Pergamon

Bioorganic & Medicinal Chemistry 10 (2002) 1181–1188

BIOORGANIC & MEDICINAL CHEMISTRY

QSAR Study on Toxicity to Aqueous Organisms Using the PI Index

Padmakar V. Khadikar,[a,*] Anjani Phadnis[b] and Anjali Shrivastava[c]

[a]*Research Division, Laxmi Fumigation and Pest Control Pvt. Ltd., 3, Khatipura, Indore 452 007, India*
[b]*Department of Chemistry, Ranjeet Singh College, Khandwa Road, Indore 452 001, India*
[c]*Department of Chemistry, Holkar Model and Autonomous College, Indore 452 001, India*

Received 16 August 2001; accepted 27 October 2001

Abstract—We have attempted to develop quantitative structure–toxicity relationships (QSTRs) to predict hydrophobicity (logP) as well as toxicity (pEC_{50} μm) of benzene derivatives using recently introduced Padmakar–Ivan (PI) index. The results have shown that both logP as well as pEC_{50} of benzene derivatives can be modelled excellently in multiparametric models in that the PI index and some indicator parameters are involved. The predictive ability of the models is discussed on the basis of the cross-validation method. The superiority of the PI index over several other topological indices is critically examined.

Introduction

A current interest in predictive toxicology is the estimation of the toxicological potential of chemicals from their calculated structural parameters.[1–6] In environmental hazards as well as in pharmaceutical drug design, one has to deal with thousands of chemical structures. They normally have very little experimental data needed for the estimation of their toxicity. In most cases, experimental data necessary for the prediction of the toxicity are not available to us. Consequently, there is a need for the development of predictive models based on parameters that can be calculated directly from the chemical structures of organic compounds acting as drugs. Our research group is actively engaged in the development of quantitative structure–activity relationships (QSARs) for predicting physicochemical, toxicological, and pharmaceutical properties of chemicals using theoretical descriptors including topological indices.[7–20] It is now well established that topological indices are quite successful in the development of QSARs and for predicting the property, activity, and toxicity of chemicals.[1–20]

A topological index is a graph-theoretical invariant which encodes quantitative information regarding the size, shape, bonding type and branching associated with the molecular architecture.[21,22] The Wiener Index (W) is the first, the oldest and most widely used topological index.[28] It is applicable to acyclic (trees) graphs only, and not to cyclic graphs. In view of this, Gutman has introduced a new index called the Szeged Index and abbreviated as Sz.[29–31] This new index is considered as the modification of the W Index to cyclic graphs, that is a cycle containing molecules. For acyclic graphs, W and Sz coincide. We have shown that the Sz is equally applicable to mono/polycyclic graphs containing acyclic side chains.[14]

Because of the coincidence of Sz with W in acyclic molecules, we have very recently introduced a W/Sz-like topological index, namely the Padmakar–Ivan (PI) index.[22–27]

Unlike the Sz Index, the newly introduced PI Index is different for acyclic and cyclic compounds and very little is known of its applicability in developing QSAR and QSPR models.[25,26] In view of this, in the present study we have used the PI Index for developing a quantitative structure–toxicity relationship (QSTR) for the toxicity of benzene derivatives (Table 1, Fig. 1). In this paper, we have attempted an estimation of logP and pEC_{50} of the benzene derivatives (Table 1) using the PI index. Our aim is not to introduce an alternate method of estimation of these properties and activities but to investigate to what extent the newly introduced PI index is useful for this purpose. Another objective of this paper is, therefore, to show distinctly the superiority of the PI index over other topological indices in explaining the toxicity of the substituted benzenes used in the present study. The results, as discussed below, show that excellent results are obtained in multiparametric models in that, in addition to the PI Index, some indicator

*Corresponding author. Tel.: +91-731-531906;
e-mail: anjalee_shrivastava@rediffmail.com

0968-0896/02/$ - see front matter
PII: S0968-0896(01)00375-3

Table 1. Benzene derivatives used, their PI, logP, pEC_{50}, values and indicator parameters IP_1, IP_2, IP_3

Comp. no.	Benzene derivative	logP	pEC_{50}	PI	IP_1	IP_2	IP_3
1	Hexachlorobenzene	5.70	6.32	126	0	1	0
2	1,2,4,5-Tetrachlorobenzene	4.75	5.51	82	0	1	0
3	1,2,4-Trichlorobenzene	4.16	4.50	66	0	1	0
4	1,2,3-Trichlorobenzene	4.04	4.53	66	0	1	0
5	1,4-Dichlorobenzene	3.57	4.39	50	1	0	0
6	1,3-Dichlorobenzene	3.57	4.24	50	1	0	0
7	1,2-Dichlorobenzene	3.45	4.38	50	1	0	0
8	Chlorobenzene	2.86	3.86	36	0	0	1
9	1,4-Dibromobenzene	3.87	4.54	50	1	0	0
10	1,3-Dibromobenzene	3.87	4.99	50	1	0	0
11	Bromobenzene	3.01	3.88	36	0	0	1
12	1-Chloro-4-bromobenzene	3.72	4.50	50	1	0	0
13	2,4,5-Trichlorotoluene	4.66	4.86	82	0	1	0
14	2,5-Dichlorotoluene	4.07	4.38	66	1	0	0
15	4-Chlorotoluene	3.35	3.88	50	0	0	1
16	4-Xylene	3.14	3.68	50	0	0	0
17	3-Xylene	3.14	3.05	50	0	0	0
18	2-Xylene	3.09	4.08	50	0	0	0
19	Toluene	2.64	3.08	36	0	0	0
20	Benzene	2.14	3.34	24	0	0	0

parameters are also used. The predictive potential of the models is discussed using a cross-validation method.[32–35] The results are discussed below.

Results and Discussion

It is well established[1] that chemicals can be grouped into structural classes for QSAR studies. However, nowadays, it has been realised that it is better to combine chemicals by toxic mechanism and try to establish theoretical correlation expressions between toxicity data and molecular descriptor, that is to develop QSTRs.[21,22] Hence, in the present investigation, we have developed QSTRs for modeling the toxicity (pEC_{50}) of benzene derivatives (Table 1). In addition, we have also attempted the modelling of hydrophobicity (logP) using the PI Index.

Many investigators have used the 1-octonol/water partition coefficient (logP) dependence QSTR as a basis for predicting the toxicity.[21,22] Conversely, it has been assumed that physiological activity (toxicity in the present case) can be modeled by logP. Otherwise toxicity data, such as *Vibrio fischeri* toxicity (pEC_{50}) can also be used. In the present investigation, we have used both logP and pEC_{50} to represent the toxicity of the benzene derivatives used and both are modelled using the PI Index. The logP and pEC_{50} data, as adopted from the earlier work,[36] are presented in Table 1. Table 1 also contains calculated values of the PI Index and assumed dummy (indicator) parameters IP_1, IP_2 and IP_3. Here, IP_1 is used as unity when two halogens are present in the aromatic nucleus. In the absence of this, IP_1 is considered zero. Similarly, IP_2 is used as unity for poly-halogen substitution (three or more halogens), in the absence of which, IP_2 is zero. Finally, IP_3 is taken as unity for mono-halogen substitution, in the absence of which, IP_3 is zero.

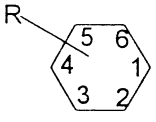

Figure 1. Benzene derivatives used in the present study.

It is worth mentioning that logarithmic expressions of P and EC_{50} are usually used because QSPR/QSAR/QSTR represent free-energy changes. For chemicals whose toxicities can be described by logP alone, it is believed that their toxic effects generally arise from membrane perturbation. For chemicals which are more reactive, some terms modelling receptor interaction or even metabolism can be incorporated into the regression expression.

A perusal of Table 1 shows that degeneracy exists in logP, pEC_{50}, and the PI Index of benzene derivatives used in the present study. The degeneracy in the PI Index is obvious because, like W and Sz, it also belongs to a first-generation topological index.[37] Balaban[37] has shown that, in spite of the observed degeneracy, the first-generation topological indices are quite useful in developing QSPR, QSAR, and QSTR. This was found to be the case with W and Sz indices and we believe that this will also be the case with the PI index.

The correlation matrix presented in Table 2 clearly shows that toxicity of benzene derivatives as expressed by logP and pEC_{50}, can be well characterized by the PI Index. That is, the mono-parametric model based on the PI Index is quite suitable for predicting both logP and pEC_{50} of the benzene derivatives used. This correlation matrix also indicates that both logP and pEC_{50} are favoured by IP_2. Note that IP_2 is an indicator parameter used to describe poly-halo-substitution in the aromatic nucleus. That means three or more halogen substitutions exhibit logP and pEC_{50} more favourably.

The correlation matrix further indicates autocorrelation and collinearity between the PI Index and IP_2. As will

be shown below, in spite of this collinearity, the multiparametric models (regression expressions) containing PI and IP_2 as the correlating parameters can be considered statistically significant.

The step-wise regression of the toxicities (logP and pEC_{50}) of the benzene derivatives resulted into four statistically significant multiparametric models for modeling logP and pEC_{50} (Table 3). We discuss these models (including monoparametric models) below.

The monoparametric models for modeling logP and pEC_{50} were found as under:

$$\log P = 0.0349\ (\pm 0.0028)PI + 1.6892 \quad (1)$$

$$pEC_{50} = 0.030\ (\pm 0.0045)PI + 2.6122 \quad (2)$$

When IP_1 is added to the aforementioned models, there is an improvement in the magnitude of correlation coefficient R but, in the resulting biparametric models, the coefficient of IP_1 terms were significantly smaller than their respective standard deviations. Such models are not statistically allowed and are not discussed here.

However, addition of the IP_2 term to the above models resulted in bi-parametric models containing PI and IP_2 terms with improved quality and statistical significance (Table 4):

$$\log P = 0.0358\ (+0.0022)PI + 0.3439\ (+0.0989)IP_2 + 1.5141 \quad (3)$$

$$pEC_{50} = 0.0314\ (+0.0039)PI + 0.4780\ (+0.1749)IP_2 + 2.3687 \quad (4)$$

It is interesting to record that, in the aforementioned eqs (3) and (4), the coefficient of PI terms in modelling logP and pEC_{50} are more or less the same. This indicates that the PI Index alone can be used for modelling toxicity of benzene derivatives used. Also, the coefficient of the added term, that is IP_2, is positive in both eqs (3) and (4). Therefore, poly-halogen-substitution is a pre-requisite for exhibition of logP as well as pEC_{50}. The larger coefficient of this term [eq (4)] indicates that such poly-substitution effect is pronounced in the exhibition of pEC_{50}. No biparametric models containing IP_1 or IP_3 in addition to the PI Index are found statistically significant. In both these cases, it was observed that their coefficients in the corresponding regression models were significantly smaller than their respective standard deviations. Such models are not allowed statistically.

Successive regression resulted into two types of better quality triparametric expressions each for modelling logP and pEC_{50}. One of these models consists of PI, IP_1, and IP_3 as the correlating parameters. The other triparametric expression contains PI, IP_1, and IP_2 as the correlating parameters. The quality of correlation presented in Table 4 shows that the latter expression is better for modelling both logP and pEC_{50}. These models are found as shown:

$$\log P = 0.0278\ (\pm 0.0028)PI + 0.5073\ (\pm 0.0879)\ IP_1 + 0.5435\ (0.1497)IP_2 + 1.7680 \quad (5)$$

Table 2. Correlation matrix for the intercorrelation of parameters presented in Table 1

	logP	pEC_{50}	PI	IP_1	IP_2	IP_3
logP	1.0000					
pEC_{50}	0.9191	1.0000				
PI	0.9471	0.8433	1.0000			
IP_1	0.0844	0.1870	−0.1275	1.0000		
IP_2	0.7495	0.6442	0.7622	−0.4237	1.0000	
IP_3	−0.3032	−0.2508	−0.3014	−0.3083	−0.2425	1.0000

Table 3. Regression expressions for modeling logP and pEC_{50} of benzene derivatives used in the present investigation

Modelling of logP	
(1)	$\log P = 0.0349\ (\pm\ 0.0028)PI + 1.6892$.
(2)	$\log P = 0.0358\ (\pm\pm 0.0022)PI + 0.3439\ (\pm 0.0989)IP_1 + 1.5141$.
(3)	$\log P = 0.0367\ (\pm 0.0024)PI + 0.3841\ (\pm 0.1061)IP_1 + 0.1526\ (\pm 0.1454)IP_3 + 1.4276$.
(4)	$\log P = 0.0278\ (\pm 0.0028)PI + 0.5073\ (\pm 0.0879)IP_1 + 0.5435(\pm 0.1497)IP_2 + 1.7680$.
(5)	$\log P = 0.0281\ (\pm 0.0023)PI + 0.6088\ (\pm 0.0813)IP_1 + 0.6384\ (\pm 0.1289)IP + 0.2767\ (\pm 0.0971)IP_3 + 1.6556$.
Modelling of pEC_{50}	
(6)	$pEC_{50} = 0.030\ (\pm 0.0045)PI + 2.6122$.
(7)	$pEC_{50} = 0.314\ (\pm 0.0039)PI + 0.4780\ (\pm 0.1749)IP_1 + 2.3687$.
(8)	$pEC_{50} = 0.0331\ (\pm 0.0042)PI + 0.5546\ (\pm 0.1865)IP_1 + 0.2905\ (\pm 0.2592)IP_3 + 2.2042$.
(9)	$pEC_{50} = 0.0233\ (\pm 0.0061)PI + 0.6431\ (\pm 0.1937)IP_1 + 0.5491\ (\pm 0.3900)IP_2 + 2.6252$.
(10)	$pEC_{50} = 0.0237\ (\pm 0.0057)PI + 0.7992\ (\pm 0.2027)IP_1 + 0.6949\ (\pm 0.3212)IP_2 + 0.4255\ (\pm 0.2419)IP_3 + 2.4524$.

Table 4. Regression parameters and quality of correlations given in Table 3

Model	Se	R^2A	R	F	P	Q
Modelling logP						
1. PI	0.2660	—	0.9471	156.827	2.532×10^{-10}	3.5605
2. PI, IP_1	0.2093	0.9327	0.9694	132.750	4.218×10^{-11}	4.6316
3. PI, IP_1, IP_3	0.2088	0.9436	0.9714	89.234	3.33×10^{-10}	4.6523
4. PI, IP_1, IP_2	0.1597	0.9608	0.9834	156.290	4.670×10^{-12}	6.1578
5. PI, IP_1, IP_2, IP_3	0.1329	0.9729	0.9892	171.458	2.510×10^{-12}	7.4432
Modelling of pEC_{50}						
1. PI	0.3661	—	0.8769	59.914	3.908×10^{-7}	2.3952
2. PI, IP_1	0.3699	0.7757	0.8941	33.863	1.177×10^{-6}	2.4171
3. PI, IP_1, IP_3	0.3672	0.7791	0.9022	23.334	4.377×10^{-6}	2.4570
4. PI, IP_1, IP_2	0.3521	0.7969	0.9105	25.847	2.253×10^{-6}	2.5859
5. PI, IP_1, IP_2, IP_3	0.3311	0.8204	0.9264	22.697	3.227×10^{-6}	2.7980

$$pEC_{50} = 0.0233\ (\pm0.0061)PI + 0.6431\ (\pm0.1937)\ IP_1 + 0.5491\ (\pm0.3900)IP_2 + 2.6252 \quad (6)$$

Eqs (5) and (6) further show that poly-halogen substitution enhances exhibition of logP and pEC_{50} and that mono-halogen substitution is less favourable.

Finally, successive regression resulted into tetra-parametric regression expressions having the highest quality. These tetra-parametric models are found as shown:

$$logP = 0.0281\ (\pm0.0023)PI + 0.6088\ (\pm0.0813)\ IP1 + 0.6384\ (\pm0.1289)IP_2 + 0.2767\ (\pm0.0971)\ IP_3 + 1.6856 \quad (7)$$

$$pEC_{50} = 0.0237\ (\pm0.0057)PI + 0.7992\ (\pm0.2027)\ IP_1 + 0.6949\ (\pm0.3212)IP_2 + 0.4255\ (\pm0.2419)\ IP_3 + 2.4524 \quad (8)$$

The results obtained in the aforementioned models [eqs (3)–(8)] indicated that halogen-substituted benzene derivatives are preferable for the exhibition of logP and pEC_{50}, and that poly-halogen substitution gives the best toxicological effect of the benzene derivatives used.

It is worth mentioning that all these bi-, tri- and tetra-parameteric models [eqs (3)–(8)] suffer from the collinarity defects due to the occurrence of PI and IP_2 terms. However, as discussed below, in spite of the collinearity defects, these models are considered statistically significant.

Randic[4] has shown that selection of descriptors to be used in QSPR, QSAR, and QSTR studies should not be delegated solely to the computers, although the statistical criteria will continue to be useful for preliminary screening of descriptors taken from a large pool. Often in an automated selection of descriptors, a descriptor will be discarded because it is highly correlated with another descriptor already selected, but what is important is not whether two descriptors parallel to one another, that is duplicate much of the same structural information, but whether they in those parts that are important for QSPR, QSAR, and QSTR correlations. If they differ in the domain which is important for the property, activity on toxicity considered, both descriptors should be retained. If they differ in parts that are not relevant for the correlation of considered property, activity or toxicity, then one of them can be discarded. In our cases, PI and IP_2 give different types of informations and are, therefore, important for toxicity consideration of the benzene derivatives used. Furthermore, IP_2 is used as a dummy parameter to account for the structural features not covered in the PI Index. Hence, in spite of the observed collinearity, PI and IP_2 can be retained in the proposed models.

Another support in favour of the retention of PI and IP_2 terms comes from the values of R_A^2; which goes on increasing as we pass from bi- to tetra-parametric models. This consistent increase in R_A^2 also indicates that, in spite of observed collinearity between PI and IP_2, the models containing both of these parameters are statistically significant. Note that, if a variable is added that does not contribute its fair share, the R_A^2 will actually decline. R_A^2 is a measure of the percentage explained variation in the dependent variable that takes into account the relationship between the number of cases and the number of independent variables in the regression model. Whereas R^2 will always increase when an independent variable is added, R_A^2 will decrease if the added variable does reduce the unexplained variation enough to offset the loss of degrees of freedom.

At this stage, it is interesting to record that we have 20 benzene derivatives, the toxicity (logP and pEC_{50}) of which is best modelled by a tetraparametric regression models. There are, therefore, 5 times more compounds than the parameters used in the most significant model [eqs (7) and (8)].

Multiple regression analysis generally requires significantly more compounds than parameters; a useful rule of thumb is 3–6 times the number of parameters under consideration. Hence, our proposed tetra-parametric models fall well within the range of thumb rule and, therefore, can be considered statistically significant.

As opposed to traditional regression models cross-validation[31–33] evaluates the validity of a model by how well it predicts the data rather than how well it fits data. The analysis uses a 'leave-one-out' scheme; a model is built with $N-1$ compounds and the Nth compound is predicted. Each compound is left out of the model derivation and predicted in turn. An indication of the performance of the model is obtained from cross-validated parameters, particularly on the basis of cross-validated r^2 which is defined as shown:

$$r_{cv}^2 = \frac{SD - PRESS}{SD} \quad (9)$$

where SD is the sum of squares deviation for each activity from the mean and PRESS (or predictive sum of squares) is the sum of the squared difference between actual and that predicted when the compound is omitted from the fitting process. Once a model is developed which has the highest cross-validated r^2, that is r_{cv}^2, this model is used to derive the conventional QSPR, QSAR and QSTR equation, and conventional R^2 and Se values.

The cross-validated parameters estimated for the aforementioned models are presented in Table 5. Since, in all the cases, PRESS < SD, this indicates that they predict better than chance and can be considered statistically significant. To be a reasonable QSPR, QSAR, and the QSTR model, PRESS/SD should be smaller than 0.4, and value of this ratio smaller than 0.1 indicates an excellent model. This ratio is smallest for models 5 and 10 (Table 5), and thus, they can be regarded as excellent models for modelling logP and pEC_{50}, respectively, of the benzene derivatives under present study.

A comparison of Se (Table 4) and S_{press} (Table 5) values shows that they are similar in all the cases. It means that these parameters are not good for deciding the uncertainty of the models. In such cases, predictive-square-error (PSE) is the best parameter for deciding the uncertainty of prediction. The smallest value of PSE indicates the best predictive potential. The PSE values reported in Table 5 support that the models 5 and 10 (Table 5) are the most appropriate models for modelling, monitoring and estimating toxicological potential (logP and pEC_{50}) of the benzene derivatives under present investigation.

In literature, a parameter Q (quality factor) is described[38–40] for deciding quality of correlation. This quality factor, Q, is defined as the ratio of correlation coefficient R to the standard error or estimation ($Q = R/Se$). That is, Q takes amount of both R and Se is deciding quality of correlation. By definition, the higher the value of R, the smaller the Se, the larger will be Q and better will the quality of correlation. These Q values for the attempted correlations are presented in Table 4, showing that Q is the highest for correlation expressed by eq. (5) in modelling logP and also for eq (10) for modelling pEC_{50}. Hence, both Q and PSE indicate that models eqs (5) and (10) (Table 5) have the highest quality as well as predictive potentials.

Also, the individual F statistics (t statistic, $p < 0.05$) for the coefficients of variables in the aforementioned best models eqs (5) and (10) were significant at $p < 0.05$ (Table 6). The predictor variables PI, IP_1, IP_2 and IP_3 accounted for a significant portion of the variance in logP and pEC_{50}. Also, each predictor contributed a significant proportion of the additional variance in the presence of other variables for each activity (i.e., logP, pEC_{50}).

The aforementioned results and discussion indicates that hydrophobicity (expressed by logP) which also represent toxicological effect of benzene derivatives can be modelled successfully by the PI Index. Likewise, toxicity expressed as pEC_{50} can also be modelled equally well by the PI Index. The PI Index and indicator parameters show that bulk parameters play an important role as toxicity to aquatic organiser, that is *V. fischeri*, an organism with the lowest lipid content.

The toxicity process involves both toxicant transport to the site of interaction on a biomacromolecule and interaction between the chemical and that site. Our results related to logP indicate that it is very important in understanding the bioconcentration mechanism of *V. fischeri* to determine whether or not logP is an ideal molecular descriptor to express the bioconcentration process between water and the animals.

The aforementioned results and discussion indicate that the PI Index can be used successfully for modelling logP of the benzene derivatives under present study. As logP has been well correlated with different topological parameters like molecular connectivity,[41] it is obvious that any so-called parameter based on graph theory with some mathematical jargon will show a similar correlation. The same has been the case here. Moreover, it is also well established that substitution by halogroup in benzene increases toxicity. Hence, as such our results do not significantly add to the existing knowledge in the present case. The results only show that the newly introduced PI Index has a significant potential in monitoring, modelling and estimating logP and pEC_{50} of the benzene derivatives.

Given the above situation, what is more important is to distinctly establish the superiority of the PI Index over other topological parameters in explaining toxicity data of substituted benzenes under present study. This was also the second objective of our study. Consequently, we now compare the results obtained from the PI Index with six other widely used topological indices. These topological indices are: Wiener Index (W)[28] Branching Index (B),[42] Randic Connecting Index (χ),[41] Balaban Index (J),[43,44] Szeged Index (Sz)[29,30] and logR.[45] The calculated values of these indices are presented in Table 7 and their relatedness is shown in Table 8.

The relative correlation potential vis-à-vis the superiority of the PI Index over the topological indices men-

Table 5. Cross-validation parameters for the models presented in Table 3

Model	PRESS	SD	PRESS/ SD	R^2_{cv}	S_{press}	PSE
For modelling logP						
1	1.2738	11.0980	0.1148	0.8852	0.2660	0.2524
2	0.7445	11.6273	0.0640	0.9360	0.2093	0.1929
3	0.6977	11.6740	0.0598	0.9402	0.2088	0.1868
4	0.4083	11.9635	0.0341	0.9659	0.1598	0.1429
5	0.2648	12.1070	0.0219	0.9781	0.1329	0.1151
For modeling of pEC_{50}						
6	3.3491	8.2452	0.4062	0.5938	0.4314	0.4092
7	2.3263	9.2680	0.2510	0.7490	0.3699	0.3410
8	2.1570	9.4373	0.2286	0.7714	0.3672	0.3284
9	2.4812	9.1131	0.2723	0.7277	0.3938	0.3522
10	1.6440	9.9503	0.1652	0.8348	0.3310	0.2867

tioned above can be established from the data presented in Tables 9 and 10.

A perusal of Table 9 shows that in modeling logP (hydrophobicity and lipophicity), the PI Index is superior to W, J, Sz and logR. Note that W is the oldest and widely used topolgoical index in QSPR and QSAR studies and that J is the most descriminating index. Our results show that the PI Index is better than the widely used W Index as well as the most discriminating, the J Index. However, our PI Index is inferior to B and χ indices in modelling logP.

The data presented in Table 10 shows that the PI Index is most suitable among all other topological indices for modelling pEC_{50}. That is, here also the PI Index is superior to the W, J, Sz, logR, B and χ indices.

Conclusion

The aforementioned results and discussion lead us to conclude that logP is an efficient parameter for expressing toxicological effects of benzene derivatives used in the present study. Both logP and pEC_{50}, ($-\log EC_{50}$; EC_{50} = concentration causing 50% inhibition of bioluminescence to *V. fischeri*) can be modelled excellently by tetraparametric models consisting of PI, IP_1, IP_2 and IP_3. Based on a cross-validation test, Q-values and *t*-test, these models are found to have high qualitative and predictive potential.

Table 6. The *t*-statistics ($p < 0.05$) for the coefficient of variables in the best eqs (5) and (10)

	T	*p*
Eq (5)	*T*(DF = 15)	
PI	12.170	0.00000
IP_1	7.486	0.00000
IP_2	4.952	0.00017
IP_3	2.851	0.01215
Eq (10)	*T*(DF = 15)	
PI	4.118	0.00091
IP_1	3.944	0.00130
IP_2	2.164	0.04705
IP_3	1.759	0.09891

Experimental

Toxicity data

The toxicity data, namely logP and EC_{50} data, were adopted[34] from the earlier work in that the benzene derivatives were diluted with 3% NaCl solution for *V. fischeri* toxicity tests. An ultrasonic approach was used to aid the dissolution of sparingly soluble compounds in water. Due to the high volatility of some chemicals, tests were always carried out in closed systems. The *V. fischeri* test instrument that measured bioluminescence was Model Toxicity Analysis DXY. Ecotoxicological descriptors were the concentration values causing a 50% inhibition of bioluminescence after 15 min exposure (15 min—EC_{50} mol/l). The 15 min—EC_{50} values were determined at $20 \pm 2\,^{\circ}C$. Bioluminescence was measured after 15 and 30 min with the toxicity analyzer. Results were similar at the two time periods; however, they used values for 15 min incubation.

In the present investigation, the aforementioned EC_{50} values were converted into pEC_{50} ($= -\log EC_{50}$) units and used as recorded in Table 1. It is worth recording

Table 7. Other topological indices: W, Sz, J, χ, B and logR for benzene derivatives (Ref. Table 1) in benzene derivatives used

Comp. no.	Topological indices						
	W	B	χ	J	Sz	LogR	PI
1	179	5.4641	5.4641	2.7600	282	87.0114	126
2	111	4.6091	4.6091	2.4620	186	35.5227	82
3	84	4.1934	4.1934	2.3462	144	26.4585	66
4	82	4.2152	4.2152	2.4133	140	25.9477	66
5	62	3.7877	3.7877	2.1924	110	19.0038	50
6	61	3.7877	3.7877	2.2306	108	18.7806	50
7	60	3.8045	3.8045	2.2794	106	18.4930	50
8	42	3.3939	3.3939	2.1229	78	12.4245	36
9	62	3.7877	3.7877	2.1924	110	19.0038	50
10	61	3.7877	3.7877	2.2306	108	18.7806	50
11	42	3.3939	3.3939	2.1229	78	12.4245	36
12	62	3.7877	3.7877	2.1924	110	19.0038	50
13	111	4.6091	4.6091	2.4620	186	35.5227	82
14	84	4.1934	4.1934	2.3462	144	26.4585	66
15	62	3.7879	3.7879	2.1924	110	19.0038	50
16	60	3.7879	3.7879	2.1924	110	19.0038	50
17	61	3.7877	3.7877	2.2306	108	18.7806	50
18	60	3.8045	3.8045	2.2794	106	18.4930	50
19	42	3.3939	3.3939	2.1229	78	12.4245	36
20	27	3.0000	3.0000	2.0000	54	7.4547	24

that logarithmic expressions are used in property, activity and toxicity studies because QSPR, QSAR, and QSTR represent free-energy changes.

Determination of logP

The logP values were also taken from the earlier work[34] wherein logarithms of the 1-octanol/water partition coefficient were calculated values from the MEDCHEM (LOGP Ver 3.55) software.

Definition and calculation of the PI Index

The definition and details of the calculation of the PI Index are already given in our earlier publications.[22–26] However, for the sake of clarity, we define the PI Index briefly here.

Let G be a molecular graph, the vertex and edge-sets of which are represented by $V(G)$ and $E(G)$, respectively. If e is an edge of G, connectivity of vertices u and v, then we write $e=uv$.

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. The meaning $n_{ev}(e|G)$ is analogues. Then, the PI Index is defined as:

$$\mathrm{PI} = \mathrm{PI}(G) = \sum_{e}[n_{eu}(e|G) + n_{ev}(e|G)] \quad (10)$$

The summation goes over all edges of G. Edges equidistant from both ends of the edge uv are not counted (taken in to account) for the calculation of PI. These PI indices in the present case are given in Table 1.

Other topological indices

The other topological indices, namely Wiener (W), Szeged (Sz), Balaban (J), Randit (χ), Branch (B), and logR indices, were calculated using the program made available by Professor Istavan Luckovits, Hungarian Academy of Sciences, Budapest, Hungary.

Indicator parameters

The indicator parameters are dummy parameters sometimes used in multiple regression analysis for accounting structural features not taken care of in the molecular descriptors used. These variables take on only two values, usually zero and one. The two values signify that the observation belongs in one of the two possible categories. The numerical values of dummy variables are not intended to reflect a quantitative ordering of cate-

Table 8. Correlation matrix exhibiting relatedness of the PI Index with other topological indices (W, Sz, J, χ, B, and logR)

	W	B	χ	J	Sz	logR	PI
W	1.0000						
B	0.9879	1.0000					
χ	0.9879	1.0000	1.0000				
J	0.9744	0.9853	0.9853	1.0000			
Sz	0.9995	0.9914	0.9914	0.9754	1.0000		
LogR	0.9998	0.9898	0.9898	0.9766	0.9998	1.0000	
PI	0.9985	0.9940	0.9940	0.9830	0.9988	0.9988	1.0000

Table 9. Regression parameters used in establishing superiority of the PI index over other topological indices in modeling logP of the benzene derivatives used

S. No.	Parameter used	Se	R	F	P	Q
1	W	0.3035	0.9345	123.990	1.666×10^{-9}	3.0790
2	B	0.2429	0.9585	203.722	2.967×10^{-7}	3.9460
3	χ	0.2429	0.9585	203.722	2.967×10^{-7}	3.9460
4	J	0.3164	0.9286	112.663	3.541×10^{-9}	2.9348
5	Sz	0.2883	0.9411	139.359	6.567×10^{-10}	3.2643
6	logR	0.2968	0.9374	130.420	1.115×10^{-9}	3.1583
7	PI	0.2660	0.9471	156.827	4.664×10^{-10}	3.5605

Table 10. Regression parameters used in establishing superiority of the PI Index over other topological indices in modelling toxocity (pEC_{50}) of the benzene derivatives used

S. no.	Parameter used	Se	R	F	P	Q
1	W	0.3691	0.8748	58.667	4.528×10^{-7}	2.3700
2	B	0.3652	0.8726	60.307	3.732×10^{-7}	2.3887
3	χ	0.3652	0.8776	60.307	3.732×10^{-7}	2.3882
4	J	0.3876	0.8608	51.524	1.110×10^{-6}	2.2208
5	Sz	0.3666	0.8766	59.720	3.997×10^{-7}	2.3911
6	logR	0.3680	0.8756	59.126	4.288×10^{-7}	2.3793
7	PI	0.3661	0.8769	59.914	3.908×10^{-7}	2.3952

gories, but they serve to identify category or class membership.

In the present study, we have used three such dummy (indicator) parameters, namely IP_1, IP_2 and IP_3. IP_1 is taken as unity for di-halogen substitution, in the absence of which IP_1 is zero. IP_2 assumes the value of unity under the condition of poly- (three or more)-halogen-substitution, otherwise IP_2 is zero. Finally, IP_3 is taken as unity for mono-halogen substitution. These values are presented in Table 1.

Statistical analysis

The statistical analyses were made as described in the literature[31–33] using of software supplied by Professor Istavan Lukovits.

Acknowledgements

The authors are thankful to Professor Ivan Gutman for introducing them to this fascinating field, namely Chemical Graph Theory and Topology and to Professor Istavan Lukovits for providing the software. They are also thankful to Professor Subhash Basak for his keen interest and fruitful help. The authors are also thankful to the referee for suggesting several modifications which has improved the quality of the paper.

References and Notes

1. Devillers, J., Balaban, A.T., Eds. *Topological Indices and Related Descriptors in QSAR and QSPR*; Gorden and Breach Science: Amsterdam, 1999
2. Basak, S. C.; Mills, D. R.; Balaban, A. T.; Gute, B. D. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 671.
3. Basak, S. C.; Granwald, G. D.; Gute, B. D.; Balasubramanian, K.; Opitz, D. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 885.
4. Randic, M.; Basak, S. C. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 614.
5. Gute, B. D.; Basak, S. C. *SAR and QSAR Environ. Res.* **1997**, *7*, 117.
6. Agrawal, V.K.; Khadikar, P.V *Bioorg. Med. Chem.* **2001**, *8,* 3035.
7. Khadikar, P. V.; Karmarkar, S.; Agrawal, V. K. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 934.
8. Khadikar, P. V.; Karmarkar, S.; Agrawal, V. K. *Nat. Acad. Sci. Lett.* **2000**, *23*, 124.
9. Karmarkar, S.; Khadikar, P. V.; Mandloi, M.; Joshi, S.; Agrawal, V. K. *Indian J. Chem.* **2001**, *40A*, 12.
10. Agrawal, V. K.; Sachan, S.; Khadikar, P. V. *Pol. J. Pharmacol.* **2000**, *52*, 39.
11. Agrawal, V. K.; Sachan, S.; Khadikar, P. V. *Acta Pharm.* **2000**, *50*, 281.
12. Mandloi, M.; Agrawal, V. K.; Mathur, K. C.; Karmarkar, S.; Khadikar, P. V. *Acta Pharm.* **2000**, *50*, 303.
13. Karmarkar, S.; Joshi, S.; Sharma, S.; Khadikar, P. V. *J. Indian Chem. Soc.* **2000**, *77*, 433.
14. Mandloi, M.; Sikarwar, A.; Sapre, N. S.; Karmarkar, S.; Khadikar, P. V. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 57.
15. Saxena, A.; Khadikar, P. V. *Acta Pharm.* **1999**, *49*, 171.
16. Karmarkar, S.; Karmarkar, S.; Joshi, S.; Das, A.; Khadikar, P. V. *J. Serb. Chem. Soc.* **1997**, *62*, 227.
17. Khadikar, P. V.; Karmarkar, S.; Sharma, S.; Seerwani, A. D.; Joshi, S. *J. Serb. Chem. Soc.* **1997**, *62*, 219.
18. Khadikar, P. V.; Karmarkar, S.; Joshi, S.; Gutman, I. *J. Serb. Chem. Soc.* **1996**, *61*, 89.
19. Khadikar, P. V.; Joshi, S. *X-ray Spectro.* **1995**, *24*, 201.
20. Agrawal, V. K.; Khadikar, P. V. *Bioorg. Med. Chem.* **2001**, *9*, 2787.
21. Trinajstic, N. *Chemical Graph Theory*, 2nd ed.; CRC: Boca Raton, FL, 1993.
22. Agrawal, V. K.; Khadikar, P. V. *Bioorg. Med. Chem.* **2001**, *9*, 3295.
23. Khadikar, P. V. *Nat. Acad. Sci. Lett.* **2000**, *23*, 113.
24. Khadikar, P. V.; Karmarkar, S.; Agrawal, V. K. *Nat. Acad. Sci. Lett.* In press.
25. Khadikar, P. V.; Karmarkar, S.; Agrwawal, V. K. *Nat. Acad. Sci. Lett.* **2000**, *23*, 165.
26. Khadikar, P. V.; Kale, P. P.; Deshpande, N. V.; Karmarkar, S.; Agrawal, V. K. *J. Math. Chem.* **2001**, *29*, 143.
27. Khadikar, P. V.; Karmarkar, S.; Agrawal, V. K. J. *Chem. Inf. Comput. Sci.* Submitted for publication.
28. Wiener, H. *J. Am. Chem. Sci.* **1947**, *69*, 17.
29. Gutman, I. *Graph Theory Notes, New York* **1994**, *27*, 9.
30. Khadikar, P. V.; Deshpande, N. V.; Kale, P. P.; Dobrynin, A.; Gutman, I.; Domotor, G. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 547.
31. Khadikar, P. V.; Kale, P. P.; Deshpande, N. V.; Karmarkar, S.; Agrawal, V. K. *Commun. Math. Comput. Chem. (MATH)* **2001**, *43*, 7.
32. Chaterjee, S.; Hadi, A. S.; Price, B. *Regression Analysis by Examples*, 3rd ed.; John Wiley & Sons: New York, 2000.
33. Lucic, B.; Trinajstic, N. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 121.
34. Lucic, B.; Amic, D.; Trinajstic, N. *J. Chem. Inf. Comput. Sci.* **2001**, *40*, 403.
35. Lucic, B.; Trinajstic, N.; Slid, S.; Karelson, M.; Kartizky, A. R. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 610.
36. Zhao, Y. H.; Cronin, M. T. D.; Deavden, J. C. *Analt. Struct.-Act. Relat* **1998**, *17*, 131.
37. Balaban, A. T. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 23.
38. Pogliani, L. *Amino Acids* **1994**, *6*, 141.
39. Pogliani, L. *J. Phys. Chem.* **1996**, *100*, 18065.
40. Institute Inc, S. A. S. *SAS/STAT User's Guide, release 6.03*; SAS Institute: Cary, NC, 1988.
41. Randic, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609.
42. Trinajstic, N. *Chemical Graph Theory*, 2nd ed.; CRC: Baca Raton, FL, 1992.
43. Balaban, A. T. *Chem. Phys. Lett.* **1982**, *189*, 399.
44. Khadikar, P. V.; Sharma, S.; Sharma, V.; Joshi, S.; Lukovits, I.; Kaveeshwar, M. *Bull. Soc. Chem. Belg.* **1997**, *106*, 767.
45. Todeschini, R.; Consonn, V. *Handbook of Molecular Descriptors*; Wiley: VCH Weinheim, Germany, 2000.