THE REACTION DISTANCE*

Jaroslav Koča

Department of Organic Chemistry, J. E. Purkyně University, 611 37 Brno

Received December 1st, 1987
Accepted February 14th, 1988

An algorithm is suggested for the calculation of the reaction distance as the minimal number of elementary steps of reorganization of valence electrons during a reaction. The basis for the formulation of this algorithm is the *synthon* isomerization graph (*SR*-graph). Chemical examples of application of the algorithm are given.

Previous papers[1,2] have dealt with the matrix and graph models of the *synthon* and its conversions, and the concept of the *reaction distance* (*RD*) between two *synthons* has been introduced[1]. The reaction distance is defined as the smallest number of elementary steps of reorganization of electrons[3] (ESRE) during a chemical reaction, or formally, as the graph distance in the $G_{FIS(A)}$ graph. In the model treated, *RD* along with the concept of the valence state of the atom plays the major role. Among the reasons for this is the fact that the model is intended particularly for chemical synthesis planning, where the following two questions are of crucial importance: what can be prepared from a given substance, and from which substance can a product of interest be synthesized. In either case, the substance sought will apparently lie "near" the given substance, spoken in reaction distance terms.

In this paper, an algorithm is formulated for the calculation of *RD* in the *synthon reaction* graph (*SR*-graph). This algorithm is based on the previously found relations for the calculation of the distance of two valence states of an atom and two states of the atomic vector[4,5].

Since *RD* is the smallest number of elementary steps of reorganization of valence electrons, the concept of elementary *SR*-graphs will be first introduced.

THEORETICAL

Elementary SR-*Graphs*

Any process of reorganization of valence electrons during a chemical reaction can be decomposed in the model into elementary steps, for which the matrix operators

* Part XIII in the series Mathematical Model of Organic Chemistry; Part XII: Collect. Czech. Chem. Commun. *53*, 3108 (1988).

α, β, γ, δ have been introduced[1]. Graph analogues of these operators are the so-called elementary *SR*-graphs, given in Table I. Each elementary *SR*-graph is the *SR*-graph of a reaction where the reaction distance between the starting and final *synthons* is unity. Elementary *SR*-graphs along with the global *SR*-graph form the starting structure for the algorithm for the calculation of the reaction distance, given below.

Reaction Distance

The reaction distance *RD* between two isomeric *synthons* $S(A)$ and $S'(A)$ has been defined as the graph distance between the corresponding vertices in the $G_{FIS(A)}$ graphs. The algorithm set up for the calculation of this quantity (the GRD algorithm) starts from the graph model of the *synthon* and from the fact that any *SR*-graph G_R can be decomposed into a finite number of elementary *SR*-graphs $G_1, G_2, \ldots, G_n$:

$$G_R = G_1 + G_2 + \ldots + G_n . \qquad (1)$$

If G_R is the *SR*-graph of the $S(A) \rightarrow S'(A)$ transformation, then the lowest integer n satisfying Eq. (*1*) is equal to the reaction distance $RD(S(A), S'(A))$. The operation of summation $(+)$ of *SR*-graphs is defined (similarly as in ref.[6]) as follows. Be $G_R = (V_R, E_R, L_R, \psi, \omega, \{-1, 1\})$ and $G'_R = (V'_R, E'_R, L'_R, \psi', \omega', \{-1, 1\})$ two *SR*-graphs, then

$$G'' = G_R + G'_R = (V''_R, E''_R, L''_R, \psi'', \omega'', \{-1, 1\}) , \qquad (2)$$

where $V''_R = V_R \cup V'_R$ (for this union, any two virtual vertices are construed as different), $E''_R = (E_R^+ \cup E_R'^+) \setminus (E_R^- \cup E_R'^-)$, $L''_R = (L_R^+ \cup L_R'^+) \setminus (L_R^- \cup L_R'^-)$, $\psi''(e) = \psi(e)$ for $e \in E_R$ and $\psi'' = \psi'(e)$ for $e \in E'_R$, $\omega''(l) = \omega(l)$ for $l \in L_R$ and $\omega''(l) = \omega'(l)$ for $l \in L'_R$, E_R^+ and L_R^+ are $\{e \in E_R \mid \psi(e) = 1\}$ and $\{l \in L_R \mid \omega(l) = 1\}$ sets, respectively, E_R^-, $E_R'^+$, $E_R'^-$, L_R^-, $L_R'^+$ and $L_R'^-$ are defined analogously. and operation $\setminus$ denotes the symmetrical difference of sets (i.e. $A \setminus B = \{x \mid (x \in A \wedge x \notin B) \wedge (x \notin A \wedge x \in B)\}$).

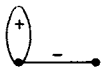
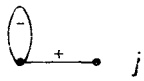
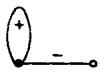
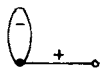
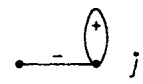
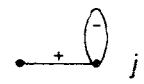
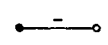
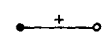
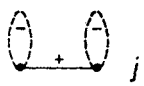
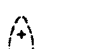
Similarly as in ref.[2], the symbol

will be used for two-electron loops and the symbol

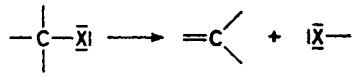
for one-electron loops.

TABLE I

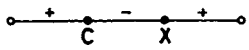
Elementary *SR*-graphs and $\alpha, \beta, \gamma, \delta$ operators ($k = 1, 2, 3$; ● nonvirtual vertices, ○ virtual vertices)

Operator	*SR*-graph	Operator	*SR*-graph
α_k^{ij}		$-\alpha_k^{ij}$	
α_k^{ii}		$-\alpha_k^{ii}$	
β_k^{ij}		$-\beta_k^{ij}$	
β_k^{ii}		$-\beta_k^{ii}$	
γ_k^{ij}		$-\gamma_k^{ij}$	
γ_k^{ii}		$-\gamma_k^{ii}$	
σ_1^{ij}		$-\sigma_1^{ij}$	
σ_1^{ii}		$-\sigma_1^{ii}$	
σ_2^{ij}		$-\sigma_2^{ij}$	
σ_2^{ii}		$-\sigma_2^{ii}$	

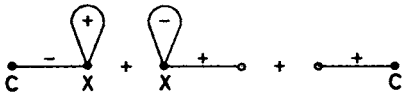
Example 1. Consider a reaction associated with the



transformation. The *SR*-graph of this transformation has the form



One of the possible decompositions,



models the *E*1 mechanism and also informs us that the *RD* between the starting and final *synthons* equals 3.

Now, have two isomeric *synthons*, $S(A)$ and $S'(A)$; G_R be the *SR*-graph of the $S(A) \rightarrow S'(A)$ transformation. The reaction distance with respect to the *SR*-graph G_R, denoted $RD(G_R)$, will be identified with the reaction distance of the two *synthons*,

$$RD(G_R) = RD(S(A), S'(A)) . \tag{3}$$

CRD *Algorithm for the Calculation of the Reaction Distance*

From the graph point of view, the calculation of *RD* is a problem of the edge coverage of the G_R graph by elementary *SR*-graphs, which in the general form is an NP-complete problem[7]. The algorithm suggested operates in two stages. In the first (steps 1 – 17) it seeks for special subgraphs of the G_R graph where the coverage can be accomplished readily, in the second, the remaining part of G_R is covered combinatorially and the alternative with the smallest number of elementary *SR*-graphs is picked out. The input is the G_R graph, the output is $RD\,(G_R)$. The following conventions are adhered to.

Describing the algorithm, the V, E, L sets in the definition of the *SR*-graph will be indexed or otherwise labelled similarly as the initial G_R graph. Hence, denoting a graph G'_R implies that $G'_R = (V'_R, E'_R, L'_R, \psi', \omega', \{-1, 1\})$. An ε-graph will be a graph that after omitting isolated vertices is a graph containing an Euler path[6]. We say that an *SR*-graph G_R satisfies condition (*) if

(i) it contains no loop,

(ii) it is an ε-graph,

(iii) for each vertex, the absolute value of the sum of evaluations of all edges incident with this vertex is lower than 2, and

(iv) the sum of evaluations of all edges is equal to 0.

The operation − between *SR* graphs will be defined as follows. Be *G1*, *G2* *SR*-graphs

$$G1 = (V1, E1, L1, \psi 1, \omega 1, \{-1, 1\})$$
$$G2 = (V2, E2, L2, \psi 2, \omega 2, \{-1, 1\}) .$$

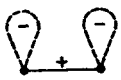
Then the difference $G1 - G2$ is the graph

$$G1 - G2 = (V, E, L, \psi, \omega, \{-1, 1\}) ,$$

where $V = V1$, $E = E1 - E2$, and $L = L1 - L2$, the sign − denoting difference of sets, and ψ and ω are restrictions of $\psi 1$ to E and $\omega 1$ to L, respectively.

The CRD *Algorithm*

1. $D = 0$.
2. Decompose graph $G_R = G1 \cup G2 \cup G3$ so that
   - $G1$ is the subgraph of graph G_R containing all vertices that carry at least one loop and all edges incident with these vertices,
   - $G2$ is the subgraph of graph G_R containing the virtual vertices of graph G_R and those edges incident with them that are not contained in $G1$, and
   - $G3$ contains all the remaining edges of graph G_R and vertices incident with them.
3. If graph $G1$ contains no edge, go to step 8. Else, decompose graph $G1$ into $G1^+$ so that graph $G1^+$ contain only edges of graph $G1$ with positive evaluation and loops with negative evaluation and graph $G1^-$ contain edges with negative evaluation and loops with positive evaluation.
4. If in $G1^+$ there exists a subgraph G' of the form



that in $G1^+$ cannot be completed to a subgraph

put $D = D + 1$, $G1^+ = G1^+ - G'$, and repeat step; else continue.

5. If in $G1^-$ there exists a subgraph G' of the form

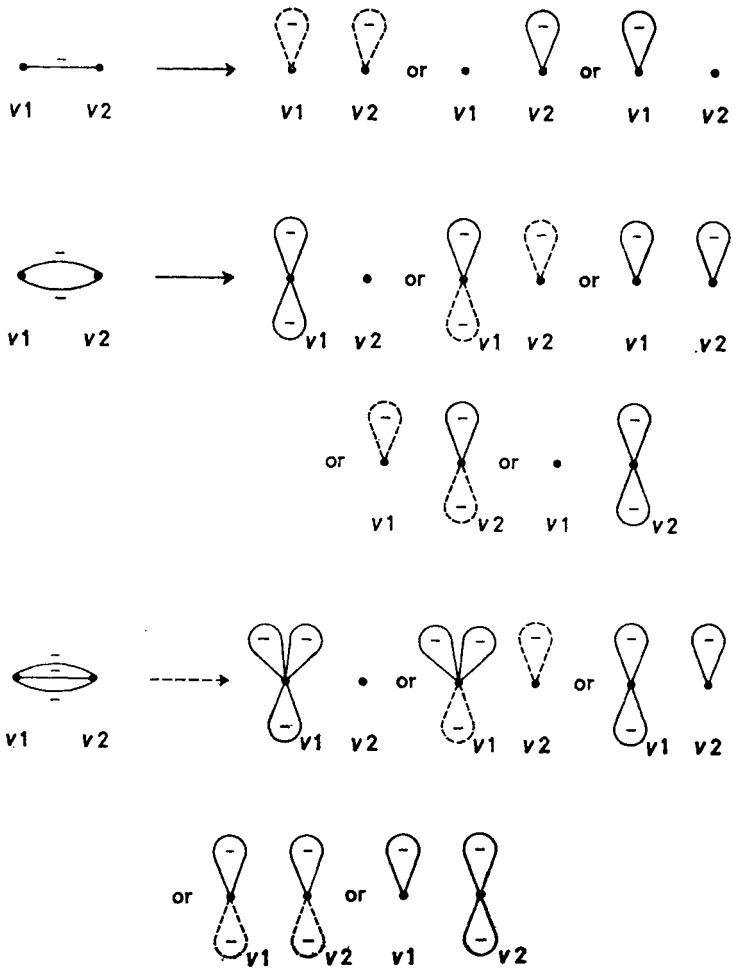
that in $G1^-$ cannot be completed to a subgraph

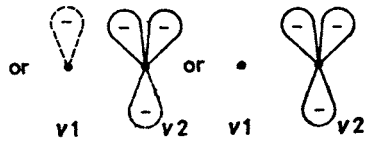
put $D = D + 1$, $G1^- = G1^- - \acute{A}'$, and repeat step 5; else continue.

6. In graph $G1^+$ choose a vertex whose degree is 1 and which is incident with at least one two-electron loop. If such a vertex exists, cancel the two-electron loop at it and the single edge incident with it, put $D = D + 1$ and return to step 6, else continue.
7. In graph $G1^-$ choose a vertex whose degree is 1 and which is incident with at least one two-electron loop. If such a vertex exists, cancel the two-electron loop at it and the single edge incident with it, put $D = D + 1$ and return to step 7, else continue.
8. Put $G' = G1^+ \cup G1^- \cup G3$. Decompose $G' = G1 \cup G3'$ so that $G1'$ and $G3'$ have the same properties as $G1$ and $G3$ in step 2.
9. Decompose graph $G3'$ into components, whose number is I. Put $G' = G1 \cup G2$, $J = 0$.
10. $J = J + 1$; if $J > I$ go to step 16, else denote the J-th component of graph $G3'$ as G'. Put $N1 = |V \cap V1'|$, $N2 = |V \cap V2|$.
11. If $(N1 > 1) \wedge [(N1 + N2) > 2] \wedge [N2 \,.\, (N1 + N2) = 2]$ go to step 10, else continue.
12. If G satisfies condition $(*)$ put $G3' = G3' - G$, $D = D + |E|$, go to step 10; else if $N1 + N2 = 0$ continue step 13 else go to step 14.
13. If there exists $h \in E$ such that graph $G - \{h\}$ satisfies condition $(*)$, put $G = G - \{h\}$, $D = D + |E|$, $G3 = (G3' - G) \cup \{h\}$. Go to step 10.
14. If $N1 = 1$ go to step 15. If in graph $G2$ there exists an edge h incident with some vertex of graph G so that graph $G \cup \{h\}$ satisfies condition $(*)$, put $G3' = G3' - G$, $G2 = G2 - \{h\}$, $D = D + |E| + 1$. Go to step 10.
15. Denote k the single vertex that graphs $G1'$ and G have in common. If the degree of k in G is greater than 1 go to step 10. Else denote the single edge in graph G incident with k as h. Put $G = G - \{h\}$. If now G satisfies condition $(*)$, put $G3' = G3' - G$, $D = D + |E|$. Go to step 10.

16. In graph G', omit each edge $h = \{v1, v2\}$ having the following properties:
    (i) $v2 \in V2$
    (ii) $v1$ is incident with no loop
    (iii) no edge of the graph $G' \cup G3'$ with evaluation opposite to that of edge h is incident with vertex $v1$.

    Denote the number of such edges N. Put $D = D + N$.
17. Put $G' = G' \cup G3'$, $D = D + |E'|$.
18. Replace each edge or two or three parallel edges $\{v1, v2\}$ in graph G' successively with some of the graphs in the following scheme so that all combinations are exhausted:





Replace edges evaluated + following the same scheme, only the evaluation of loops in it will be +. A graph containing no edge is always obtained. There are $3^{n1} \cdot 5^{n2} \cdot 7^{n3}$ ways for this treatment, where $n1$, $n2$ and $n3$ are the numbers of single edges, parallel pairs of edges and parallel groups of three edges, respectively. For each combination, calculate the number N as

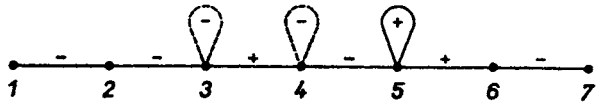
$$N = A^+ + A^- + B^+ + B^- - C_1 - C_2 - E_1 - E_2 ,$$

where A^+ is the number of two-electron loops with evaluation + in $G' - G2$, A^- is the number of two-electron loops with evaluation − in $G' - G2$, B^+ and B^- are the numbers of one-electron loops with evaluation + and −, respectively, in $G' - G2$, $C_1 = \min\{A^+, A^-\}$, $C_2 = \min\{B^+, B^-\}$, $E_1 = \min\{F(A^+ - A^-), F(B^- - B^+) + 2\}$, $E_2 = \min\{F(A^- - A^+), F(B^+ - B^-) + 2\}$, $F(x) = x$ for $x \geqq 0$, $F(x) = 0$ for $x < 0$.

19. Put $RD = D + Nmin$, where $Nmin$ is the lowest N from step 18. The RD obtained is the reaction distance of the *synthons* studied.
20. End of algorithm.

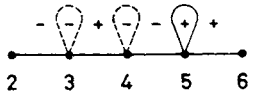
For proving correctness of the algorithm it would be necessary to demonstrate that the distance found by it is really the shortest, and moreover, that a path having the length of the calculated distance exists. The second part of the proposition follows implicit from the phenomenological assumption made in ref.[4], according to which graphs of conversion of valence states of atoms are continuous. For proving the first part of the proposition it is necessary to demonstrate that arrangements of the reaction graph made by the algorithm (as far as step 18) are the most economical. Those edges of the reaction graph not affected by the treatment then are processed combinatorially and out of the possibilities, the minimum is picked out. No exact proof of correctness of the algorithm has been made but no formal discrepancies were observed in any of the particular cases of chemical systems treated.

Example 2. Assume that a reaction leading from the initial to the final states of the *synthon* is modelled by the reaction graph



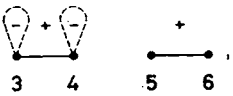
This graph will be treated in the steps of the CRD algorithm as follows:

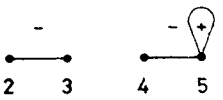
1. $D = 0$

2. $G1$: 

, $G2$: 

, $G3$: 

3. $G1'$: 

, $G1''$: 

4. $G1^{+}$: 

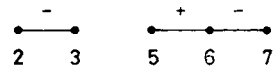
, $D = 1$

5. No change

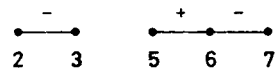
6. No change

7. $G1^{-}$: 

, $D = 2$

8. G': 

, $G1'$: $\{\emptyset\}$

$G3'$: 

9. G': 

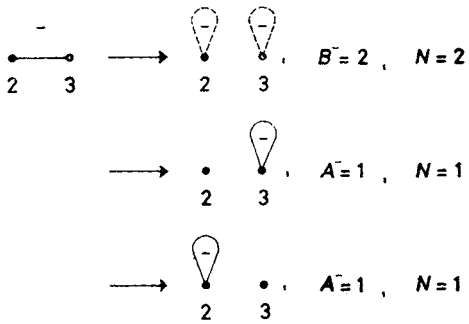
10. $J = 1$, $N1 = 0$, $N2 = 1$
11. No change
12. No change
14. No change
10. $J = 2$, $N1 = 0$, $N2 = 0$
11. No change

12. $G3'$:  ; $D = D + 2 = 4$

10. $J = 3$

16. G': $(\emptyset)$, $D = D + 1 = 5$

17. G':  , $D = D + 1 = 6$

18. 

19. $RD = D + N_{min.} = D + 1 = 7$

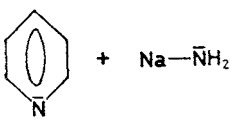
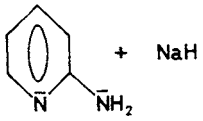
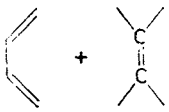
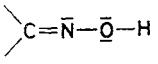
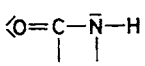
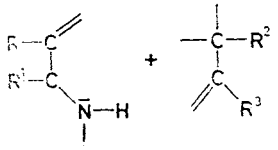
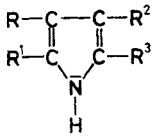
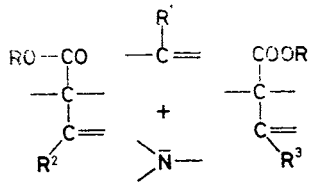
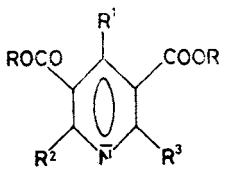
Examples of pairs of isomeric synthons along with their *RD* obtained by the CRD algorithm are given in Table II. The algorithm was implemented in FORTRAN on a PDP 11/34 computer and in PL/1 on an EC 1033 computer[8]. Experience showed that even in the most unfavourable cases the computation times for actual chemical systems ($RD < 13$) are reasonably low (less than 5 s CPU on the EC 1033 for $RD = 12$).

CONCLUSIONS

The reaction distance was given attention for several reasons. First, an alternative to the chemical distance[9] *CD*, it models the chemical reaction considerably better from the kinetic point of view. From this standpoint, *RD* appears to be a necessary auxiliary tool in building up deductive models in reaction mechanism suggestions. Second, as mentioned, the model is designed to serve for chemical synthesis purposes, where substances "near to" the given product or starting substance will be sought as potential precursors for retrosynthesis or successors for forward synthesis.

TABLE II

Pairs of isomeric *synthons* $S(A)$ $S'(A)$ and their RD (steps that do not concern the skeleton shown are not included)

$S(A)$	$S'(A)$	RD	Comment
		4	Chichibabin reaction
		6	Diels–Alder reaction
		8	Beckmann rearrangement
		12	Knorr sythesis
		21	Hantsch synthesis

An interesting problem, mathematical by nature but chemical in effect, is that associated with the principle of minimal chemical distance $(PMCD)$[10]. This problem is not solved by the *synthon* model because $PMCD$ is a problem of indexing rather than distance, and so the principle of minimal reaction distance $(PMRD)$ remains in general the same problem as $PMCD$. Anyway, RD will be an important heuristics in the computer planning of chemical syntheses.

Thanks are due to Dr L. Matyska (Institute of Pure Chemicals, Lachema, Brno) for CRD *algorithm implementation and testing on an* EC 1033 *computer in* PL/1 *language.*

REFERENCES

1. Koča J.: Collect. Czech. Chem. Commun. *53*, 1007 (1988).
2. Koča J.: Collect. Czech. Chem. Commun. *53*, 0000 (1988).
3. Kratochvíl M., Koča J., Kvasnička V.: Chem. Listy *79*, 807 (1985).
4. Koča J., Kratochvíl M., Kunz M., Kvasnička V.: Collect. Czech. Chem. Commun. *49*, 1247 (1984).
5. Koča J., Kratochvíl M., Matyska L., Kvasnička V.: Collect. Czech. Chem. Commun. *51*, 2637 (1986).
6. Kvasnička V., Kratochvíl M., Koča J.: Collect. Czech. Chem. Commun. *48*, 2284 (1983).
7. Kučera L.: *Kombinatorické algoritmy*. SNTL, Prague 1983.
8. Matyska L.: *Program* CRD. J. E. Purkyně University, Brno 1987.
9. Dugundji J., Ugi I.: Fortschr. Chem. Forsch. *39*, 19 (1973).
10. Jochum C., Gasteiger J., Ugi I., Dugundji J.: Z. Naturforsch. *37b*, 1205 (1982).
11. Hrnčiar P.: *Organická chémia*. SPN, Bratislava 1982.

Translated by P. Adámek.