Turkish Journal of Chemistry

Volume 26 | Number 3 Article 5

1-1-2002

Application of Quantitative Structure-Retention Relationships (QSRR) to a Set of Organic Bromo and Nitrile Derivatives

GUSTAVO P. ROMANELLI

JUAN CARLOS AUTINO

EDUARDO A. CASTRO

Turk J Chem
26 (2002) , 335 – 343.

Application of Quantitative Structure-Retention Relationships (QSRR) to a Set of Organic Bromo and Nitrile Derivatives

Gustavo P. ROMANELLI, Juan Carlos AUTINO
LADECOR, Departamento de Química, Facultad de Ciencias Exactas, Universidad Nacional de La Plata, Calles 47 y 115, 1900 La Plata-ARGENTINA
Eduardo A. CASTRO*
CEQUINOR, Departamento de Química, Facultad de Ciencias Exactas, Universidad Nacional de la Plata, C.C. 962, Calles 47 y 115, 1900 La Plata-ARGENTINA
e-mail: castro@dalton.quimica.unlp.edu.ar

Received 01.02.2001

The retention times are studied for two sets of 1-bromo-2-aryiloxyetanes and 3- aryloxypropiononitriles derivatives by means of Quantitative Structure-Retention Relationships (**QSRR**). Five quantum mechanical molecular descriptors are used to calculate the regression equations. The fitting polynomials are computed in several-variable at first, second, and third order equations. Results are only significant when resorting to several-variable formulae which seems to point out the rather complex physical chemistry nature of the property under study.

Key Words: QSRR - Quantum mechanical molecular descriptors - Retention Index - Bromo and Nitrile Derivatives

Introduction

The conception that there exists a close relationship between bulk properties of compounds and the molecular structure of those compounds is quite rooted in chemistry. This idea allows one to provide a clear connection between the macroscopic and the microscopic properties of matter, and thus has been firmly established as one of the central foundations of chemistry. Therefore, it is the basic tenet of chemistry to attempt to identify these assumed relationships between molecular structure and physical chemistry properties and then to quantify them.

Today there are two main approaches for computing physical and chemistry properties. One of them is the direct computation for each molecule through the implementation of quantum mechanical and/or statistical mechanical means[1−4]. Although the number of properties one can calculate in this way is growing steadily, it is still fairly limited, and the molecules are of modest size. The second approximation is the Quantitative Structure Activity/Property Relationships (QSAR/QSPR) theory. It is much more empirical

*Corresponding author

in nature, but it does not present the aforesaid limitations of the first method. In fact, this method provides more flexibility for computing physical, chemical and biological properties within the confines of requiring experimental data for a training set[5,6]. QSAR/QSPR methods are unquestionably of great relevance in modern chemistry and biochemistry. The key concept is to transform searches for compounds with derived properties using chemical intuition and experience into a mathematically quantified and computerised form. Once a satisfactory correlation between structure and property is found, any number of compounds, including those not yet synthesised, can be readily screened on the computer in order to select chemical structures with the desired properties. It is then possible to select the most promising compounds to synthesise and test in the laboratory. Accordingly, the QSAR/QSPR approach conserves resources and accelerates the process of development of new molecules for use as drugs, materials, or additives or for any other purpose. The recent exponential growth in the number of papers dealing with QSAR/QSPR studies clearly demonstrates the rapid progress in this area.

The retention index (**RI**) concept, first proposed by E. Kovats[7], has turned out to be a very useful tool for the presentation and interpretation of chromatographic data. The main advantage of RIs is the possibility of their calculation by numerous methods[8] for comparison with experimental data from chromatographic and/or chromato-spectral methods of organic compound identifications. The calculation of RIs is very important for the formation of GC databases[9] because the number of experimentally measured values is not as large as the number of standard mass spectra[10]. The realisation of this possibility demands the attainment of a satisfactory degree of precision of calculated RI values, so that the search for new methods of calculating RIs with high precision and quality control of these data is very topical at present.

Quantitative Structure-chromatographic Retention Relationships (QSRR) are among the most extensively studied manifestations of Linear Free-Energy Relationships (LFER). These are the statistically derived relationships between the structures of solutes and their chromatographic retention[8]. Using QSRR, the chromatographic column can be regarded as a "free-energy transducer", translating differences in chemical potentials of solutes resulting from differences in their structures to the chromatographic RIs. If statistically significant QSRR are derived and if these equations approximate the experimental RIs for a structurally representative set of model solutes, it is possible to define the dominant factors which determine the interactions of solute molecules with the chemical entities forming the chromatographic system[11,12]. In other words, QSRR analysis can provide an insight into the molecular mechanism of chromatographic retention phenomenon in a given HPLC system[13].

One main approach to QSRR employs as independent variables in correlation equations the structure descriptors provided by the computational chemistry. Having good QSRR equations with such descriptors, one can predict RIs for any given structural formula. It is also possible to assign physical sense to the more commonly used theoretical descriptors. This, in turn, facilitates interpretation of the separation mechanism operating in a given HPLC system[14].

Quantum-chemical methods and molecular modelling techniques enable the definition of a large number of molecular and local quantities characterising the reactivity, shape, and binding properties of complete molecules as well as of molecular fragments and substituents. Because of the large well-defined physical information content encoded in many theoretical descriptors, their use in the design of a training set in a QSAR/QSPR study presents two main advantages: a) the compounds and their various fragments and substituents can be characterised on the basis of their molecular structure only; and b) the proposed mechanism of action can be directly accounted for in terms of the chemical reactivity of the compounds under

study[15]. Consequently, the derived QSAR/QSPR model will include information regarding the nature of the intermolecular forces involved in determining the biological activity or physical property of the compounds in question.

In a previous paper[16] we applied improved QSRR to study the molecular mechanism of stationary phases for HPLC in order to complement the identification of the main structural factors determining the interactions of analyte molecules with chemical components of chromatographic systems. We used two sorts of independent variables: structural descriptors and linear free-energy relationship descriptors and the combined employment of them allowed us to get quite satisfactory second-order regression equations.

The aim of this paper is to resort to a well known set of molecular descriptors to predict retention times (t_r) of several organic bromo and nitrile derivatives through the employment of QSRR and to compare the theoretical results with available experimental data. A central issue of this study is to find out which are the more suitable molecular descriptors for this physical property and then to discuss the main driving factors determining the intimate mechanism/s governing the chromatographic retention phenomena.

Molecular Sets

3-Aryloxypropiononitriles

3-Aryloxypropiononitriles **(1)** are important reactive intermediates in the synthesis of compounds of technological and biopharmacological use, such as enzymatic inhibitors[17], juvenile hormones for plague control[18,19], biocides[20,21] , compounds of current use in the leather industry[22] and wine pigments with antiprotozoal activity[23].

(1) Ar-O-CH_2-CH_2-CN

Experimental retention times were measured by one of us (GPR) for a set of 14 molecules[24] with Ar = C_6H_5, 2-BrC_6H_4, 2-IC_6H_4, 2-MeC_6H_4, 2-MeOC_6H_4, 2-$NO_2C_6H_4$, 3-ClC_6H_4, 3-$NO_2C_6H_4$, 4-MeOC_6H_4, 4-$NO_2OC_6H_4$, 2-Br, 4-Cl-C_6H_3, 2-Br, 4-Me- C_6H_3, 1- naphthyl and 2-naphthyl.

1-Bromo-2-Aryloxyethanes

1-Bromo-2-Aryloxyethanes **(2)** are also important reactive intermediates in the synthesis of compounds of technological and pharmacological use[17,19,21,25−32].

(2) Ar-O-CH_2-CH_2-Br

Experimental retention times were also experimentally determined by one of us (GPR) for a set of 14 molecules[24] with Ar = C_6H_5, 2-BrC_6H_4, 2-IC_6H_4, 2-MeC_6H_4, 2-MeOC_6H_4, 2-$NO_2C_6H_4$, 3-ClC_6H_4, 3-$NO_2C_6H_4$, 4-MeOC_6H_4, 4-$NO_2OC_6H_4$, 2-Br, 4-Cl-C_6H_3, 2-Br, 4-Me-C_6H_3, 1-naphthyl and 2-naphthyl.

Molecular Descriptors and Theoretical Calculations

The molecular descriptors chosen for computing retention times of several organic bromo and nitrile derivatives through the employment of QSRR relate as directly as possible to the key physical chemistry property studied here (i.e., retention time). They are van der Waals-surface-bounded molecular volume (V), the

log of the octanol-water partition coefficient (log p), polarisability (P), solvent-accessible surface bounded molecular volume (SAG), molar refractivity (RM) and molecular mass (M).

Calculation of log p is carried out using atomic parameters derived by Ghose and coworkers[33]. Computation of molar refractivity was made via the same method as log p.

Ghose and Crippen presented atomic contributions to the refractivity[34]. The solvent-accessible surface bounded molecular volume and van der Waals-surface-bounded molecular volume calculations are based on a grid method derived by Bodor et al.[35], using the atomic radii of Gavezotti[36]. The polarisability was estimated from an additivity scheme given by Miller[37], where different increments are associated with different atom types[38].

The five quantum-chemical descriptors were computed with the aid of the Software Chem Plus[38] and the calculations were run on a PC. We have made a complete regression analysis resorting to linear, quadratic and cubic relationships in several independent variables. Computations were carried out by means of the Mathematica® software[39].

Table 1. Retention time (t_r) and logarithm of the retention time (log t_r) corresponding to the 1-bromo-2-aryloxyethanes and 3-aryloxypropiononitriles.

3-Aryloxypropiononitriles

Compound number	Ar	t_r[2]	log t_r
1.1	C_6H_5	10.70	1.03
1.2	2-BrC_6H_4	16.52	1.22
1.3	2-$CH_3C_6H_4$	11.60	1.06
1.4	2-$CH_3OC_6H_4$	14.00	1.15
1.5	2-IC_6H_4	16.53	1.22
1.6	2-$NO_2C_6H_4$	21.55	1.33
1.7	3-$NO_2C_6H_4$	24.50	1.39
1.8	3-ClC_6H_4	12.75	1.11
1.9	4-$NO_2C_6H_4$	22.40	1.35
1.10	4-$CH_3OC_6H_4$	13.86	1.14
1.11	2-Br,4-ClC_6H_4	22.20	1.35
1.12	2-Br,4-$CH_3C_6H_4$	22.40	1.35
1.13	1-naphthyl	23.37	1.37
1.14	2-naphthyl	23.88	1.38

1-Bromo-2-aryloxyetanes (2)

Compound number	Ar	t_r[4]	log t_r
2.1	C_6H_5	10.00	1.00
2.2	2-BrC_6H_4	15.75	1.20
2.3	2-$CH_3C_6H_4$	10.90	1.04
2.4	2-$CH_3OC_6H_4$	12.40	1.09
2.5	2-IC_6H_4	14.40	1.16
2.6	2-$NO_2C_6H_4$	20.14	1.30
2.7	3-$NO_2C_6H_4$	23.80	1.38
2.8	3-ClC_6H_4	11.18	1.05
2.9	4-$NO_2C_6H_4$	21.70	1.34
2.10	4-$CH_3OC_6H_4$	11.90	1.08
2.11	2-Br,4-ClC_6H_4	20.61	1.31
2.12	2-Br,4-$CH_3C_6H_4$	21.00	1.32
2.13	1-naphthyl	22.92	1.36
2.14	2-naphthyl	22.40	1.35

Results and Discussion

The two molecular sets with their corresponding experimental retention time and log of the retention time (log t_r) are given in Table 1.

We made complete calculation of all possible one, two, three, four and five fitting equations of log t_r versus molecular descriptor/s at first, second, and third order polynomials. All the one variable equations, independently of the polynomial order, give very poor results, as can be seen in Table 2.

Table 2. Regression coefficients for one variable equations.

Molecular		r	
Descriptor	Linear Equation	Quadratic Equation	Cubic Equation
Bromo derivatives			
SAG	0.6803	0.6803	0.6824
V	0.7202	0.7250	0.7251
Log p	0.3250	0.3691	0.3734
RM	0.7104	0.7314	0.7322
P	0.5482	0.5552	0.5810
M	0.2372	0.7071	0.7618
Nitrile derivatives			
SAG	0.8001	0.8059	0.8196
V	0.7390	0.7461	0.7490
Log p	0.3451	0.4154	0.4191
RM	0.7740	0.8059	0.8222
P	0.5785	0.5964	0.6173
M	0.5785	0.5964	0.6173

However, when turning to several variables fitting relationships, results improved noticeably, even with first order equations. Some representative examples are given below for both molecular sets and complete results are available upon request from one of us (EAC):

Nitriles derivatives

$$\begin{array}{l} logt_r = 0.1292 + 0.02184RM \\ n = 14 r = 0.7740 EV = 0.007177 \end{array} \quad (1)$$

$$\begin{array}{l} logt_r = -1.2868 + 0.02049M + 2.682410^{-5}M^2 - 5.290610^{-8}M^3 \\ n = 14 r = 0.7957 EV = 0.07958 \end{array} \quad (2)$$

$$logt_r = -0.07813 + 0.05404RM - 0.07250P \textbf{eq. 3} n = 14 r = 0.8822 EV = 0.0402662 \quad (3)$$

$$\begin{array}{ll} logt_r = & 1.0768 + 0.5600RM - 1.4622P - 1.350010^{-3}M - 4.875510^{-3}RM^2 + \\ & +3.355110^{-2}P^2 + 7.291210^{-6}M^2 \\ n = 14 & r = 0.9739 EV = 1.580410^{-3} \end{array} \quad (4)$$

$$\begin{array}{ll} logt_r & = 7.2189 - 0.04338V - 1.0511logp + 1.1282RM - 2.1129P + 3.635010^{-5}V^2 \\ & +0.2322(logp)^2 - 0.1043RM^2 + 0.04960P^2 \\ n = 14 & r = 0.9851 EV = 0.0113 \end{array} \quad (5)$$

Bromo derivatives

$$\begin{array}{l} logt_r = 0.07638 + 0.02119RM \\ n = 14 r = 0.7104 EV = 0.01043 \end{array} \quad (6)$$

$$\begin{array}{l} logt_r = -0.913764 + 1.9510^{-3}M + 7.3810^{-5}M^2 + 1.856810^{-7}M^3 \\ n = 14 r = 0.7618 EV = 0.1061 \end{array} \quad (7)$$

$$\begin{array}{l} logt_r = -0.419585 + 0.087196RM + -0.147614P \\ n = 14 r = 0.9081723 EV = 0.0402662 \end{array} \quad (8)$$

$$\begin{array}{ll} logt_r = & 4.5212 + 0.09771RM - 0.7926P + 0.01176M + 3.420410^{-5}RM^2 + \\ & +0.01422P^2 - 2.132910^{-5}M^2 \\ n = 14 & r = 0.9902 EV = 7.0610^{-4} \end{array} \quad (9)$$

$$\begin{array}{ll} logt_r = & 1.2591 + 0.02473V + 0.5038 logp - 0.1612RM - 0.4711P + 2.018010^{-5}V^2 - \\ & -0.08661(logp)^2 + 2.595410^{-3}RM^2 + 5.798510^{-3}P^2 \\ n = 14 & r = 0.9869 EV = 0.001336 \end{array} \quad (10)$$

Thus, we note that significant results can be obtained when using two or more variables in any order. Naturally, fitting equations improve when resorting to higher-order (second and third order) polynomials.

Statistical parameters meet the usual numerical requirements to satisfy the standard acceptance conditions, so that previous equations can be considered valid means to predict the retention time of the present molecular sets. Thus, the chosen quantum mechanical descriptors are suitable to computer retention times, and when using higher-order relationships, one has an appropriate quantitative tool to analyse this property. Moreover, these results seems to show that the physical-chemistry under study (t_r) cannot be rationalised by way of just only one quantum-chemical descriptor, which may mean the rather complex nature of the underlying physical-chemistry phenomenon.

In Figures 1 and 2 we present some comparisons between theoretical and experimental results, where we can see the satisfactory agreement between both sets of data. There is a significant difference between the R values of Figures 1 and 2, but they correspond to different molecular sets and different fitting equations.

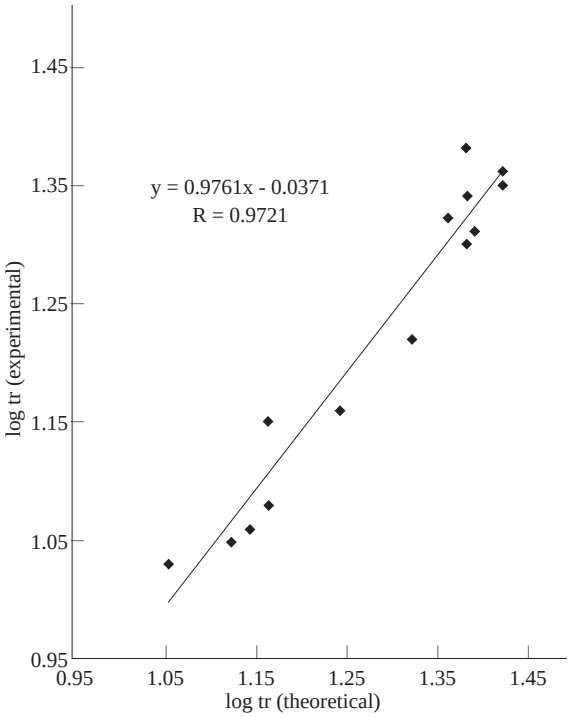

Figure 1. log tr (exp.) vs log tr (theo.) for 3-aryloxypropiononitriles (Eq. 4)

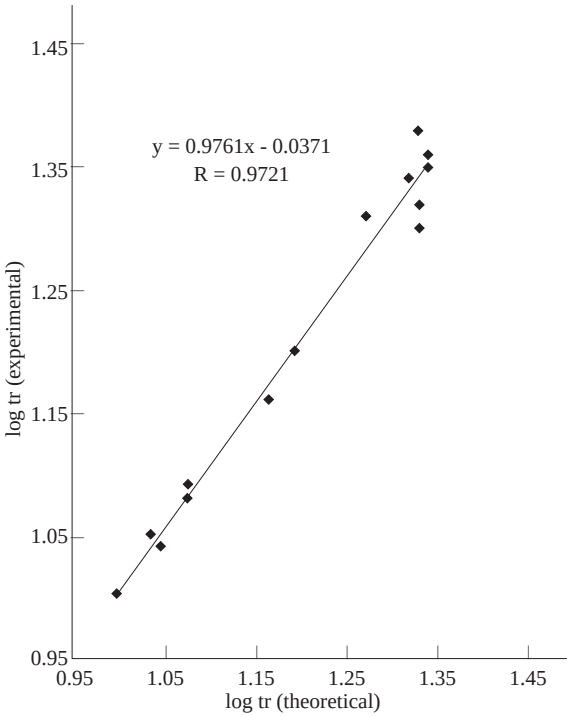

Figure 2. log tr (exp.) vs log tr (theo.) for 1-Bromo-2-Aryloxyethanes (Eq. 9)

Conclusions

One of the widely used data reduction techniques, multiple regression analysis often gives valuable insights into structure-property relationships. However, most often a direct interpretation of the results emerging from such analysis is rather difficult. It is generally understood that QSRR correlations at best suggest a parallel between the quantities involved (evaluators and responses) and do not necessarily reflect a cause-effect relationship[40]. The physical-chemistry properties studied in this work via the aforesaid quantum-mechanical descriptors M, SAG, V, log p, RM and P are dependent upon the structure in general terms and also are dependent on more subtle quantities, some of which are directly related to the chosen descriptors. This study represents the first attempt to thoroughly analyse the influence of the intimate molecular structure on retention time. The present results suggest a dependence of the quantum-mechanical chosen descriptor on retention time. In order to advance a step further, the next question to be addressed should be the degree of correlation between the variables. It is well known that the degree of interrelatedness depends upon the polynomials, and it is always present[41].

Work along this line of research is being carried out in our laboratories and results will be presented elsewhere in the near future.

References

1. E. Clementi, **"Computational Aspects for Large Chemical Systems"**, Lecture Notes in Chemistry, Vol. 19, Springer-Verlag, Berlin, 1980.
2. M. Dupuis, Ed., **"Supercomputer Simulations in Chemistry"**, Lectures Notes in Chemistry, Vol 44, Springer-Verlag, Berlin, 1986.

3. A. K. Rappé and C. J. Casewit, **"Molecular Mechanics across Chemistry"**, University Science Books, Sausalito, CA, 1997.
4. J. M. Goodman, **"Chemical Applications of Molecular Modelling"**, Royal Society of Chemistry, London, 1998.
5. M. Karelsson, V. Lobanov and A. R. Katritzky, **Chem. Rev. 96**, 1027, (1996).
6. C. Hansch and A. Leo, **"Exploring QSAR. Fundamentals and Applications in Chemistry and Biology"**, ACS Professional Reference Books, American Chemical Society, Washington, D.C.,1995.
7. E. Kováts, **Z. Anal. Chem. 181**, 351, (1961).
8. R. Kaliszan, **"Quantitative Structure-Chromatographic Retention Relationships"**, Wiley and Sons, New York, p 303, 1987.
9. I. G. Zenpurich, **Fresenius. J. Anal. Chem. 365**, 305, (1999).
10. McLafferty and D. B. Stanffer, **"Important Peak Index of the Registry of Mass Spectral Data"**, Cornell University and Palisade Corp., New York, p. 4074.
11. R. Kaliszan, in: **"High Performance Liquid Chromatography"**, P.R. Brown and R.A. Hartwick, Eds., Wiley and Sons, New York, 1989.
12. R. Kaliszan, M. Markuszewski, P. Haber, A. Nasal, T. Cserháti, E. Forgács, R.M. Gadzala-Kopciuch and B. Buszewski, **Chem. Anal.** (Warsaw) **43** 547, (1998).
13. R. Kaliszan, **J. Chromatogr. 656**, 417, (1993).
14. R. Kaliszan, **Anal. Chem. 64**, 619 A, (1992).
15. M. Cochi, M. C. Menziani, P. G. De Benedetti and G. Cruciani, **Chemom. Intell. Lab. Sys. 14**, 209, (1992).
16. G. P. Romanelli, L. F. R. Cafferata and E. A. Castro, **Khim. Fiz. 19**, 105, (2000).
17. J. F.Eggler, L.S. Melvin Jr., and A. Marfat, **Eur. Pat. Appl.**, EP 313,296, 1989; **Chem. Abstr. 111**, 194595, (1989).
18. W. S. Bowers, **U.S. Pat.**, 4,656,189, 1987, **Chem. Abstr. 107**, 115493, (1987).
19. W. S. Bowers, **Ger. Offen. Pat.**, 2,639,671, 1977; **Chem. Abstr. 87**, 68153, (1977).
20. M. Hayashi, K. Wada, and K. Munakata, **Agric. Biol. Chem.**, **47**, 2653, (1983).
21. K. Munekata, **Jpn. Kokai Tokkyo Koho**, JP 82 72, 938, 1982; **Chem. Abstr. 97**, 127279, (1982).
22. A. Minagawa and H. Matsuda, **Macromol. Chem., Rapid Commun.**, **2**, 449, (1981).
23. D. H. Barton, L. Cottier, K. Freund, F. Luini, P. D. Magnus and I. Salazar, **J. Chem. Soc., Perkin Trans. I**, 499, (1976).
24. G. P. Romanelli, Doctoral Thesis, La Plata University, 1994.
25. J. Lichtenberger, J. Coré and R. Geyer, **Bull. Soc. Chem. Fr.**, 997, (1962).
26. J. Clayton and F. Peacock, **S. African** 6908, 464 (1971); **Chem. Abstr. 77**, 15624, (1972).
27. I. Cervena, J. Metys and P. Miroslav, **Czech. CS.** 268,400 (1988), **Chem Abstr. 114**, 228955, (1991).
28. P. M. Cardirola, H. V. der Goot and H. Timmerman, **Eur. J. Chem. 27**, 571, (1992).
29. J. M. Yanni and D. A. Walsh, *U.S.* 4,950,674 (1990); **Chem. Abstr. 115**, 49413, (1991).
30. J. E. Arrowsmith, **Eur. Pat. Appl.** EP. 286,278 (1988); **Chem Abstr. 110**, 57291, (1989).

31. C. Malen, G. De Manteuil, and P. Colpaert, **Eur. Pat. Appl.** EP. 445, 026; **Chem. Abstr. 115**, 256155, (1991).
32. N. Ito, K. Takeuchi, M. Abe, and I. Tsumeo, **Fr. Demande FR** 2,520,737 (1982); **Chem. Abstr. 100**, 52587, (1984).
33. V. N. Viswanadhan, A. K. Ghose, G. N. Revankar, and R. K. Robins, **J. Chem. Inf. Comput. Sci. 29**, 163, (1989).
34. K. Ghose and G. M. Crippen, **J. Chem. Inf. Comput. Sci. 27**, 21, (1987).
35. N. Bodor, Z. Gabanyi, and C. Wong, **J. Am. Chem. Soc. 111**, 3783, (1989).
36. A. Gavezotti, **J. Am. Chem. Soc. 100**, 5220, (1983).
37. K. J. Miller, **J. Am. Chem. Soc. 112**, 8533, (1990).
38. **Chem Plus Extension for HyperChem© Modelling for Windows**^{TM}, Hypercube, Inc. Gainesville, Florida, 1994.
39. P. T. Tan , **"A Physicist's Guide to Mathematica© "**, Academic Press, New York, 1997.
40. M. Randic, **J. Mol Struct.** (THEOCHEM) **232**, 45, (1991).
41. M. Randic, **J. Comp. Chem. 14**, 363, (1993).
42. G. P. Romanelli, J. C. Autino, A. A. Vitale and A. B. Pomilio, **J. Chem. Research** (S), 386, (1993).