Pergamon

SCIENCE DIRECT®

Bioorganic & Medicinal Chemistry 11 (2003) 5045–5050

BIOORGANIC & MEDICINAL CHEMISTRY

QSAR Study on Bioconcentration Factor (BCF) of Polyhalogented Biphenyls Using the PI Index

Padmakar V. Khadikar,[a,*] Shalini Singh,[a] Dheeraj Mandloi,[b] Sheela Joshi[c] and Amrit V. Bajaj[c]

[a]*Research Division, Laxmi Fumigation & Pest Control Pvt. Ltd., 3 Khatipura, Indore 452007, India*
[b]*Institute of Engineering & Technology, D.A. University, Indore 452017, India*
[c]*School of Chemical Sciences, D.A. University, Indore 452017, India*

Received 8 May 2003; accepted 28 August 2003

Abstract—Attempt has been made to estimate the accuracy, predictive power, and domain of application of the PI (Padmakar–Ivan) index for modeling bioconcentration factor (BCF) of polyhalogenated biphenyls. Relative potential of PI index is investigated by comparing the results obtained using this index with those obtained from Wiener (W) and Szeged (Sz) indices. In addition, attempt has also been made to model hydrophobicity/lipophilicity (logP) of the polyhalogenated biphenyls using these indices. It was observed that these distance-based topological indices gave better results for modeling log BCF than logP.

Introduction

Bioconcentration is the process of accumulation of chemicals by organisms through nondietary routes.[1–3] In aquatic ecosystems, the bioconcentration factor (BCF) of an organic chemical is defined as the ratio of its concentration in a target organism to that in water at steady state.[1,4] Devillers[1] observed that the most common method for estimating BCF consists of establishing correlation between BCF and hydrophobicity (logP). The regression equations have the following general from:

$$\log\mathrm{BCF} = a\ \log\mathrm{P} + b \qquad (1)$$

Consequent to above, BCF-QSAR equations derived from logP are widely available in the literature. In some cases parabolic relationship is observed between logBCF and logP.

In our recent publications,[5–7] we observed that the Padmakar–Ivan (PI) index[8–16] is quite useful for modeling logP (hydrophobicity/lipophilicity) of chemicals and also that PI can be successfully used in place of logP.

The aforementioned results prompted us to undertake present investigation in that we have used PI index in place of logP for modeling logBCF of polyhalogenated biphenyls. The polyhalogenated biphenyls considered in the present study are quite different from those used in our earlier study.[5] Here, we have considered both bromo- and chloro-substituted biphenyls.

Since the set of compounds used herein is different from the set we used earlier we have also attempted the use of PI index for modeling hydrophobicity/lipophilicity (logP) of this new set. This is, therefore, another objective of the present investigation. We have investigated the relative potential of Wiener (W),[17] Szeged (Sz),[18–20] and PI[8–16] indices for modeling both logP and logBCF of the new set of polyhalogenated biphenyls. This, therefore, constitutes the third objective of the present study. The results are discussed below. Our present study is centered on this set of biphenyles (Table 1) as log BCF are available for this set only. These logBCF along with logP were adopted from Devillers.[3]

Results and Discussion

The obtained results are listed in Tables 1–3. Table 1 contains information regarding structural details of polyhalogenated biphenyls together with the values of their hydrophobicity/lipophilicity (logP) and bioconcentration factor (logBCF). Topological indices, namely W, Sz

*Corresponding author. Tel.: +91-731-253-1906 fax: +91-7662-242175; e-mail: pvkhadikar@rediffmail.com; (P.V. Khadikar). dheeraj_m@sancharnet.in (D. Mandloi).

0968-0896/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2003.08.028

Table 1. Polyhalogenated biphenyls, their abbreviations, logP and logBCF

Compd	Biphenyl	Abbreviation	logP	logBCF
1	Decachloro-	DCBP	8.27	3.92
2	4,4′-Dibromo-	4,4′-DBBP	5.72	4.24
3	2,5-Dichloro-	2,5-DCBP	5.16	4.00
4	3,5-Dichloro-	3,5-DCBP	5.37	3.79
5	2,2′,4,4′,6,6′-Hexabromo-	HBBP	7.20	4.66
6	2,2′,4,4′,5,5′-Hexachloro-	HCBP	6.90	4.84
7	2,2′,3,3′,4,4′,5,5′-Octachloro-	OCBP	7.67	4.33
8	2,2′,5,5′-Tetrabromo-	TetBBP	6.50	4.97
9	2,2′,3,3′-Tetrachloro-	TetCBP	6.18	4.69
10	2,2′,5,5′-Tetrachloro-	TetCBP	6.09	4.63
11	2,3′,4′,5-Tetrachloro-	Tett3CBP	6.23	4.62
12	2,4,6-Tribromo-	TriBBP	6.03	3.88
13	2,2′,5-Trichloro-	TriCBP	5.60	4.30
14	2,4,5-Tribromo-	Tri2BBP	5.90	3.78
15	2,4′,5-Trichloro-	Tri2CBP	5.79	4.63
16	2,4,5-Trichloro-	Tri3CBP	5.90	4.26

Table 2. Wiener (W), Szeged (Sz) and PI indices of polyhalogenated biphenyls (ref Table 1)

Compd	W	Sz	PI
1	907	1483	494
2	315	555	198
3	280	518	198
4	298	522	198
5	555	935	330
6	573	963	330
7	738	1224	408
8	412	700	260
9	408	692	260
10	412	700	260
11	426	728	260
12	348	700	228
13	346	594	220
14	354	612	228
15	362	626	228
16	354	612	228

Table 3. Comparison of observed and estimated logBCF of polyhalogenated biphenyls (ref Table 1)

Compd	logBCF (obsd)	LogBCF estimated using					
		W		Sz		PI	
		Est.	Res.	Est.	Res.	Est.	Res.
1	3.92	3.80	0.12	3.81	0.11	3.80	0.12
2	4.24	4.10	0.14	4.10	0.14	3.98	0.26
3	4.00	3.90	0.10	3.99	0.01	3.98	0.02
4	3.79	4.00	−0.21	4.00	−0.21	3.98	−0.19
5	4.66	4.80	−0.14	4.75	−0.09	4.78	−0.12
6	4.84	4.83	0.01	4.76	0.09	4.78	0.06
7	4.33	4.59	−0.26	4.53	−0.20	4.58	−0.25
8	4.97	4.53	0.44	4.48	0.49	4.53	0.44
9.	4.69	4.51	0.18	4.46	0.23	4.53	0.16
10	4.63	4.53	0.10	4.48	0.15	4.53	0.10
11	4.62	4.57	0.05	4.53	0.09	4.53	0.09
12	3.88	4.26	−0.38	4.48	−0.60	4.29	−0.41
13	4.30	4.26	0.04	4.22	0.08	4.29	0.01
14	3.78	4.29	−0.51	4.27	−0.49	4.29	−0.51
15	4.63	4.33	0.30	4.31	0.32	4.29	0.34
16	4.26	4.29	−0.03	4.27	−0.01	4.29	−0.03

Res., Residue, that is the difference between observed and estimated logBCF.

and PI indices, calculated for the set of compounds under study are also presented in Table 2.

The data presented in Tables 1 and 2 show that degeneracy is present in logP, logBCF, as well as in all the three topological indices (W, Sz, PI) used. However, compared to topological indices the degeneracy present in logP and logBCF in very little. The presence of degeneracy in the topological indices is due to the fact that they belong to first-generation topological indices.[21,22]

When logBCF are plotted against logP and PI index independently, in both the cases we obtained a parabolic relationship (Figs 1 and 2). Consequently, we have performed and compared the results obtained from linear as well as non-linear statistics. In case of linear relationships better results are obtained by splitting the data set into two categories: A and B. Such a splitting is done on basis of the logP values. The polyhalogenated biphenyls with logP smaller than 6 constitute the category A, while with logP larger than 6 constitute category B. This means that the set of 16 polyhalogentaed biphenyls exhibit famalial relationship. Earlier, such type of famalial correlations was observed by Rouvary and El-Basil,[23] Randic[24] as well as by us.[7] Randic[24] emphasized the importance of graph recognition in studies of this type and pointed out that chemical graphs can belong to families in the same way that plants belong to genera.

Modeling of logP

The regression analysis[25] indicated that logP can be modeled successfully using all the three topological indices (W, Sz, PI) and that in each case excellent statistics are obtained. The models are found as below:

$$\log P = 4.1639 + 0.0048\ W \quad (2)$$
$$n = 16,\ \ Sy = 0.8434,\ \ r = 0.9703,\ \ Q = 1.1505$$

$$\log P = 3.9536 + 0.0031\ Sz \quad (3)$$
$$n = 16,\ \ Sy = 0.8434,\ \ r = 0.9711,\ \ Q = 1.1514$$

$$\log P = 3.5843 + 0.0099\ PI \quad (4)$$
$$n = 16,\ \ Sy = 0.8417,\ \ r = 0.9703,\ \ Q = 1.1528$$

Here and hereafter, n represent the number of compounds, Sy is standard deviation in y means activity (logP in our case), r is correlation coefficient, and Q is the quality factor.[26] This quality factor Q is defined as the ratio of correlation coefficient (r) to the standard deviation (Sy) i.e., $Q = r/Sy$. This factor accounts for the predictive power of the model.

Based on Q values, the aforementioned results show that predictive power of W, Sz, and PI indices in modeling logP is more or less similar. This is obvious as the differences in the correlation coefficients are so minimal, that a test of significance performed using Q factor indicated the equivalence in the predictive power of

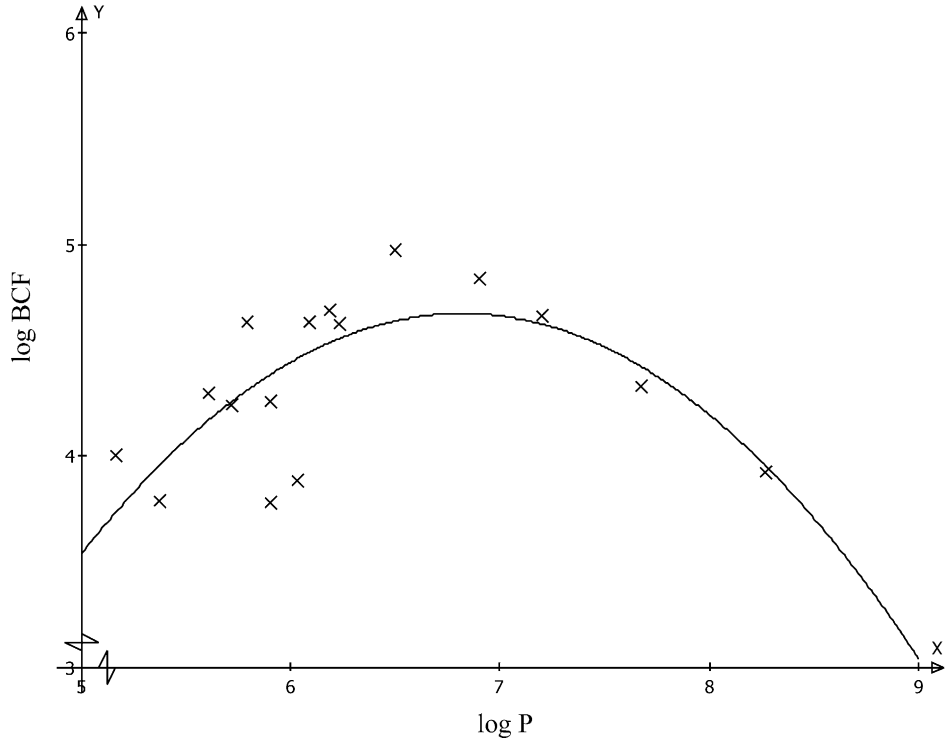

Figure 1. Correlation of logBCF with logP.

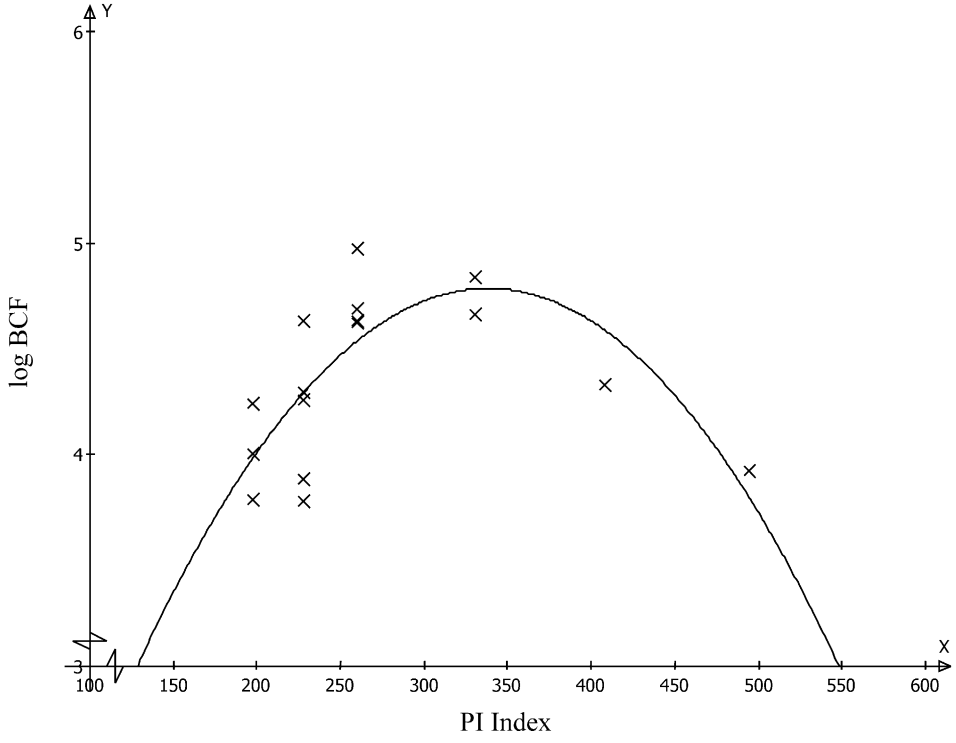

Figure 2. Correlation of logBCF with PI index.

these topological indices. Under such condition, no *t*-test is needed. Hence, we can conclude that PI, W, and Sz indices are capable of modeling logP of any type of polyhalogented biphenyls.

The statistics in the splitted category, that is under categories A and B resulted into slight improvement in the correlation coefficient (changes from 0.97 to 0.98).

In view of the above and in accordance with Devillers,[1,3] we have attempted parabolic correlations for modeling logP using W, Sz, and PI indices. The following parabolic models were obtained:

$$\log P = 2.81 + 0.0102\ W - 4.74 \times 10^{-6} W^2 \quad (5)$$
$$n = 16,\ \ Se = 0.1758,\ \ r = 0.9809,\ \ Q = 5.5796.$$

$$\log P = 2.47 + 0.0066\ Sz - 1.8000 \times 10^{-6} Sz^2 \quad (6)$$
$$n = 16,\ \ Se = 0.1635,\ \ r = 0.9836,\ \ Q = 6.0159$$

$$\log P = 1.96 + 0.0207\ PI - 1.6100 \times 10^{-5} PI^2 \quad (7)$$
$$n = 16,\ \ Se = 0.1766,\ \ r = 0.9808,\ \ Q = 5.5538$$

The aforementioned results show that slightly better statistics is obtained under parabolic correlations and that all the proposed models have similar predictive potential.

Modeling of bioconcentration factor (log BCF)

Figures 1 and 2 show that the correlation of logBCF with logP is similar to that of the correlation of logBCF with PI index. In both the cases parabolic correlations are observed. Like modeling of logP, here also we carried out: (i) simple linear regression, (ii) regression under splitted categories A and B, and (iii) parabolic regression for modeling logBCF. The results are discussed below.

Simple linear regressions gave the following models for modeling logBCF:

$$\log BCF = 4.2592 + 0.0002\ W \quad (8)$$
$$n = 16,\ \ Sx = 171.2729,\ Sy = 0.3878,$$
$$r = 0.0858,\ Q = 0.2213$$

$$\log BCF = 4.2860 + 0.0001\ Sz \quad (9)$$
$$n = 16,\ \ Sx = 267.4396,\ \ Sy = 0.3878,$$
$$r = 0.0546,\ \ Q = 0.1408$$

$$\log BCF = 4.2254 + 0.0004\ PI \quad (10)$$
$$n = 16,\ \ Sx = 82.1284,\ \ Sy = 0.3878,$$
$$r = 0.0944,\ \ Q = 0.2434$$

These results show that none of the topological indices are capable for modeling logBCF independently. In view of this failure we have attempted modeling of logBCF using logP which gave the following regression equation:

$$\log BCF = 4.3259 + 0.0022\ \log P \quad (11)$$
$$n = 16,\ \ Sx = 2.5027,\ \ Sy = 0.3760,$$
$$r = 0.0190,\ \ Q = 0.0050$$

It means that even logP is incapable of modeling logBCF of polyhalogented biphenyls in monoparametric regression.

The regressions under splitted conditions and after considering compounds **5**, **14** and **15** as outlyers gave very much improved statistics. The PI index gave results according to the following regression equations:

$$\log BCF(A) = 6.2753 - 0.0047\ PI \quad (12)$$
$$n = 4,\ \ Sx = 100.8034,\ \ Sy = 0.4834,$$
$$r = -0.9842,\ \ Q = 2.0360$$

$$\log BCF(B) = 1.9061 + 0.0103\ PI \quad (13)$$
$$n = 9,\ \ Sx = 26.8514,\ \ Sy = 0.3327,$$
$$r = 0.8335,\ \ Q = 2.5053$$

In view of the above, and following Devillers,[1,3] we have finally attempted parabolic correlations for modeling logBCF of all the 16 polyhalogenated biphenyls. The results obtained are found as under:

$$\log BCF = 1.4300 + 0.0116\ W - 9.9000 \times 10^{-6}\ W^2 \quad (14)$$
$$n = 16,\ Se = 0.2670,\ r = 0.7674,\ Q = 2.8742$$

$$\log BCF = 1.2500 + 0.0072\ Sz - 3.6900 \times 10^{-6}\ Sz^2 \quad (15)$$
$$n = 16,\ Se = 0.2999,\ r = 0.6959,\ Q = 2.3204$$

$$\log BCF = 0.1490 + 0.0274\ PI - 4.0500 \times 10^{-6}\ PI^2 \quad (16)$$
$$n = 16,\ Se = 0.2739,\ r = 0.7532,\ Q = 2.7490$$

$$\log BCF = -11.3000 + 4.6700\ \log P - 0.3430 \log P^2 \quad (17)$$
$$n = 16,\ Se = 0.2954,\ r = 0.7048,\ Q = 2.3859$$

Higher order polynomical, that is third order polynomial gave slightly better statistics such that r is improved to 0.7980, 0.7193, 0.7776, 0.7276, respectively, when W, Sz, PI, and logP are used as the correlating parameters for modeling logBCF.

The above results (eqs 14–17) indicate that under parabolic regression the predictive power of logP is not much different from that of W, Sz, and PI respectively.

In order to support our finding we have estimated logBCF from the best second order parabolic regression based on W, Sz, and PI indices (eqs 14–17). The results obtained are presented in Table 3. The residue, that is the difference between observed and estimated logBCF, supports our findings.

Conclusion

From the present study, we conclude that the topological indices W, Sz, and PI can be successfully used for modeling hydrophobicity/lipophilicity(logP) and bioconcen-

tration factor (logBCF) of the new set of polyhalogented biphenyls. Slightly better results are obtained upon splitting the compounds into two category or by performing parabolic regression. Interestingly, we observed that under splitted conditions and under second order parabolic correlations similar results are obtained.

Experimental

Hydrophobicity/lipophilicity (logP)

Hydrophobicity/lipophilicity (logP) for the set of 16 polyhalogenated biphenyls was adopted from the literature.[2]

Bioconcentration factor (logBCF)

The experimental logBCF values as reported by Devillers.[3]

Topological indices

All the topological indices are computed from the hydrogen-suppressed graphs of the compounds under study. The details of calculations are available in the literature.[25–30] However, below we give the expressions used for the calculations of topological indices used.

Wiener index (W). The Wiener index (W)[17] of a graph G is just the sum of distances of all pairs of vertices of G.

$$W = W(G) = 1/2 \sum d(v, u|G) \tag{18}$$

where d(v|G) is called the distance number (minimum distances) of vertex v and is defined as:

$$d(v|G) = \sum_{u \varepsilon V(G)} d(v, u|G) \tag{19}$$

Szeged index (Sz). The Szeged index (Sz)[18–20] of a graph G is defined as:

$$Sz = Sz(G) \sum_{e \varepsilon E(G)} \{n_1(e|G).n_2(e|G)\} \tag{20}$$

where n1(e|G) and n2(e|G) counts the vertices of G close to the vertices u and v, respectively. The vertices equidistant from both the ends of an edge are not taken into account.

PI index. The PI index[8–16] in its definition relies on the notation of edge types in the chemical graph of the molecule to be characterized and is considered as the modification of Szeged index.[18–20] The PI index is defined as:

$$PI = PI(G) = \sum_{e} [n_{eu}(e|G) + n_{ev}(e|G)] \tag{21}$$

where $n_{eu}(e|G)$ is the number of edges lying closer to the vertex u than the vertex v. The meaning of n_{ev} is analogous. Edges equidistant from both ends of the edge uv are not counted (taken into account) for the calculation of PI index.

Regression analysis. All the regression analyses[25] were carried out using Equation grapher with regression analyzer software.

Acknowledgements

The authors are thankful to Professor Istvan Lukovits for providing softwares. One of the authors (P.V.K.) is thankful to Professor Ivan Gutman for introducing him to this fascinating field of Topology and Graph Theory. The authors are also thankful to the Referees for their excellent suggestions thus improving the quality of our paper.

References and Notes

1. Bintein, S.; Devillers, J.; Karcher, W. *SAR-QSAR Environ. Res.* **1993**, *1*, 29.
2. Barron, M. G. *Environ. Sci. Technol.* **1990**, *24*, 1612.
3. Devillers, J.; Bintein, S.; Domine, D. *Chemosphere* **1996**, *33*, 1047.
4. Hamelink, J. L., Current Biocorrelation Test Methods and Theory. In *Aquatic Toxicology and Hazard Evaluation*; Mayer, F. L., Hamelink, J. L., Eds.; ASTM STP: 1977; Vol. 634, p 149.
5. Khadikar, P. V.; Singh, S.; Shrivastava, A. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 1125.
6. Khadikar, P. V.; Agrawal, V. K.; Karmarkar, S. *Bioorg. Med. Chem.* **2002**, *10*, 3499.
7. Agrawal, V. K.; Singh, J.; Khadikar, P. V. *Bioorg. Med. Chem.* **2002**, *10*, 3981.
8. Khadikar, P. V. *Nat. Acad. Sci. Lett.* **2000**, *23*, 113.
9. Khadikar, P. V.; Karmarkar, S.; Agrawal, V. K. *Nat. Acad. Sci. Lett.* **2000**, *23*, 124.
10. Khadikar, P. V.; Karmarkar, S.; Agrawal, V. K. *Nat. Acad. Sci. Lett.* **2000**, *23*, 165.
11. Khadikar, P. V.; Karmarkar, S.; Agrawal, V. K. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 934.
12. Khadikar, P. V.; Kale, P. P.; Deshpande, N. V.; Karmarkar, S.; Agrawal, V. K. *J. Math. Chem.* **2001**, *29*, 143.
13. Agrawal, V. K.; Khadikar, P. V. *Bioorg. Med. Chem.* **2001**, *9*, 3035.
14. Khadikar, P. V.; Karmarkar, S.; Singh, S.; Shrivastava, A. *Bioorg. Med. Chem.* **2002**, *10*, 3163.
15. Khadikar, P. V.; Phadnis, A.; Shrivastava, A. *Bioorg. Med. Chem.* **2002**, *10*, 1181.
16. Khadikar, P. V.; Bajaj, A. V.; Mandloi, D. *Indian J. Chem.* **2002**, *41A*, 2065.
17. Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
18. Gutman, I. *Graph Theory Notes New York* **1994**, *27*, 9.
19. Khadikar, P. V.; Deshpande, N. V.; Kale, P. P.; Dobrynin, A.; Gutman, I.; Domotor, G. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 547.
20. Khadikar, P. V.; Kale, P. P.; Deshpande, N. V.; Karmarkar, S.; Agrawal, V. K. *Commun. Math. Comput. Chem. (MATCH)* **2001**, *43*, 7.
21. Balaban, A. T. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 23.

22. Khadikar, P. V.; Sharma, S.; Sharma, V.; Joshi, S.; Lukovits, I.; Kaveeshwar, M. *Bull. Soc. Chim. Belg.* **1997**, *106*, 767 and references therein.
23. Rouvary, D. H.; El-Basil, S. *J. Mol. Str. (Theorem)* **1988**, *165*, 9.
24. Randic, M. *Theor. Chem. Acta* **1983**, *62*, 485.
25. Chatterjee, S.; Hadi, A.; Price, B. *Regression Analysis by Examples*, 3rd ed.; Wiley: New York, 2000.
26. Pogliani, L. *Amino Acids* **1994**, *6*, 141.
27. Todeschini, R.; Consonni, V. *Hand Book of Molecular Descriptors*; Wiley-VCH: Weinheim, Germany, 2000.
28. Diudea, M. V., Ed. *QSPR/QSAR Studies by Molecular Descriptors*; Nova Science: Huntington, NY, 2000.
29. Devillers, J., Balaban, A. T., Eds. *Topological Indices and Related Descriptors in QSAR and QSPR*; Gordon and Breach: New York, 2000.
30. Kier, L. B.; Hall, L. H. *Molecular Structure Description*; Academic: New York, 1999.