & (\text{if } e \in E_2), \end{cases} \quad (4a)$$

$$\psi(l) = \begin{cases} -1 & (\text{if } l \in L_1), \\ +1 & (\text{if } l \in L_2). \end{cases} \quad (4b)$$

This has been done in accordance with the usual convention<sup>1</sup>, an edge/loop annihilated (created) in a course of the process (1) is evaluated by the integer -1 (+1). Hence, the reaction graph assigned to the process (1) is determined as an ordered

5-tuple

$$R = (\tilde{V}, \tilde{E}, \tilde{L}, \psi, \{-1, +1\}), \quad (5a)$$

where

$$\begin{aligned} \tilde{E} &= E_1 + E_2, \\ \tilde{L} &= L_1 + L_2, \end{aligned} \quad (5b)$$

and  $\tilde{V} \subseteq V$  is composed of those vertices that are incident with edges/loops from  $\tilde{E}$  and  $\tilde{L}$ , respectively. Finally, the process ( $I$ ) is now represented by

$$G_1 + R = G_2. \quad (6)$$

The reverse transformation process

$$G_2 \Rightarrow G_1 \quad (7)$$

is expressed by a similar graph-theory "equation"

$$G_2 + \bar{R} = G_1, \quad (8)$$

where the reaction graph  $\bar{R}$  is almost identical with the original  $R$ , but now the corresponding mapping  $\bar{\psi}$  is simply determined by  $\bar{\psi} = -\psi$ . The reaction graph  $\bar{R}$  is determined as the ordered 5-tuple

$$\bar{R} = (\tilde{V}, \tilde{E}, \tilde{L}, \bar{\psi}, \{-1, +1\}). \quad (9)$$

The RGs  $R$  and  $\bar{R}$  are different only in the evaluation of edges and loops, these entities are evaluated in  $\bar{R}$  by the opposite sign interchanges already used in  $R$ .

The reaction matrix<sup>1</sup> of a reaction graph  $\mathbf{R}$  is a square symmetric matrix  $\mathbf{R}$ , the rows and columns of which are labeled by the vertices of the RG, the element being determined by

$$r_{ii} = (\text{the multiplicity}^1 \text{ of the loop } [i, i]) \times \psi([i, i]), \quad (10a)$$

$$r_i^j = (\text{the multiplicity of the edge } [i, j]) \times \psi([i, j]). \quad (10b)$$

This relatively abstract definition of entries of  $\mathbf{R}$  has the following simple interpretation

$r_{ii}$  = the number of loops created ( $r_{ii} < 0$ )  
 [annihilated ( $r_{ii} < 0$ )] incident with  
 the  $i$ -th vertex,

$r_i'$  = the number of edges created ( $r_i' < 0$ )  
 [annihilated ( $r_i' < 0$ )] incident with  
 the  $i$ -th and  $j$ -th vertices.

Now, using the concept of the reaction matrix, we give the second alternative determination of the RGs  $R$  and  $\bar{R}$ ,

$$R = (\bar{V}, \mathbf{R}) \quad \text{and} \quad \bar{R} = (V, {}^-\mathbf{R}), \quad (11)$$

where the reaction matrices  $\mathbf{R}$  and  ${}^-\mathbf{R}$  are simply related by

$${}^-\mathbf{R} = -\mathbf{R}. \quad (12)$$

Since the MGs  $G_1$  and  $G_2$  are isomeric, the total numbers of created and annihilated edges/loops in the corresponding RG are exactly balanced, this requirement immediately implies

$$M = P = 2n, \quad (13a)$$

$$M^{(+)} + P^{(+)} = M^{(-)} + P^{(-)} = n, \quad (13b)$$

for an integer  $n = 1, 2, \dots$ . The entries  $M$  and  $P$  in (13a) are the total number of edges and loops, respectively, in the reaction graph  $R$ ,

$$M = |\tilde{\mathcal{E}}| \quad \text{and} \quad P = |\tilde{\mathcal{L}}|, \quad (14)$$

and the relations

$$M = M^{(+)} + M^{(-)}, \quad (15a)$$

$$P = P^{(+)} + P^{(-)}, \quad (15b)$$

are their splitting into the parts of edges/loops with  $+1$  and  $-1$  evaluations, respectively.

The individual vertices appearing in RGs assigned to a "realistic" chemistry are presented in Fig. 1. The number of edges (loops) incident with a given vertex is denoted by  $n_e$  ( $n_l$ ), this integer is called the edge (loop) degree (related to the vertex).

The total degree  $n = n_e + n_l$  of the vertex can be simply split as  $n^{(+)} + n^{(-)}$ , where  $n^{(+)}$  and  $n^{(-)}$  correspond to the total number of edges and loops evaluated by +1 and -1, respectively. In our "realistic" chemistry we shall consider only those vertices that are satisfying the following two restrictions (Fig. 1)

$$1 \leq n \leq 4, \quad (16a)$$

$$0 \leq |n^{(+)} - n^{(-)}| \leq 1. \quad (16b)$$

We consider only the vertices (atoms) for which, at most, two edges (bonds) are created and/or annihilated.

In general, an arbitrary graph must contain either none or an even number of vertices of the odd edge degree. The same should be satisfied also for the RGs, these vertices of the odd degree are appearing in chemical transformations in which the charged atoms contribute in a formation/break of bonds with simultaneous change of their valences. Therefore, we restrict the possible form of RGs: The reaction graphs are composed either of none or, at most, two vertices of the odd edge degree. For our forthcoming considerations will be very worthwhile to discuss only the chemical transformations that are expressed by connected RGs. It means, we study only the chemical transformations of two isomeric MGs for which it is impossible to separate this process in two or more elementary processes. In the opposite case, if the RG is disconnected (*i.e.* it contains two or more components that are not linked by an edge), the transformation process is separated in a class of independent transformations.

Summarizing, the RGs corresponding to "realistic" chemical transformations

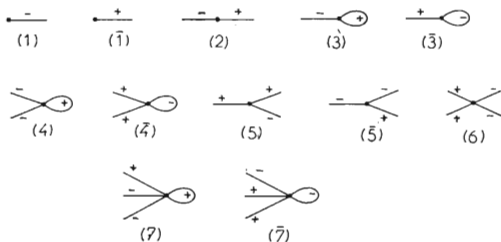


FIG. 1

The vertices of a reaction graph that are satisfying the conditions (16a) and (16b)

satisfy the following two requirements, where the first one is postulated to achieve a considerable simplification of the present theory (but still covering both real and important chemical transformation), and the second one means that we merely study the chemical transformations between two isomeric MGs:

- (i) A reaction graph is connected and composed of only vertices listed in Fig. 1, where either none or, at most, two vertices of the odd edge degree are used.
- (ii) A reaction graph contains an even total number of edges and loops [Eq. (13a)], and furthermore, the total numbers of edges and loops evaluated by  $-1$  and  $+1$ , respectively, are equal [Eq. (13b)].

In the forthcoming discussions under the term "reaction graph" we shall automatically understand a RG satisfying both the requirements.

In order to suggest a proper and transparent classification scheme of RGs we introduce the following concept: An Euler alternating walk (EAW) on a RG is a sequence of subsequently incident edges/loops with alternating edge/loop evaluation (for uniqueness we postulate that the first edge/loop is evaluated by  $-1$ ), and it contains all the edges and loops exactly ones. In a closed EAW the initial and terminal vertices coincide, an open EAW is not closed. An EAW is represented by oriented walk (this is achieved by requiring that the first edge and loop, respectively, must be evaluated by the integer  $-1$ ) which visits all the edges and loops exactly ones, the vertices may be intersected by the walk, in general, more than one. The following very important property of RGs is satisfied<sup>8</sup>:

*Theorem.* (1) A RG contain a closed EAW if and only if it contains the vertices of even edge degree.

(2) A RG contains an open EAW if and only if it contains exactly two vertices of odd edge degree, the EAW then starts and ends at these two vertices.

In general, one can obtain for a RG many different EAWs, we shall solve the very serious problem to canonize these alternative possibilites. Simultaneously, we canonize also the direction of the chemical transformation (1) and (7), that is we formally classify the direction in these two alternative possibilities (1) and (7) as the canonical direction.

Let us have a vocabulary  $\mathfrak{B}$  composed of fully ordered symbols, for simplicity, this vocabulary is identified with some part of Roman alphabet,

$$\mathfrak{B} = \{I < J < K < L \dots\}. \quad (17)$$

An arbitrary EAW is unambiguously expressed *via* a finite string of symbols belonging to the vocabulary (17). As was mentioned above, an EAW is uniquely determined as a sequence of all edged and loos, these edges/loops are subsequently inciden-

mined as a sequence of all edged and loops, these edges/loops are subsequently is with their common vertices. Going successively through all these vertices we get a sequence of vertices unambiguously determining the given EAW. The obtained sequence of vertices may be alternatively expressed by a string constructed over the vocabulary  $\mathfrak{B}$ , it contains  $2n + 1$  symbols. In particular, traversing the EAW, the vertices are subsequently labeled by the symbols from the vocabulary  $\mathfrak{B}$ , the first vertex is labeled by the symbol  $I$ , the second (if it is still unlabeled) by  $J$ , and so on. If we during this process meet a vertex already labeled, we use in the formation of string its label assigned in the preceding stage of traversing. The resulting string of symbols

$$S = M_1 M_2 \dots M_{2n+1}, \quad (18)$$

where  $M_1, M_2, \dots, M_{2n+1} \in \mathfrak{B}$ , unambiguously determines the given EAW. The pair  $M_1 M_2$  determined the first edge (or loop if  $M_1 = M_2$ ) evaluated by  $-1$ , the next pair  $M_2 M_3$  determines the second edge/loop evaluated by  $+1$ , and finally, the last pair  $M_{2n} M_{2n+1}$  determines the last edge/loop of the given EAW evaluated by  $+1$ . If  $M_1 = M_{2n+1}$ , then the EAW is closed, in the opposite case (*i.e.*  $M_1 \neq M_{2n+1}$ ) the EAW is open.

Let us denote all the possible strings that can be formed for a RG as  $S_1, S_2, \dots, S_p$ . Here we have to note that not all of these strings corresponds to different EAWs on the given RG. In particular, if an EAW is closed, then we have usually few possibilities in choosing the first vertex, a closed EAW may be represented by few different strings. Hence, in order to get a canonical indexing (or labeling) by the symbols of  $\mathfrak{B}$  of the RG, we postulate that the canonical indexing of the reaction graph is realized by an EAW which produces the smallest string (let us have two string  $S = M_1 M_2 \dots M_{2n+1}$  and  $S' = M'_1 M'_2 \dots M'_{2n+1}$ , if  $M_j < M'_j$  for an index  $1 < j < 2n + 1$  and  $M_k = M'_k$  for all  $k = 1, 2, \dots, j - 1$ , then we say that the string  $S$  is smaller than  $S'$ , formally  $S < S'$ , this process is called the lexicographical ordering of strings). In the case when we have obtained two or more smallest identical strings for different indexing of vertices, it is obvious that this is caused by an existence of topological equivalence among the vertices forming a subset of  $\bar{V}$ . Therefore, for highly symmetric RGs the canonical indexing of its vertices should be accompanied with the proper permutations of these indices, applying such a permutation to a fixed indexing of the RG we get another indexing which is determined by the same smallest string as the first one.

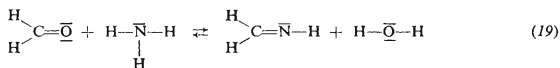
Now we turn our attention to the problem to canonize the direction of chemical transformations. Let a chemical transformation (1) be determined by a RG  $R$ , its reverse counterpart (7) is determined by  $\bar{R}$ . The canonical indexing of  $R$  ( $\bar{R}$ ) is defined by the smallest string  $S$  ( $\bar{S}$ ). Then the canonical direction is determined in accordance with the property which transformation (1) or (2) produces a smaller string. In parti-

cular, if  $S < \bar{S}$  ( $\bar{S} < S$ ), we postulate that the transformation  $G_1 \Rightarrow G_1$ ) corresponds to the canonical direction. Getting  $S = \bar{S}$ , we say that both these directions are canonically equivalent.

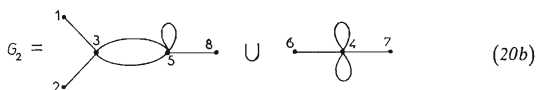
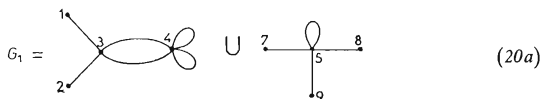
The most frequently appearing chemical reactions<sup>9</sup> in the canonical form are summarized in Table I, the corresponding RGs are listed in Fig. 2 to 5.

### Illustrative Example

We shall study the following simple model chemical transformation (classified as the addition-elimination reaction in which the carbonyl group is replaced by the imino group)



The corresponding MGs are



where we have used a trial indexing of vertices. The RGs  $R$  and  $\bar{R}$  ascribed to the transformations  $G_1 \Rightarrow G_2$  and  $G_2 \Rightarrow G_1$ , respectively, are determined „diagrammatically” as

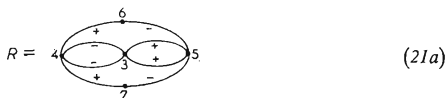




TABLE I  
Canonical reactions

Graph	String	Reaction	Automorphisms	Other notations	
				<i>cf. ref.</i> <sup>9</sup>	<i>cf. ref.</i> <sup>11</sup>
Reaction graphs involving 2 or 3 vertices ( <i>cf. Fig. 2</i> )					
2-1	IIJ	$I + J \Rightarrow I - J$	(IJ)	$\bar{R}_{23}$	$R_1$
3-1	IJK	$I - J + K \leftrightarrow I + J - K$	(IJK)	$\bar{R}_{11}$	$R_2$
3-2	IIJKK	$\bar{I} + J - K \Rightarrow I - J + \bar{K}$	(IJK)	$\bar{R}_{12}$	$R_3$
3-3	IIJKI	$\bar{I} + J - K \Rightarrow J - I - K$	(IJK), (IKJ)	$\bar{R}_{21}$	$R_T$
3-4	IIJKKJI	$\bar{I} + J = K \leftrightarrow I = J + \bar{K}$	(IJK)	$\bar{R}_{25}$	
Reaction graphs involving 4 vertices ( <i>cf. Fig. 3</i> )					
4-1	IJKLI	$I - J + K - L \leftrightarrow I - L + J - K$	(IJKL), (IJLK) (KLIJ), (LKJI)	$\bar{R}_1$	$R_8$
4-2	IIJKLI	$\bar{I} + \bar{J} + K - L \Rightarrow L - I - J - K$	(IJKL), (IJLK)		
4-3	IIJKKLI	$\bar{I} + J - K - L \leftrightarrow J - I - L + \bar{K}$	(IJKL), (IJLK)		
4-4	IIJKL	$\bar{I} + J - K + L \Rightarrow I - J + K - L$	(IJKL), (ILKJ)	$\bar{R}_3$	
4-5	IJKIL	$J - I - K + L \Rightarrow J - K + I - L$	(IJKL), (IKJL)	$\bar{R}_{20}$	
4-6	IIJKILL	$\bar{I} - L + J - K \Rightarrow J - I - K + \bar{L}$	(IJKL)	$\bar{R}_5, \bar{R}_{43}$	
4-7	IIJKILJJK	$\bar{I} - L + J = K \Rightarrow \bar{L} - J - I - K$	(IJKL)	$\bar{R}_{14}$	$\bar{R}_{14}$
4-8	IIJKKJILL	$\bar{I} - L + J = K \Rightarrow I = J + \bar{K} + \bar{L}$	(IJKL)	$\bar{R}_{36}$	
4-9	IJKLIJKLI	$I = J + K = L \leftrightarrow I = L + J = K$	(IJKL), (KLIJ)	$\bar{R}_{42}$	

Reaction graphs involving 5 vertices (cf. Fig. 4)

5-1	<i>I</i> <i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> <i>J</i>	$I + J - K + L - M \Rightarrow I - J - M + K - L$	( <i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> )	$R_9$
5-2	<i>I</i> <i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> <i>M</i>	$I + J - K + L - M \Rightarrow I - J + K - L + \bar{M}$	( <i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> )	$R_{12}$
5-3	<i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>J</i> <i>K</i> <i>M</i> <i>I</i>	$I = J + L - K - M \rightarrow J = K + L - I - M$	( <i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> ), ( <i>I</i> <i>J</i> <i>K</i> <i>M</i> <i>L</i> )	$R_{10}$
5-4	<i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>J</i> <i>K</i> <i>M</i> <i>I</i>	$L - J - I - K - M \Rightarrow L - I - M + K - N$	( <i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> ), ( <i>I</i> <i>K</i> <i>J</i> <i>M</i> <i>L</i> )	$R_6$
5-5	<i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> <i>I</i>	$J - I - K + L - M \Rightarrow L - I - M + J - K$	( <i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> ), ( <i>I</i> <i>K</i> <i>J</i> <i>M</i> <i>L</i> )	$R_7$
5-6	<i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> <i>I</i>	$I + J - K + L - M \Rightarrow M - I - J + K - L$	( <i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> ), ( <i>I</i> <i>M</i> <i>L</i> <i>K</i> <i>J</i> )	$R_{10}$
5-7	<i>I</i> <i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>J</i> <i>M</i> <i>M</i>	$I - M + K - J - L \Rightarrow I = J + K - L + \bar{M}$	( <i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> ), ( <i>I</i> <i>K</i> <i>L</i> <i>K</i> <i>M</i> )	$R_{37}$

Reaction graphs involving 6 vertices (cf. Fig. 5)

6-1	<i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> <i>N</i> <i>I</i>	$I - J + K - L + M - N \leftrightarrow I - N + J - K + L - M$	( <i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> <i>N</i> ), ( <i>J</i> <i>I</i> <i>N</i> <i>M</i> <i>L</i> <i>K</i> ) ( <i>K</i> <i>L</i> <i>M</i> <i>N</i> <i>J</i> ), ( <i>L</i> <i>K</i> <i>J</i> <i>I</i> <i>N</i> <i>M</i> )	$R_{11}$
6-2	<i>I</i> <i>I</i> <i>J</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> <i>N</i> <i>I</i>	$I + \bar{J} + K - L + M - N \Rightarrow N - I - J - K + M - L$	( <i>M</i> <i>N</i> <i>I</i> <i>J</i> <i>K</i> <i>L</i> ), ( <i>N</i> <i>M</i> <i>L</i> <i>K</i> <i>J</i> <i>I</i> ) ( <i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> <i>N</i> ), ( <i>J</i> <i>I</i> <i>N</i> <i>M</i> <i>L</i> <i>K</i> )	$R_2$
6-3	<i>I</i> <i>J</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> <i>N</i> <i>I</i>	$I + J - K - L + M - N \Rightarrow N - I - J + \bar{K} + L - M$	( <i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> <i>N</i> )	
6-4	<i>I</i> <i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>L</i> <i>M</i> <i>N</i> <i>I</i>	$I + J - K + \bar{L} + M - N \Rightarrow N - I - J + K = L - M$	( <i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> <i>N</i> ), ( <i>I</i> <i>N</i> <i>M</i> <i>L</i> <i>K</i> <i>J</i> ) ( <i>L</i> <i>K</i> <i>J</i> <i>I</i> <i>N</i> <i>M</i> ), ( <i>L</i> <i>M</i> <i>N</i> <i>I</i> <i>J</i> <i>K</i> )	
6-5	<i>I</i> <i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> <i>N</i> <i>N</i>	$I - N + J - K + L - M \Rightarrow J - I - M + K - L + \bar{N}$	( <i>I</i> <i>J</i> <i>L</i> <i>K</i> <i>M</i> <i>N</i> ), ( <i>I</i> <i>M</i> <i>L</i> <i>K</i> <i>J</i> <i>N</i> )	$\bar{R}_{13}, \bar{R}_{15}$
6-6	<i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> <i>N</i> <i>L</i> <i>I</i>	$J - I - K + M - L - N \Rightarrow I = L + J - K + M - N$	( <i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> <i>N</i> ), ( <i>I</i> <i>K</i> <i>J</i> <i>L</i> <i>M</i> <i>N</i> ) ( <i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>N</i> <i>M</i> ), ( <i>I</i> <i>K</i> <i>J</i> <i>L</i> <i>N</i> <i>M</i> ) ( <i>L</i> <i>M</i> <i>N</i> <i>I</i> <i>J</i> <i>K</i> ), ( <i>L</i> <i>M</i> <i>N</i> <i>I</i> <i>K</i> <i>J</i> ) ( <i>L</i> <i>N</i> <i>M</i> <i>I</i> <i>J</i> <i>K</i> ), ( <i>L</i> <i>N</i> <i>M</i> <i>I</i> <i>K</i> <i>J</i> ) ( <i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> <i>N</i> )	$\bar{R}_{16}$ $\bar{R}_3, \bar{R}_{25}$
6-7	<i>I</i> <i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> <i>N</i>	$I + J - K + L - M + N \Rightarrow I - J + K - L + M - N$	( <i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>M</i> <i>N</i> ), ( <i>J</i> <i>I</i> <i>M</i> <i>N</i> <i>K</i> <i>L</i> ) ( <i>J</i> <i>I</i> <i>M</i> <i>N</i> <i>L</i> <i>K</i> ), ( <i>J</i> <i>I</i> <i>L</i> <i>K</i> <i>M</i> <i>N</i> )	$\bar{R}_{31}$
6-8	<i>I</i> <i>J</i> <i>K</i> <i>L</i> <i>J</i> <i>M</i> <i>N</i> <i>I</i>	$I = J + K - L + M - N \Rightarrow L - I - N + K - J - M$		

Using these RGs, the transformations  $G_1 \Rightarrow G_2$  and  $G_2 \Rightarrow G_1$  are expressed by the „equations”

$$G_1 + R = G_2, \quad \bar{R} \quad (22a)$$

$$G_2 + \bar{R} = G_1. \quad (22b)$$

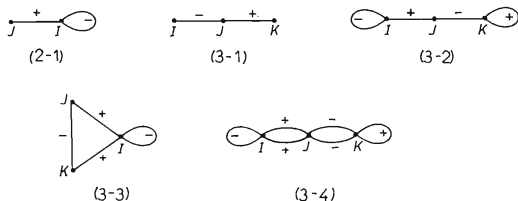


FIG. 2

The reaction graphs involving 2 and 3 vertices (Table I)

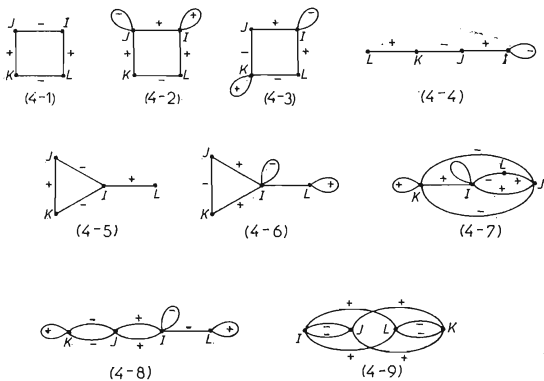


FIG. 3

The reaction graphs involving 4 vertices (Table I)

The reaction matrices  $R$  and  ${}^{-}R$  corresponding to the RGs  $R$  and  $\bar{R}$  are simply determined as square symmetric matrices the rows and columns of which are labeled

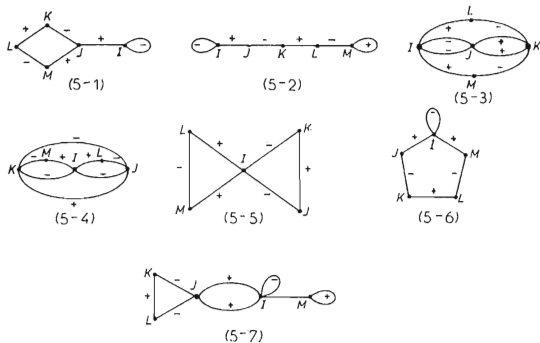


FIG. 4

The reaction graphs involving 5 vertices (Table I)

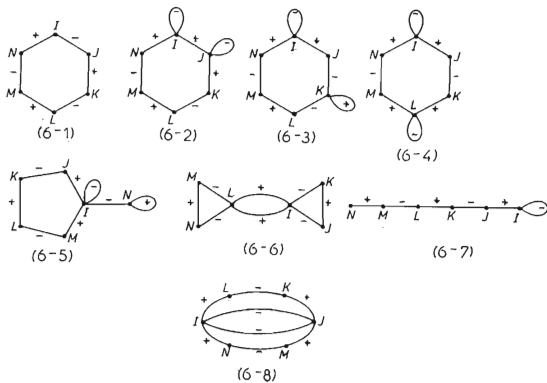


FIG. 5

The reaction graphs involving 6 vertices (Table I)

by the integer labels use in (21a) and (21b)

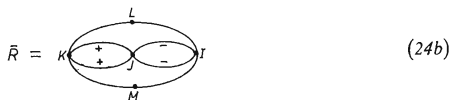
$$R = \begin{matrix} & 3 & 4 & 5 & 6 & 7 \\ \begin{pmatrix} 0 & -2 & 2 & 0 & 0 \\ -2 & 0 & 0 & 1 & 1 \\ 2 & 0 & 0 & -1 & -1 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \end{pmatrix} & \begin{matrix} 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix} \end{matrix} \quad (23a)$$

and

$${}^{-}R = -R. \quad (23b)$$

Since the atoms indexed by 6 and 7 are topologically equivalent, we can construct from the reaction matrix  $R$   $5!/2! = 60$  other reaction matrices by  $R' = P^T R P$ , where  $P$  is a symmetric matrix corresponding to a permutation of indices (3, 4, 5, 6, 7), the permutations differing only in a transposition of the indices 6 and 7 are taken as equivalent. All the matrices  $R, R', R'', \dots$  correspond to the RGs that are mutually isomorphic. Therefore, it is highly desirable to have in the mathematical model of organic chemistry a formal tool (algorithm) to be carried out a canonical indexing of the RG, and simultaneously, to determine a canonical direction of the transformation of two isomeric MGs  $G_1$  and  $G_2$  (i.e. to decide whether  $G_1 \Rightarrow G_2$  or  $G_2 \Rightarrow G_1$  is classified as the canonical direction). In order to solve these two very serious problems we shall use the method described in the previous section.

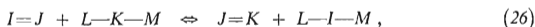
Since the RGs corresponding to standard and most frequently appearing chemical reactions are described by the RGs with relatively uncomplicated "topology", we can get the canonisation of the RGs by making use of their simple visual inspection (Fig. 6). In general, mainly for the computer implementation of this problem, we use an algorithm to construct the Euler walk on a graph. The chosen algorithm should be slightly modified, we construct the EAW. The canonical form of RGs (21a) and (21b) is



We see that these graphs are fully topologically equivalent [e.g. the graph (24a) may be constructed from the graph (24b) if the vertices  $I$  and  $K$  are mutually replaced]. The corresponding smallest string which determines the EAWs must be equivalent,

$$S = \bar{S} = IJKLIJKMI. \quad (25)$$

Hence, we conclude that the chemical transformation described by the RG  $R$  has no preferred direction, or in other words, both the directions are canonically equivalent. Following the RG (24a), the assigned canonical chemical reaction is



where the double-side arrow indicates that both the directions are canonically equivalent. Finally, the canonical form of the reaction matrix  $R$  is

$$R = \begin{pmatrix} I & J & K & L & M \\ \left( \begin{array}{ccccc} 0 & -2 & 0 & 1 & 1 \\ -2 & 0 & 2 & 0 & 0 \\ 0 & 2 & 0 & -1 & -1 \\ 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \end{array} \right) & \begin{array}{l} I \\ J \\ K \\ L \\ M \end{array} \end{pmatrix} \quad (27)$$

This matrix may be simply constructed by using the string (25). Let  $M_1M_2$  be a pair of juxtaposed symbols of the string  $S$ , its position (reading from the left to right) is denoted by  $p \in \langle 1, 2n \rangle$ . Then the vertices  $M_1$  and  $M_2$  are linked by an edge (or loop if  $M_1 = M_2$ ) evaluated by  $(-1)^p$ .

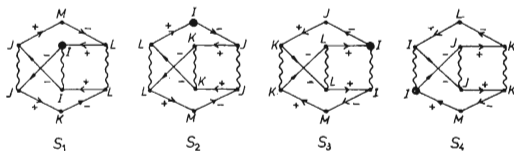


FIG. 6

The different EAWs for the reaction graph  $R_{5-3}$  (Fig. 4), only the last one gives the smallest string  $S_4$ . In order to visualize the EAWs we have split the vertices composed of four edges (vertex 6 in Fig. 1) into two subvertices connected by a vertical wavy line. The heavy dot (always indexed by  $I$ ) represents the starting vertex for a given EAW  $S_1 = IJKLIJMLI$ ,  $S_2 = IJKLMJ \cdot KLI$ ,  $S_3 = IJKLIMKLI$ ,  $S_4 = IJKLIJKMI$ .

Formally, let us denote the canonical reaction matrix (27) as  $\mathbf{R}(I, J, K, L, M)$ , we have explicitly displayed the order of labels. Using this matrix, the reaction matrix (23a) is determined by

$$\mathbf{R} = \mathbf{P}^T \mathbf{R}(I, J, K, L, M) \mathbf{P}, \quad (28a)$$

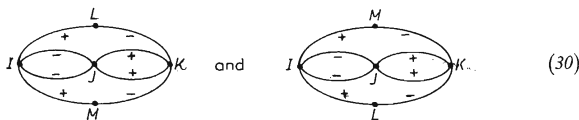
where  $\mathbf{P}$  is a permutation square matrix corresponding to the permutation

$$P = \begin{pmatrix} I & J & K & L & M \\ J & I & K & L & M \end{pmatrix}. \quad (28b)$$

One can see, from the diagram (24a), that the RG  $R$  has two topologically equivalent vertices<sup>1</sup>  $L$  and  $M$ . The original RG  $R(I, J, K, K, M)$  is fully equivalent with its counterpart  $R(I, J, K, M, L)$ . Using the graph-theory terminology<sup>1</sup>, the RG  $R(I, J, K, L, M)$  has two automorphisms realized by the following two permutations

$$P_1 = \begin{pmatrix} I & J & K & L & M \\ I & J & K & L & M \end{pmatrix} \quad \text{and} \quad P_2 = \begin{pmatrix} I & J & K & L & M \\ I & J & K & M & L \end{pmatrix} \quad (29)$$

or diagrammatically



These two RGs have the same reaction matrix, in particular, we get

$$\begin{aligned} \mathbf{R}(I, J, K, L, M) &= \mathbf{R}(I, J, K, M, L) \\ &= \mathbf{P}_2^T \mathbf{R}(I, J, K, L, M) \mathbf{P}_2. \end{aligned} \quad (31)$$

In order to avoid possible misunderstandings we shall present together with the canonical form of the given RG also the permutation providing its automorphisms, see Table I [we use an abbreviated form of the permutations, *e.g.* the permutations (29) are presented as  $P_1 = (IJKLM)$  and  $P_2 = (IJKML)$ ].

#### DECOMPOSITION OF REACTION GRAPHS

In order to decompose a RG into more elementary RGs (*e.g.* to indicate its reaction mechanism *via* the possible mechanistic steps) that are mutually edge/loop disjoint

we use formally the sum operation extensively applied in preceding sections of this communication (see Appendix). Let us have a pair of RGs

$$R_1 = (\bar{V}, \bar{E}, \bar{L}_1, \psi_1, \{-1, +1\}), \quad (32a)$$

$$R_2 = (\bar{V}_2, \bar{E}_2, \bar{L}_2, \psi_2, \{-\gamma, +1\}). \quad (32b)$$

The operation sum of two RGs  $R_1$  and  $R_2$

$$R = R_1 + R_2 = (\bar{V}, \bar{E}, \bar{L}, \psi, \{-1, +1\}) \quad (33a)$$

is determined by

$$\bar{E} = \bar{E}_1 + \bar{E}_2 = E_1 \cup E_2, \quad (33b)$$

$$\bar{L} = \bar{L}_1 + \bar{L}_2 = L_1 \cup L_2, \quad (33c)$$

where we have to remember the RGs  $R_1$  and  $R_2$  are edge/loop disjoint [*i.e.*  $E_1 \cap E_2 = \emptyset$  and  $L_1 \cap L_2 = \emptyset$ , see comment below the relation (A3)]. The vertex set  $\bar{V}$  is composed of those vertices of  $\bar{V}_1$  and  $\bar{V}_2$  that are incident with edges/loops of  $\bar{E}$  and  $\bar{L}$ , respectively. Finally, the mapping  $\psi$  is

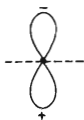
$$\psi(x) = \begin{cases} \psi_1(x) & (\text{for } x \in R_1), \\ \psi_2(x) & (\text{for } x \in R_2), \end{cases} \quad (33d)$$

where the term  $x$  expresses either an edge or loop of  $R_1$  and  $R_2$ , respectively.

The decomposition of a RG  $R$  denotes its splitting into other RGs, *e.g.*

$$R = R_1 + R_2 + \dots + R_p, \quad (34)$$

where the RGs on the right hand side are usually classified as the elementary reaction graphs (roughly speaking, these RGs are composed of two or, at most, three vertices). For better visualisation of the decomposition we have found that there is very useful to use a notion of the so-called virtual loops



(35)



Since they are presented as a pair of loops with the opposite evaluation, their actual impact on the resulting RG is completely cancelled.

As an illustrative example of the above concept we have listed in Fig. 7 a few alternative possibilities how to decompose the reaction graph  $R_{4-1}(I, J, K, L)$ ,

$$\begin{aligned}
 R_{4-1}(I, J, K, L) &= \bar{R}_{2-1}(I, J) + R_{2-1}(I, L) + \bar{R}_{2-1}(K, L) \\
 &\quad + R_{2-1}(K, J) \\
 &= R_{3-2}(K, J, I) + R_{3-2}(I, L, K) \\
 &\quad \bar{R}_{2-1}(J, I) + R_{2-1}(J, K) + R_{2-1}(L, I) \\
 &\quad \bar{R}_{2-1}(L, K) \\
 &= R_{3-2}(L, I, J) + R_{3-2}(J, K, L)
 \end{aligned} \tag{36}$$

and many other possibilities.

Let us consider a transformation  $G_1 \Rightarrow G_2$ , using the RG  $R$  this relation is "algebraized" as

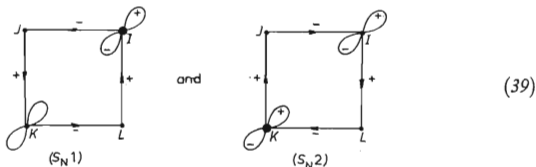
$$G_1 + R = G_2, \tag{37}$$

and furthermore, we assume that  $R$  is decomposed into its elementary components by (34). A particular reaction mechanism of  $G_1 \Rightarrow G_2$  may be expressed by making use of the following sequence

$$(\dots((G_1 + R_{\alpha_1}) + R_{\alpha_2}) + \dots + R_{\alpha_p}) = G_2, \tag{38}$$

where the indices  $(\alpha_1, \alpha_2, \dots, \alpha_p)$  are a permutation of the integers  $(1, 2, \dots, p)$ . Although the individual components in (34) are mutually commuting, their preselected ordering is of the great importance if we try to affect an actual chemical mechanism. The first elementary is determined by  $G_1 + E_{\alpha_1} = G_2^{(1)}$ , the second by  $G_2^{(1)} + R_{\alpha_2} = G_2^{(2)}$ , and finally, the last  $p$ -th step by  $G_2^{(p-1)} + R_{\alpha_p} = G_2^{(p)} = G_2$ . From the decomposition (34) we can form, in general,  $p!$  different mechanisms that are differing in a sequence of application of elementary RGs on the substrate graph  $G_1$ .

For example, let us consider the mechanisms  $S_{N1}$  and  $S_{N2}$  assigned to the RG  $R_{4-1}(I, J, K, L)$ ; the reaction graph can be presented in two alternative ways

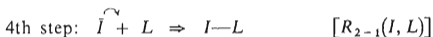
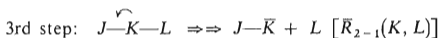
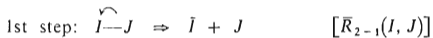


where the heavy dot denotes the starting vertex for the oriented path of elementary processes. Both the mechanism are expressed by sequence of elementary RGs, where the first one starts at the heavy dot.

*S<sub>N</sub>1 Mechanism:*

$$(((G_1 + \bar{R}_{2-1}(I, J)) + R_{2-1}(K, J)) + \bar{R}_{2-1}(K, L)) + R_{2-1}(I, L) = G_2 \quad (40a)$$

which corresponds to the sequence of four elementary chemical reactions.



*S<sub>N</sub>2 Mechanism:*

$$(((G_1 + R_{2-1}(Q, J)) + \bar{R}_2\bar{\psi}_1(I, J)) + R_2\bar{\psi}_1(I, L)) + \bar{R}_{2-1}(Q, L) = G_2 \quad (40b)$$

where

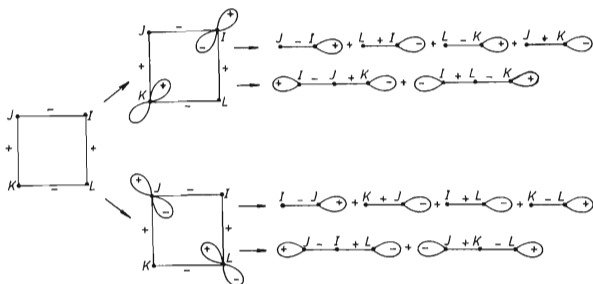
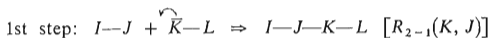
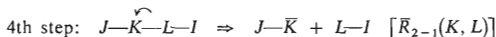
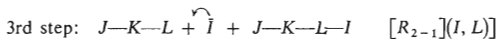
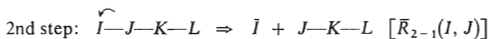
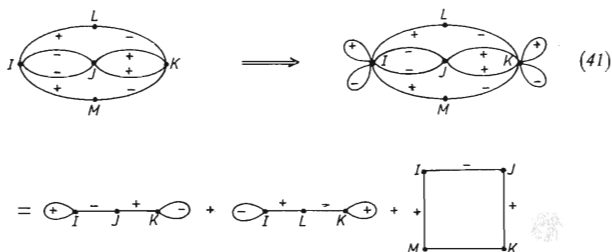


FIG. 7

The decomposition of the reaction graph  $R_{4-1}$  (Fig. 3). We use the concept of virtual loops



As another we shall study the RG (24a) (denoted by the label 5-3 in Fig. 4) corresponding to the addition-elimination reaction (19). Individual mechanistic steps of this reaction are assigned to the following decomposition of the RG  $R = R_{5-3}$ , (IJ, K, L, M)

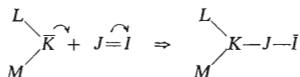


Algebraically

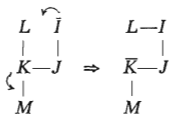
$$R_{5-3}(I, J, K, L, M) = R_{3-2}(K, J, I) + R_{3-2}(I, L, K) + R_{4-1}(J, I, K, M). \quad (42)$$

This decomposition of the RG  $R_{5-3}$  corresponds to three mechanistic steps.

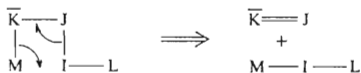
1st step: Application of  $R_{3-2}(K, J, I)$



2nd step: Application of  $R_{3-2}(I, L, K)$



3rd step: Application of  $R_{4-1}(J, I, K, M)$ , a pericyclic process involving 4-centre activated complex, gives



To conclude this section we have to stress, we are far from claiming that the described mechanisms are the only possible mechanistic explanations of the studied reaction operators.

#### APPENDIX

The  $sum^1 A + B$  (often called the disjunctive sum) of two sets  $A$  and  $B$  is defined as a set formed by the elements of  $A$  and  $B$  with the exception of those that belong at the same time to  $A$  and  $B$ . In more abstract form, using the logical operators "and" ( $\wedge$ ) and "or" ( $\vee$ ), the sum  $A + B$  is determined as follows

$$A + B = \{x; (x \in A \wedge x \notin B) \vee (x \notin A \wedge x \in B)\}. \quad (A)$$

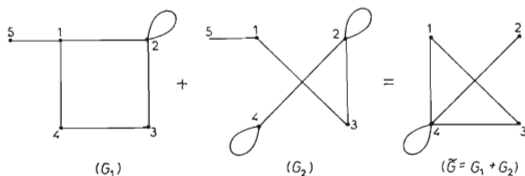


FIG. 8

An illustrative example of the concept of sum of two graphs

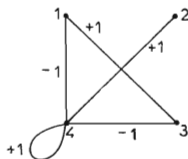


FIG. 9.

An illustrative example of the reaction graph assigned to the situation presented in Fig. 8

For example, for sets  $A = \{1, 2, 3, 4\}$  and  $B = \{3, 4, 5, 6\}$  we get  $A + B = \{1, 2, 5, 6\}$ . The sum of sets satisfies

$$A + B = B + A, \quad (A2a)$$

$$A + (B + C) = (A + B) + C, \quad (A2b)$$

$$A + A = \emptyset, \quad (A2c)$$

$$A + \emptyset = A, \quad (A2d)$$

where  $\emptyset$  is the empty set. A set "equation"  $A + X = B$  can be "solved" with respect to the set  $X$  as

$$X = A + B. \quad (A3)$$

We emphasize, if the sets  $A$  and  $B$  are disjoint ( $A \cap B = \emptyset$ ), then the sum  $A + B$  is simply equal to the union  $A \cup B$ , i.e.  $A + B = A \cup B$ , for  $A \cap B = \emptyset$ .

Let us consider a pair of isomeric graphs<sup>1</sup>  $G_1 = (V, E_1, L_1, \mathfrak{B})$ .

The sum of molecular graphs  $G_1$  and  $G_2$ , denoted by  $G_1 + G_2$ , is determined as

$$\tilde{G} = G_1 + G_2 = G_2 + G_1 = (\tilde{V}, \tilde{E}, \tilde{L}, \varphi, \mathfrak{B}), \quad (A4)$$

where the sets  $\tilde{E}$  and  $\tilde{L}$  are

$$\tilde{E} = E_1 + E_2, \quad (A5a)$$

$$\tilde{L} = L_1 + L_3, \quad (A5b)$$

and the resulting vertex set  $\tilde{V}$  is composed of only those vertices that are incident with edges and loops belonging to  $\tilde{E}$  and  $\tilde{L}$ , respectively. The concept of the sum of graphs is illustrated in Figs 8 and 9. By using the property (A3), the equation (A4) may be solved with respect to either the graph  $G_1$  or  $G_2$ , we arrive at

$$G_1 = G_2 + \tilde{G}, \quad (A6a)$$

$$G_2 = G_1 + \tilde{G}. \quad (A6b)$$

#### REFERENCES

1. Kvasnička V.: This Journal, in press.
2. Kvasnička V.: This Journal, in press.
3. Dugundji J., Ugi I.: Topics Curr. Chem. 39, 19 (1973).
4. Brandt J., Bauer J., Frank R. M., von Scholley A.: Chem. Scripta 18, 53 (1981).
5. Balaban A. T., Farcasiu D., Banica R.: Rev. Roumanine Chim. 11, 1225 (1966).
6. Sinanoglu O.: J. Amer. Chem. Soc. 97, 2309 (1975).
7. Nemes I., Vidoczy J., Botar L., Gal D.: Theor. Chim. Acta 45, 215, 225 (1977).
8. Euler L.: Comment. Acad. Petropolutan. 8, 128 (1736).
9. Bart J. C. J., Garagnani E.: Z. Naturforsch. 31b, 1646 (1976); 32b, 455, 465, 678 (1977).
10. Halmos P. R.: Naive Set Theory. Van Nostrand, Princeton 1960.
11. Kratochvil M., Koča J., Kvasnička V.: Chem. Listy, in press.