
TOPOLOGICAL ASPECTS OF CHEMICAL REACTIVITY. ON THE SIMILARITY OF MOLECULAR STRUCTURES

Robert PONEC

*Institute of Chemical Process Fundamentals,
Czechoslovak Academy of Sciences, 165 02 Prague 6-Suchdol*

Received February 28th, 1986

A topological similarity index is introduced, allowing to characterize quantitatively the extent of reorganization of electron density in the course of chemical reactions. The possible application of this new index for the description and the classification of chemical reactions is discussed on the basis of certain analogy of similarity criteria with the so-called least-motion principle.

There is probably no other concept that contributed to the development of chemistry so remarkably as the ill-defined, qualitative concept of similarity. As a typical example of the intuitive use of this concept one can mention, *e.g.*, the Mendeleev periodic law, the disclosure of which was closely connected with the effort to systematize the similarities in the properties of elements. From the intuitively interpreted meaning of similarity arises also one of the most fruitful chemical principles — the principle of analogy — on the basis of which a number of fundamental concepts was introduced. As an example let us mention, *e.g.*, the concepts of isostericity, isoelectronicity, aromaticity *etc.*, each of which reflects different aspects of similarity *e.g.* in spatial arrangement, electronic structure, physicochemical properties *etc.* From the similarity of certain structural fragments originates also the fundamental concept of LFER theory — the concept of reaction series and the substituent effect.

In spite of the unquestionable contribution of the intuitively interpreted concept of similarity to the development of chemistry, there appear still more numerous indications suggesting that further exploitation of a number of qualitative concepts requires a deeper specification of their exact meaning. This is probably the reason why numerous papers have recently appeared attempting to solve these questions by means of different statistical techniques¹⁻³, topological and information theoretical approaches⁴⁻⁷ or in terms of different quantum chemical characteristics⁸⁻¹⁴.

From this broad spectrum of theoretical tools used for the specification of different aspects of molecular similarity we restrict ourselves in this study only to the specification of similarity in electronic structure from the point of view of possible implications for the quantitative theory of chemical reactivity. Some aspects of this problem were also discussed in the study of Fratev¹⁰.

THEORETICAL

The idea of quantitative employment of similarity in electronic structure for the classification of similarity in chemical properties is probably due to Polansky¹¹ who

attempted to justify theoretically Clar's classical idea of local benzenoid regions in condensed aromatic hydrocarbons¹⁵. For this purpose Polansky introduced coefficient r_L characterizing the similarity of a given local benzenoid fragment L in a molecule to the same isolated standard in terms of the corresponding density matrices P and P_L (Eq. (1)).

$$r_L = \frac{1}{2N_L} \text{Tr} (PP_L) \quad (1)$$

Because of the normalizing factor $1/2N_L$ in which N_L denotes the number of electrons in a fragment L , the value of this coefficient varies between 0-1 measuring thus the extent of similarity.

Despite of the fact that the final results of our study are closely related to the above formula by Polansky, the basic idea of our approach does not originate from his considerations, but rather from the recently published work by Carbo¹² and Jenkins¹³. To characterize molecular similarity of two molecules A and B these authors introduced a similarity index r_{AB} defined by Eq. (2) in terms of density functions ϱ_A, ϱ_B .

$$r_{AB} = \frac{\int \varrho_A \varrho_B \, dv}{\left[\int \varrho_A^2 \, dv \right]^{1/2} \left[\int \varrho_B^2 \, dv \right]^{1/2}} \quad (2)$$

This approach was originally proposed for the purpose of rational prediction of physical and biological properties, but it is apparent that the idea itself is general enough to be extended for the classification of chemical reactivity as well. In this respect an immediate possibility appears to consider the reactant and the product of the reaction as molecules A and B in the similarity of which we are interested. Unfortunately it appears that the calculation of the corresponding indices by integration of Eq. (2) is unsatisfactory for many reasons. The most important drawback originates from the fact that the similarity index calculated according to Eq. (2) is not invariant with respect to mutual position and distance of both molecules. The analysis of the invariance problems reveals that this unsatisfactory feature is due to the presence of generally multicenter integrals of type $I_{\mu\nu\lambda\sigma}$ (Eq. (3)) in which the integration is performed over orbitals centered on both molecules A and B.

$$I_{\mu\nu\lambda\sigma} = \int \chi_{\mu}^A \chi_{\nu}^A \chi_{\lambda}^B \chi_{\sigma}^B \, dv \quad (3)$$

From the form of Eq. (3) it is also clear that these invariance problems cannot be in any way eliminated by introducing usual quantum chemical approximations, *e.g.*, of CNDO type¹². Any such approximate solution can, more or less, only reduce the

numerical difficulties but it still leaves the essence of the problem substantially untouched. Fortunately it appears that the invariance problems, escaping the satisfactory solution in the framework of usual quantum chemical techniques can be completely resolved by applying certain approximations introduced in our recent paper on topological aspects of chemical reactivity¹⁶.

In the following part the nature of the necessary topological approximations will be described in detail. For this purpose let us start with a simple example of general chemical reaction in which the reactant A is transformed into the product B. The structure of these molecules can be described in the sense of the overlap determinant method by the bonding functions Φ_A , Φ_B constructed in the usual way from individual bonds forming the irreducible core in the reaction. The individual bonds are described in terms of linear combinations of atomic orbitals χ or χ' (Eq. (4)) in which the primes distinguish again between the orbitals of the reactant and the product.

$$\varphi_i^A = \sum_{\mu} a_{\mu i} \chi_{\mu} \quad (4a)$$

$$\psi_j^B = \sum_{\nu} b_{\nu j} \chi'_{\nu} \quad (4b)$$

Up to now there is thus close similarity with the original overlap determinant method. At this stage, however, the first slight modification appears. This modification concerns the actual choice of expansion coefficients. Whereas in the case of the original overlap determinant method it was desirable to choose these coefficients as simply as possible in the form of classically localized bonds, for the purpose of quantitative specification of molecular similarity a more realistic description is preferred. From that reason we use, whenever possible, numerically more reliable characterization in terms of delocalized molecular orbitals, *e.g.*, of the HMO type. After having determined the expansion coefficients, the density functions ϱ_A and ϱ_B are introduced in a usual way (Eq. (5)).

$$\varrho_A(r) = 2 \sum_i^{\text{occ}} \sum_{\mu, \nu} a_{\mu i} a_{\nu i} \chi_{\mu}(r) \chi_{\nu}(r) = \sum_{\mu, \nu} p_{\mu\nu}^A \chi_{\mu}(r) \chi_{\nu}(r) \quad (5a)$$

$$\varrho_B(r) = 2 \sum_j^{\text{occ}} \sum_{\mu, \nu} b_{\mu j} b_{\nu j} \chi'_{\mu}(r) \chi'_{\nu}(r) = \sum_{\mu, \nu} p_{\mu\nu}^B \chi'_{\mu}(r) \chi'_{\nu}(r) \quad (5b)$$

Combining Eqs (3) and (5) one obtains the final Eq. (6) that represents just an exact transcription of the original Eq. (3) in the framework of the topological approximation.

$$r_{AB} = \frac{\sum_{\mu\nu} \sum_{\lambda\sigma} p_{\mu\nu}^A p_{\lambda\sigma}^B \int \chi_{\mu} \chi_{\nu} \chi'_{\lambda} \chi'_{\sigma} dv}{\left[\sum_{\mu\nu} \sum_{\lambda\sigma} p_{\mu\nu}^A p_{\lambda\sigma}^A \int \chi_{\mu} \chi_{\nu} \chi_{\lambda} \chi_{\sigma} dv \right]^{1/2} \left[\sum_{\mu\nu} \sum_{\lambda\sigma} p_{\mu\nu}^B p_{\lambda\sigma}^B \int \chi'_{\mu} \chi'_{\nu} \chi'_{\lambda} \chi'_{\sigma} dv \right]^{1/2}} \quad (6)$$

Problems with the invariance of this expression can be now completely avoided with the aid of the so-called assigning tables describing the mutual relation of basis sets χ and χ' . The form of these tables can be alternatively expressed in the matrix notation (Eq. (7)).

$$\chi' = \tau \chi$$

$$\begin{pmatrix} \chi'_1 \\ \chi'_2 \\ \vdots \\ \chi'_n \end{pmatrix} = \begin{pmatrix} \tau_{11} & \tau_{12} & \cdots & \tau_{1n} \\ \tau_{21} & \tau_{22} & \cdots & \tau_{2n} \\ \vdots & & & \\ \tau_{n1} & & \cdots & \tau_{nn} \end{pmatrix} \begin{pmatrix} \chi_1 \\ \chi_2 \\ \vdots \\ \chi_n \end{pmatrix} \quad (7)$$

The detailed construction of these tables can be found in the original literature¹⁶ and need not be therefore repeated. For our purposes it is sufficient only to note that the matrices τ are well approximated by diagonal matrices with the elements τ_{ii} equal to ± 1 in dependence on whether there is a change of nodal properties of the corresponding orbitals during the reaction. Owing to the special diagonal form the matrix τ fulfills not only the obvious identity $\tau^T = \tau$ but also the orthogonality relation $\tau^T = \tau^{-1}$ allowing to transform the expression for the density function ρ_B in the form (8) suggesting the way of eliminating the invariance problems.

$$\rho_B(r) = \sum_{\alpha\beta} \sum_{\lambda\sigma} \tau_{\alpha\lambda}^{-1} p_{\lambda\sigma}^B \tau_{\sigma\beta} \chi_\alpha(r) \chi_\beta(r) \quad (8)$$

All complicating multicenter integrals $I_{\mu\nu\alpha\beta}$ can be eliminated by introducing the ZDO-like approximation (9).

$$\int \chi_\mu \chi_\nu \chi_\alpha \chi_\beta \, dv \approx \delta_{\mu\alpha} \delta_{\nu\beta} \quad (9)$$

The original equation (6) is thus finally reduced to the simple form (10)

$$r_{AB} = \frac{\sum_{\mu\nu} p_{\mu\nu}^A \bar{p}_{\mu\nu}^B}{[\sum_{\mu\nu} p_{\mu\nu}^2]^A^{1/2} [\sum_{\mu\nu} \bar{p}_{\mu\nu}^2]^B^{1/2}}, \quad (10)$$

where the expression for matrix elements $\bar{p}_{\mu\nu}^B$ is given by the usual similarity relations (11).

$$\bar{p}_{\mu\nu}^B = \sum_{\lambda\sigma} \tau_{\mu\lambda}^{-1} p_{\lambda\sigma}^B \tau_{\sigma\nu} \quad (11)$$

The Eq. (10) can be further simplified by taking into account the idempotency relation (12)

$$\sum_{\mu} \sum_{\nu} p_{\mu\nu}^2 = 2 \sum_{\mu} p_{\mu\mu} = 2N, \quad (12)$$

allowing to rewrite it in the final form (13) in which the more compact matrix notation is simultaneously used. (N denotes the number of electrons in the system).

$$r_{AB} = \frac{1}{2N} \text{Tr}(\mathbf{P}_A \bar{\mathbf{P}}_B) \quad (13a)$$

$$\bar{\mathbf{P}}_B = \boldsymbol{\tau}^{-1} \mathbf{P}_B \boldsymbol{\tau} \quad (13b)$$

Comparing Eq. (13a) with Eq. (1) the close similarity of both expressions is immediately seen. In the light of this similarity, Eq. (13) can be regarded as a generalized form of the original Polansky expression. The main advantage arising from this generalization concerns the possibility of exploitation of considerations based on similarity in the classification of chemical reactions. In the following part such a possibility will be demonstrated with a specific example of concerted pericyclic reactions.

RESULTS AND DISCUSSION

As a simple example let us discuss in detail the electrocyclic transformation of 1,3-butadiene to cyclobutene. The structure of both molecules is described by bonding functions Φ_A and Φ'_B . In the sense of the above discussion the function Φ_A describing the butadiene is constructed from HMO π molecular orbitals. On the other hand, the function Φ'_B describing the cyclobutene is described classically in terms of localized bonds (Eq. (14)).

$$\begin{aligned} \Phi_A &= |\varphi_1 \bar{\varphi}_1 \varphi_2 \bar{\varphi}_2| \\ \Phi'_B &= |\pi'_{23} \bar{\pi}'_{23} \sigma'_{14} \bar{\sigma}'_{14}| \end{aligned} \quad (14)$$

The corresponding density matrices \mathbf{P}_A and \mathbf{P}_B are then given by Eq. (15).

$$\mathbf{P}_A = \begin{pmatrix} 1 & 0.894 & 0 & -0.447 \\ 0.894 & 1 & 0.447 & 0 \\ 0 & 0.447 & 1 & 0.894 \\ -0.447 & 0 & 0.894 & 1 \end{pmatrix}_x \quad (15a)$$

$$\mathbf{P}_B = \begin{pmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}_x \quad (15b)$$

The next step now consists in transforming the product density matrix \mathbf{P}_B from the basis of atomic orbitals $\boldsymbol{\chi}'$ into the basis $\boldsymbol{\chi}$ serving simultaneously to the descrip-

tion of the matrix P_A . This transformation is described by the similarity relations (13b). The form of the corresponding matrices τ in the case of the allowed conrotatory and forbidden disrotatory cyclization is given by Eq. (16).




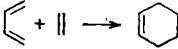
$$\tau_{\text{con}} = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & -1 \end{pmatrix} \quad (16a)$$

$$\tau_{\text{dis}} = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix} \quad (16b)$$

Using these matrices the similarity indices for both alternative reaction mechanisms can be calculated and their values are given in Eq. (17).

$$r_{\text{con}} = 0.723, \quad r_{\text{dis}} = 0.500 \quad (17)$$

TABLE I
Calculated values of the similarity index for several selected pericyclic reaction

Reaction	r_{con}	r_{dis}
	0.723	0.50
	0.658	0.758
	0.794	0.737
Reaction	$r_{\text{s+a}}$	$r_{\text{s+s}}$
$2 \parallel \rightarrow \square$	0.5	0.5
	0.574	0.574

These values immediately suggest that the similarity of the reaction partners is larger for the conrotatory than for the disrotatory cyclization. This surprising conclusion can be given an interesting interpretation arising from certain parallelism with the so-called least-motion principle¹⁷⁻²¹. This principle, originally formulated by Rice and Teller^{20,21}, connects the condition of an easy course of the reaction with the requirement of minimum changes in both the spatial arrangements and electronic configurations of the reacting molecules. All attempts to apply this principle has been so far restricted only to the first of these conditions whereas the second was completely ignored. The above introduced similarity index represents an interesting characteristics just from the point of view of this second, usually neglected, condition since it quantitatively characterizes the extent of reorganization of the electron configurations. In the sense of such an interpretation the large value of similarity index suggests small requirements on reorganization and *vice versa*. In the above case of butadiene cyclization it means that the least-motion criterion, characterized by the similarity index, prefers the conrotatory reaction that is also allowed by the Woodward-Hoffmann rules. Such a parallel between the Woodward-Hoffmann approach and the similarity criterion is not restricted only to the above case of the butadiene cyclization, but is probably more general. As demonstrated by the data summarized in Table I, its validity can be extended for the whole class of electrocyclic reactions at least. On the other hand it probably fails to discriminate between the allowed and forbidden mechanisms for cycloaddition reactions.

Despite of this limitation we believe that the proposed similarity index can contribute to deeper understanding of the factors underlying different approaches to the formulation of selection rules in chemical reactivity, and also that it opens the possibility of further quantitative exploitation of the similarity concept in chemistry.

REFERENCES

1. Anderberg M. R.: *Cluster Analysis for Application*. Academic Press, New York 1973.
2. Jurs P. C. and Isenhour Th. L.: *Chemical Applications of Pattern Recognition*. Wiley, New York 1975.
3. Wold S., Sjöström M.: *Chemometrics, Theory and Applications* (B. Kowalski, Ed.). Am. Chem. Soc. Symp. Ser. 52, 243 (1977).
4. Motoc I., Balaban A. T.: *Rev. Roum. Chim.* 26, 593 (1981).
5. Bonchev D.: *Match* 67, 65 (1979).
6. Balaban A. T.: *Pure Appl. Chem.* 55, 199 (1983).
7. Bonchev D., Mekenyan O., Trinajstić N.: *Int. J. Quant. Chem.* 17, 845 (1980).
8. Fratev F., Polansky O. E., Bonchev D., Enchev V.: *J. Mol. Struct. (THEOCHEM)* 88, 105 (1982).
9. Fratev F., Bonchev D., Enchev V.: *Croat. Chim. Acta* 53, 545 (1980).
10. Fratev F., Polansky O. E., Mehlhorn A., Monev V.: *J. Mol. Struct.* 56, 245 (1979).
11. Polansky O. E., Derflinger G.: *Int. J. Quant. Chem.* 1, 379 (1967).
12. Carbo R., Leyda L., Arnau M.: *Int. J. Quant. Chem.* 17, 1185 (1980).

13. Bowen-Jenkins P. E., Cooper L. D., Richards W. G.: *J. Phys. Chem.* **89**, 2195 (1985).
14. Graovac A., Gutman I., Randić M., Trinajstić N.: *J. Am. Chem. Soc.* **95**, 6267 (1973).
15. Clar E.: *The Aromatic Sextet*. Wiley, London 1972.
16. Ponec R.: *This Journal* **49**, 455 (1984).
17. Ehrenson S.: *J. Am. Chem. Soc.* **96**, 3784 (1974).
18. Hine J.: *J. Org. Chem.* **31**, 1236 (1966).
19. Tee O. S.: *J. Am. Chem. Soc.* **91**, 7144 (1969).
20. Rice F. O., Teller E.: *J. Chem. Phys.* **6**, 439 (1938).
21. Rice F. O., Teller E.: *J. Chem. Phys.* **7**, 199 (1939).

Translated by the author.