

GRAPH-THEORETICAL INTERPRETATION OF UGI'S CONCEPT OF THE REACTION NETWORK

V. KVASNIČKA and J. POSPÍCHAL

Department of Mathematics, Slovak Technical University, 81237 Bratislava, Czechoslovakia

Received 10 November 1989

Abstract

The concept of a reaction network, initially suggested by Ugi and coworkers, in the framework of the graph-theoretical model of organic chemistry is elaborated. The reaction network for a pair of isomeric educt molecular (G_E) and product molecular graphs (G_P) is determined as an oriented graph. Its edge, beginning at a graph-vertex G_{i-1} and ending at a graph-vertex G_i , corresponds to a feasible transformation (chemical reaction) constrained by a condition of descending chemical distance from the product graph G_P , i.e. $CD(G_{i-1}, G_P) > CD(G_i, G_P)$. In the reaction network, an oriented path which begins at G_E and ends at G_P corresponds to the decomposition of the overall transformation $G_E \Rightarrow G_P$ into a sequence of "elementary" transformations $G_0 = G_E \Rightarrow G_1 \Rightarrow G_2 \dots \Rightarrow G_{i-1} \Rightarrow G_i = G_P$ that may be assigned to intermediates of the overall transformation.

1. Introduction

The graph-theoretical model [1–6] of organic chemistry offers very simple yet sufficiently diverse formal tools for the description of chemical structures and reactions [7]. A similar idea was first conceived by Ugi and Dugundji [8,9] in the framework of their famous matrix model of constitutional chemistry.

The graph-theoretical model can, loosely speaking, be understood as an alternative formulation of the Ugi and Dugundji matrix model; instead of matrices, it employs the notions and concepts of graph theory. Moreover, this transfer from matrices to graphs allows the use of the very rich and flexible formal tools of the graph theory. Accordingly, many theoretical and algorithmic problems of the model can be formulated and considered very effectively and transparently by making use of the theoretical machinery of graph theory. In particular, different graph metrics (e.g. chemical distance, reaction distance, etc.) can be introduced in a straightforward way.

The purpose of this communication is to formulate a graph-theoretical approach for the construction of the so-called reaction network. For a pair of educt and product compounds (represented by an educt graph G_E and a product graph G_P , respectively) that are mutually related by an overall transformation (chemical reaction) $G_E \Rightarrow G_P$, we

construct an oriented graph (called the reaction network) whose vertices correspond to intermediates of the transformation [7]. In order to overcome a possible combinatorial explosion in the construction of intermediates belonging to the reaction network, the process of its construction should be controlled by a mandatory requirement of decreasing chemical distances for the intermediates in going from the educt graph to the product graph. Such a criterion corresponds closely to the principle of minimal structural change, which is well-known in organic chemistry [10,11]. Ugi and Dugundji [8] initially called attention to this principle in the form of their principle of minimal chemical distances as a heuristic rule. It took into account only those intermediates of the transformation $G_E \Rightarrow G_P$ that are closer (from the standpoint of the chemical metrics) to the product graph.

The reaction network constructed in such a way has the following simple chemical interpretation: An oriented path of the reaction network which begins at the educt graph and ends at the product graph represents a synthetic route going through intermediates that are increasingly more and more "similar" to the target product graph. A similar intuitively formulated approach has recently been used by Ugi et al. [12–14].

2. Basic concepts

Let us postulate that a graph G is a formal structure [15,16] composed of a nonempty *vertex set* $V(G) = \{v_1, v_2, \dots\}$, *edge set* $E(G) = \{e_1, e_2, \dots\}$, and a mapping $\phi: E(G) \rightarrow \{0, 1, 2, 3, \dots\}$. An *edge* of $E(G)$ incident with two vertices $v_1, v_2 \in V(G)$ will be denoted by $\{v_1, v_2\}$; if $v_1 = v_2$, then this edge is often called a *loop*. In our forthcoming considerations, we shall always assume that the graph G does not contain multiple edges. The mapping ϕ evaluates each edge of $E(G)$ by a positive integer: it can be extended outside the edge set $E(G)$ in such a way that an edge $e \notin E(G)$ has a zero evaluation. The nonnegative integer $\phi(e)$, assigned to an edge e , will be called the *multiplicity*, i.e. the edges of zero multiplicity do not belong to $E(G)$. Hence, the graph G is fully determined by the following ordered triple

$$G = (V, E, \phi), \quad (1)$$

where $V = V(G)$ and $E = E(G)$. The chemical interpretation of the above-introduced graph is very simple: The vertices correspond to atoms, the edges correspond to bonds with the same multiplicity as the edges, and loops correspond to lone electron pairs.

Two graphs $G_1 = (V_1, E_1, \phi_1)$ and $G_2 = (V_2, E_2, \phi_2)$ are *isomorphic* ($G_1 \approx G_2$) iff there exists a 1:1 mapping (called the *isomorphism*)

$$\chi: V(G_1) \rightarrow V(G_2) \quad (2)$$

that preserves adjacency of vertices and evaluation of edges. That is, for each pair of adjacent vertices $v_1, v_2 \in V(G_1)$, the mapped pair of vertices $\chi(v_1), \chi(v_2)$ is also adjacent and the evaluations of the corresponding edges are of the same value,

$$\{v_1, v_2\} \in E(G_1) \Leftrightarrow \{\chi(v_1), \chi(v_2)\} \in E(G_2), \quad (3a)$$

$$\phi(\{v_1, v_2\}) = \phi(\{\chi(v_1), \chi(v_2)\}). \quad (3b)$$

Two graphs G_1 and G_2 are isomeric [7] ($G_1 \cong G_2$) iff they have the same number of vertices and the sums of evaluations of edges are identical, formally,

$$|V(G_1)| = |V(G_2)|, \quad (4a)$$

$$\sum_{e \in E(G_1)} \phi_1(e) = \sum_{e \in E(G_2)} \phi_2(e), \quad (4b)$$

where $|X|$ denotes the number of elements (cardinality) of a set X .

A graph G' is called a *subgraph* of the graph G ($G' \subseteq G$) iff $V(G') \subseteq V(G)$, $E(G') \subseteq E(G)$, and

$$\forall e \in E(G'): \phi'(e) \leq \phi(e). \quad (5)$$

The concept of isomerism determined over a universe of graphs can be formally considered as a realization of an "equivalence" between abstract elements. Hence, the universe of graphs can be decomposed into disjoint families of mutually isomeric graphs. The *family of isomeric graphs* \mathcal{F}_{pq} is composed of all possible graphs determined as follows:

$$\mathcal{F}_{pq} = \{G = (V, E, \phi); |V| = p \text{ and } \sum_{e \in E(G)} \phi(e) = q\}. \quad (6)$$

A *common subgraph* of two graphs G_1 and G_2 is a graph \tilde{G} which is simultaneously isomorphic to the subgraphs $G'_1 \subseteq G_1$ and $G'_2 \subseteq G_2$, i.e. $\tilde{G} \cong G'_1 \cong G'_2$. A *maximal common subgraph* of G_1 and G_2 , denoted by $G_1 \cap G_2$, is a common subgraph with the largest possible sum of edge evaluations,

$$\sum_{e \in E(G_1 \cap G_2)} \phi(e) = \max, \quad (7)$$

where ϕ is the mapping of the maximal common subgraph $G_1 \cap G_2$. The *chemical distance* between two isomeric graphs G_1 and G_2 from \mathcal{F}_{pq} is determined as follows [7] (cf. also refs. [17,18]),

$$CD(G_1, G_2)$$

$$= \sum_{e \in E(G_1)} \phi_1(e) + \sum_{e \in E(G_2)} \phi_2(e) - 2 \sum_{e \in E(G_1 \cap G_2)} \phi(e) = 2[q - \sum_{e \in E(G_1 \cap G_2)} \phi(e)]. \quad (8)$$

In our recent publications [7, 19], we have demonstrated that the chemical distance thus defined is a *metric*, i.e. it satisfies all three attributes of an abstract metric: symmetry, positive semidefiniteness, and triangle inequality. We have to emphasize that $CD(G_1, G_2) = 0$ iff $G_1 \approx G_2$.

A *group of automorphisms* $\Gamma(G) = \{\alpha_1, \alpha_2, \dots\}$ of the graph G is composed of 1:1 mappings

$$\alpha: V(G) \rightarrow V(G) \quad (9)$$

that conserve the adjacency of vertices and evaluation of the corresponding edges,

$$\{v_1, v_2\} \in E(G) \Rightarrow \{\alpha(v_1), \alpha(v_2)\} \in E(G), \quad (10a)$$

$$\phi(\{v_1, v_2\}) = \phi(\{\alpha(v_1), \alpha(v_2)\}). \quad (10b)$$

Two distinct vertices $v_1, v_2 \in V(G)$ are called *topologically equivalent* iff there exists a nontrivial automorphism $\alpha \in \Gamma(G)$ that transforms the vertex v_1 into another one v_2 , $\alpha(v_1) = v_2$. This means that if the graph G has a nontrivial group of automorphisms (i.e. $|\Gamma(G)| > 1$), then the vertex set $V(G)$ may be decomposed into disjoint subsets of topologically equivalent vertices.

3. Reaction graphs

The notion of a *reaction graph* has been introduced by the present authors [20] as a proper graph-theoretical tool [7] to express a *transformation* $G_1 \Rightarrow G_2$, where $G_1, G_2 \in \mathcal{F}_{pq}$. The same term "reaction graph" was used previously by Balaban and coworkers [26]; in their approach it corresponds to an analogy of the concept "reaction network", specified below. The term "reaction graph" we have used [20] as a graph-theoretical counterpart of Ugi's concept of the reaction matrix [8]. The reaction graph is unambiguously constructed for a given pair of graphs G_1, G_2 on the basis of their maximal common subgraph $G_1 \cap G_2$; the approach is closely related to Ugi's [8,9] *principle of minimum chemical distance* as a very important heuristic tool for the construction of reaction graphs. In this section, we shall give a more general presentation of reaction graphs; they will be defined independently of a graph G as a formally simple tool used to transform one graph into another one.

The *reaction graph* G_R is determined by a nonempty vertex set $V(G_R) = \{w_1, w_2, \dots\}$, edge set $E(G_R) = \{f_1, f_2, \dots\}$, and a mapping

$$\psi: V(G) \rightarrow \{0, \pm 1, \pm 2, \dots\} \quad (11a)$$

that assigns to each edge $f \in E(G_R)$ a nonzero integer restricted by the following condition:

$$\sum_{f \in E(G_R)} \psi(f) = 0. \quad (11b)$$

The mapping ψ and the condition (11b) have a very instructive chemical interpretation: An edge of G_R evaluated by a positive/negative integer corresponds to a created/annihilated edge (edges) in a graph G . The relation (11b) expresses a condition of strict "stoichiometry" for the transformation induced by the reaction graph G_R , i.e. the educt and product graphs are isomeric. In a similar way as for graphs, the mapping ψ may be extended outside of $E(G_R)$, the edges $f \notin E(G_R)$ are evaluated by $\psi(f) = 0$.

Now let us focus our attention on how to specify the reaction graph G_R so that, when applied onto a graph G , it produces a new graph G' isomeric to the parent one (i.e. $G, G' \in \mathcal{F}_{pq}$ or, in other words, the transformation $G \Rightarrow G'$ is strictly "stoichiometric"). Since the vertex sets of G and G_R are, in general, different, we have to introduce a 1:1 mapping of $V(G_R)$ onto a subset $\tilde{V}(G) \subseteq V(G)$,

$$\omega: V(G_R) \rightarrow \tilde{V}(G). \quad (12)$$

This mapping specifies the correspondence between vertices of the reaction graph G_R and vertices of a subgraph of G induced by the subset $\tilde{V}(G)$. In our forthcoming considerations, we shall use an inverse mapping $\omega^{-1}: \tilde{V}(G) \rightarrow V(G_R)$ extended outside of $\tilde{V}(G)$ to the whole vertex set $V(G)$, where $\omega^{-1}(v) = z$, for each $v \in V(G) \setminus \tilde{V}(G)$ and z is a "virtual" vertex which does not belong to $V(G_R)$. The resulting graph G' formed from the graph G by making use of the reaction graph G_R and the mapping ω is formally determined as an ordered triple

$$(G, \omega, G_R) = G'. \quad (13)$$

Its vertex set $V(G')$ is equal to the vertex set $V(G)$; the edge set $E(G')$ is determined through the mapping ϕ' as follows:

$$\phi'(\{v_1, v_2\}) = \phi(\{v_1, v_2\}) + \psi(\{\omega^{-1}(v_1), \omega^{-1}(v_2)\}). \quad (14)$$

We call the transformation (13) *feasible* iff the mapping ϕ' is nonnegative,

$$\phi'(\{v_1, v_2\}) \geq 0. \quad (15)$$

In the opposite case, when for a pair of vertices $v_1, v_2 \in V(G)$ we have $\phi'(\{v_1, v_2\}) < 0$, the reaction graph G_R specified by the mapping ω is inapplicable to the graph G . Hence, the edge set $E(G')$ of the graph G' produced by a feasible transformation (G, ω, G_R) is determined by

$$E(G') = \{\{v_1, v_2\}; \phi'(\{v_1, v_2\}) > 0\}. \quad (16)$$

The chemical distance between the graph G and the resulting graph G' is bounded by

$$\sum_{f \in E(G_R)} |\psi(f)| \geq CD(G, G') \geq 0. \quad (17)$$

The left-hand side inequality immediately follows from the fact that the notion of chemical distance was defined via the maximal common subgraph of G and G' ; therefore, its even integer value should be minimal with respect to all possible reaction graphs that can be applied to the original graph G . The right-hand side inequality implies that the chemical distance $CD(G, G')$ is nonnegative, a property which is in accordance with the fact that the chemical distance is a metric.

Let us consider two feasible transformations

$$(G, \omega_1, G_R) = G_1, \quad (18a)$$

$$(G, \omega_2, G_R) = G_2. \quad (18b)$$

The resulting graphs G_1 and G_2 are produced by the application of the same reaction graph G_R to the original graph G ; the transformations (18a,b) are different only in mappings ω_1 and ω_2 , where $\omega_1 \neq \omega_2$. We say that the mappings ω_1 and ω_2 are *equivalent* ($\omega_1 \equiv \omega_2$) iff the produced graphs G_1 and G_2 are isomorphic,

$$\omega_1 \equiv \omega_2 \Leftrightarrow G_1 \approx G_2. \quad (19)$$

It is easy to show that the mappings ω_1 and ω_2 are equivalent iff they are related by

$$\omega_1 = \alpha \omega_2 \beta, \quad (20a)$$

where α and β are automorphisms of the graph G and the reaction graph G_R , respectively, $\alpha \in \Gamma(G)$, $\beta \in \Gamma(G_R)$, formally.

4. Reaction network

In our recent communications [7,21], we have introduced the so-called graph of chemical distances \mathcal{G}_{pq}^{CD} . Its vertex set $V(\mathcal{G}_{pq}^{CD})$ is identical with the family of isomeric graphs \mathcal{F}_{pq} . Two vertices (graphs) $G_1, G_2 \in V(\mathcal{G}_{pq}^{CD})$ are adjacent iff their chemical distance equals two, i.e. $CD(G_1, G_2) = 2 \Leftrightarrow \{G_1, G_2\} \in E(\mathcal{G}_{pq}^{CD})$. This requirement follows immediately from our definition of the chemical distance (8). Its value between two nonisomorphic but isomeric molecular graphs is even, i.e. its possible minimum is equal to two. Moreover, we have demonstrated that the chemical distance between an arbitrary pair of graphs $G_1, G_2 \in \mathcal{F}_{pq}$ is equal to twice the graph distance between the same graphs in \mathcal{G}_{pq}^{CD} . Let us select two vertices-graphs G_E (educt graph) and G_P (product

graph) in \mathcal{G}_{pq}^{CD} , then all paths that are connecting these graphs may be formally considered as a synthetic step-by-step route for the fixed educt and product graphs; the graphs lying on a path are, loosely speaking, interpreted as intermediates of the synthesis route. This simple and transparent interpretation of paths in the graph of chemical distances does not correspond in many cases to the usual meaning of intermediates as relatively stable compounds. Many graphs from a path connecting G_E and G_P may correspond to compounds with, for example, quite exotic valence states of atoms [7]. In order to remove this drawback of \mathcal{G}_{pq}^{CD} used as a background for the construction of synthetic routes of the overall transformation [7] $G_E \Rightarrow G_P$, we suggest another approach called the *reaction network*, involving only stable (or relatively stable) intermediate compounds. First, we have to introduce a set \mathcal{G}_R of preselected non-isomorphic reaction graphs; this set is composed of those reaction graphs that are typical of simple synthetic steps. The reaction network is determined as a directed graph, its vertices are some graphs (including G_E and G_P) from \mathcal{F}_{pq} . A pair of these graphs G_1 and G_2 is connected by an oriented edge beginning at G_1 and ending at G_2 if (i) there exists a reaction graph $G_R \in \mathcal{G}_R$ and a mapping ω such that

$$(G_1, \omega, G_R) = G_2 \quad (21)$$

is a feasible transformation, and (ii) the chemical distances of G_1 and G_2 from the educt graph G_E and the product graph G_P are restricted by

$$CD(G_1, G_P) > CD(G_2, G_P), \quad (22a)$$

or, equivalently,

$$CD(G_1, G_E) < CD(G_2, G_E). \quad (22b)$$

Moreover, we shall postulate that the educt (product) graph G_E (G_P) is incident only with outgoing (incoming) oriented edges and the remaining graphs are incident, at least, with one incoming and one outgoing oriented edge. An oriented path beginning at G_E and ending at G_P has the following chemical interpretation: It represents a synthetic route to G_P from G_E via intermediates satisfying the *principle of minimal structural change* [10,11] (often used in organic chemistry as an important heuristic), expressed in our graph-theoretical model by the conditions (22) of decreasing chemical distance [7] between successive intermediates and the product graph G_P .

Each oriented edge of the reaction network can be evaluated by a positive integer expressing a lowering of the chemical distance from the product graph. In particular, let us consider an oriented edge beginning at G_1 and ending at G_2 , and let these graphs be related by a feasible transformation (21) restricted by the conditions (22). Then, this edge will be evaluated by a positive integer

$$\Delta CD(G_1, G_2) = CD(G_1, G_P) - CD(G_2, G_P). \quad (23)$$

This reflects an "order" of fitness of the graph G_2 to the product graph G_p achieved by a feasible transformation (21). Its increased value indicates a longer "jump" toward the product graph G_p in the synthetic route going from an intermediate graph G_1 to another intermediate graph G_2 . It is easy to demonstrate that an upper bound of $\Delta CD(G_1, G_2)$ is

$$\sum_{f \in E(G_R)} |\psi(f)| \geq \Delta CD(G_1, G_2), \quad (24)$$

i.e. the maximal possible value of $\Delta CD(G_1, G_2)$ is equal to the number of all edges that are created and annihilated by the transformation (21).

Let us consider an oriented path in the reaction network which begins at the educt graph G_E and ends at the product graph G_p , and let this path be composed of a sequence of $(n + 1)$ vertices-graphs

$$G_0 = G_E \rightarrow G_1 \rightarrow G_2 \dots G_{n-1} \rightarrow G_n = G_p. \quad (25)$$

This path contains n oriented edges. The sum of their evaluation exactly equals the chemical distance between G_E and G_p ,

$$\sum_{i=1}^n \Delta CD(G_{i-1}, G_i) = CD(G_E, G_p). \quad (26)$$

Introducing the inequality (24) into (26), we obtain

$$\sum_{i=1}^n \sum_{f \in E(G_R^{(i)})} |\psi_i(f)| \geq CD(G_E, G_p), \quad (27)$$

where $G_R^{(i)} \in G_R$ is a reaction graph assigned to the i th oriented edge in (25). The sum of edge evaluation is equal to the chemical distance $CD(G_E, G_p)$ only if the given path (25) in the reaction network is going through the "shortest" corresponding path in the graph of chemical distances G_{pq}^{CD} .

Recently, Ugi et al. [12–14] have suggested the so-called bilateral approach (cf. also ref. [7]) for a construction of the reaction network; in the appendix, we present a very simple and transparent depth-first search algorithm for its construction.

5. Illustrative example – Fischer indole synthesis

The theory of reaction networks outlined in the previous section will be illustrated by the Fischer indole synthesis [22] (see fig. 1). Here, an arylhydrazone of a ketone is treated with a catalyst (usually zinc chloride), whereupon elimination of ammonia takes place and an indole is formed. A simplified mechanistic step-by-step decomposition [22] of the overall Fischer synthesis into simple [1, 3] and [3, 3] sigmatropic

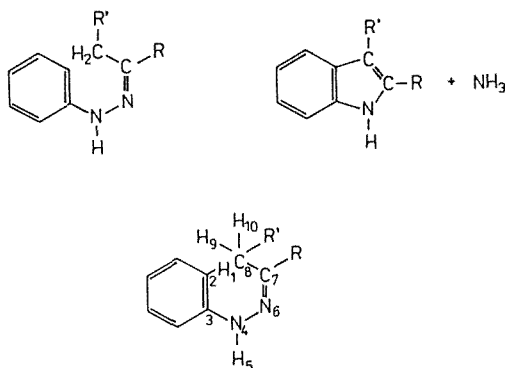


Fig. 1. The Fischer indole synthesis. The bottom scheme represents an indexing of atoms that are involved in the reaction.

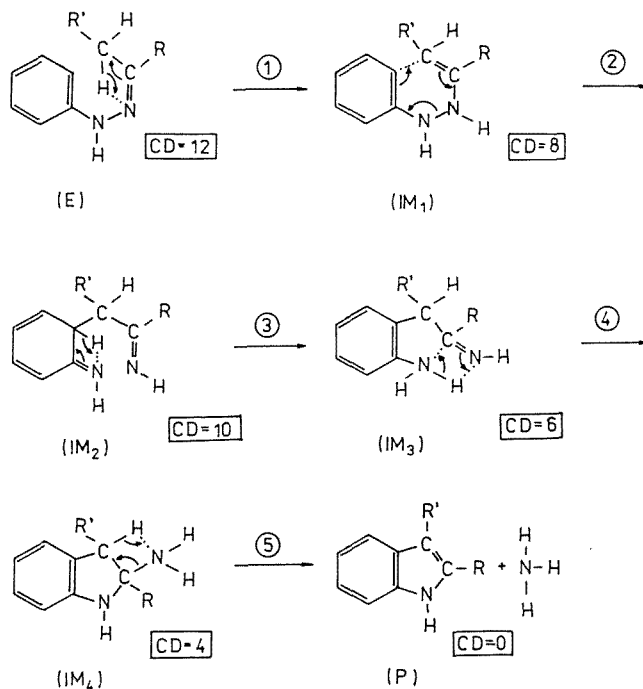


Fig. 2. A mechanistic decomposition of the Fischer indole synthesis; the steps 1, 3, 4, and 5 (2) are simple $[1, 3]$ ($[3, 3]$) sigmatropic rearrangements. The numbers in boxes express the chemical distance of a given compound from the product compound. In some special cases, in particular for E, IM_3 , and IM_4 , there exist pairs of topologically equivalent hydrogens; the selection of one of them in our reaction mechanism was made arbitrarily.

rearrangements is outlined in fig. 2, where the protonated forms were ignored. There is much evidence for this mechanism. In particular, the intermediate IM_3 can be detected spectroscopically: the intermediate IM_4 was isolated, and ^{15}N labeling experiments show that the β -nitrogen (farther from the benzene ring) is eliminated as ammonia.

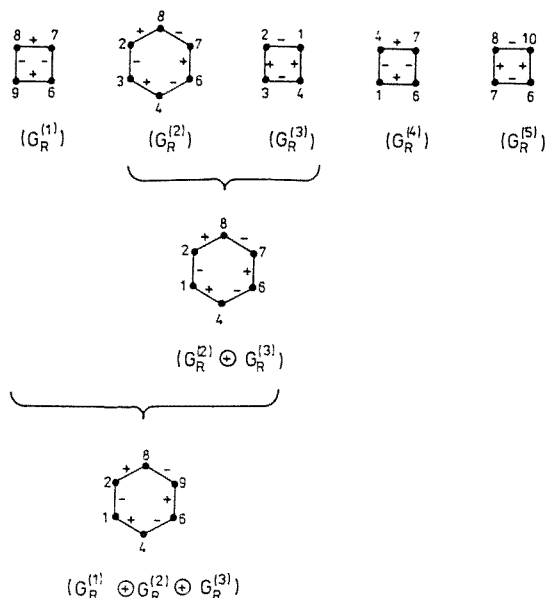


Fig. 3. The reaction graphs assigned to single mechanistic steps of the Fischer indole synthesis presented in fig. 2. The indexing of vertices (atoms) is given in the bottom scheme of fig. 1. The hexagonal reaction graphs in the second and third rows correspond to the "sum" of $G_R^{(1)}$, $G_R^{(2)}$ and $G_R^{(1)}$, $G_R^{(2)}$, $G_R^{(3)}$, respectively.

Figure 3 shows the reaction graphs that are assigned to the respective single mechanistic steps, where the square (hexagonal) reaction graphs correspond to $[1, 3]$ ($[3, 3]$) sigmatropic rearrangement.

The maximal common subgraphs of educt and product compounds are given in fig. 4. We see that the maximal common subgraph denoted by MCS_B is in contradiction to the above-mentioned experimental observation that the β -nitrogen of the educt compound is eliminated as ammonia. In fact, the right (product) compound contains an N-H bond from ammonia which is matched to a similar bond from the left (educt) incident with α -nitrogen. The reaction graphs of the Fischer indole synthesis constructed on the basis of maximal common subgraphs given in fig. 4 are presented in fig. 5. Here, it is interesting to note that the "sum" of the reaction graphs in fig. 3 (first row) gives the reaction graph A corresponding to the maximal common subgraph MCS_A . Hence, we may say, loosely speaking, that the proposed mechanism of the Fischer indole synthesis (fig. 2) from a global standpoint satisfies the so-called principle of minimal chemical distance [7-9], but the partition of the mechanism into single

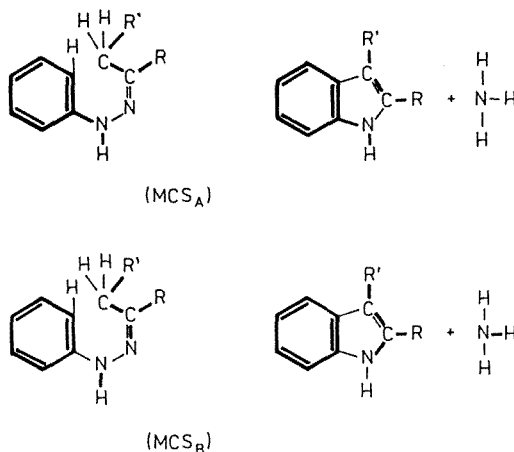


Fig. 4. Two maximal common subgraphs of the educt and product molecular graphs (compounds).

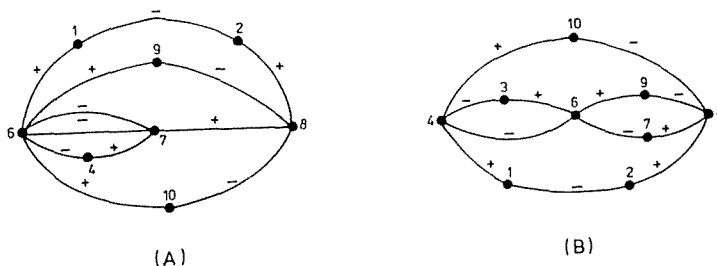


Fig. 5. The reaction graphs of the Fischer indole synthesis constructed by making use of the maximal common subgraphs given in fig. 4.

mechanistic steps obviously violates this principle. The chemical distances of educt and intermediate compounds from the product compound are not monotonically decreasing from $CD(\text{educt}, \text{product}) = 12$ to zero chemical distance. There exists a singular intermediate IM_2 with $CD(\text{IM}_2, \text{P}) = 10$ slightly higher than $CD(\text{IM}_1, \text{P}) = 8$ corresponding to a precursor IM_1 of the intermediate IM_2 (see CD 's in fig. 2). In order to overcome this discrepancy of the proposed mechanism with the requirement of a monotonically decreasing chemical distance (a heuristic rule) in going successively from the educt to the product via intermediates, we may use an artificial procedure of merging the second and third mechanistic steps into a single step. This is formally expressed by the "sum" of the reaction graphs $G_R^{(2)}$ and $G_R^{(3)}$; the merge is expressed by the reaction graph $G_R^{(2)} \oplus G_R^{(3)}$ displayed in the second row in fig. 3. Moreover, we may continue this merging of reaction graphs; for example, the first three reaction graphs that produce the crucial intermediate IM_3 (see fig. 2) may be merged into one hexagonal reaction graph, denoted by $G_R^{(1)} \oplus G_R^{(2)} \oplus G_R^{(3)}$ in fig. 3. This means that two $[1, 3]$ and one $[3, 3]$ sigmatropic rearrangements are merged into a single synchronous electrocyclic process.

The final transition from the intermediate IM_3 to the product (indole) via an intermediate can be carried out in two different ways. The first reaction path is identical to that shown (going via IM_4) in fig. 2, the second reaction path involves another intermediate denoted by IM'_4 (see fig. 6). The whole reaction network of the Fischer indole synthesis is of great complexity; here, we have presented only the part relevant for the elucidation of the reaction mechanism.

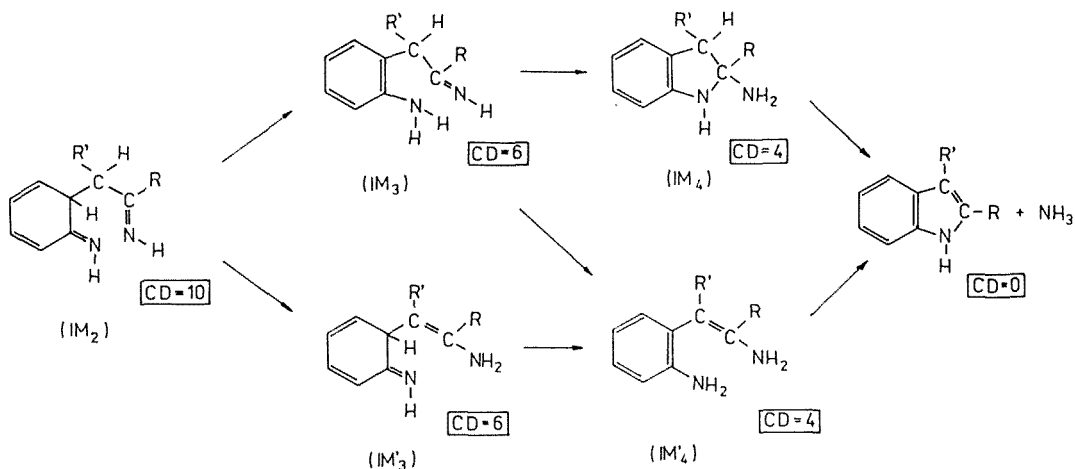


Fig. 6. Three alternative reaction paths going from the intermediate IM_2 to the product P (indole and ammonia). All these paths satisfy the criterion of decreasing chemical distance. The paths going through the intermediate IM'_3 or IM'_4 do not contain the spectroscopically observed intermediate IM_3 and the isolated intermediate IM_4 .

6. Summary

The concept of reaction networks offers a very important theoretical tool for generating reaction paths that connect the educt and product molecular graphs (compounds). Its efficient application needs an extremely fast algorithm for the calculation of chemical distances between a pair of isomeric molecular graphs. Recently, this problem was successfully solved by Nicholson et al. [23] and also by the present authors [24]. An original idea of Levi [25] was used: the problem of a maximal common induced subgraph of two graphs G_1 and G_2 may be reduced to the problem of finding a maximal clique in the product $G_1 \times G_2$. Since the chemical distance is defined via the maximal common subgraph, which is not induced, in general, this approach should be modified in such a way that maximal cliques of the product of line graphs $\hat{G}_1 \times \hat{G}_2$ are involved.

A conceptionally similar approach for the construction of reaction networks was elaborated by Ugi and his associates [12–14]. They studied, mainly, the Streith synthesis and a prebiotic formation of adenine from five molecules of hydrogen cyanide.

The obtained results give very serious conclusions for a mechanistic elucidation of these two mechanisms. Similarly, in section 5 we have studied the Fischer indole synthesis, where it was demonstrated that our theoretical approach also produces many valuable conclusions for an elucidation of mechanisms of organic syntheses. The network constructed on the presented theoretical analysis contains potential reaction pathways, only one of which has actually become reality. What is very important is that all the obtained conclusions are deduced on the basis of the requirement that the chemical distance decreases when we are going through a reaction path from the educt to the product via intermediates. This requirement is a heuristic, which in some cases can be violated, but the scope of this violation is limited. It seems likely that this criterion of decreasing chemical distance offers results that are acceptable for organic chemists in their contemplations about elucidating mechanisms of complex overall chemical transformations.

Acknowledgement

We thank Professor Milan Kratochvíl for his valuable suggestions.

Appendix

We present here a depth-first search algorithm for the construction of a reaction network in the symbolic Pascal-like form.

```

i := 0;  $\mathcal{U}_0 :=$  a set composed of the educt graph  $G_E$ ;
repeat if  $|\mathcal{U}_i| > 0$  then
  begin  $G_i :=$  an arbitrary graph of  $\mathcal{U}_i$ ;
        remove the graph  $G_i$  from the set  $\mathcal{U}_i$ ;
         $CD_i :=$  chemical distance between  $G_i$  and  $G_P$ ;
        if  $CD_i = 0$  then write  $(G_0, G_1, G_2, \dots, G_i)$ 
          {output of a synthetic route
            $G_0 = G_E \Rightarrow G_1 \Rightarrow G_2 \dots G_i = G_P$ }
        else if  $i < i_{\max}$  then
          begin  $i := i + 1$ ;
                 $\mathcal{U}_i :=$  the set of all graphs  $G \in \mathcal{F}_{pq}$  that are
                  produced from  $G_{i-1}$  by feasible trans-
                  formations and their chemical distances
                  from  $G_P$  are smaller than  $CD_{i-1}$ ;
          end;
        end else  $i := i - 1$ ;
until  $i = 0$ ;
  
```

Note: The constant i_{\max} represents the maximal depth of the searching tree. If the variable i is greater than this constant, then the depth-first search construction of the reaction network is coming back at the higher level $i - 1$.

References

- [1] A.T. Balaban (ed.), *Chemical Applications of Graph Theory* (Academic Press, London, 1986).
- [2] N. Trinajstić, *Chemical Graph Theory*, Vols. 1 and 2 (CRC Press, Boca Raton, FL, 1983).
- [3] R.B. King (ed.), *Chemical Applications of Topology and Graph Theory* (Elsevier, Amsterdam, 1983).
- [4] I. Gutman and O. Polansky, *Mathematical Concepts in Organic Chemistry* (Springer-Verlag, Berlin, 1985).
- [5] N. Trinajstić (ed.), *Mathematics and Computational Concepts in Chemistry* (Ellis Harwood, Chichester, 1986).
- [6] N.S. Zefirov and C.I. Kutchanov (eds.), *Applications of Graph Theory in Chemistry* (Nauka, Novosibirsk, 1988) (in Russian).
- [7] J. Koča, M. Kratochvil, V. Kvasnička, L. Matyska and J. Pospíchal, *Synthon Model of Organic Chemistry and Synthesis Design* (Springer-Verlag, Berlin, 1989).
- [8] J. Dugundji and I. Ugi, *Topics Curr. Chem.* 39(1973)19.
- [9] I. Ugi, J. Bauer, J. Brandt, J. Friedrich, J. Gasteiger, C. Jochum and W. Schubert, *Angew. Chem. Int. Ed. Engl.* 18(1979)111.
- [10] W. Hüchel, *Theoretische Grundlagen der organischen Chemie* (Akademie Verlag, Leipzig, 1952), Band 1.
- [11] G.W. Wheland, *Advanced Organic Chemistry* (Wiley, New York, 1962).
- [12] E. Fontain, J. Bauer and I. Ugi, *Z. Naturforsch.* 42b(1987)889.
- [13] J. Bauer, E. Fontain and I. Ugi, *Anal. Chim. Acta* 210(1988)123.
- [14] I. Ugi, J. Bauer, R. Baumgartner, E. Fontain, D. Forstmeyer and S. Lohberger, *Pure Appl. Chem.* 60(1988)1573.
- [15] F. Harary, *Graph Theory* (Addison-Wesley, Reading, Mass., 1969).
- [16] M. Bezhad and G. Chartrand, *Introduction to the Graph Theory* (Allyn and Bacon, Boston, Mass., 1971).
- [17] M.A. Johnson, in: *Graph Theory and its Applications to Algorithms and Computer Science*, ed. Y. Alavi, G. Chartrand, L. Lesniak, D.R. Lick and C.E. Wahl (Wiley, New York, 1985), p. 457.
- [18] V. Baláž, J. Koča, V. Kvasnička and M. Sekanina, *Časopis pěst. mat.* 111(1986)431.
- [19] V. Baláž, V. Kvasnička and J. Pospíchal, *Časopis pěst. mat.* 114(1989)155.
- [20] V. Kvasnička, M. Kratochvil and J. Koča, *Collect. Czech. Chem. Comm.* 48(1983)2284.
- [21] V. Kvasnička and J. Pospíchal, *J. Math. Chem.* 3(1989)161.
- [22] J. March, *Advanced Organic Chemistry*, 3rd Ed. (McGraw-Hill, New York, 1985), p. 1032.
- [23] V. Nicholson, C.-C. Tsai, M. Johnson and M. Naim, in: *Graph Theory and Topology in Chemistry*, ed. R.B. King and D.H. Rouvray (Elsevier, Amsterdam, 1987), p. 226.
- [24] V. Kvasnička and J. Pospíchal, *Rep. Mol. Theory* (in press).
- [25] G. Levi, *Calcolo* 9(1972)1.
- [26] A.T. Balaban, D. Farcasiu and R. Banica, *Rev. Roum. Chim.* 11(1966)1205.