

# The Concept of Molecular Similarity and the Angle of Total $\Pi$ -Electron Energy

Lemi TÜRKER

*Middle East Technical University, Department of Chemistry,  
Ankara-TURKEY*

Received 18.7.1995

The concept of molecular similarity is developed and a novel similarity index based on the cosine of the angle of total  $\Pi$ -electron energy is proposed to interrelate structural variations to operant properties of molecules isoconjugate with alternant hydrocarbons.

## Introduction

Molecular similarity constitutes the basis of structure-activity relationships. The idea behind it is that structurally related compounds should exhibit similar behavior as long as the property being considered is structure dependent. In the literature there exist many studies based on the use of a reference structure to compare the molecular shapes which are quantitatively or qualitatively correlated with certain activity<sup>1-7</sup>. Recently, quantitative structure-activity relationship (QSAR) studies involving topological indices have been quite popular<sup>8,9</sup>. The topological indices are certain numerical quantities based on various invariants or characteristics of molecular graphs<sup>8,10,11</sup>.

In the present study, within the constraints of the Hückel (HMO) and perturbation molecular orbital (PMO) theories a formula for the angle of total  $\Pi$ -electron energy of an hetero system, isoconjugate with an alternant hydrocarbon is derived which is proposed as a molecular similarity index.

## Theory

Which chemical compounds are chemically similar and how similarity can be determined are very difficult to answer without considering the general principles of the perturbation theory. Actually, chemical behavior of compounds relative to each other is determined by relatively small differences in energy or some other chemical property. Hence, the similarity of molecules is a relative concept. Namely, a logical assertion such as if  $A \simeq B$  and  $A \simeq C$  then  $B \simeq C$  among compounds A, B and C may not be always valid. Hence, in a scientific work, then concept of “operative resemblance” is more meaningful than “absolute resemblance”.

Now, consider a set of N compounds  $S$ , such that

$$S = \{S_i : i \in N\}, \quad (1)$$

where each member  $S_i$  is characterized by a set  $K_i$  which possesses property classes <sup>12</sup>  $M_j, M_k$ , etc. such as

$$K_i = M_j \cup M_k \cup \dots \quad (2)$$

and

$$M_i \cap M_j = \phi \quad \text{for } i \neq j \quad (3)$$

If  $K_i \cap K_l = \phi$  for compounds  $S_i$  and  $S_l$  then definitely they do not have any common property. Such compounds are called "nontolerant compounds" and do not have any operative or absolute resemblance <sup>12</sup>. On the other hand, if  $K_i \cap K_l = T$  such that  $T \subseteq K_i$  and  $T \subseteq K_l$  then sets  $K_i$  and  $K_l$  are tolerant. Set  $T$  may include a collection of operant properties which are required for a certain activity to occur.

For the quantitative treatment of an operant property, many descriptive parameters of physico-chemical or topological origin may be used. The success of a QSAR study is of course based on how descriptive the parameters used are.

Within the constraints of HMO approach, the angle of total  $\pi$ -electron energy defined as <sup>13</sup>

$$Q = \arccos E/2 (ne)^{1/2} \quad (4)$$

is thought to be a good descriptive parameter for operant properties dictated by topological as well as quantum-chemical factors because  $E$ , the total  $\pi$ -electron energy <sup>14</sup>, is related to molecular orbital energies and the  $(ne)$  term (where  $n$  and  $e$  are half the number of atoms and the number of bonds involved in conjugation, respectively) reflects the gross topology of the system. Since, in HMO treatment the molecular orbital energies are interrelated to adjacency of atoms, then  $Q$  inherently should reflect the fine topological characteristics of the system.

Now, consider an alternant hydrocarbon,  $G(2n, e)$ , and its heteroanalog  $G'(2n, e)$ . Let  $X_1, X_2, \dots, X_n$  and  $Y_1, Y_2, \dots, Y_n$  be occupied molecular orbital energies of  $G$  and  $G'$ , respectively. In perturbation molecular orbital theory,  $E$  of the perturbed system,  $E'$ , is expressed as <sup>14</sup>.

$$E' = E + \sum_{j=1}^n \sum_p b_j c_{jp}^2 \delta\alpha_p + 2 \sum_{j=1}^n \sum_{pr} b_j c_{jp} c_{jr} \delta\beta_{pr}, \quad (5)$$

where  $b_j$  is the occupancy number of the  $j^{th}$  molecular orbital,  $c_{jp}, c_{jr}$  are then eigenvectors of the centers subject to perturbational effects and whereas  $\delta\alpha_p$  and  $\delta\beta_{pr}$  are the perturbations in Coulomb and resonance integrals <sup>14</sup>, respectively. Inserting  $E = \sum_{j=1}^n b_j X_j$ ,  $\delta\alpha_p = h_p \beta$  and  $\delta\beta_{pr} = k_{pr} \beta$  into Eq.5 one obtains

$$E' = \sum_{j=1}^n b_j X_j + \sum_{j=1}^n \sum_p b_j c_{jp}^2 h_p \beta + 2 \sum_{j=1}^n \sum_{pr} b_j c_{jp} c_{jr} k_{pr} \beta, \quad (6)$$

where  $h_p$  and  $k_{pr}$  are heteroatom parameters and  $\beta$  is the resonance integral. expressing  $E'$  and  $X_j$  in  $\beta$  units and assuming the perturbed and unperturbed systems are all closed shell systems i.e.,  $b_j = 2$ , Eq. 6 can be simplified. Further simplification is possible by neglecting the effect of perturbations on the bond order between centres  $p$  and  $r$  (the last term in Eq. 6). Thus,

$$E' = 2 \sum_{j=1}^n X_j + 2 \sum_{j=1}^n \sum_p c_{jp}^2 h_p \beta. \quad (7)$$

Now, let  $E_n$  be an  $n$ -dimensional Euclidean linear space<sup>15</sup> and vectors  $A$  and  $B$  defined as

$$A(Y_1, Y_2, \dots, Y_n) \quad \text{and} \quad B(1, 1, \dots, 1),$$

Then, the following equations hold for the scalar product<sup>16</sup> of these vectors:

$$(A, B) = \sum_{i=1}^n Y_i \quad (8)$$

$$(A, B) = (n \sum_{i=1}^n Y_i^2)^{1/2} \cos Q', \quad (9)$$

where  $Q'$  is the angle between vectors  $A$  and  $B$ . If  $Y_1$  through  $Y_n$  stand for the occupied molecular orbital energies of  $G'$ , the perturbed system of  $G$ , then it is quite easily proved that

$$E' = 2(n \sum_{i=1}^n Y_i^2)^{1/2} \cos Q'; \quad (10)$$

Since, the molecular orbital energies of  $G'$  is given as<sup>14</sup>.

$$Y_i = X_i + \sum_p c_{ip}^2 h_p \quad (11)$$

Inserting Eq. 11 into Eq. 10 then combining Eqs.10 and 7 then finally substituting  $E = 2 \sum_{i=1}^n X_i$  and  $\sum_{i=1}^n X_i^2 = e$ , (see ref. 10) one obtains

$$\cos Q' = \frac{E + 2 \sum_{i=1}^n \sum_p c_{ip}^2 h_p}{2(n(e + 2 \sum_{i=1}^n X_i \sum_p c_{ip}^2 h_p + \sum_{i=1}^n (\sum_p c_{ip}^2 h_p)^2)^{1/2}}. \quad (12)$$

Applying the pairing theorem<sup>17</sup> for perturbed centers

$$2 \sum_{i=1}^n c_{ip}^2 = \sum_{i=1}^{2n} c_{ip}^2 = 1 \quad (13)$$

and assuming the perturbation is monocentric in type, Eq. 12 can be simplified into Eq. 14, we obtain

$$\cos Q' = \frac{E + h_p}{2(n(e + 2h_p \sum_{i=1}^n c_{ip}^2 + h_p^2 \sum_{i=1}^n c_{ip}^4)^{1/2}} \quad (14)$$

## Results and Discussion

Eqs. 12 and 14 enable one to calculate the angle of total  $\pi$ -electron energy of a perturbed system via molecular orbital characteristics of its unperturbed isoconjugate system (an alternant hydrocarbon). Note that as  $h_p \rightarrow 0$ , Eqs. 12 and 14 yield  $\cos Q$  of the unperturbed system. Hence,

$$\lim_{h_p \rightarrow 0} \frac{\cos Q'}{\cos Q} = 1. \quad (15)$$

Thus, Eq.15 encouragingly implies that the ratio  $\cos Q'/\cos Q$  can be used as an index of molecular similarity in QSAR studies such that the ratio closer to unity more similar the compounds are. Consequently, their operant properties dictated by topological and quantum-chemical factors tend to be identical.

Although, the scope of application of the index proposed above should be greater than it is foreseen, presently  $pK_a$  values of some monoazaacenes and benzacenes are considered as examples for the applicability and success of the similarity index. Table 1. tabulates  $pK_a$  values of some monoazaacenes and benzacenes and the respective  $\cos Q'/\cos Q$  ratios. Note that  $\cos Q$  stand for the parent hydrocarbon (unperturbed system). The compounds in the table fall into two sets of isomeric azabenzenoids. The similarity index predicts that the basicity order should be  $2 > 1$  and  $3 > 4 > 5$  as compared to the corresponding isoconjugate hydrocarbon anions, which of course are very basic <sup>17,18</sup> ( $pK_a > 25$ ). As it is seen in Table 1, the predicted order of basicity is closely followed by the experimental order, implying that the similarity index is quite promising, at least for  $pK_a$  studies.

**Table 1.**  $pK_a$  values and  $\cos Q'/\cos Q$  ratios of some monoazaacenes and benzacenes

No	compound	$pK_a^{19}$	$\cos Q'/\cos Q$
1	Quinoline	4.94	1.0068443
2	Isoquinoline	5.14	1.0058731
3	5,6-Benzquinoline	5.15	1.0033085
4	7,8-Isobenzquinoline	4.25	1.0047277
5	7,8-benzquinoline	4.25	1.0052566

$pK_a$  values in  $H_2O$  (20°C).  $h_p = 0.5$  is used <sup>20</sup>

## Conclusion

The similarity index proposed above is based on the comparison of cosines of the angle of total  $\pi$ -electron energies of the perturbed and unperturbed systems. Although, the PMO approach has been used above to develop the idea of the molecular similarity based on  $\cos Q$  values, one is not necessarily restricted to used Eq. 12 (or Eq. 14) to calculate  $\cos Q'$  value because alternatively it can be calculated by solving the secular determinant for the perturbed system, getting molecular orbital energies and using Eq. 10.

One the other hand, the success of the index within a set of compounds isoconjugate with the same parent hydrocarbon is based on how realistic heteroatom parameters are selected. Because of that the success within a set of isomeric compounds for which the same heteroatom parameters are used, should be greater.

## References

1. A. J. Hopfinger, **J. Am. chem. Soc.**, 7196-7201 (1980).
2. I. Motoc and O. Dragomir, **Match**, **12**, 117-126 (1981).
3. J. E. Amoore, **Nature**, **214**, 1095-1099 (1967).
4. Z. Simon, I. I. Badilescu and T. Racovitan, **J. Theor. biol.**, **66**, 485-489 (1979).

5. W. P. Purcell and B. Testa, "Biological Activity and chemical Structure" (Ed. J. A. Keverling Buisman) p. 269, Elsevier, Amsterdam (1977).
6. L. Türker, E. Şener, I. Yalçın, U. Akbulut and I. Kayalıdere, **Scientia Pharmaceutica**, 58, 107-113 (1990).
7. E. Şener, H. Turgut, I. Yalçın, I. Ören, L. Türker, N. Çelebi and A. Akın, **Int. J. Pharmaceutics**, 110, 109-115 (1994).
8. A. T. Balaban, I. Motoc, D. Boncev and O. Mekenyan (Ed. F. L. Boschke) Topics in Current Chem. 114, pp 21-25, Springer-Verlag, Berlin, (1983).
9. L. B. Kier and L. H. Hall, "Molecular Connectivity in Chemistry and Drug Design", Academic Press, New York (1986).
10. I. Gutman and O. E. Polonsky, "Mathematical Concepts in Organic Chemistry" Springer-Verlag, Berlin (1986).
11. H. Hosoya, **J. Chem. Docum.**, 12, 181-185 (1972).
12. J. A. Schreider, "Equality, Resemblance and Order", Mir Pub. Moscow (1975).
13. L. Türker, **Match**, 28, 261-276 (1992).
14. E. Heilbronner and H. Bock, "The HMO Model and Its application" Wiley, London (1976).
15. N. V. Efimov and E. R. Rozendron, "Linear algebra and Multidimensional Geometry," Mir Pub., Moscow (1975).
16. H. D. Ikramov, "Linear algebra", Mir Pub., Moscow (1983).
17. M. J. S. Dewar, "The Molecular Orbital Theory of Organic Chemistry", McGraw-Hill, New York (1969).
18. M. J. S. Dewar and R. C. Dougherty, "The PMO Theory of Organic Chemistry", Plenum-Rosetta, New York (1975).
19. A. Albert, R. Goldacre, J. Phillips, **J. Chem. Soc.** 2, 2240- 2249 (1948).
20. A. Streitwieser, Jr., "Molecular Orbital Theory for Organic Chemists", Wiley, New York (1961).