CLASSIFICATION SCHEME OF CHEMICAL REACTIONS*

Vladimír KVASNIČKA

Department of Mathematics,
Slovak Institute of Technology, 812 37 Bratislava

Received January 10th, 1983

The main concept of the suggested classification scheme of chemical reactions is the so-called reaction core. The reaction core is determined as an ordered triple, $\mathfrak{R} = (R, G, \varphi$, where R is a reaction graph, G is an intact molecular subgraph containing the edges and loops that are common for both substrate and product molecular graphs mutually related by the above reaction graph, and finally, φ is an evaluation of vertices from G by appropriate atomic symbols. An arbitrary chemical reaction, independently of its mechanistic stepwise realization, may be unambiguously characterized by this concept of reaction core.

In our recent communication[1] (part III of this series) we have introduced the concept of reaction graphs for systematic and unambiguous representation of huge classes of chemical reactions in organic chemistry. In short, a reaction graph is a digrammatic representation (closely related to the reaction matrices of Ugi and Dugundji[2]) of the electron-flow processes of an individual chemical transformation of some concrete substrate molecules into a resulting product molecules (both represented by molecular graphs). It was demonstrated by Bart and Garagnani[3] (see also the recent work of Arens[4]) that such a few electron-flow patterns (in our graph-theory approach represented by reaction graphs with great heuristic and classification value) are very useful for description of the almost all chemical transformations in organic chemistry. Unfortunately, the concept of reaction graphs offers very rough and vague information about the chemical transformations. Therefore, in order to get a more precize specification of chemical transformations we have to enlarge the concept of reaction graphs into the so-called reaction cores. A reaction core is a more deep specification of the relevant substructure in the chemically transformed molecular graphs. The pertinent reaction graph is accompanied by further necessary informations specifying the bonds and electron lone pairs that are remaining unchanged during the course of chemical transformation, and the chemical type of participating atoms.The similar ideas have been used in an intuitive form by Arens[4], Balaban[5], Hendrickson[6], Zefirov[7], Brandt and coworkers[8], and Roberts[10]. The purpose

* Part IV in the series Mathematical Model of Organic Chemistry; Part III: This Journal *48*, 2284 (1983).

of present communication is to give formally precize specification of the very serious concept of reaction cores and to outline its potential applicability to classify the chemical transformations in a way closely related to up-to-date methodological trends of organic chemistry. Perhaps, it may be necessary at this point to emphasize[9] that the formalism conveys no information concerning the mechanism of the reactions, *e.g.* it say nothing whether a reaction corresponds to a concerted or stepwise process. Furthermore, the suggested formalism also has no stereochemical implications.

THEORETICAL

Let us have a pair of isomeric[9] molecular graphs G'_1 and G'_2, and we shall study the following chemical transformation process

$$G'_1 \Rightarrow G'_2 \,. \tag{1}$$

Usually, these molecular graphs are different only in small parts of whole graphs, their environments for both used molecular graphs are identical. We propose that the molecular graphs G'_1 and G'_2 may be written as a union[9] of two edge/loop disjoint subgraphs,

$$G'_1 = \tilde{G} \cup G_1 \,, \tag{2a}$$

$$G'_2 = \tilde{G} \cup G_2 \,. \tag{2b}$$

Applying the concept of reaction graphs[1], the process (*1*) is "algebraized" by

$$G'_1 + R = G'_2 \,, \tag{3}$$

or, by making use of the decompositions (*2a*) and (*2b*), it may be simplified as follows

$$G_1 + R = G_2 \,. \tag{4}$$

Passing from (*3*) to (*4*) we have taken into an explicit account only those parts of G'_1 and G'_2 that are relevant for the transformation (*1*), the presentation of the "environment" subgraph $\tilde{G}$ was fully suppressed. In organic chemistry this formal procedure has the following simple counterpart: A chemical reaction is schematically determined only by the smallest skeleton of atoms which has to be specified for its unambiguous determination. Embedding the skeleton by an "environment" we arrive at an actual form of the given chemical reactions.

The reaction graph in (*4*) can be decomposed in two edge/loop disjoint subgraphs,

$$R = R_+ + R_- \,, \tag{5}$$

where the subgraphs $R_+(R_-)$ is composed of only those edges/loops of R that are evaluated by $+1(-1)$. It means, the reaction subgraph $R_+(R_-)$ is composed of the annihilated (created) edges and loops in the molecular subgraph $G_1(G_2)$. Introducing (5) in (4) and after simple algebraic manipulations (see Appendix in ref.[1]) we arrive at

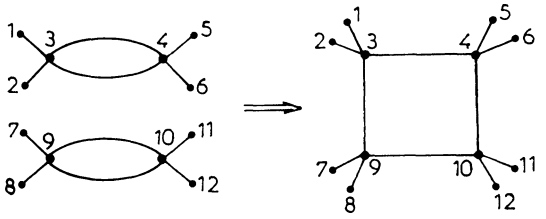
$$G_1 + R_- = G_2 + R_+ = G \,. \tag{6}$$

The resulting molecular subgraph G is composed of those edges and/or loops that are common for both subgraphs G_1 and G_2, in our forthcoming considerations it will be called the intact molecular subgraph[8]. If we have determined, *a priori*, the intact molecular subgraph G and a pertinent reaction graph R, then the subgraphs G_1 and G_2 are simply determined as

$$G_1 = G + R_- \,, \tag{7a}$$

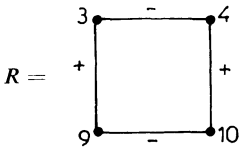
$$G_2 = G + R_+ \,. \tag{7b}$$

Example. We shall study the cycloaddition of two ethylene molecules, graphically

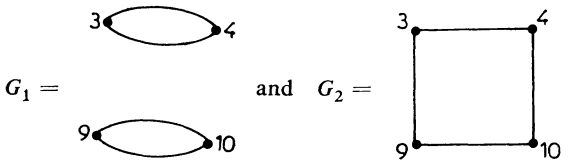


(G_1') (G_2')

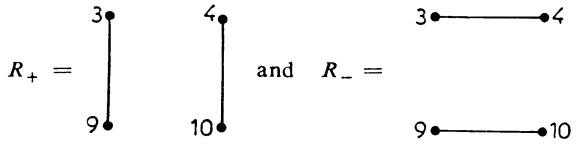
The corresponding reaction graph R is



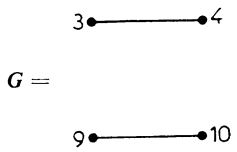
Removing the irrelevant environment (with respect to the reaction graph R) in G_1' and G_2' we arrive at the following two molecular subgraphs G_1 and G_2 that are composed of edges relevant for the studied cycloaddition transformation of two ethylene molecules,



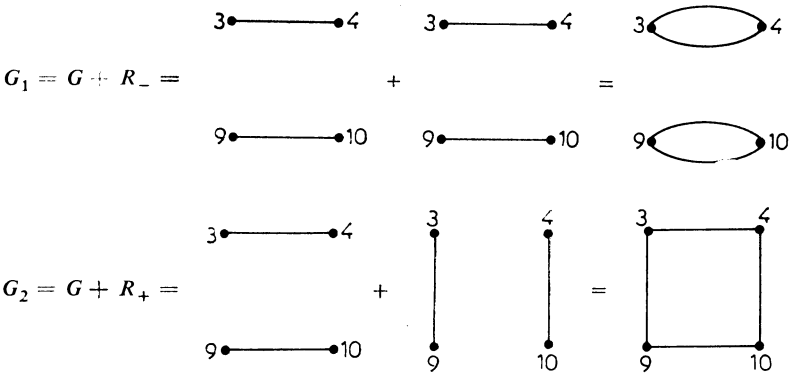
The decomposition of R into R_+ and R_- is



Then the intact molecular sugraph G is



The edges of G are common for the molecular subgraphs G_1 and G_2. Finally, the molecular subgraphs G_1 and G_2 are determined by



After this illustrative example of basic concepts and notions introduced above, we now turn our attention to the general determination of chemical reactions. Let us have a reaction graph[1]

$$R = (V, \tilde{E}, \tilde{L}, \psi, \{-1, +1\}) . \tag{8}$$

We introduce, with respect to this reaction graph, an intact molecular subgraph

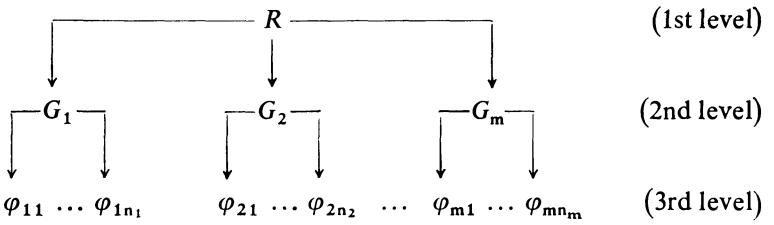
$$G = (V, E, L) . \tag{9}$$

This intact molecular subgraph is composed of the same vertices as the initial reaction graph R, in some special cases the edge and loop sets of G may be simultaneously empty (*i.e.* $E = L = \emptyset$). Finally, let the vertices of G be evaluated, using a mapping

φ, by atomic symbols. The the reaction core $\mathfrak{R}$ is determined as the ordered 3-tuple

$$\mathfrak{R} = (R, G, \varphi) . \tag{10}$$

In general, we may assign to a preselected (parent) reaction graph R a class of intact molecular subgraphs $G_1, G_2, \ldots, G_m$. And furthermore, an arbitrary molecular subgraph G_i may be evaluated by different mappings $\varphi_{i1}, \varphi_{i2}, \ldots, \varphi_{in_i}$. Schematically



Employing this scheme, we construct many different reaction cores with respect to a fixed reaction graphs R,

$$\mathfrak{R}_{ij} = (R, G_i, \varphi_{ij}) , \tag{11}$$

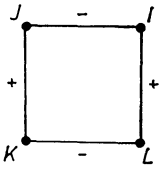
where $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n_i$. A hierarchic classification scheme of chemical reactions is suggested. At the first (highest) level the reaction core is characterized by the reaction graph. At the second level the different intact molecular subgraphs are assigned to the preselected reaction graph. And finally, the last third level specifies by atomic symbols the vertices of intact molecular subgraphs.

The reaction core specified above may be simply presented in an unambiguous linear string form, the property of which is of great importance for the proper machine coding of chemical reactions. In the preceding part[1] of this series we have suggested simple and straightforward scheme to canonize the vertices of reaction graphs, and moreover, traversing the given reaction graph through an Euler alternating walk we get its linear string notation composed of vertex labels. The same canonical labeling of vertices is immediately applicable also for the assigned intact molecular subgraphs. Such vertex-labeled intact molecular subgraphs are simply linearly determined by the list of its edges and loops that are, of course, lexicographically ordered. Finally, the mapping φ which evaluates the vertices of intact molecular subgraphs by atomic symbols is linearly expressed *via* a string of atomic symbols (or their equivalents which more deeply describe the valence condition of atoms) in the same order as the vertices of intact subgraphs are labelled. For instance, the reaction core corresponding to the studied cycloaddition reaction is

$$\# \text{ IJKLI } \# \text{ (I, L) (J, K) } \# \text{ CCCC } \# .$$

The string placed between the first and second delimiters # corresponds to the canonical string notation of the 4-vertex reaction graph R (see graph 4–1 in Fig. 1 and Table II in ref.[1]). The string occured between the second and third delimiters specifies the intact molecular subgraph G in a form of lexicographically ordered list of two edges (I, L) and (J, K). Finally, the last string placed between the third and fourth delimiters specifies the mapping, φ the atomic symbols are given in the same order as in the intact molecular subgraph, at the present case all vertices are carbon atoms.

Illustrative example — cyclic reaction graph involving four vertices. The outlined theory will be now illustrated by simple example of cyclic reaction graph involving four vertices (atoms)

$R =$ 

(12)

This reaction graph was classified[1] as R_{4-1}, its string notation is IJKLI and the corresponding "parent" chemical reaction is

$$\text{I—J} + \text{K—L} \rightarrow \text{I—L} + \text{J—K}. \qquad (13)$$

The assigned pertinent intact molecular subgraphs are listed in Fig. 1. Using these intact subgraphs we can introduce the following classification of four-centre chemical reactions that are described by the reaction graph (*12*):

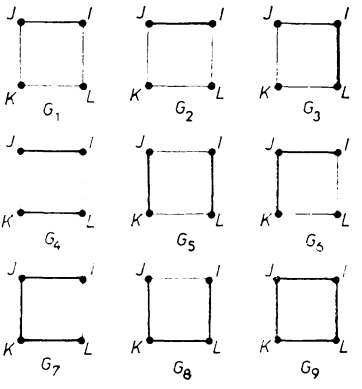
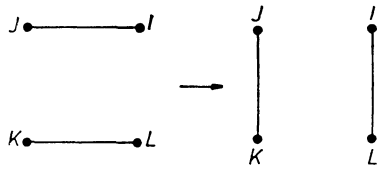


FIG. 1

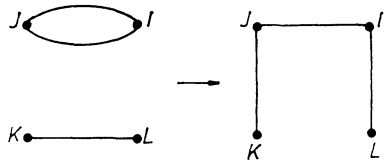
Intact molecular subgraphs assigned to the cyclic reaction graph involving four vertices. The edges of intact subgraphs are denoted by heavy lines

1) Substitution reactions (intact subgraph G_1)

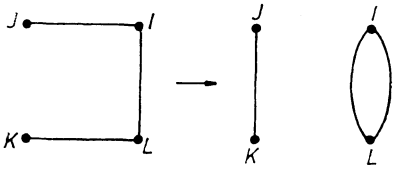


$$[\mathfrak{R}_{1j} = (R, G_1, \varphi_{1j})] \qquad (14)$$

2) Addition and elimination reactions (intact subgraphs G_2 and G_3)

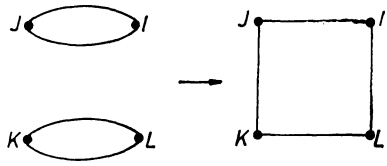


$$[\mathfrak{R}_{2j} = (R, G_2, \varphi_{2j})] \qquad (15a)$$

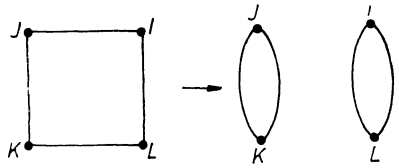


$$[\mathfrak{R}_{3j} = (R, G_3, \varphi_{3j})] \qquad (15b)$$

3) Cycloaddition and cyclofragmentation reactions intact subgraphs G_4 and G_5)

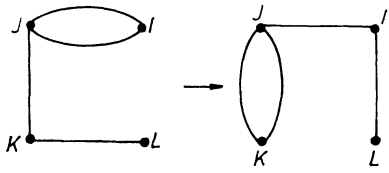


$$[\mathfrak{R}_{4j} = (R, G_4, \varphi_{4j})] \qquad (16a)$$



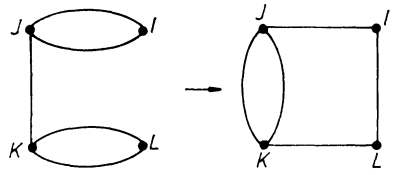
$$[\mathfrak{R}_{5j} = (R, G_5, \varphi_{5j})] \qquad (16b)$$

4) [1,3] sigmatropic rearrangement (intact subgraph G_6)

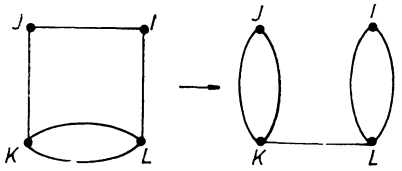


$$[\mathfrak{R}_{6j} = (R, G_6, \varphi_{6j})] \qquad (17)$$

5) Electrocyclic and retro-electrocyclic reactions (intact subgraphs G_7 and G_8)

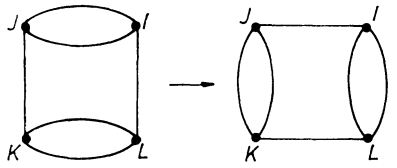


$$[\mathfrak{R}_{7j} = (R, G_7, \varphi_{7j})] \qquad (18a)$$



$$[\mathfrak{R}_{8j} = (R, G_8, \varphi_{8j})] \qquad (18b)$$

6) Resonance of two VB structures (intact subgraph G_9)



$$[\mathfrak{R}_{9j} = (R, G_9, \varphi_{9j})] \qquad (19)$$

This last process does not represent a genuine chemical reaction.

The mappings φ_{ij} presented on the r.h.s. of classified reactions (*14*) to (*19*) evaluate the vertices by atomic symbols. This mapping cannot be introduced by an arbitrary way, each vertex (atom) has an obligatory minimum formal valency indicated by the number of edges (bonds) incident with this vertex. For instance, the vertex *J* in (*18a*) may not be evaluated as oxygen atom, it may be, *e.g.*, carbon or nitrogen atom.

The authors thanks Dr M. Kratochvíl for his encouragement and critical reading of manuscript.

REFERENCES

1. Kvasnička V., Kratochvíl M., Koča J.: This Journal *38*, 2284 (1983)
2. Dugundji J., Ugi I.: Topics Curr. Chem. *39*, 19 (1973).
3. Bart J. C. J., Garagnani E.: Z. Naturforsch. *31b*, 1649 (1976); *32b*, 455, 465, 678 (1977).
4. Arens J. F.: Rec. Trav. Chim. Pays-Bas *98*, 155, 395, 471 (1979).
5. Balaban A. T.: Rev. Roum. Chim.: *17*, 875 (1967).
6. Hendrickson J. B.: Angew. Chem., Int. Ed. Engl. *13*, 47 (1974).
7. Zefirov N. S., Trach S. S.: Zh. Org. Khim. *11*, 225, 1785 (1975); *12*, 7, 697 (1976).
8. Brandt J., Bauer J., Frank R. M., von Scholey A.: Chem. Scr. *18*, 53 (1981).
9. Kvasnička V.: This Journal *48*, 2097 (1983).
10. Roberts D. C.: J. Org. Chem. *43*, 1473 (1978).

Translated by the author.