

A Novel PI Index and Its Applications to QSPR/QSAR Studies

Padmakar V. Khadikar* and Sneha Karmarkar

Research Division, Laxmi Pest & Fumigation Pvt. Ltd., 3 Khatipura, Indore 452 007, India

Vijay K. Agrawal

QSAR & Computer Chemical Laboratories, Department of Chemistry, A.P.S. University,
Rewa 486 003, India

Received October 10, 2000

A novel topological index, PI (Padmakar–Ivan index), is derived in this paper. The index is very simple to calculate and has disseminating power similar to that of the Wiener (W) and the Szeged (Sz) indices. The comprehensive studies show that the proposed PI index correlates highly with W and Sz as well as with physicochemical properties and biological activities of a large number of diversified and complex compounds. The proposed PI index promises to be a useful parameter in the QSPR/QSAR studies. The stability of each model is demonstrated by applying cross-validation test. Furthermore, more favorable comparison with other representative indices such as the Randic index is also made in order to establish the predictive ability of the PI index. The results have shown that in several cases the PI index gave better results.

INTRODUCTION

Application of graph theory to chemistry and to structure–property–activity (QSPR/QSAR) relationships has led to the emergence of several critical graph-theoretical indices.¹ First applications of graph-theoretical invariants in studies of structure–property relationships (QSPR) was proposed by Wiener² in 1947. However, it was after Randic³ proposed a topological index for characterization of molecular branching that dramatic expansion of studies in the area started.

Molecular graphs represent chemical molecules which are composed of two entities: graph vertices representing atoms and graph edges representing bonds.^{1,4} The common feature of graph theory-based structure descriptors is that they are derived by counting graph edges, vertices, or both. Rather simple arithmetic operations are carried out to get numerical indices. These indices are supposed to comprise information on properties/activities of the molecules.

There are two main areas of applications of structural descriptors based on molecular graphs: (i) determination of chemical structures for purposes of chemical documentation and (ii) determination of mathematical properties of chemical structures, which could next be related to physicochemical and/or biological properties of chemical compounds. The present study is mainly centered at this second application of graph theoretical descriptors.

Regardless of the descriptors used in the development of QSPR/QSAR models, all of them share in common a basic approach: molecules are represented by vectors constructed in turn by molecular parameters, which are supposed to contain relevant information about molecular structure.^{1–4}

Consequent to above, the use of topological indices in QSPR as well as QSAR studies have become of major interest in recent years, and especially the QSPR/QSAR

models have become a powerful tool for predicting numerous physicochemical properties and biological activities of organic compounds acting as drugs as well as for molecular design. Recently, several review papers on topological indices have been published.^{5–12}

So far, a great number of topological indices have been proposed. In addition to the Wiener² (W) and the Randic³ (χ) indices mentioned earlier, the other indices included are the Hosoya's index¹² (Z), Balaban's index¹³ (J), Bonchev's index¹⁴ (ID), Schultz's index¹⁵ (MTI), modification of Wiener index,¹⁶ etc. Among these indices, the most successful was the Randic³ molecular connectivity index (χ), which was further developed and applied extensively by Randic,³ Kier and Hall, and their co-workers.^{7,8}

Despite the large achievement in this field being attained, existing topological index approaches to QSPR/QSAR may need further improvement. Therefore, it is desirous to find new topological indices for modeling properties/activities of organic compounds acting as drugs. Consequently, a new topological index was introduced recently by Gutman,¹⁷ which he along with one of the present authors (P.V.K.) and co-workers named it the Szeged index and abbreviated it as Sz.¹⁸ This new index is supposed to be a modification of the Wiener index for cyclic compounds. For trees W and Sz coincides.^{17,18}

Very recently, we have made a comparative study¹⁹ using Wiener, Szeged, and molecular connectivity indices and reported that W and Sz are closely related to ${}^m\chi_R$ and that the W, the Sz, and the ${}^1\chi_R$ indices have similar modeling potential. ${}^m\chi_R$ gives slightly better results than both W and Sz. Other connectivity indices, namely ${}^0\chi_R$ and ${}^2\chi_R$, correlate poorly with the activity (log P).

In our earlier²⁰ work we reported a comparative study of the W, the Sz, and the MTI indices of cycloalkanes and demonstrated that these indices are closely related quantities, being equally capable of predicting the physicochemical

* Corresponding author phone: 91-731-531-906; e-mail: pvkhadikar@yahoo.com.

properties of cycloalkanes, while W yielded slightly better results than both Sz and MTI in developing quantitative structure–property relationships.

A few basic mathematical properties of Sz were established,^{21–43} and its certain chemical, biochemical, and environmental applications were reported.^{44–59}

The aforementioned study on the applicability of Sz index has indicated that (i) the Szeged index is basically meant for cyclic graphs; (ii) for trees Wiener and Szeged indices coincide; and (iii) all the drug molecules which were considered were graphs of fixed ring size carrying zero, one, or more tree-like attachments.

Balaban⁶⁰ has emphasized two characteristics of topological indices. First, until now no single topological index could be used universally in optimal correlations; therefore, more than hundreds of topological indices are in existence. Second, whereas any constitution of a normal covalent compound may be converted into various topological indices according to definite procedures the reverse is not true: retrieval of the chemical constitution from the numerical value of the topological index is not generally impossible but in special cases yes!

In view of the above, in the present communication we have introduced a new topological index and named it Padmakar–Ivan index and abbreviated it as PI. This newly proposed topological index, PI, had an advantage over the Szeged index (Sz) in that PI does not coincide with the Wiener index (W) for acyclic (trees) molecules. Thus, unlike the Sz index, the PI index is different for acyclic and cyclic graphs. The derived PI index is very simple to calculate and has a discriminating power similar to that of the W and Sz indices. In many cases the PI index gives better results than both the W and the Sz. Furthermore, the PI index is easy to compute, is fast on computers, and is also conceptionally simple and easy to follow on paper. The details are discussed below.

DEFINITION OF THE PI INDEX

Let G be a simple molecular graph without directed and multiple edges and without loops, the vertex and edge-sets of which are represented by $V(G)$ and $E(G)$, respectively. If e is an edge of G connecting the vertices u and v , then we write $e = uv$. The number of vertices of G is denoted by n .

The distance between a pair of vertices u , w of G is denoted by $d(u, w|G)$.

Let e be an edge of G connecting the vertices u and v , $e = uv \in E(G)$.

We define for $e = uv$ two quantities $n_{eu}(e|G)$ and $n_{ev}(e|G)$. $n_{eu}(e|G)$ is the number of edges lying closer to the vertex u than the vertex v , and $n_{ev}(e|G)$ is the number of edges lying closer to the vertex v than the vertex u . Edges equidistant from both ends of the edge uv are not counted.

The PI index, $PI = PI(G)$ of the graph G is defined as

$$PI = PI(G) = \sum_{e \in E(G)} [n_{eu}(e|G) + n_{ev}(e|G)] \quad (1)$$

The summation goes over all edges of G .

METHODOLOGY

We have done comprehensive studies on physicochemical properties and biological activities of a series of diversified

compounds, (i) alkanes, (ii) cycloalkanes, (iii) polychlorinated biphenyls (PCBs), and (iv) substituted benzoyl alcohols, and observed that the PI index promises to be a useful topological index in the QSPR/QSAR studies. The proposed models were finally tested by a cross-validation test. The results are discussed below.

The values of the PI index were computed for each series of compounds using a carbon hydrogen suppressed structure. In addition to the PI index, we have also computed the W and Sz indices with a view to study their relative correlation potential in developing QSPR as well as QSAR models. The index values were arranged according to ascending order, and the resulting data were analyzed to identify the best model. Subsequently, the topological indices (W, Sz, PI) were correlated with the concerned property/activity, and resulting correlations were subjected to regression analysis using the method of least squares.⁶¹ The needed properties and activities were adopted from earlier studies.^{62–65}

The computation of W and Sz indices were made using the Wiener-1 program made available to us by Prof. Istvan Lukovits, Hungarian Academy of Sciences, Budapest, Hungary. Similarly regression analyses were performed using the Regress-1 program supplied by Prof. Lukovits. All calculations were made on a Power Macintosh 9600/233 computer, and the results are discussed in the following sections.

RESULTS AND DISCUSSION

Case of Alkanes. The alkanes represent an especially attractive class of compounds at the starting point for the application of molecular modeling techniques. Many properties of alkanes vary in a regular manner with molecular mass and extent of branching, and because the alkanes are nonpolar, a number of complexities that arise with more polar compounds are avoided. Hence, complexities due to polarity, polarizability, and hydrogen bonding are avoided in alkanes. Thus, the physicochemical properties of alkanes are dominated by their inherent structural features, such as molecular dimension or shape. The molecular size has an impact and influence on most properties (activities), although other factors such as branching and steric factors, also have smaller influences. In the present study we are mainly concerned with the size effect and then with the effect due to branching.

Consequent to above we have first discussed unbranched alkanes, and the study was then extended to a large number of alkanes consisting of branched and unbranched structures. We first developed QSPR models based on boiling points (bp) and then considered several other properties of alkanes.^{66,67}

The boiling point (bp) of a compound is related directly to the chemical structure of the molecules. Pioneering work in applying QSPR to the bp of alkanes was done by Wiener.² Since then, there have been very extensive efforts to apply structural information to fit experimental bp. Most of this work was centered at homologous and congeneric series of compounds.

To achieve our goal we first discuss how PI indices of alkanes are computed and then establish a relationship among the W, Sz, and PI indices. Such a study will help us in developing quantitative structure–property relationships

Table 1. PI Indices of Nonbranched Alkanes, Their Physicochemical Parameters, and Log P Values^{58a}

compd no.	nonbranched alkane	<i>n</i>	PI	W	bp (°C)	log P octanol/water	MR	MV	PR	α
1	methane	1	0	0	-164.00	1.53	11.31	61.50	111.70	4.48
2	ethane	2	0	1	-88.60	2.06	15.94	78.00	151.50	6.32
3	propane	3	2	4	-42.10	2.60	20.58	94.50	191.20	8.15
4	butane	4	6	10	-0.50	3.14	25.21	111.00	231.00	9.99
5	pentane	5	12	20	36.10	3.67	29.84	127.50	270.80	11.83
6	hexane	6	20	35	68.70	4.21	34.47	144.00	310.60	13.66
7	heptane	7	30	56	98.40	4.74	39.11	160.50	350.40	15.50
8	octane	8	42	84	125.70	5.28	43.74	177.00	390.20	17.34
9	nonane	9	56	120	150.77	5.82	48.37	193.60	430.00	19.17
10	decane	10	72	165	174.12	6.35	53.01	210.10	469.70	21.01
11	undecane	11	90	220	196.80	6.89	57.64	226.60	509.50	22.85
12	dodecane	12	110	286	216.30	7.42	62.27	243.10	549.30	24.28
13	tridecane	13	132	364	235.40	7.96	66.90	259.60	589.10	26.52
14	tetradecane	14	156	455	263.70	8.50	71.54	276.10	628.90	28.36
15	pentadecane	15	182	560	270.10	9.03	76.17	292.60	668.70	30.19
16	hexadecane	16	210	680	280.00	9.57	80.80	309.10	708.40	32.03
17	heptadecane	17	240	816	292.00	10.10	85.44	325.60	748.20	33.87
18	octadecane	18	270	969	308.00	10.64	90.07	342.10	788.00	35.70
19	nonadecane	19	305	1140	320.00	11.18	94.70	358.60	825.80	37.40
20	eicosane	20	342	1330	342.70	16.54	141.03	524.70	1225.60	55.91

^a PI – Padmakar–Ivan index, W – Wiener index, bp – boiling point, log P – logarithm of partition coefficient (P) in octanol/water, MR – molar refraction, MV – molar volume, PR – parachore, α – polarizability.

(QSPR) as well as quantitative structure–activity relationships (QSAR) for the alkane series.

The PI index is defined (eq 1) in such a manner that the following results are immediate for acyclic (tree) graphs.

Result 1. If *G* is an acyclic graph containing *n* vertices then

$$PI = n^2 - 3n - 2 = (n-1)(n-2) \quad (2)$$

That is, the PI index is completely degenerate in an isomeric series of acyclic structures.

Result 2. By definition, PI = 0, for acyclic graphs when *n* = 1 and 2.

Result 3. Recall that the Wiener index *W* of a path graph⁴ *P_n* with *n* vertices is given by

$$W = W(P_n) = (1/6)n(n^2-1) \quad (3)$$

The relationship between PI and *W* for acyclic graph (*n*-alkanes) records is as follows:

$$PI = 6/n[(n-2)/(n-1)]W = [(6n-12)/(n^2-n)]W \quad (4)$$

This relation (4) establishes that no path graph exists with the same values of PI and *W* indices. Also, that all acyclic graphs with *n* = 3 and 8 give *W* = 2PI, otherwise *W* and PI are different such that *W* > PI.

As stated in the Introduction we have tested the above relationships for unbranched alkanes and extended the study to a large series of unbranched and branched alkanes.⁶⁸ The PI indices so calculated for the series of unbranched alkanes are presented in Table 1. This table also records *W*, bp, and log *P* of unbranched alkanes.

The perusal of Table 1 confirms the aforementioned findings.

To develop QSPR/QSAR models we have adopted a simple as well as curvilinear regression analyses.⁶¹

The regression fit employing simple regression analysis for the data presented in Table 1 gave the following

relationship between PI and *W* (or *S_z*) for unbranched alkanes.

$$PI = 17.6693 (\pm 4.4863) + 0.2630 (\pm 0.0082)W \quad (5)$$

$$Se = 14.9074, \quad R = 0.9913, \quad Q = 0.00665$$

This shows an outstanding correlation between *W* and PI. Recall that for alkanes *W* = *S_z*.

In eq 5, *Q* is the quality factor⁶⁸ and is defined as the ratio of correlation coefficient (*R*) to the standard error of estimation (*Se*).

It is possible to think of a countless variety of correlations. However, the curvilinear correlation with three fitting parameters improved the quality of the above correlation (eq 5) so that the *R* value changed from 0.9913 to 0.9988. This curvilinear correlation for the unbranched alkanes considered is found as follows

$$PI = A_0 + A_1W + A_2W^2 \quad (6)$$

where *A*₀ = 6.7560(± 2.0814); *A*₁ = 0.3672(± 0.0106); *A*₂ = -9.0940 E-05(± 8.8345 E-06); *Se* = 5.7018; *R* = 0.9988; and *Q* = 0.1752.

The regression analysis for the correlation of *W* and PI with bp of unbranched alkanes in the present study indicated that the best results are obtained in curvilinear analysis. The regression data for such curvilinear correlations involving three and five parameters are found as follows.

Correlation of *W* with bp:

$$Se = 63.8808 (27.3518); \quad R = 0.9104 (0.9907); \\ Q = 0.0143 (0.0362) \quad (7)$$

Correlation of PI with bp:

$$Se = 48.4011 (22.1731); \quad R = 0.9496 (0.9940); \\ Q = 0.0196 (0.0448) \quad (8)$$

The values in the parentheses in the above eqs 7 and 8 are the regression parameters for the five fitting parameters.

Table 2. Correlation Matrix for the Correlation of Physicochemical Properties and Biological Activities of Alkanes and Their Correlation with Wiener (W) and PI Indices^a

	W	PI	log P	MR	MV	PR	α
W	1.00000						
PI	0.98037	1.00000					
log P	0.85860	0.93971	1.00000				
MR	0.85860	0.93971	1.00000	1.00000			
MV	0.85934	0.94018	1.00000	1.00000	1.00000		
PR	0.85859	0.93970	1.00000	1.00000	1.00000	1.00000	
α	0.85919	0.94014	0.99997	0.99997	0.99997	0.99997	1.00000

^a WEnDash EnDash – Wiener index; PI – PI index; log P – logarithm of partition coefficient in octanol/water system; MR – molar refraction; MV – molar volume; PR – parachor, α – polarizability.

The aforementioned results show that the correlations of W and PI with bp are the best modeled by curvilinear analysis and also that the results are statistically outstanding.

For developing a QSAR model for the same unbranched alkanes we have considered log P (logarithm of octanol–water partition coefficient P) to represent their physiological activity. The hydrophobic constant (log P) is used to rationalize interactions of small ligands with various macromolecules in the fields of biochemistry,⁶⁹ medicinal chemistry,⁷⁰ and environmental sciences.⁷¹ Considerable experience in the use of log P in the study of QSAR by regression analysis clearly indicated that measured log P should be used whenever possible. Nevertheless, when the number of compounds is great and the structural variations limited, simple economics makes it desirable to measure log Ps for only the key structures and to calculate the remainder where any group not interactive in the measured solutes can be assumed to be negligible.

Here also, in developing QSAR modeling using log P, curvilinear correlations gave better results. The regression parameters for curvilinear correlations of W and PI with log P for the unbranched alkanes considered, involving three and five parameters, are given below.

Correlation of W with log P

$$\text{Se} = 0.8068 (0.2881); \quad R = 0.9616 (0.9975); \\ Q = 1.1919 (2.4623) \quad (9)$$

Correlation of PI with log P

$$\text{Se} = 0.6609 (0.2526); \quad R = 0.9711 (0.9976); \\ Q = 1.4694 (3.9493) \quad (10)$$

The values in the parenthesis are for five fitting parameters.

Here also, the correlation of W and PI with log P of unbranched alkanes considered is outstanding and PI gives slightly better results than W.

We have now extended our study for a large series of unbranched and branched alkanes.⁷² For developing QSPR models for this series of alkanes we have considered four different properties, molar refraction (MR); molar volume (MV); parachor (PR); and polarizability (α), while the QSAR model was developed using log P as the physiological activity.

The correlation matrix for the aforementioned quantities is shown in Table 2.

A perusal of Table 2 shows that collinearity between PI and log P, MR MV, PR, and α is excellent and considerably

Table 3. Regression Expression for Modeling log P, MR, MV, PR, and α of Alkanes Using W (=Sz) and PI Indices^a

log P = 0.0032 ($\pm 4.4655 \text{ E-}04$)W + 4.9875	(1)
log P = 0.0185 (± 0.0016)PI + 3.9970	(2)
MR = 0.0275 (± 0.0039)W + 41.2116	(3)
MR = 0.1603 (± 0.0137)PI + 32.6505	(4)
MV = 0.0979 (± 0.0137)W + 167.9644	(5)
MV = 0.5721 (± 0.0489)PI + 137.4289	(6)
PR = 0.2355 (± 0.0331)W + 368.4591	(7)
PR = 1.3764 (± 0.1181)PI + 249.9426	(8)
α = 0.0109 (± 0.0015)W + 16.3113	(9)
α = 0.0636 (± 0.0054)PI + 12.9180	(10)

^a Terms involved are defined in the footnote of Table 2.

Table 4. Quality of Correlations for Modeling log P, MR, MV, PR, and α of Alkanes Using W (=Sz) and PI Indices^a

correlation	Se	R	Q	F
1	1.9569	0.8586	0.43875	50.492
2	1.3054	0.9397	0.71985	135.920
3	16.9143	0.8586	0.05076	50.492
4	11.2831	0.9397	0.08328	135.307
5	60.1944	0.8395	0.01394	50.825
6	40.1002	0.9402	0.02344	137.083
7	145.2508	0.8586	0.00591	50.487
8	96.8948	0.9397	0.00970	135.903
9	6.6912	0.8592	0.12840	50.759
10	4.4568	0.9401	0.21093	136.982

^a Refer to Table 3. Se = standard error of estimation; R = usual correlation coefficient; Q = quality factor; F = F-ratio.

better than W. This shows that compared to W, PI will prove to be a better topological index for modeling properties/activities of alkanes.

We have used Table 2 for obtaining QSPR/QSAR models using simple as well as curvilinear regression analyses. The regression models using simple regression analyses are shown in Table 3, and their qualities are demonstrated in Table 4.

The quality parameters presented in Table 4 confirm the superiority of the PI index over W for modeling property/activity. Better results are obtained by subjecting the data to curvilinear analyses. These data are not shown in the table but can be made available upon request to one of the authors (P.V.K.).⁶⁸

CASE OF CYCLOALKANES

In this case also we first developed QSPR/QSAR models based on boiling points (bp) and log P of cycloalkanes and then extended the study considering a larger series of branched and unbranched cycloalkanes.⁷²

For cycloalkanes structure-boiling point studies though highly desirable are scarce.⁷³ Such studies suffered from problems both with the data and with the topological indices. The diversities in the structures, due to strain, is overwhelming so as to render the include/exclude decision everything but clear-cut. Furthermore, a small fraction of all possible cycloalkanes even of small size has been synthesized, and mostly their boiling points were determined during distillation and as such bp are of low precision. Also, the problem of stereochemistry has to be dealt with in cycloalkanes. In the majority of cases better QSPR/QSAR results are obtained by splitting the data depending upon the ring size of cycloalkanes.

Table 5. PI, W, Sz, Log P, and Other Physicochemical Properties for Some Monocyclic Graphs (Unbranched Cycloalkanes)^a

compd no.	unbranched cycloalkane	n	PI	W	Sz	log P	MR	MV	PR	α
1	cyclopropane	3	6	3	3	1.61	13.83	53.20	120.10	5.48
2	cyclobutane	4	8	8	16	2.14	18.44	70.90	160.10	7.31
3	cyclopentane	5	20	15	20	2.68	23.05	88.70	200.20	9.14
4	cyclohexane	6	24	27	54	3.22	27.67	106.40	240.20	10.96
5	cycloheptane	7	42	42	63	3.75	32.28	124.10	280.20	12.79
6	cyclooctane	8	48	64	128	4.29	36.69	141.90	320.30	14.62
7	cyclononane	9	72	90	144	4.82	41.50	159.60	360.30	16.41
8	cyclodecane	10	80	125	250	5.36	46.11	177.40	400.40	18.28
9	cycloundecane	11	110	165	275	5.90	50.72	195.10	440.40	20.11
10	cyclododecane	12	120	216	432	6.43	55.34	212.80	480.40	21.93
11	cyclotridecane	13	156	273	468	6.97	59.95	230.60	520.50	23.76
12	cyclotetradecane	14	168	343	686	7.50	64.54	248.30	560.50	25.59
13	cyclopentadecane	15	210	420	735	8.04	69.17	266.00	600.60	27.42
14	cyclohexadecane	16	224	512	1024	8.58	73.78	283.80	640.60	29.25
15	cycloheptadecane	17	272	612	1058	9.11	78.39	301.50	680.60	31.07
16	cyclooctadecane	18	288	729	1458	9.65	83.01	319.30	720.70	32.90
17	cyclononadecane	19	342	855	1539	10.18	87.67	337.60	760.70	34.73
18	cyclodecadecane	20	360	1000	2000	10.72	93.23	354.70	800.80	36.56
19	cycloundecadecane	21	420	1155	2100	11.00				
20	cyclododecadecane	22	440	1331	2662	11.50				

^a PI – Padmakar–Ivan index, W – Wiener index, bp – boiling point, log P – logarithm of partition coefficient (P) in octanol/water, MR – molar refraction, MV – molar volume, PR – parachore, α – polarizability.

Before discussing the QSPR/QSAR models for cycloalkanes let us first drive the relationship among their W, Sz, and PI indices.

In the case of normal (unbranched) cycloalkanes with n vertices the Wiener index, W, is given by the following expression:⁷⁴

$$W = \begin{cases} n^3/8 & \text{when } n = \text{even} \\ n(n^2-1)/8 & \text{when } n = \text{odd} \end{cases} \quad (11)$$

Also for such cyclic graphs PI is given by

$$PI = \begin{cases} n(n-2) & \text{for } n = \text{even} \\ n(n-1) & \text{for } n = \text{odd} \end{cases} \quad (12)$$

Comparing eqs 11 and 12 and noting that n_{eu} and n_{ev} for a unicyclic graph with n vertices coincides with n_{eu} and n_{ev} for a unicyclic graph with $n+1$ vertices, we have the following.

(i) For a cyclic graph with even n

$$PI = 8/n^2(n-2)W \quad (13)$$

(ii) For a cyclic graph with odd n

$$PI = 8 [(n-1)/(n^2-1)]W = 8\{1/(n+1)\} W \quad (14)$$

Similarly Szeged index (Sz) for unbranched cycloalkanes is given by the following equations:

$$Sz = \begin{cases} n^3/4 & \text{when } n = \text{even} \\ [n(n-1)^2]/4 & \text{when } n = \text{odd} \end{cases} \quad (15)$$

Comparing eq 15 with the eq 14 we obtained the following.

(i) For even n

$$PI = 4/n^2(n-2)Sz \quad (16)$$

(ii) For odd n

$$PI = 4[(n-1)/(n-1)^2]Sz = 4\{1/(n-1)\}Sz \quad (17)$$

The aforementioned relationships lead us to derive the following corollaries.

Corollary 1. There exists only one unbranched cyclic graph with even n ($n = 4$) where PI coincides with W.

Corollary 2. There exists only one unbranched cyclic graph with odd n ($n = 7$) where also PI coincides with W.

Corollary 3. There exists only one unbranched cyclic graph with ($n = 3$) where PI is found to be twice that of W and Sz.

Corollary 4. Except for the above cases and except when $n = 5$, the PI index is always smaller than W.

Corollary 5. There exists only one unbranched cyclic graph with $n = 5$ where PI coincides with Sz.

Corollary 6. There exists only two unbranched cyclic graphs with $n = 4$ and 9 where $Sz = 2PI$.

Corollary 7. There exists only one unbranched cyclic graph with $n = 13$ where $Sz = 3PI$.

For investigating relationships among W, Sz, and PI we have considered a set of 20 unbranched cycloalkanes as shown in Table 5. This table also records W, Sz, PI, and log P for these compounds.

The simple linear regression of the data presented in Table 5 demonstrated the following relationship of PI with W and Sz.

$$PI = 36.0769 (\pm 7.1018) + 0.3367 (\pm 0.0125)W \quad (18)$$

$$Se = 22.6335, \quad R = 0.9879, \quad Q = 0.0437$$

$$PI = 40.1998 (\pm 9.1954) + 0.1723 (\pm 0.0084)Sz \quad (19)$$

$$Se = 29.6254, \quad R = 0.9791, \quad Q = 0.0331$$

Equations 18 and 19 show that correlation of PI with W and Sz of unbranched cycloalkanes in the present study is outstanding and that correlation between PI and W is slightly better than the correlation of PI with Sz.

Table 6. Correlation Matrix for the Correlation of Physicochemical Properties and Biological Activities of Cycloalkanes and Their Correlations with Wiener (W), Szeged (Sz), and PI Indices^a

	W	Sz	PI	log P	MR	MV	PR	α
W	1.00000							
Sz	0.99653	1.0000						
PI	0.98775	0.97719	1.00000					
log P	0.94037	0.93149	0.97699	1.00000				
MR	0.94208	0.93355	0.97777	0.99996	1.00000			
MV	0.94036	0.93148	0.97699	1.00000	0.99996	1.00000		
PR	0.94037	0.93150	0.97700	1.00000	0.99996	1.00000	1.00000	
α	0.94042	0.93155	0.97703	1.00000	0.99996	1.00000	1.00000	1.00000

^a Abbreviated as in Table 2.

The curvilinear analyses of the data gave better results such that three parametric regression changes gave a correlation coefficient as high as 0.998. This curvilinear correlation is shown as follows.

$$PI = 17.2466 (\pm 4.0944) + 0.4913 (\pm 0.0202)W - 1.3240 \text{ E-}04 (\pm 0.00002)W^2 \quad (20)$$

$$Se = 10.6741, \quad R = 0.9975, \quad Q = 0.0935$$

$$PI = 18.8272 (\pm 6.7498) + 0.2634 (\pm 0.0170)Sz - 3.9989 \text{ E-}05 (\pm 7.0919 \text{ E-}06)Sz^2 \quad (21)$$

$$Se = 17.9935, \quad R = 0.9928, \quad Q = 0.0552$$

Thus, curvilinear regressions also demonstrate that correlation between PI and W is better than correlation between PI and Sz.

The QSPR models based on boiling points of unbranched cycloalkanes and their W, SZ, and PI employing three parametric regressions are found as follows.

$$bp = -20.7175 (\pm 11.9368) + 3.8645 (\pm 0.5329)W - 0.0173 (\pm 0.0042)W^2 \quad (22)$$

$$Se = 16.4256, \quad R = 0.9852, \quad Q = 0.0590$$

$$bp = -13.1503 (\pm 10.2416) + 1.9958 (\pm 0.2328)Sz - 0.0046 (\pm 0.0013)Sz^2 \quad (23)$$

$$Se = 21.1139, \quad R = 0.9755, \quad Q = 0.0463$$

$$bp = -44.9564 (\pm 15.035) + 5.4586 (\pm 0.9175)PI - 0.0315 (\pm 0.0105)PI^2 \quad (24)$$

$$Se = 53.8378, \quad R = 0.9863, \quad Q = 0.0713$$

The *Q*-values, therefore, suggest that the PI index is the better index for modeling bp of unbranched cycloalkanes used.

We have now extended our study considering a larger series consisting of both branched and unbranched cycloalkanes.⁷² The same physicochemical properties and biological activity which were used for alkane series were used in the present case also. The correlation matrix for these parameters is given in Table 6.

A perusal of Table 6 shows that better QSPR/QSAR models can be obtained using the PI index and that a

Table 7. Regression Expressions for Modeling log P, MR, MV, PR, and α of Cycloalkanes using W, Sz, and PI Indices

log P = 0.0085 ($\pm 7.7127 \text{ E-}04$)W + 3.5585	(1)
log P = 0.0044 ($\pm 4.2691 \text{ E-}04$)Sz + 3.6465	(2)
log P = 0.0239 (± 0.0013)PI + 2.7760	(3)
MR = 0.0738 (± 0.0066)W + 30.5229	(4)
MR = 0.0379 (± 0.0036)Sz + 31.2760	(5)
MR = 0.2068 (± 0.0111)PI + 23.7796	(6)
MV = 0.2822 (± 0.0235)W + 117.7410	(7)
MV = 0.1447 (± 0.0141)Sz + 120.6525	(8)
MV = 0.7914 (± 0.0432)PI + 91.8453	(9)
PR = 0.6371 (± 0.0576)W + 265.7885	(10)
PR = 0.3267 (± 0.0319)Sz + 272.3615	(11)
PR = 1.7865 (± 0.0975)PI + 207.3336	(12)
α = 0.0291 (± 0.0026)W + 12.1293	(13)
α = 0.0149 (± 0.0015)Sz + 12.4294	(14)
α = 0.0816 (± 0.0044)PI + 9.4602	(15)

Table 8. Quality of Correlations for Modeling Log P, MR, MV, PR, and α of Cycloalkanes Using W (=Sz) and PI Indices^a

correlation	Se	R	Q	F
1	1.0033	0.9404	0.93730	122.274
2	1.0729	0.9315	0.86820	104.912
3	0.6290	0.9770	1.55325	355.792
4	8.5471	0.9421	0.11022	126.230
5	9.1347	0.9335	0.10219	108.520
6	5.5430	0.9560	0.17246	347.972
7	33.2049	0.9404	0.02832	122.254
8	35.5076	0.9315	0.02623	104.904
9	20.8196	0.9770	0.04692	335.672
10	74.9447	0.9404	0.01254	122.293
11	80.1440	0.9313	0.01162	104.932
12	46.9857	0.9770	0.02079	335.846
13	3.4209	0.9440	0.27595	122.397
14	3.6582	0.9316	0.25466	105.023
15	2.1442	0.9770	0.45564	336.255

^a Refer to Table 7.

correlation potential in modeling these parameters follows the following sequence.

$$PI > W > Sz$$

The QSPR as well as QSAR modeling employing simple regression analysis are presented in Table 7, and their qualities are summarized in Table 8. The quality data presented in Table 8 confirms the aforementioned sequence.

Still better results are obtained using curvilinear correlations; such data are not given in the table but can be made available upon request to one of the authors (P.V.K.).

DEFORMATION MODES OF CYCLOALKANES

In principle, the infrared (IR) spectra have very appealing properties for QSAR research: they are generated in the

Table 9. Deformation Modes (δ_{CH_2} , Δ_{max} , and PI Indices of Unbranched Cycloalkanes^a

compd no.	cycloalkanes	<i>n</i>	δ_{CH_2} , cm ⁻¹	Δ_{max} , cm ⁻¹	PI	W	Sz
1	cyclopentane	5	1455			20	15
2	cyclohexane	6	1450			24	27
3	cycloheptane	7	1464	1453	11	42	42
4	cyclooctane	8	1477	1450	27	48	64
5	cyclononane	9	1487	1444	43	72	90
6	cyclodecane	10	1483	1445	33	80	125
7	cyclododecane	12	1470	1447	23	120	216
8	cyclotridecane	13	1461	1447	14	156	273
9	cyclotetradecane	14	1463	1446	17	168	343
10	cyclopentadecane	15	1460	1449	11	210	420
11	cyclohexadecane	16	1461	1448	13	224	512
12	cycloheptadecane	17	1460	1450	10	272	612

^a *n* = ring size; δ_{CH_2} = CH₂ deformation mode; Δ_{max} = difference in doublet of deformation mode.

Table 10. Correlation Matrix for the Correlation of the PI, W, and Sz with Deformation Modes (δ_{CH_2} , Δ_{max})^a

	W	Sz	PI	δ_{CH_2}	Δ_{max}
W	1.00000				
Sz	0.99377	1.00000			
PI	0.97899	0.99225	1.00000		
δ_{CH_2}	-0.68164	-0.67776	-0.67188	1.00000	
Δ_{max}	-0.57450	-0.58273	-0.57055	0.97533	1.00000

^a Abbreviated as Table 9.

range of low energy molecular interaction⁷⁵ that play a fundamental role in molecular recognition, and they are extremely specific fingerprints of the molecules. This particularly applies to the fingerprint region (1500–600 cm⁻¹).⁷⁶

A very limited number of QSAR studies employing IR spectral data have been reported.⁷⁷ In this section we discovered relationships among W, Sz, PI, and deformation modes (δ_{CH_2} , Δ_{max}) for unbranched cycloalkanes shown in Table 9. Here, Δ_{max} stands for the difference⁶³ between the doublets of δ_{CH_2} .

The correlation matrix for proposing QSAR models for modeling deformation modes of cycloalkanes is given in Table 10.

The QSAR models based on regression analyses of the data (Table 9) are shown in Table 11.

A perusal of Table 11 gave the following results: (i) δ_{CH_2} are better modeled than Δ_{max} using W, Sz, and PI indices; (ii) in the case of δ_{CH_2} the best results are obtained using the PI index; (iii) use of W and PI for modeling Δ_{max} gives similar results; (iv) both δ_{CH_2} and Δ_{max} increases with decrease in the magnitude of W, Sz, and PI indices; and (v) the proposed models are not excellent but are statistically fairly good.

The low *r*-values (≈ -0.68 for δ_{CH_2} and ≈ -0.60 for Δ_{max}) may be attributed to the reasons mentioned in the introduc-

tory part of this section. That is, ring size is responsible for the results mentioned above. Consequently, better results could be obtained by splitting the data based on the ring size of cycloalkanes. Indeed, better results are obtained by considering (i) low and middle size unbranched cycloalkanes as the class and (ii) high size cycloalkanes as the another class. Under such a situation the following results are obtained.

Modeling of δ_{CH_2}

index	class	Se	<i>R</i>	<i>Q</i>
W-index	class 1	15.2665	0.8945	0.0586
	class 2	1.2247	-0.5259	0.4294
Sz-index	class 1	15.2665	0.8317	0.0545
	class 2	1.2247	-0.5607	0.4578
PI-index	class 1	15.2665	0.9413	0.0617
	class 1	3.8341	-0.7583	0.1978

These results show that the best results are obtained using the PI index i.e., it is the best topological index for modeling δ_{CH_2} .

Modeling of Δ_{max}

index	class	Se	<i>R</i>	<i>Q</i>
W-index	class 1	13.4040	0.7219	0.0539
	class 2	2.7386	-0.7008	0.2559
Sz-index	class 1	13.4040	0.6154	0.0459
	class 2	2.7386	-0.5444	0.1988
PI-index	class 1	13.9523	0.8746	0.0627
	class 2	4.7610	-0.8683	0.1824

Once again the results show that the PI index is the best topological index among the three for modeling Δ_{max} .

MODELING OF POLYCHLORINATED BIPHENYLS IN ENVIRONMENT

Vapor pressure determines the volatility of a chemical compound. It governs the exchange rates for a chemical across an air–water interface. QSPR for the estimation of vapor pressure from descriptors derived solely from the molecular structure offer a highly promising alternative which is more general and practically suitable for the prediction of the vapor pressures of new chemical products. Alternatively, estimation of vapor pressure offers a method of modeling compounds in the environment.

Recently, we have used W and Sz indices successfully for modeling polychlorinated biphenyls (PCBs) in the environment.^{46,78} These polychlorinated biphenyls are presented in Figure 1.

In the present study we introduce modeling of PCBs using the PI index and compare the results with those obtained by using the W and Sz indices. The data needed for this purpose are presented in Table 12.

Table 11. Modeling of Deformation Modes in Unbranched Cycloalkanes (δ_{CH_2} and Δ_{max}) Using W, Sz, and PI Indices

S.N.	regression equation	Se	<i>R</i>	<i>F</i>	<i>Q</i>
1	$\delta_{CH_2} = -0.0313 (\pm 0.01280)W + 1475.9345$	7.7380	-0.6778	5.948	0.0876
2	$\delta_{CH_2} = -0.0167 (\pm 0.0070)Sz + 1475.8009$	7.4278	-0.6719	5.760	0.0905
3	$\delta_{CH_2} = -0.0784 (\pm 0.0318)PI + 1478.4296$	7.3378	-0.6816	6.073	0.0929
4	$\Delta_{max} = -0.0367 (\pm 0.0166)W + 30.6014$	9.9046	-0.6150	4.865	0.0621
5	$\Delta_{max} = -0.0193 (\pm 0.0091)Sz + 30.3512$	10.0648	-0.5983	4.455	0.0562
6	$\Delta_{max} = -0.0907 (\pm 0.0417)PI + 33.3228$	9.9531	-0.6100	4.740	0.0613

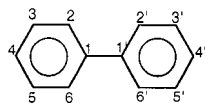


Figure 1. General structure of polychlorinated biphenyls (PCBs) used in the present study.

Table 12. Polychlorinated Biphenyls (PCBs), Their PI Indices, and Vapor Pressures (log VP) at 25 °C and 100 °C^a

compd no.	PCB	log VP		PI	W	Sz
		25 °C	100 °C			
1	-Biphenyl (BP)	0.004		144	198	360
2	2-Cl-BP	0.185	2.451	170	238	426
3	3-Cl-BP	-0.141		170	246	438
4	4-Cl-BP	-0.757		170	250	450
5	2,2'-Cl ₂ -BP	-0.873		198	287	499
6	3,3'-Cl ₂ -BP	-1.589	1.326	198	301	527
7	4,4'-Cl ₂ -BP	-2.580	1.067	198	315	555
8	2,5'-Cl ₂ -BP	-1.111	1.679	198	296	523
9	2,3,4-Cl ₃ -BP	-1.860	1.037	228	357	626
10	2,4,6-Cl ₃ -BP	-1.907	1.420	228	348	600
11	2,2',5,5'-Cl ₄ -BP	-2.304	0.913	260	412	700
12	2,2',4,5,5'-Cl ₅ -BP	-2.955	0.389	294	488	824
13	2,2',4,4',6,6'-Cl ₆ -BP	-2.763	0.452	330	560	927
14	2,2',3,3',5,5',6,6'-Cl ₈ -BP	-4.540		408	705	1152
15	2,2',3,3',4,4',5,5',6,6'-Cl ₁₀ -BP	-7.276	2.992	494	876	1483

^a Refer to Figure 1. PCB – polychlorinated biphenyls, log VP – log of vapor pressure, PI – Padmakar–Ivan index, W – Wiener index, Sz – Szeged index.

Table 13. Correlation Matrix for the Correlation of W, Sz, PI, and Log VP (Both at 25 °C and 100 °C) of Polychlorinated Biphenyls (PCBs)

	W	PI	Sz	log VP (at 25 °C)
W	1.0000			
PI	0.9980	1.0000		
Sz	0.9991	0.9973	1.0000	
log VP (at 25 °C)	-0.9591	-0.9544	-0.9623	1.0000

	W	PI	Sz	log VP (at 100 °C)
W	1.0000			
PI	0.9992	1.0000		
Sz	0.0002	0.9990	1.0000	
log VP (at 100 °C)	-0.9682	-0.9637	-0.9732	1.0000

The correlation matrix used for proposing QSPR model is shown in Table 13.

The data presented in Table 13 show that the correlation among the W, Sz, and PI indices of PCBs is outstanding and that Sz will prove to be the best topological index among the three topological indices for modeling PCBs in the environment.

Regression expressions for modeling PCBs at 25 °C and 100 °C using W, Sz, and PI are presented in Table 14.

The regression parameters (Table 14) show that both at 25 °C and 100 °C, Sz gives the best result.

Table 14. Regression Equations and Their Quality for Modeling Log VP of Polychlorinated Biphenyls (PCBs) Using W, Sz, and PI Indices

regression no.	regression equation	Se	R	F
At 25 °C				
1	log VP = -0.0097 (±7.9453 E-04)W + 1.7712	0.5677	-0.9591	149.150
2	log VP = -0.0060 (±4.7130 E-04)Sz + 1.9996	0.5452	-0.9623	162.834
3	log VP = -0.0186 (±0.0016)PI + 2.4960	0.5984	-0.9109	132.912
At 100 °C				
4	log VP = -0.0075 (±6.863 E-04)W + 3.9234	0.3853	-0.9682	119.820
5	log VP = -0.0046 (±3.8217 E-04)Sz + 4.0489	0.3539	-0.9732	143.471
6	log VP = -0.0146 (±0.0014)PI + 4.5715	0.4111	-0.9637	104.253

The observed high collinearity among W, Sz, and PI indicates that the new PI index duplicates much of the information content in W and Sz. Randic⁷⁹ has stated that one should particularly be aware of a common pitfall in regression analysis in discarding descriptors that are highly intercorrelated. He further stated that by discarding one of the descriptors which mostly duplicates another we may be discarding a descriptor that nevertheless may carry useful structural information in the parts in which it does not parallel with the other descriptors.

Following Randic⁷⁹ we explore the novelty of the PI index by considering its combination with other indices in multiple regressions using two variables. The results obtained are presented in Table 15. This table also records some of the results obtained from multiple regression analysis using three descriptors.

The results presented in Table 15 show that two pairs of two descriptors each at 25 °C exhibit similar potential for modeling PCBs. However, at 100 °C the multiple regression involving Sz and PI indices is better suited for modeling PCBs. In the triparametric multiple correlation the *R*-value was found to be the same as that of the bivariate correlations. However, the *F*-value in the latter case was found to be lower. Hence, despite an observed high collinearity between Sz and PI their combination gave the best model for modeling PCBs in the environment. This shows that the PI index carries useful structural information not present in the Sz index. A detailed study on this aspect is under way.

In support of our aforementioned findings we have obtained respective regression expressions, and these are shown in Table 16.

In the case of multiple regression expressions at 25 °C (Table 16) containing all three topological indices W, Sz, and PI we observed that the correlations of W and PI terms are lower in magnitude than their standard deviations. Such a correlation, therefore, is not statistically allowed and should be discarded.

Similarly, the regression expression having W, Sz, and PI as correlating parameters at 100 °C is also to be discarded for similar reasons.

The results recorded in Table 15 show that the combination of Sz and PI is the best for modeling PCBs in environment.

INHIBITORY INFLUENCE OF SUBSTITUTED BENZOYL ALCOHOLS ON *A. NIGER*

Finally, we study the role of the PI index in modeling inhibition influence of substituted benzoyl alcohols (Figure 2) on *A. niger* and then compared the results to those obtained by using W and Sz.

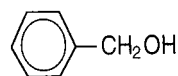
The substituted benzoyl alcohols used, their inhibitory activities (log 1/C), and the W, Sz, and PI indices are given in Table 17.

Table 15. Quality Factor (Q) and Other Quality Parameters for Two- and Three- Descriptor Regression Expressions for Modeling Log VP of Polychlorinated Biphenyls (PCBs) at 25 °C and 100 °C

regression no.	descriptors used	Se	R	F	Q
At 25 °C					
1	W, PI	0.5841	0.9600	70.584	1.6435
2	PI, Sz	0.5471	0.9650	81.249	1.7638
3	W, Sz, PI	0.5695	0.9653	50.045	1.6949
At 100 °C					
4	W, PI	0.3799	0.9730	62.231	2.5612
5	PI, Sz	0.7100	0.9918	211.521	4.7228
6	W, Sz, PI	0.2265	0.9919	121.282	4.3792

Table 16. Two- and Three-Descriptors Regression Expressions for Modeling Log VP of Polychlorinated Biphenyls (PCBs) at 25 °C and 100 °C

at 25 °C	
1.	$\log VP = -0.0165(\pm 0.0128)W + 0.0131(\pm 0.0245)PI + 1.2426$
2.	$\log VP = 0.0190(\pm 0.0020)PI + 0.0121(\pm 0.0064)Sz + 1.4474$
3.	$\log VP = 0.0062(\pm 0.0218)W + 0.0155(\pm 0.0241)PI - 0.0145(\pm 0.0116)Sz + 1.6839$
at 100 °C	
4.	$\log VP = -0.0268(\pm 0.0174)W + 0.0377(\pm 0.0341)PI + 2.2086$
5.	$\log VP = 0.0648(\pm 0.0163)PI - 0.0247(\pm 0.0051)Sz + 1.5755$
6.	$\log VP = -0.0018(\pm 0.0124)W + 0.0666(\pm 0.0217)PI - 0.0241(\pm 0.0065)Sz + 1.4819$

**Figure 2.** General structure of substituted benzoyl alcohol used in the present study.**Table 17.** A Group of Substituted Benzoyl Alcohols and Their PI, W, and Sz Indices and Inhibitory Influence on *A. niger*^a

compd no.	R	PI	log 1/C	W	Sz
1	H	66	1.51	94	148
2	4-chloro-	84	2.07	127	199
3	2,4-di chloro-	104	3.07	158	244
4	3,4-di chloro-	104	3.07	160	248
5	2,4,5 trichloro-	126	3.32	196	301
6	3,4,5 trichloro-	146	3.63	197	303
7	2-bromo-	84	2.15	121	187
8	4-bromo-	84	2.27	127	199
9	4-iodo-	84	2.75	127	199
10	4-methyl-	84	1.79	127	199
11	2,4-dimethyl-	104	2.14	158	244
12	3,5-dimethyl, 4-chloro-	146	3.05	197	303
13	3,5-dimethyl, 4-iodo-	146	3.42	197	303
14	2-nitro-	126	2.49	215	305
15	4-nitro-	126	2.00	233	341
16	4-cyano-	104	1.67	170	260
17	2-hydroxy-	104	1.39	158	236
18	3-hydroxy-	104	1.39	164	248
19	4-hydroxy-	104	1.39	170	260

^a Refer to Figure 2.

The correlation matrix for obtaining various QSAR models is shown in Table 18.

A perusal of Table 18 shows the presence of high collinearity among W, Sz, and PI of substituted benzoyl alcohols. The data show that PI will prove to be a better index for modeling inhibitory activities (log 1/C) of substituted benzoyl alcohols.

The regression expressions and their qualities as shown in Table 19 indeed confirm the aforementioned finding that the PI index is the most appropriate among the three indices

Table 18. Correlation Matrix for the Intercorrelation of W, Sz, and PI Indices of Substituted Benzoyl Alcohols and Their Correlation with Inhibitory Influence on *A. niger*^a

	W	Sz	PI	log 1/C
W	1.0000			
Sz	0.9940	1.0000		
PI	0.9037	0.9289	1.0000	
log 1/C	0.3894	0.4435	0.6043	1.0000

^a Log 1/C.**Table 19.** Regression Equations and Their Qualities for Modeling Inhibition of *A. niger* by Substituted Benzoyl Alcohols Using W, Sz, and PI Indices

S.N.	regression equations	Se	R	F
1	$\log 1/C = 0.0079(\pm 0.0045)W + 1.0556$	0.7047	0.3894	3.039
2	$\log 1/C = 0.0064(\pm 0.0031)Sz + 0.7507$	0.6857	0.4435	4.163
3	$\log 1/C = 0.0190(\pm 0.0061)PI + 0.3131$	0.6096	0.6043	9.779

Table 20. Quality Factor (Q) and Other Quality Parameters for Two- and Three- Descriptor Regression Expressions for Modeling Inhibition of *A. niger* by Substituted Benzoyl Alcohols Number

regression no.	descriptors used	Se	R	F	Q
1	W, Sz	0.6014	0.6469	5.758	1.0756
2	PI, W	0.5580	0.7060	7.966	1.2652
3	PI, Sz	0.5760	0.6830	6.994	1.1857
4	W, Ip ₁	0.3529	0.8943	31.945	2.5341
5	Sz, Ip ₁	0.3494	0.8965	32.754	2.5658
6	PI, Ip ₁	0.3300	0.9100	137.995	2.7575
7	PI, W, Sz	0.559	0.7270	5.606	1.3005
8	PI, W, Ip ₁	0.3400	0.9090	23.755	2.6735
9	PI, Sz, Ip ₁	0.3400	0.9090	23.749	2.6735
10	W, Sz, Ip ₁	0.3601	0.8970	20.586	2.4909

for modeling inhibitory activities (log 1/C) of substituted benzoyl alcohols.

In the present case also we applied the Randic⁷⁹ approach of the combination of two descriptors for searching the best QSAR model. In addition we have also obtained results from multivariate regressions using all three descriptors. The results are shown in Table 20. Here, in addition to the W, Sz, and PI indices we have also used a dummy parameter (indicator parameter) Ip₁. This dummy parameter Ip₁ is taken as unity when halogen is present in the benzoyl alcohol; otherwise its value was zero.

None of the binary combinations of topological indices, except for a combination of PI and Ip₁, gave better results. This shows that the PI index along with Ip₁ is the most suitable combination for proposing the QSAR model. This shows that the presence of halogen favors inhibitory activity of benzoyl alcohols. The respective regression expressions are given in Table 21.

All three-parametric multiple correlations are to be discarded on the grounds that were discussed in the case of PCBs.

CROSS-VALIDATION TEST

To judge the validity of the predictive power of the various QSPR/QSAR models discussed so far a cross-validation test^{80–82} was applied to the original data, and the resulting PRESS (predicted residual sum of squares) for each regression equation (QSPR/QSAR model) was calculated according

Table 21. Regression Equations for Two- and Three-Descriptor Models for Modeling Inhibition of *A. niger* by Substituted Benzoyl Alcohols

1	$\log 1/C = -0.0877(\pm 0.0355)W + 0.0684(\pm 0.0252)Sz - 0.3833$
2	$\log 1/C = -0.170(\pm 8.4046 \text{ E-}03)W + 0.0430(\pm 0.0130)PI + 0.5470$
3	$\log 1/C = -0.0120(\pm 7.1348 \text{ E-}03)Sz + 0.0440(\pm 0.0160)PI + 0.7200$
4	$\log 1/C = 0.0090(\pm 0.0023)W + 1.1694(\pm 0.1625)Ip_1 + 0.2621$
5	$\log 1/C = 0.0065(\pm 0.0016)Sz + 1.1293(\pm 0.1605)Ip_1 + 0.1446$
6	$\log 1/C = 0.0151(\pm 0.0033)PI + 1.0006(\pm 0.1537)Ip_1 + 0.2081$
7	$\log 1/C = -0.0530(\pm 0.0380)W + 0.0300(\pm 0.0310)Sz + 0.0320(\pm 0.0170)PI + 0.0410$
8	$\log 1/C = 4.0187 \text{ E-}04 (\pm 6.1112 \text{ E-}03)W + 0.0140(\pm 9.6000 \text{ E-}03)PI + 1.0080 (\pm 0.1900)Ip_1 + 0.2020$
9	$\log 1/C = 1.8368 \text{ E-}04 (\pm 4.7785 \text{ E-}03)Sz + 0.0150(\pm 0.0110)PI + 1.0040(\pm 0.1800)Ip_1 + 0.2020$
10	$\log 1/C = -0.0068 (\pm 0.0259)W + 0.0112(\pm 0.0184)Sz + 1.0991(\pm 0.2019)Ip_1 + 0.0734$

to the following equation

$$PRESS = \sum_i [(Y_i - \hat{Y}_i)^2 / (1 - h_{ii})^2]$$

where Y_i and \hat{Y}_i are the response (property/activity) values of observations i ($i = 1, 2, 3, \dots, n$), observed and calculated by the best equation.

The total PRESS value is a good measure of the predictive power of the model. If the total PRESS is smaller than the sum of the squares of the regression values (SSY), the model can be regarded as predicting better than chance and can be considered statistically significant.

The ratio of PRESS/SSY can also be used to calculate approximate confidence intervals of prediction of observations. To be a reasonable QSPR/QSAR model PRESS/SSY should be smaller than 0.4, and the value of this ratio lower than 0.1 indicates an excellent model.

The cross-validation test was used to check the validation of QSPR/QSAR models because this technique seems to substantially reduce the probability of chance correlation relative to multiple regression.

If the PRESS value is transformed in a dimensionless term by relating the two initial sum of squares, we obtain the cross-validation correlation coefficient namely, R^2_{cv} i.e., the component to the fraction of unexplained variants over then total variants.

$$R^2_{cv} = 1 - \frac{PRESS}{SSY}$$

PRESS and R^2_{cv} have good properties which render them appropriate for statistically testing with critical distribution. However, for practical purposes of end users in various branches of chemistry, the use of the square root of PRESS/N seems to be more directly related to the uncertainty of the prediction, since it has the same meanings as the actual Y values. The ratio (PRESS/N) is called predictive squared error (PSE).

The uncertainty in the prediction (s_{PRESS}) is determined by the following expression

$$s_{PRESS} = \sqrt{\frac{PRESS}{N-K-1}}$$

where N is the number of samples and K is the number of variables involved in the model.

The cross-validated parameters, mainly PRESS/SSY, R^2_{cv} , and PSE (in log units), were calculated for the various models discussed in all the aforementioned sections for the variety of models and are presented in Table 22.

Based on the cross-validation test (Table 22) we arrived at the following results.

(i) All the models (regression expressions) presented in Table 3 have PRESS/SSY smaller than 0.4; therefore, all are reasonable models. Also, that for models 2, 4, 6, 8, and 10 this ratio is almost constant and very close to 0.13. This indicates that these models (2, 4, 6, 8, and 10) are excellent models. Record that in all these excellent models the PI index is involved. Thus, the cross-validation test indicates that the PI index is the best topological index for modeling properties/activities of alkanes. Other cross-validation parameters also support these results.

(ii) Except for model 10 all 15 models presented in Table 7 have PRESS/SSY nearer to 0.13. Thus, they are also excellent models according to the cross-validation test.

(iii) Table 11 records regression equations for modeling δCH_2 and Δ_{max} for a series of cycloalkanes. They in fact represent QSAR models for modeling IR spectra of unbranched cycloalkanes.

The cross-validation parameters for the first three models (Table 11) indicate that none of the topological indices used (W, Sz, PI) are suitable statistically for modeling deformation modes (δCH_2) of the unbranched cycloalkanes used. However, for the remaining three models the cross-validation test suggests that the W, Sz, and PI indices are excellent indices for modeling Δ_{max} .

(iv) Table 14 records results for modeling polychlorinated biphenyls (PCBs) in the environment. Excellent cross-validation parameters are obtained for all six QSPR models suggesting that W, Sz, and PI are equally good for modeling PCBs in the environment. However, slightly better results are involved using Sz indices.

(v) In the case of PCBs (Table 16) a cross-validation test gave better results for the binary combination of topological indices. Thus, despite outstanding collinearity among W, Sz, and PI they give better results because they have different information content.

(vi) A cross-validation test for models presented in Table 19 indicates excellent modeling ability of W, Sz, and PI indices for the inhibitory action of substituted benzoyl alcohols on *A. niger*. However, W gave slightly better results than Sz and PI.

(vii) The cross-validation test (Table 21) indicates that none of the binary combination of W, Sz and PI resulted in a better QSAR model for modeling inhibition of *A. niger* by substituted benzoyl alcohols. However, when the indicator parameter IP_1 (used for the presence of halogen) is coupled with each of the topological indices then resulting binary combinations gives excellent results. So the cross-validation test also supports for the presence of halogen for inhibitory activity of benzoyl alcohols against *A. niger*.

Table 22. Cross-Validated Parameters (PRESS/SSY, R^2_{cv} , PSE) for the Proposed QSPR/QSAR Models^a

regression equation	PRESS/SSY	R^2_{cv}	PSE (log units)
Table 3			
1	0.3565	0.6435	0.2687
2	0.1324	0.8676	0.0929
3	0.365	0.6435	1.2054
4	0.1324	0.8676	1.0296
5	0.3542	0.6458	1.7567
6	0.1313	0.8687	1.5803
7	0.3565	0.6435	2.1392
8	0.1324	0.8676	1.9634
9	0.3546	0.6454	0.8026
10	0.1314	0.8686	0.6262
Table 7			
1	0.1306	0.8692	-0.0471
2	0.1525	0.8475	-0.0179
3	0.0476	0.9524	-0.2498
4	0.1268	0.8732	0.8835
5	0.1474	0.8526	0.9122
6	0.0460	0.9540	0.6793
7	0.1309	0.8691	1.4727
8	0.1525	0.8475	1.5018
9	0.0477	0.9523	1.2700
10	0.6916	0.3084	1.8262
11	0.1525	0.8475	1.8554
12	0.0475	0.9524	1.6235
13	0.1367	0.8693	0.4857
14	0.1523	0.8477	0.5148
15	0.0477	0.9523	0.2828
Table 11			
1	0.8231	0.1769	0.7506
2	0.7848	0.2152	0.7538
3	0.8478	0.1522	0.7485
4	0.3557	0.6433	0.9078
5	0.2058	0.7942	0.9148
6	0.3122	0.7678	0.9099
Table 14			
1	0.0872	0.9128	-0.2334
2	0.0798	0.9202	-0.2401
3	0.0978	0.9022	-0.2056
4	0.0668	0.9332	-0.5030
5	0.0558	0.9442	-0.5392
6	0.0667	0.9333	-0.4461
Table 16			
1	0.0850	0.9150	-0.2335
2	0.0738	0.9262	-0.2619
3	0.0733	0.9265	-0.2634
4	0.0562	0.9448	-0.5375
5	0.0165	0.9835	-0.7948
6	0.0165	0.9835	-0.7953
Table 19			
1	0.1788	0.8212	0.1766
2	0.2449	0.7551	-0.1880
3	0.2590	0.7410	-0.2398
Table 21			
1	0.7172	0.2828	-0.5519
2	0.9958	0.0042	-0.2904
3	0.8564	0.1436	-0.2768
4	0.2564	0.7496	-0.4896
5	0.2442	0.7558	-0.4944
6	0.4105	0.7895	-0.5204
7	0.8920	0.9080	-0.3037
8	0.2104	0.7896	-0.5205
9	0.2106	0.7894	-0.5203
10	0.2948	0.7572	-0.4963

^a PRESS = sum of squares of residual, SSY = sum of squares of regression, R^2_{cv} = cross-validation correlation coefficient, and PSE = predictive square error.

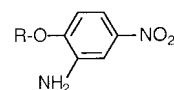
DISCRIMINATING POTENTIAL OF PI, $^1\chi$, AND $^1\chi^v$ INDICES

The aforementioned results and discussion show that the PI index cannot discriminate isomeric alkanes and alkyl or

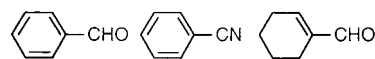
other groups substituted compounds with the same number of vertices (n). For example, for butane and isobutane with $n = 4$ both have the same values of $PI = 6$. As we see in Table 17, compounds 2, 7–10, and compounds 3, 4, 11, 16–19 have the values of $PI = 84$ and 104, respectively. This is obvious as like the Wiener index (W) and Szeged indices (Sz), the PI index also belongs to the first-generation topological index.^{83,84} In this regards the discriminating ability of the PI index proposed by us is not obviously improved in comparison with the Wiener index and Szeged indices. However, this was not our main objective of the present study. Our main objective was to modify Szeged index (Sz) and to propose Szeged-like topological index which can be applicable both to acyclic and cyclic graphs. Recall that Wiener index is not applicable to cyclic graphs and that for acyclic graphs Wiener and Szeged indices coincide.

In view of the above more favorable comparison will be with the discriminating ability of the original Randic index³ ($^1\chi$) and the valence connectivity index^{7,8} ($^1\chi^v$). Both these indices belong to second generation topological indices.^{83,84} Hence, by definition no or very little degeneracy should be observed in these indices. However, there are cases in that high degeneracy is observed both in $^1\chi$ as well as $^1\chi^v$. Some such cases are summarized below.

I. Cases in that $^1\chi$ Exhibit Degeneracy.^{7,8} (1) 3,3-Dimethylpentane and 2,4-dimethylpentane both have $^1\chi = 3.121$. (2) 1,3-Dimethylcyclohexane and 1,4-dimethylcyclohexane have $^1\chi = 3.788$. (3) 3-Methyl-, 1-butanol, isopentanol, secondary pentanol, isoamyl alcohol, and 2-pentanone all have $^1\chi = 2.770$. (4) 2-Methyl-1-propanol, 2-butanol, isopentone, and butanone have $^1\chi = 2.270$. (5) 3-Pentanol and 2-methylbutanol both have $^1\chi = 2.808$. (6) In the case of substituted nitroaniline



bromo-, chloro-, fluoro-, and methyl-substituted nitroanilines have the same value of $^1\chi = 5.109$. (7) 4-Methyl- and 3-methylphenols have $^1\chi = 2.545$. (8) The compounds



have $^1\chi = 3.932$. (9) 2-Methyl-1-chloro-, 2-methyl-1-bromo-, and 2-methyl-1-iodopropane have $^1\chi = 2.270$. (10) Chloro-, bromo-, and iodoteterylbutyl have the same value of $^1\chi = 2.000$. (11) 2-Hexanol and 3-methylpentanol show the same value of $^1\chi = 3.307$. (12) Ethylamine, ethyl alcohol, and ethyl chloride have $^1\chi = 1.414$. (13) Diethylamine, diethyl ether, and diethylsulfide have $^1\chi = 2.414$. (14) 1,2-Dichloroethylene, propyl alcohol, 1,2-dibromoethylene, propylamine, and propionitrile all have $^1\chi = 1.914$.

II. Cases in that $^1\chi^v$ Exhibit Degeneracy.^{7,8} Like the aforementioned degeneracy in $^1\chi$, there are several cases in that $^1\chi^v$ exhibited degeneracy. Some such cases are summarized below: (1) In the case of alkanes, the alkanes viz. i-hexene, *trans*-2-hexene, *trans*-3-hexene, 2-methyl-1-pentene, 3-methyl-1-pentene, 4-methyl-1-pentene, 2-methyl-2-pentene, 3-methyl-transd-2-pentene, 4-methyl-*trans*-2-pentene, 2-methyl-1-butene, 2,3-dimethyl-1-butene, 3,3-dimethyl-1-butene, and 2,3-dimethyl-2-butene all exhibit the same

Table 23. Comparison of Predictive Ability of the PI Index with Other Representative Topological Indices (Regression Parameters and Quality of Correlations): A Case of PCBs^a

S.N.	topological index used	regression parameters and quality of correlations				
		A	B	Se	R	Q
1	Padmakar—Ivan index: PI	-0.0198	2.9286	0.1486	-0.9798	6.5936
2	Randic index: ${}^0\chi$	-0.9539	8.3344	0.2339	-0.9816	4.1967
	1χ	-1.6924	9.9457	0.1690	-0.9678	5.7266
	2χ	-1.3618	5.8440	0.2101	-0.9880	4.7003
3	Kier & Hall index: ${}^0\chi^v$	-1.8378	13.0438	0.1660	-0.9781	5.8922
	$1\chi^v$	-3.1816	14.0977	0.1660	-0.9780	5.8916
	$2\chi^v$	-2.8012	6.7540	0.1745	-0.9354	5.3605
4	Balaban index: J	-10.0439	14.1437	0.1137	-1.0000	8.7959
	J _{HET}	-6.1758	12.2263	0.0266	-0.9902	37.2256
5	Schultz index MTI	-0.0027	2.3644	0.1660	-0.9870	5.9458
6	Harary index: H	-0.2954	6.6589	0.1486	-0.9420	6.3392
7	extended EAS	-0.2097	5.8965	0.1660	-0.9725	5.8584
	adjacency index:					
8	Zagreb group M ₁	-0.1207	7.3246	0.1245	-0.9263	7.4402
	indices M ₂	-0.0374	1.3376	0.1274	-0.7409	5.8155
9	information index I	16.0300	41.7548	0.1620	-1.0000	6.1728
10	Path index	-0.1207	7.3246	0.1432	-1.0000	6.9852
	(molecular ID number)					
11	topological I-index	-1.2375	9.2925	0.0870	-0.9672	11.1172
12	P ₂ - index	-0.5089	9.0482	0.1650	-1.0000	6.0606
13	P ₃ - index	-0.2200	3.1770	0.1654	-0.9908	5.9903
14	mean Wiener index: W _m	-0.1582	5.5827	0.1511	-1.0000	6.6181
15	mean square	-0.0194	2.2123	0.1473	-0.7898	5.3618
	Wiener index: W _{ms}					
16	root-mean-square	-7.0219	25.4683	0.1607	-0.9217	5.7355
	Wiener index: W _{rms}					

^a A - slope, B - intercept, Se - standard error of estimation, R - correlation coefficient, Q - quality factor.

value for $1\chi^v = 2.5236$. (2) 1-Butene, *trans*-2-butene, and 2-methylpropane have $1\chi^v = 1.5236$. (3) 1-Pentene, *trans*-2-pentene, 2-methylbutene, 3-methyl-1-butene, and 2-methyl-2-butene all have $1\chi^v = 2.0236$. (4) 3- and 4-Methoxy substituted phenyl diethyl phosphates have $1\chi^v = 2.433$. Similarly, 3-nitro- and 4-nitro- derivatives have the same value of 2.358 for $1\chi^v$. (5) In the case of substituted phenols the following cases of degeneracy were observed: (i) 2-methyl,4-chlorophenol and 3-methyl,4-chlorophenol have $1\chi^v = 3.064$; (ii) 2,6-dimethyl-4-chloro-, 3,5-dimethyl-4-chloro-, and 2,5-dimethyl-4-chlorophenols have $1\chi^v = 3.480$; (iii) 2,4,6-trichloro- and 2,4,5-trichloro-phenols have $1\chi^v = 3.684$; and (iv) 2,6-dimethyl-, and 2,5-dimethyl-phenols have $1\chi^v = 2.968$.

The aforementioned cases, therefore, indicate that in spite of the fact that both 1χ and $1\chi^v$ belong to second generation topological indices,^{83,84} there are cases wherein high degeneracy is observed. Hence, the observed degeneracy in the PI index is not surprising. Furthermore, the cases discussed in this paper show that in spite of observed degeneracy, the PI index gave statistically significant results. Also that in some cases excellent results were obtained using the PI index.

The above referred comparison of discriminating ability of PI, 1χ and $1\chi^v$ indices prompted us to undertake more favorable comparison of the results obtained from the PI index and other related topological indices. We do so by taking the example of monitoring polychlorinated biphenyls (PCBs) in the environment as discussed earlier (Table 12).

COMPARISON OF THE PI INDEX WITH OTHER REPRESENTATIVE INDICES

In view of the above and in an attempt to make a more favorable comparison of the PI index with other representa-

tive indices, we have considered the relative correlation ability of the PI index with other representative indices mentioned in Table 23 in modeling log VP of PCBs at 25 °C.

All the referred indices (Table 23) have been extensively described in the literature, and, therefore, the details of the calculations of these indices are not given here. However, below we give the respective expression used for their calculations.

(1) Randic Indices³ (${}^0\chi$, 1χ , and 2χ). The Randic index $1\chi = 1\chi(G)$ of G was introduced by Randic in 1975 as the connectivity index. This index is one of the most widely used topological indices in QSPR/QSAR analyses. In its original form it is defined as below

$$1\chi = 1\chi(G) = \sum_{ij} [\delta_i \delta_j]^{-1/2}$$

where δ_i is the valence of a vertex i, equal to the number of bonds connected to the atom i, in G, representing the graph of a compound. The meaning of δ_j is analogous. The expressions for computing ${}^0\chi$ and 2χ are analogous.

(2) Kier and Hall Valence Connectivity Indices (${}^0\chi^v$, $1\chi^v$, and $2\chi^v$). In an attempt to include multiple bonds and heteroatoms in the Randic index,³ Kier and Hall^{7,8,85} proposed the valence values of atoms according to the following equation

$$\delta_i^v = Z^v - h_i$$

where Z^v is the number of valence electrons of atom i and h_i is the number of hydrogens bonded to it. Thus, the Randic connectivity index was modified by Kier and Hall replacing δ_i with δ_i^v .

The Randic/Kier/Hall approach has been applied successfully to a variety of physicochemical and biological activities.

(3) Balaban Index (J). The Balaban index,^{13,60} $J = J(G)$, was introduced by Balaban in 1982 as the average distance sum connectivity index. It is defined as

$$J = \frac{M}{\mu+1} \sum_{\text{all edges}} (d_i d_j)^{-0.5}$$

where M is the number of edges of G ; μ is the cyclomatic number of G ; and d_i is the distance sum where $i = 1, 2, 3, \dots, N$. The cyclomatic number equals the minimum number of edges that must be removed from G in order to transform it to the related acyclic graph.

(4) Balaban Heteroatom Index¹³ (J_{HET}). This is an extension of Balaban index (J) to molecules containing heteroatoms. In the case of heteroatoms differentiation is made between the atoms of different kinds by modifying the corresponding elements of the distance matrix \mathbf{D} . For instance, the following modification was suggested for the diagonal elements

$$(\mathbf{D})_{ii} = 1 - (Z_c/Z_i)$$

where $Z_c = 6$ and Z_i is determined by the number of all electrons of atom i or namely Z_i is the atomic number of given elements.

The off-diagonal elements of the modified distance matrix for heteroatom systems are given by the following equation

$$(\mathbf{D})_{ij} = \sum_r k_r$$

where the summation is over r bonds.

The bond parameter k_r is given by the following expression

$$k_r = 1/w_r X(Z_c)^2/(Z_i + Z_j)$$

where w_r is the bond weight with values of 1, 1.5, 2, and 3 for a single, aromatic, double, and triple bond, respectively.

(5) Schultz Index (MTI). The Schultz index,⁸⁶ $\text{MTI} = \text{MTI}(G)$ of G , was introduced by Schultz in 1989 as the molecular topological index and is defined by the following expression

$$\text{MTI} = \sum_{i=1}^N e_i$$

where e_i ($i = 1, 2, 3, \dots, N$) represents the elements of the row matrix of order N

$$V(\mathbf{A} + \mathbf{D}) = (e_1 \ e_2 \ e_3 \ \dots \ N)$$

where V is the valence row matrix, \mathbf{A} is the adjacency matrix, and \mathbf{D} is the distance matrix.

(6) Harary Index (H). The Harary index,⁸⁷⁻⁹⁰ $H = H(G)$ of G , was introduced by Plasvic et al. in 1991 in honor of Prof. Frank Harary on his 70th birthday. It is defined as

$$H = 1/2 \sum_{i=1}^N \sum_{j=1}^N (\mathbf{D}_{ij})^{-2}$$

where \mathbf{D}^{-2} is the matrix whose elements are the squares of the reciprocal distances in G .

(7) Extended Adjacency Indices ($\text{EA}\Sigma$ and EA_{max}). The extended adjacency matrix indices^{91,92} $\text{EA}\Sigma$ and EA_{max} were introduced by Yang et al. in 1994. These indices are based on the extended adjacency matrices of molecules in that they influence factors of heteroatoms and multiple bonds which were considered. These indices possess high discriminating power and correlate with a number of physicochemical properties and biological activities of organic compounds. They are defined as

$$\text{EA} = \{g_{ij}\}$$

where $g_{ij} = a_{ij}[(v_i/v_j) + (v_j/v_i)]/2$ are the elements of the EA matrix and a_{ij} denotes elements of the \mathbf{A} matrix.

$\text{EA}\Sigma$ is the sum of the absolute eigen values of the EA matrix, and EA_{max} is the maximum of the absolute eigen values of the EA matrix.

(8) Zagreb Group Indices⁹³ (M_1 and M_2). These indices are based on π -electron energy, in which two terms appear in the approximate formula for the total π -energy which are used separately as topological indices. They are defined as

$$M_1(G) = \sum_{j=1}^N D_j^2$$

$$M_2(G) = \sum_{i,j} D_i D_j$$

The symbol D_i stands for the valency of the vertex i . The sum over in eq () is over all vertices of G , while the sum in eq () is over all edges.

(9) Information Indices (I). Bonchev and Trinajstić⁹³ applied information theory to the problem of characterizing molecular structure. The information content of the system i with N elements is defined by the following relationship

$$I = N \log_2 N - \sum_{j=1}^n N_j \log_2 N_j$$

where n is the number of different sets of elements, N_j is the number of elements in the j th set of elements, and summation is over all sets of elements. The logarithm is taken at the base 2 for measuring information contents in bytes.

(10) Topological I-Index. Topological I-index⁹⁴ of a graph G is based on the topological distances from a given vertex in the edge weighted graph of the organic molecule and is defined as

$$I = \frac{\sum_{r=1}^N n_r g_r}{\sum_{j=1}^N n_r}$$

where n_r is the number of r th kind of vertices for which g_r is the topological distance from the root in the edge weighted graph and the topological distance d_{ij} between the vertices i and j is defined as the distance associated with a minimum

weight. The weights in the edge-weighted graphs correspond to k values of the Huckel parameter^{95,96} for the heteroatom.

(11) Path Numbers (P_n) as Molecular Descriptors (Atomic I_D Number). Path numbers^{97,101} have been used as molecular descriptors to compare structures. Paths or self-avoiding walks are enumerated first for individual atoms in a structure—this yields the atomic path sequence $a_1, a_2, a_3, \dots, a_n$. The first number gives the number of paths of length one for a considered atom. Next is the number of paths of length two and so on.

If path numbers for all atoms are added (and divided by two, since each path involves two atoms as terminate), one obtains a path sequence for a molecule as a whole: m_1, m_2, \dots, m_n . Such a sequence can be preceded by m_0 , the number of atoms in a molecule (graph).

If paths (weighted or otherwise) for each atom are added, one obtains a single number representation of each atom (vertex) in a graph G ; such atomic parameters are called an atomic identification number or an atomic I_D number^{102,103} and be used in QSPR/QSDAR studies.

(12) Molecular I_D Number. Similar to an atomic I_D number^{104,105} is the molecular I_D number defined as the sum of all paths (weighted or nonweighted) in a molecule (graph). This index carries considerable structural information and is successfully used in QSPR/QSAR analyses.

(13) Mean Wiener Index^{102,110} (W_m). The mean value of topological distances is defined^{106,107} as

$$W_m = W_m(G) = \{1/N(N-1)\} \sum_{ij} d_{ij}$$

where n is the number of vertices (atoms). The summation is taken over all the elements of distance matrix, D .

The total number of elements in a triangular off-diagonal submatrix is equal to $1/2\{N(N-1)\}$. Therefore the mean Wiener index (mean value of the Wiener number) is given by the following expression:

$$W_m(G) = 2W/N(N-1)$$

(14) Mean Square Wiener Index (W_{ms}). The mean square value of the Wiener number is defined^{102,107-110} as the mean of the square of the elements of the off-diagonal submatrix.

$$W_{ms} = W_{ms}(G) = \{1/N(N-1)\} \sum_{ij} d_{ij}^2$$

(15) Root-Mean-Square Wiener Index (W_{rms}). The root-mean-square value of the Wiener number^{109,110} is defined as the square root of the mean of the square of the elements of the off-diagonal submatrix:

$$W_{rms} = W_{rms}(G) = [\{1/N(N-1)\} \sum_{ij} d_{ij}^2]^{0.5}$$

All the aforementioned topological indices were computed for the set of PCBs used in the earlier section and are used for modeling log VP of PCBs. The results are summarized in Table 23 in that the regression parameters and quality of correlations are mentioned. Here, we have modified the earlier regression by taking care of the outliers.

Trinajstić⁹³ while discussing topological indices and their applications to structure—property and structure—activity

relationships recommended that the Randic connectivity index³ is the most used topological index, and this index and its variants be used in QSPR/QSAR studies at a priority level. Also that the information index, though complicated, can also be used easily for QSAR studies. Finally, he recommended that the Balaban index¹³ (J and J_{HET}) has favorable features for use in QSPR/QSAR studies and stated that preliminary results in this direction are promising.

In light of the PI index (Table 23) we observed that the original Randic index³ ($^1\chi$) has a comparatively weaker ability than the PI index. However, both $^0\chi$ and $^2\chi$ have greater predictive potential than the PI index.

With regards to Balaban and information theoretical indices the recommendations of Trinajstić is found to be true even after the introduction of the PI index.

In addition to the above, the Schultz (MTI)-, molecular I_D number-, P_2 -, P_3 -, and W_m -indices are found to be superior to the PI index, while the remaining topological indices, $^0\chi^v$, $^1\chi^v$, $^2\chi^v$, H , $EA\Sigma$, M_1 , M_2 , topological I-index, W_{ms} , and W_{rms} , have inferior predictive ability than the newly introduced PI index.

This comparison leads us to infer that the PI index in spite of its merits and demerits serves as a convenient descriptor of molecular structure and that it fulfills the standard strategy in devising QSAPR or QSAR schemes which consist of the following two steps: (i) representation of structure by a given topological index and (ii) finding the quantitative relationship between topological indices and physicochemical properties (QSPR) and biological activities (QSAR) of the series of compounds under investigation.

However, there is a need and scope to modify the PI index to overcome some of its demerits; attempts in this direction are underway, and the results will be published soon.

CONCLUSIONS

From the aforementioned results and discussion we arrived at the following conclusions: (i) The proposed PI index is very simple to calculate and has discriminating power similar to that of the W and S_z indices. (ii) The PI index is easy to compute, is fast on computers, and is also conceptionally simple and easy to follow on paper. (iii) Unlike the Szeged index, the PI index does not coincide with the Wiener index for acyclic graphs. (iv) Like the W and S_z indices, the PI index belongs to the first-generation topological indices. (v) The PI index exhibits low to high degeneracy. However, there are cases in that both $^1\chi$ and $^1\chi^v$ (second generation topological index) exhibit degeneracy to the extent of the PI index. (vi) In spite of the observed degeneracy the PI index with fairly good structural selectivity and correlation ability is a suitable parameter for modeling physicochemical properties as well as physiological activities of diversified and complex compounds. It promises to be a useful parameter in QSAR/QSPR studies. (vii) In majority of the cases considered in the present study, we observed that in some cases the PI index gave better results than the W and S_z indices. However, there are some examples also wherein W or S_z gave slightly better results than PI. (viii) The cross-validation method has established the aforementioned relative predictive ability of the PI index.

ACKNOWLEDGMENT

Authors are thankful to Prof. I. Luckovits, Hungarian Academy of Sciences, Budapest, Hungary for providing Wiener-1 and Regress-1 Softwares for computation of some of the topological indices and for making regression analyses. They are also thankful to the referees for suggesting several interesting and fruitful suggestions.

REFERENCES AND NOTES

- Trinajstić, N. *Chemical Graph Theory*; CRC Press: Boca, Raton, FL, 1983; 2nd ed. 1992.
- Wiener, H. Structural determination of paraffin boiling points. *J. Am. Chem. Soc.* **1947**, *69*, 17–20.
- Randić, M. On characterization of molecular branching. *J. Am. Chem. Soc.* **1975**, *97*, 6609–6615.
- Gutman, I.; Polansky, O. E. *Mathematical Concepts in Organic Chemistry*; Springer-Verlag: Berlin, 1986.
- Pogliani, L. Modelling properties with higher-level molecular connectivity-descriptors. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 104–111.
- Wang, L. S.; Zhi, Z. L. *Molecular Connectivity and Molecular Structure—Activity*; Environmental Scientific Press of China: Beijing, 1999; p 12.
- Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure—Activity Relationship*; Wiley: New York, 1986.
- Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
- Xu, L. *Chemometrical Method*; Scientific Press of China: Beijing, 1985; p 304.
- Livingstone, D. J. *Quantitative Structure—Activity Relationships. In Similarity Models in Organic Chemistry, Biochemistry and Related Fields*; Zalewski, R. I.; Krygowski, T. M., Shorter, J., Eds.; Elsevier: Amsterdam, 1999; pp 557–627.
- Ferguson, A. M.; Heritage, T.; Jonathan, P.; Pack, S. E.; Philips, L.; Rogan, J.; Shaith, P. EVA—a new theoretically based molecular descriptors for use in QSAR/QSPR analysis. *J. Comput.-Aided Mol. Design*, **1997**, *11*, 143–152.
- Hosoya, H. Topological index. A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332–2339.
- Balaban, A. T. Highly discriminating distance-based topological indices. *Chem. Phys. Lett.* **1982**, *89*, 399–404.
- Bonchev, D.; Trinajstić, N. Information theory, distance matrix, and molecular branching. *J. Chem. Phys.* **1977**, *67*, 4517–4533.
- Schultz, H. P. Topological organic chemistry; I. Graph theory and topological indices. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 227–228.
- Randić, M.; Guo, X. F.; Oxley, T.; Krishnapriyam, H. Wiener index: Source of novel graph invariants. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 709–716.
- Gutman, I. A formula for the Wiener number of trees and its extension to graphs containing cycles. *Graph Theory Notes New York* **1994**, *27*, 9–15.
- Khadikar, P. V.; Deshpande, N. V.; Kale, P. P.; Dobrynin, A.; Gutman, I.; Domotor, G. The Szeged index and an analogy with the Wiener index. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 547–550.
- Mandloi, M.; Sikarwar, A.; Sapre, N. S.; Karmarkar, S.; Khadikar, P. V. A comparative QSAR study using Wiener, Szeged and molecular connectivity indices. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 57–64.
- Das, A.; Domotor, G.; Gutman, I.; Joshi, S.; Karmarkar, S.; Khaddar, D.; Khaddar, T.; Khadikar, P. V.; Popovic, L.; Sapre, N. S.; Sapre, N.; Schirhetti, A. A comparative study of the Wiener, Schultz, and Szeged indices of cycloalkanes. *J. Serb. Chem. Soc.* **1997**, *62*, 235–239.
- Gutman, I.; Khadikar, P. V.; Rajput, P. V.; Karmarkar, S. The Szeged index of polyacenes. *J. Serb. Chem. Soc.* **1995**, *60*, 759–764.
- Gutman, I.; Khadikar, P. V.; Khaddar, T. Wiener and Szeged indices of benzenoid hydrocarbons containing a linear polyacene fragment. *Commun. Math. Comput. Chem. (MATCH)* **1997**, *35*, 105–116.
- Khadikar, P. V.; Kale, P. P.; Deshpande, N. V.; Agrawal V. K. On the Szeged index of monocyclic graphs. *J. Indian Chem. Soc.* **2000**, *77*, 449–452.
- Khadikar, P. V.; Karmarkar, S.; Agrawal V. K.; Mandloi, M.; Joshi, S. On the Szeged index of [N]-cycles and mono-cycles containing acyclic side chains. *Nat. Acad. Sci. Lett.* **2000**, *23*, 50–56.
- Gutman, I.; Popovic, L.; Khadikar, P. V.; Karmarkar, S.; Joshi, S.; Mandloi, M. Relations between Wiener and Szeged indices of monocyclic molecules. *Commun. Math. Comput. Chem. (MATCH)* **1997**, *35*, 91–103.
- Dobrynin, A.; Gutman, I. On a graph invariant related to the sum of all distances in a graph. *Publ. Inst. Math. (Beograd)* **1994**, *56*, 18–22.
- Dobrynin, A.; Gutman, I. Solving a problem connected with distances in graphs. *Graph Theory Notes, New York* **1995**, *28*, 21–23.
- Dobrynin, A.; Gutman, I.; Domotor, G. A Wiener-type graph invariant for some bipartite graphs. *Appl. Maths. Lett.* **1995**, *8*, 57–62.
- Gutman, I.; Klavzar, S. An algorithm for the calculation of the Szeged index of benzenoid hydrocarbons. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1011–1014.
- Klavzar, S.; Rajapakse, A.; Gutman, I. On the Szeged and Wiener index of graphs. *Appl. Math. Letter* **1996**, *9*, 45–49.
- Dobrynin, A.; Gutman, I. On the Szeged index of unbranched catacondensed benzenoid molecules. *Croat. Chem. Acta* **1996**, *69*, 845–856.
- Gutman, I.; Klavzar, S. A method for calculating Wiener numbers of benzenoid hydrocarbons. *ACH Model Chem.* **1996**, *133*, 389–399.
- Zerovnik, J. Computing the Szeged index. *Croat. Chem. Acta* **1996**, *69*, 837–843.
- Gutman, I. Wiener numbers of benzenoid hydrocarbons containing a linear polyacene fragment. *Indian J. Chem.* **1997**, *36A*, 644–648.
- Dobrynin, A. Graphs having the maximal value of the Szeged index. *Croat. Chem. Acta* **1997**, *70*, 819–825.
- Dobrynin, A.; Gutman, I. Congruence relations for the Szeged index of hexagonal chains. *University Beograd. Publ. Elektrotehn. Fak. Ser. Mater.* **1997**, *8*, 106–113.
- Dobrynin, A. The Szeged index for the compliments of hexagonal chains. *Commun. Math. Comput. Chem. (MATCH)* **1997**, *35*, 227–242.
- Dobrynin, A.; Gutman, I. Szeged index of some polycyclic bipartite graphs with circuits of different size. *Commun. Math. Comput. Chem. (MATCH)* **1997**, *35*, 117–128.
- Diudia, M. V.; Parv, B.; Topan, M. Derived Szeged and Cluj indices. *J. Serb. Chem. Soc.* **1997**, *62*.
- Lukovits, I. Szeged index: Formulas for fused bicyclic graphs. *Croat. Chem. Acta* **1997**, *70*, 863–871.
- Chepoi, V.; Klavzar, S. The Wiener index and the Szeged index of benzenoid systems in linear time. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 752–755.
- Zerovnik, J. Szeged index of symmetric graphs. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 77–80.
- Gutman, I.; Dobrynin, A. The Szeged index—a success story. *Graph Theory Notes New York* **1998**, *34*, 37–44.
- Agrawal V. K.; Bano S.; Mathur K. C.; Khadikar P. V.; Karmarkar S.: Novel substitution constant for QSAR studies. *Nat. Acad. Sci. Lett.* **1999**, *22*, 55–59.
- Khadikar, P. V.; Karmarkar, S.; Joshi, S.; Gutman, I. Estimation of the protonation constants of salicylaldehydic acids by means of the Wiener topological index. *J. Serb. Chem. Soc.* **1996**, *61*, 89–95.
- Karmarkar, S.; Karmarkar, S.; Joshi, S.; Das, A.; Khadikar, P. V. Novel application of Wiener vis-à-vis Szeged indices in predicting polychlorinated biphenyls in the environment. *J. Serb. Chem. Soc.* **1997**, *62*, 227–234.
- Sachan, S.; Bano, S.; Agrawal, V. K.; Khadikar, P. V.; Srivastava, A. K. QSAR studies on antiasthmatic agents. *Nat. Acad. Sci. Lett.* **1999**, *22*, 119–124.
- Agrawal, V. K.; Bano, S.; Mathur, K. C.; Khadikar, P. V. Topological effect of substitution on antitubercular activities of quinolones. *Proc. Nat. Acad. Sci.* **1999**, *69A*, 145–154.
- Agrawal, V. K.; Bano, S.; Mathur, K. C.; Khadikar, P. V. Quantitative structure—activity relationships studies on antitubercular N₁-cyclopropyl quinolones using Szeged index (Sz). *Indian J. Chem.* **1999**, *38A*, 1203–1208.
- Agrawal, V. K.; Bano, S.; Mathur, K. C.; Khadikar, P. V. Novel application of Wiener vis-à-vis Szeged indices: Antitubercular activities of quinolones. *Proc. Indian Acad. Sci., (Chem. Sec.)* **2000**, *112*, 137–146.
- Agrawal, V. K.; Sachan, S.; Khadikar, P. V. Topological potential of Zwitterion analogues of cimetidine as H₂-receptor antagonists. *Pol. J. Pharmacol.* **2000**, *52*, 39–46.
- Karmarkar, S.; Joshi, S.; Sharma, V.; Khadikar, P. V. Correlation potential of Wiener vis-à-vis Szeged indices: antiinflammatory activities of phenols. *J. Indian Chem. Soc.* **2000**, *77*, 433–432.
- Karmarkar, S.; Khadikar, P. V.; Mandloi, M.; Joshi, S.; Agrawal, V. K. Topological estimation of proton-ligand formation constants: a multivariate analysis. *Indian J. Chem.* **2001**, *40A*, 12–22.
- Karmarkar, S.; Saxena, A.; Verma, R. G.; Karmarkar, S.; Mathur, K. C.; Mathur, S.; Singh, S.; Khadikar, P. V. Determining environmental behaviour and biological activity of RDX and related compounds. *Poll. Res.* **2000**, *19*, 337–344.
- Joshi, S.; Karmarkar, S.; Khadikar, P. V. Environmental pollution by polyhexes. *Sci., Cult.* **2000**, *66*, 288–290.

- (56) Khadikar, P. V.; Kale, P. P.; Deshpande, N. V.; Karmarkar, S.; Agrawal, V. K. Szeged indices of hexagonal chains. *Commun. Match. Comput. Chem. (MATCH)* (in press).
- (57) Sharma, V.; Karmarkar, S.; Thukral, G. V.; Khadikar, P. V. Estimation of anti-inflammatory activity of phenols by Szeged index (Sz). *Nat. Acad. Sci. Lett.* **2000**, *23*, 81–85.
- (58) Agrawal, V. K.; Sachan, S.; Khadikar, P. V. QSAR studies on antihistaminic activity of some thiazolidine-4-ones. *Acta Pharm.* **2000**, *50*, 281–290.
- (59) Mandloi, M.; Agrawal, V. K.; Mathur, K. C.; Karmarkar, S.; Khadikar, P. V. QSAR analysis of indolealkylamine receptor binding using Szeged index. *Acta Pharm.* **2000**, *50*, 303–313.
- (60) Balaban, A. T. Chemical graphs XXXIV. Five new topological indices for the branching of tree-like graphs[1]. *Theor. Chim. Acta* **1979**, *53*, 355–375.
- (61) Chatterjee, S.; Hadi, A. S.; Price, B. *Regression Analysis by Examples*, 3rd ed.; Wiley: New York, 2000.
- (62) *Handbook of Chemistry and Physics*, 76 ed.; Linde, D. R., Ed.; CRC Press: Boca Raton, FL, 1996.
- (63) Avram, M.; Mateescu, Gh. D. *Infrared Spectroscopy: Applications in Organic Chemistry*; Robert E. Krieger Pub: Huntington, New York, 1978; pp 119–200.
- (64) Rouvary, D. H.; Tatong, W. Novel applications of topological indices. 3. Prediction of the vapour pressure in polychlorinated biphenyls. *Intern. J. Environ. Stud.* **1989**, *33*, 247–257.
- (65) Carter, D. V.; Charlson, P. T.; Fenton, A. H.; Housey, J. R.; Lessel, B. Inhibition influence of substituted benzoyl alcohols on *A. niger*. *J. Pharm. Pharmacol. Suppl.* **1978**, *10*, 1495–1500.
- (66) Karizky, A. R.; Lobanov, V. S.; Karelson, M. Normal boiling points of organic compounds: Correlation and prediction by a quantitative structure–activity relationship. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 38–49.
- (67) Ivanciuc, O.; Ivanciuc, T.; Balaban, A. T. Quantitative structure–activity relationship study of normal boiling points of halogen/oxygen-sulphur containing organic compounds using CODESSA program. *Tetrahedron* **1998**, *54*, 9129–9142.
- (68) Pogliani, L. Molecular connectivity-descriptors of the physical properties of the α -amino acids. *J. Phys. Chem.* **1994**, *98*, 1494–1499.
- (69) Sangster, J. Octanol–water partition coefficients of simple organic compounds. *J. Phys. Chem. Ref. Data* **1989**, *18*, 1111–1215.
- (70) Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; Wiley: New York, 1979.
- (71) *Trends in Medicinal Chemistry '90 Proc Xth Int Sym Med. Chem Jerusalem*; Sarel, S., Mechoulam, R., Agrant, I., Eds.; Blackwell Scientific Pub: Melbourne, Paris, 1990.
- (72) Further data on statistical analyses performed are available from P. V. Khadikar upon request.
- (73) Rucker, G.; Rucker, C. On topological indices, boiling points, and cycloalkanes. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 788–802.
- (74) Gutman, I.; Yeh, Y. N.; Lee, S. L.; Luo, Y. L. Some recent results in the theory of Wiener number. *Indian. J. Chem.* **1993**, *32A*, 651–661.
- (75) Albrecht-Buchler, G. In defence of Nonmolecular Cell Biology. *Int. Rev. Cytol.* **1990**, *120*, 191–241.
- (76) Silverstein, R. M.; Bassler, G. C. *Spectroscopic Identification of Organic Compounds*, 2nd ed.; Wiley International Edition: New York, 1968.
- (77) Benighil, R.; Passerini, L.; Livingstone, D. J.; Johnson, M. A.; Giuliani, A. Infrared spectra information and their correlation with QSAR descriptors. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 558–562.
- (78) Mathur, K. C.; Singh, S.; Mathur, S.; Khadikar, P. V. Modelling polychlorinated biphenyls (PCBs) in environment. *Poll. Res.* **1999**, *18*, 405–409.
- (79) Randic, M. Novel molecular descriptors for structure–property study. *Chem. Phys. Lett.* **1993**, *21*, 478–483.
- (80) Lucic, B.; Trinajstić, N. A new efficient approach for variable selection based on multiple regression: Prediction of gas chromatographic retention times and response factors. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 610–621.
- (81) Lucic, B.; Trinajstić, N. Multivariate regression out performs several robust architectures of neural networks in QSAR modelling. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 121–132.
- (82) Zhang, W.; Tropsha, A. Novel variable selection quantitative structure–property relationship approach based on the K-nearest neighbor principle. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 185–194.
- (83) Balaban, A. T. Using real numbers as vertex invariants for third generation topological indices. *J. Chem. Inf. Comput. Sci.* **1992**, *30*, 23–28.
- (84) Khadikar, P. V.; Sharma, S.; Sharma, V.; Joshi, S.; Lukovits, I.; Kaveeshwar, M. QSAR study of the effect of benzohydroxamic acids on DNA synthesis. *Bull. Soc. Chem. Belg.* **1997**, *106*, 767–772.
- (85) Kier, L. B.; Hall, L. H. Recent advances in molecular connectivity for biological SAR analysis. In *IUPAC Pesticide Chemistry Human Welfare and the Environment*; Miyamoto, J., Ed.; Pergamon Press: 1983; pp 351–356.
- (86) Schultz, H. P. Topological organic chemistry. 1. Graph theory and topological indices of alkanes. *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 227–228.
- (87) Plavšić, D.; Nikolić, S.; Trinajstić, N.; Mihaljčić, Z. On the Harary index for the characterization of chemical graphs. *J. Math. Chem.* **1993**, *12*, 235–250.
- (88) Mihaljčić, Z.; Trinajstić, N. A graph-theoretical approach to structure–property relationships. *J. Chem. Educ.* **1992**, *69*, 701–712.
- (89) Joshi, S.; Singh, S.; Agrawal, V. K.; Mathur, K. C.; Karmarkar, S.; Khadikar, P. V. Novel estimation of the edge-shift in X-ray absorption discontinuity by the Harary index. *Nat. Acad. Sci. Lett.* **1999**, *22*, 159–167.
- (90) Karmarkar, S.; Thukral, G. S.; Agrawal, V. K.; Mathur, K. C.; Khadikar, P. V. Estimation of antituberculosis activity of the quinolones by Harary index. *Nat. Acad. Sci. Lett.* **1999**, *22*, 207–211.
- (91) Yang, Yi-Qiu; Xu, Lu.; Hu Chang-Yu. Extended adjacency matrix indices and their applications. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1140–1145.
- (92) Yao, Yu-Yuan; Xu, Lu.; Yang, Yi-Qiu; Yuan Xiu-Shun. Study on structure–activity relationships of organic compounds: Three new topological indices and their applications. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 590–594.
- (93) Trinajstić, N. *Chemical graph theory*; CRC Press: Boca Raton, FL, 1983; Vol. II, Chapter IV.
- (94) Lall, R. S. Structure activity relationship of organo-phosphorous insecticides. *Asian J. Chem.* **1990**, *2*, 37–42.
- (95) Baryz, M.; Jashari, G.; Lall, R. S.; Srivastava, Y. K.; Trinajstić, N. In *Chemical applications of topology and graph theory*; King, R. B., Ed.; Elsevier: Amsterdam, 1983; pp 222–230.
- (96) Streitwieser, A., Jr. *Molecular orbital theory for organic chemists*; Wiley: New York, 1961.
- (97) Platt, J. R.; Influence of neighbour bonds on additive bond properties in paraffins. *J. Chem. Phys.* **1947**, *15*, 419–420.
- (98) Randić, M. Nonempirical approach to structure–activity studies. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1984**, *11*, 137–153.
- (99) Randić, M. Graph theoretical approach to structure–activity studies: Search for optimal antitumor compounds. In *Molecular basis of cancer, Part A: Macromolecular structure carcinogenes and oncogenes*; Alan R. Liss Inc.: New York, 1985; pp 309–318.
- (100) Randić, M. On the characterization of three dimensional structure. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1988**, *15*, 201–208.
- (101) Randić, M. In search of structural invariants. *J. Math. Chem.* **1992**, *9*, 97–146.
- (102) Randić, M.; Jerman-Blazic, B.; Grossman, S. C.; Rouvary, D. H. A rational approach to the optimal design of drugs. *Math. Comput. Modelling* **1988**, *9*, 571–582.
- (103) Randić, M.; Jerman-Blazic, B.; Rouvary, D. H.; Seybold, P. G.; Grossman, S. C. The search for active substructure in structure–activity studies. *Int. J. Quantum Chem., Quantum Biol. Symp.* **1987**, *14*, 245–260.
- (104) Randić, M. On molecular investigation numbers. *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 164–175.
- (105) Carter, S.; Trinajstić, N.; Nikolić, S. On the use of ID numbers in drug research: A QSAR study of neuroleptic pharmacophores. *Med. Sci. Res.* **1988**, *16*, 185–186.
- (106) Lall, R. S. Topology and physical properties of *n*-alkanes. *Current Sci.* **1981**, *50*, 668–670.
- (107) Lall, R. S. On structural analysis of chemical mutations. *Indian J. Biochem. Biophys.* **1988**, *25*, 364–367.
- (108) Khadikar, P. V.; Joshi, S. Novel applications of topological indices: I-correlation of edge-shifts in X-ray absorption with Wiener indices. *X-ray Spectro* **1995**, *24*, 201–204.
- (109) Mathur, K. C.; Singh, S.; Mathur, S.; Khadikar, P. V. Modelling polychlorinated biphenyls (PCBs) in environment. *Poll. Res.* **1999**, *18*, 405–409.
- (110) Karmarkar, S.; Khadikar, P. V.; Agrawal, V. K.; Mathur, K. C.; Mandloi, M.; Joshi, S. Topological estimation of proton-ligand formation constants of potential antitumor agents: Salicylhydroxamic acids. *Proc. Indian Acad. Sci. (Chem. Sec.)* **2000**, *112*, 43–49.