Pergamon

Bioorganic & Medicinal Chemistry 10 (2002) 3163–3170

BIOORGANIC & MEDICINAL CHEMISTRY

Use of the PI Index in Predicting Toxicity of Nitrobenzene Derivatives

Padmakar V. Khadikar,[a,*] Sneha Karmarkar,[a] Shalini Singh[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, A.P.S. University, Rewa 480003, India*
[c]*Department of Chemistry, Govt. Holkar Model and Autonomous College, Indore 452 001, India*

Received 20 March 2002; accepted 7 June 2002

Abstract—The PI Index is a Szeged-like topological index developed very recently. It has useful applications in chemistry which are yet to be investigated thoroughly. Herein, we report quantitative structure–toxicity relationship (QSTR) study using the PI Index. We have used 41 monosubstituted nitrobenzene for this purpose. The results have shown that the PI Index alone is not an appropriate index for modelling toxicity of nitrobenzene derivatives. Combinations of the PI Index with other distance-based topological indices resulted into statistically significant models and excellent results are obtained in pentaparametric models. The predictive potential of the models is discussed on the basis of cross-validation method, as well as by estimating root-mean-square error (RMS).

Introduction

An important aspect of modern toxicology research is the prediction of toxicity of chemicals from their molecular structures.[1–8] Since Hansch and Fujita[9] developed the quantitative structure–activity relationship (QSAR) method, it has been successfully applied to drug and agrochemical designing as well as to the prediction of toxicological end points. Quantitative structure–activity/toxicity relationship (QSAR/QSTR) models have emerged as useful tools to handle the data gap in toxicology and pharmacology.[10–13] QSAR/QSTR models can be used to estimate complex properties of chemicals from simpler experimental or computed properties.[14,15] In view of the fact that most chemicals in commerce and environmental pollutants have very little test data, it would be desirable if we could develop toxicologically relevant QSTRs from properties that can be calculated directly from chemical structures. Graph theoretical descriptors, namely topological indices, are such structural descriptors for this purpose. There is a recent upsurge of interest in the topological indices to QSAR/QSTR studies.[10–17] A topological index is a numerical descriptor of molecular structure and is sensitive to such key constitutional features as size, shape, symmetry, branching and heterogenicity of utmost environment in the molecules.[16,17]

A plethora of topological indices are reported in the literature.[16–18] However, very few of them are successful in QSAR/QSTR studies. The Wiener Index (W) is the first, oldest, and widely used topological index in such studies.[19] Even today, the W Index is successfully used in QSAR/QSTR studies.[20] However, this index is not applicable to cyclic compounds and compounds containing cycle(s) with tree-like (acyclic) side chain(s). Consequently, a new topological index, the Szeged Index (Sz) was introduced by Gutman,[21,22] The Szeged Index is considered as a modification of the Wiener Index to cyclic compounds. In the case of acyclic compounds, Sz and W coincides.

Consequent to the coincidence of the Sz Index with the W Index of acyclic compounds we have very recently

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

0968-0896/02/$ - see front matter © 2002 Published by Elsevier Science Ltd.
PII: S0968-0896(02)00211-0

introduced another topological index, namely the PI (Padmakar–Ivan) Index.[23–28] The PI Index is considered as the modification of the Sz Index so that now the PI Index is different for acyclic and cyclic compounds. Our earlier communication[24] stressed that the PI Index has useful applications in chemistry, which is yet to be investigated thoroughly. In this article, we will use the PI Index and its combinations with different distance-based topological indices,[30–45] in predicting toxicity of nitrobenzenes.

Results and Discussion

The present study was primarily undertaken to determine the extent to which the newly introduced PI Index can be used to predict the toxicity of a series of nitrobenzene derivatives (Table 1). Experimental values of toxicity (IGC_{50}, mmol L^{-1}) were adopted[29] and converted into their $p(IGC_{50})$ units and are given in Table 1. These values are of the toxicity on nitrobenzene derivatives to *Tetrahymena pyriformis*.

Table 1. Structural details of nitrobenzene derivatives, their toxicity and molecular descriptors

Compd	*R*	$p(IGC_{50})$	W	Sz	PI	J	MRI	V_w
1	H	0.355	88	142	66	2.2215	0.4735	1.009
2	2-NH_2	0.077	188	278	126	2.4000	0.3439	1.108
3	2-CH_3	0.479	114	180	84	2.3945	0.3279	1.163
4	2-Cl	0.676	114	180	84	2.3945	0.3833	1.174
5	2-Br	0.863	114	180	84	2.6049	0.3833	1.215
6	2-CH_2OH	−0.155	197	287	126	2.4185	0.2667	1.231
7	2-C_6H_5	0.301	352	586	228	1.9566	0.5562	1.727
8	2-$CONH_2$	−0.721	290	404	176	2.6166	0.2662	1.283
9	2-CHO	0.174	150	228	104	2.4513	0.2814	1.189
10	3-NH_2	0.026	197	278	126	2.0640	0.3439	1.108
11	3-OH	0.506	156	240	94	2.3270	0.3074	1.077
12	3-CH_3	0.572	117	192	84	2.2720	0.3279	1.163
13	3-Cl	0.836	117	192	84	2.2720	0.3833	1.174
14	3-Br	1.215	117	192	84	2.2720	0.3833	1.215
15	3-CN	0.451	156	240	104	2.3270	0.3074	1.586
16	3-CH_2OH	−0.220	206	305	126	2.2932	0.2667	1.231
17	3-C_6H_5	1.569	370	622	228	2.1555	0.5552	1.727
18	3-$CONH_2$	−0.193	305	434	176	2.4647	0.2667	1.283
19	3-CHO	0.140	156	240	104	2.3270	0.2814	1.189
20	3-NO_2	0.762	197	296	126	2.0640	0.4376	1.196
21	3-OCH_3	0.670	156	240	104	2.3270	0.3035	1.244
22	3-$COCH_3$	0.317	197	296	126	2.0640	0.2824	1.338
23	4-CH_3	0.796	120	192	84	2.4470	0.3279	1.163
24	4-C_2H_5	0.804	162	252	104	2.2401	0.2866	1.317
25	4-OCH_3	0.544	162	252	104	2.2401	0.3035	1.244
26	4-OC_2H_5	0.829	215	323	126	2.1948	0.3599	1.398
27	4-OC_4H_9	1.420	358	502	176	2.0888	0.1932	1.706
28	4-F	0.253	120	192	84	2.4470	0.3833	1.055
29	4-Cl	0.559	120	192	84	2.4410	0.3833	1.174
30	4-Br	0.461	120	192	84	2.4450	0.3833	1.215
31	4–1N	0.569	162	252	104	2.2401	0.3074	1.186
32	4-CH_2CN	0.132	215	323	116	2.1948	0.2667	1.340
33	4-CH_2Cl	1.180	162	252	104	2.2401	0.2882	1.328
34	4-CH=NOH	0.678	280	406	150	2.1515	0.2400	1.301
35	4-NHC_6H_5	1.886	552	876	294	1.7814	0.3748	1.834
36	4-CH_2OH	0.101	215	323	116	2.1948	0.2667	1.231
37	4-$COOCH_3$	0.398	262	388	138	2.2958	0.2635	1.419
38	4-$COOC_2H_5$	0.710	331	475	176	2.2722	0.2381	1.573
39	4-$CONH_2$	0.179	320	464	176	2.3478	0.2667	1.283
40	4-CHO	0.203	162	252	104	2.2401	0.2814	1.189
41	4-NO_2	1.301	206	314	126	2.2930	0.4356	1.196

The calculated values of the PI Index and other supporting topological indices are summarized in Table 1. The correlation matrix[46,47] for $p(IGC_{50})$ and the five topological indices used in this study is given in Table 2 which shows that including the PI Index none of the topological indices show collinearity with the toxicity $p(IGC_{50})$ of nitrobenzene derivatives used. This means that no simple regression model using the present set of topological indices is possible to model the toxicity. However, the data presented in the correlation matrix (Table 2) does show that $p(IGC_{50})$ can be modeled successfully in multiparameteric models and that the multiparametric models should involve the combinations of the PI Index with J, MRI and V_w. Furthermore, both the W and Sz indices are highly linearly correlated with the PI Index. Therefore, their combination with the PI Index will result into a multiparametric modeling in that the resulting models may suffer from the defect due to collinearity. However, our earlier results[20,24,26,38] showed that the Sz Index contain some hether to unknown information not present in the W and PI indices. Consequently, simultaneous presence of the W and Sz indices may result into a statistically significant model. As discussed below, this is found to be true in the present case also.

The data presented in Table 2 (correlation matrix) also show that information content (MRI) and steric parameter (V_w) play a dominant role in the exhibition of $p(IGC_{50})$. It means in multiparametric models these parameters will prominently occur.

The aforementioned results promoted us to undertake multi-parametric regressions in that the PI Index is to be combined with the remaining four topological indices in different ways, so as to arrive at statistically significant models for modeling the toxicity of nitrobenzene derivatives.

We have attempted several combinations of the PI Index, that is starting from bi-parameter to hexa-parametric regressions. The results have shown that no statistically significant bi- and tri-parametric models could be obtainable for modelling toxicity of the nitrobenzenes used. However, we obtained a statistically significant tetra-parametric (Table 3) model consisting of the combination of PI, J, MRI and V_w. This model is found as under:

$$\begin{aligned} {}_{p}(IGC_{50}) = {} & -0088\ (\pm 0.0020) \\ & \mathrm{PI} - 1.1054\ (\pm 0.4018) \\ & \mathrm{J} + 2.7647\ (\pm 0.6526) \\ & \mathrm{MRI} - 2.7717\ (\pm 0.5351)\ V_w - 0.2887 \end{aligned} \quad (1)$$

No other tetra-parametric models resulted having better quality than the model expressed by eq 1. It is worth mentioning that use of the Sz Index instead of the PI Index in eq 1 resulted into a model with a considerable lower (poor) statistics. This explains superiority of the PI Index over the Sz Index. Recall that the PI Index is a modified parameter of the Sz Index.

Table 2. Correlation matrix for the intercorrelation of molecular description used and their correlation with toxicity

	$PIGC_{50}$	PI	Sz	W	J	MRI	V_w
$PIGC_{50}$	1.0000						
PI	0.2807	1.0000					
Sz	0.3556	0.9891	1.0000				
W	0.2784	0.9825	0.9905	1.0000			
J	−0.5796	−0.5513	−0.5944	−0.5550	1.0000		
MRI	0.5099	0.1578	0.1311	0.0242	0.0325	1.0000	
V_w	0.5678	0.8484	0.8773	0.8511	0.5250	0.1716	1.0000

Table 3. Quality of proposed multiparametric models

Model	Parameters used	SE	R^2_A	R^2	R	F	P	Q
1 (eq 1)	PI,J,MRI, V_w	0.3054	0.6790	0.7119	0.8438	21.624	1.527×10^{-10}	2.7629
2 (eq 2)	PI,W,J,MRI,V_w	0.2661	0.7564	0.7876	0.8875	25.215	4.68×10^{-9}	3.3352
3 (eq 3)	PI,Sz,J,MRI V_w	0.2554	0.77567	0.8044	0.8969	27.957	3.887×10^{-11}	3.5118

SE, standard error and estimation; R^2_A, adjustable R^2; R^2, coefficient of variance; R, correlation coefficient; F, F-ratio; p, probability; Q, quality factor.

In the aforementioned eq 1, the coefficients of PI, J and V_w are negative, while that of the MRI term is positive. The earlier reported[24] high collinearity of PI with both the W and Sz indices indicates that, like W and Sz, the PI Index contributes towards size and shape. Hence, negative coefficient of PI in eq 1 indicates that decrease in size and shape of nitrobenzene favours the exhibition of $_P(IGC_{50})$.

The Balaban Index (J)[33–36] is a highly discriminating descriptor, whose value do not substantially increase with the molecular size and the number of rings present. Furthermore, J is considered as the average distance sum connectivity index. The set of compounds under present investigation are all monocyclic molecules, any change in size and shape will then be attributed to tree-like acyclic side chain. Therefore, changes in J values are due to connectedness. Negative coefficient of J term in eq 1 indicates that decrease in connectedness favour $_P(ICG_{50})$ exhibition.

The van der Waals volume[43–45] is a steric parameter and is directly related to the geometric locus around molecules where repulsive and attractive interactions with approaching atoms balance each other. The negative coefficient of V_w in eq 1 shows that bulky substituents in the aromatic nucleus causes a decrease in the toxicity of the nitrobenzenes. Therefore, for a nitrobenzene to produce lowermost fatility rates the substitution of bulky substituent should be preferred.

Finally, the molecular redudency index (MRI)[37–42] is an information theoretic topological index. The positive coefficient of MRI term in eq 1 suggests increase information content favours the exhibition of $_P(IGC_{50})$.

Addition of the Wiener Index (W) as an additional parameter resulted in a pentaparametric model with improved quality (Table 3). The model is found as under:

$$\begin{aligned} {}_P(IGC_{50}) = {} & -0.0332\ (\pm0.0072) \\ & PI + 0.0129\ (\pm0.0037) \\ & W - 0.8467\ (\pm0.3579) \\ & J + 5.0086\ (\pm0.8596) \\ & MRI + 2.3491\ (\pm0.4817)\ V_w - 0.6530 \end{aligned} \quad (2)$$

A comparison of eq 2 with eq 1 indicates that the coefficients of the PI and MRI terms are increased while that of J and V_w are decreased. Furthermore, the added W term in eq 2 has positive coefficient. The opposite signs of PI and W indices in the model eq 2 may be attributed to the collinearity defects. However, here collinearity defect is not that serious as the coefficients of both the W and PI terms are significantly higher than their respective standard deviations. Furthermore, addition of the W Index to eq 1 resulted in an increase in the value of R^2_A (Table 3). It is worthwhile to record that, if a variable is added that does not contribute its fair share, the R^2_A value will actually decline. Hence, in spite of a change in sign, that is opposite signs of PI and W (due to collinearity defect) eq 2 is statistically significant.

In order to investigate the relative correlation potential of W and Sz indices, we tried another pentaparametric model in that we have used the Sz Index in place of W. Use of Sz in place of W gave the following penta-parameter model in that quality is slightly improved (Table 3) This improved model is found as under:

$$\begin{aligned} {}_P(IGC_{50}) = {} & -0.0346\ (\pm0.0067) \\ & PI + 0.0096\ (\pm0.0024) \\ & Sz - 0.6356\ (\pm0.3559) \\ & J + 3.4419\ (\pm0.5712) \\ & MRI + 1.9272\ (\pm0.4945)\ V_w - 0.2373 \end{aligned} \quad (3)$$

Like the earlier model, here PI and Sz also possessing high collinearity are present in the proposed model eq 3.

Here also, the opposite signs of the coefficients of PI and Sz terms may likewise be attributed to collinearity defect. However, because the coefficients of PI and Sz terms are significantly larger than their respective standard deviations, we can safely say that, in spite of the observed collinearity between the PI and Sz indices, the proposed model eq 3 is statistically significant.

It is worth mentioning that the Sz Index is a modification of the W Index for cyclic compounds and also that high collinearity existed between the W and Sz indices in the present case also. However, since the compounds under present study are all monocyclic molecules containing one aromatic ring with various tree-like (acyclic) side chains, any change in Sz will be due to variation in the tree-like (acyclic) side chains in that coincidence of W and Sz is well known. Therefore, a better model being obtained (eq 3) by replacing W by Sz indicates that Sz contains some additional structural information not present in the W Index. Research in this direction is under way and will be published elsewhere. However, at present, we can say that the better quality of this model is principally due to the presence of a cyclic component (aromatic nucleus) for which only Sz is applicable and not W. This indicates that use of Sz in place of W further favours exhibition of $_{P}(ICG_{50})$. Other interpretations of eq 3 is the same as discussed for eq 2. No other higher parametric correlations gave better results than eq 3.

In order to give an exact idea with regards to the correlation potentials of the aforementioned proposed models, we recorded quality of these models in Table 4. The data show that there is a favourable increase in R, R^2, R^2_A values and a favourable decrease in Se (Table 3). In addition, we have calculated the quality factor (Q).[46,47] The data show that Q-values go on increasing favourably. Note that this quality factor Q is defined[46,47] as the ratio of R to Se ($Q = R/Se$) and takes account of correlation coefficient and standard error of estimation for deciding the quality of the model.

Finally, we have used the cross-validation method for discussing predictive potential of the proposed models.[48–50] Cross-validation evaluates the validity of a model by how well it predicts data rather than how well it fits data. The analysis uses a 'leave-one-out' scheme; a model is built with N-1 compound 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-validation parameters, particularly cross validated r^2 (r^2_{cv}). This r^2_{cv} defined as:

$$r^2_{cv} = \frac{SD - PRESS}{SD} \quad (4)$$

where SD is the sum of squares deviation for each activity from the mean. PRESS (Predictive Sum of squares) is the sum of the squared differences between the actual and that predicted when the compound is omitted from the fitting process.

Table 4 records the aforementioned cross-validation parameters viz., PRESS, SD, and r^2_{cv}. In addition, in Table 4 we have also recorded S_{press} and PSE values are also recorded. S_{press} accounts for uncertainty of prediction; while PSE is the predictive square error. This PSE is more directly related to the uncertainty of the prediction. The parameter PSE is more useful when S_{press} coincides with Se.

The perusal of the data presented in Table 4 indicates that in all the three cases PRESS < SD. Hence, the proposed models predicts better than chance and are thus statistically significant. The ratio PRESS/SD accounts for approximate confidence intervals of predictions of new observation (compounds). To be a reasonable QSAR/QSTR model, PRESS/SD should be smaller than 0.4, the value of this ratio smaller than 0.1 indicates an excellent model. The ratio of PRESS/SD, in our models goes on decreasing ultimately giving the lowest value (0.2432) for the model expressed by eq 4, indicating it be the excellent model for modeling $_{P}(ICG_{50})$ of nitrobenzenes under present investigation.

It is interesting to record that in the present case S_{press} coincides with Se. That is, both are incapable to deciding predicting potential of the proposed models. We have, therefore, used PSE for this purpose. The data show that PSE goes on decreasing and attains lowest value (0.2354) for the model expressed by eq 3. This indicates that uncertainty of predictions is the lowest for this model.

Finally, r^2_{cv} is found to be highest for the model expressed by eq 3, suggesting once again that it has the highest predictive potential. As usual, here also $r^2_{cv} < R^2$.

Further confirmation in favour of our results is obtained by estimating $_{P}(IGC_{50})$ values from eqs 2 and 3, and comparing them with the observed values of $_{P}(IGC_{50})$, Such a comparison is shown in Table 5. The data show that the values of $_{P}(IGC_{50})$ estimated using eq 3 are more closer to the observed values of $_{P}(IGC_{50})$. The residue, that is the difference between observed and calculated $_{P}(ICG_{50})$ also confirms that the model based

Table 4. Cross-validation parameters for the proposed models (ref Table 3 and the text)

Model	PRESS	SD	PRESS/SD	R^2_{cv}	S_{press}	PSE
1 (eq 1)	3.2644	8.0673	0.4047	0.5954	0.3054	0.2856
2 (eq 2)	2.4068	8.9249	0.2697	0.7303	0.2661	0.2452
3 (eq 3)	2.2170	9.1148	0.2432	0.7568	0.2554	0.2354

PRESS, predictive sum of squares; SD, sum of square deviation; r^2_{cv}, cross-validated r^2; S_{press}, uncertainty of prediction; PSE, predictive square error.

Table 5. Observed and estimated toxicity of nitrobenzene derivatives from the best models

Compd	*R*	$_P(IGC_{50})$ (Obs)	$_P(IGC_{50})$ estimated from			
			eq 2		eq 3	
			Est.	Res.	Est.	Res.
1	H					
2	2-NH_2	0.077	−0.104	0.1814	−0.140	0.2168
3	2-CH_3	0.479	0.384	0.0949	0.429	0.0502
4	2-Cl	0.676	0.687	−0.0114	0.641	0.0353
5	2-Br	0.863	0.784	0.0793	0.720	0.1433
6	2-CH_2OH	−0.155	−0.101	−0.0537	−0.094	−0.062
7	2-C_6H_5	0.301	1.524	−0.2230	1.485	−0.1844
8	2-$CONH_2$	−0.721	−0.601	−0.1199	−0.728	0.0074
9	2-CHO	0.174	−0.033	0.2070	0.051	0.1233
10	3-NH_2	0.026	0.297	−0.2705	0.074	−0.078
11	3-OH	0.506	0.349	0.1574	0.465	0.0412
12	3-CH_3	0.572	0.527	0.0453	0.622	−0.0498
13	3-Cl	0.836	0.830	0.0060	0.834	0.0023
14	3-Br	1.215	0.926	0.2887	0.913	0.3023
15	3-CN	0.451	0.273	0.1779	0.329	0.1224
16	3-CH_2OH	−0.220	0.121	−0.3412	0.159	−0.3785
17	3-C_6H_5	1.569	1.589	−0.0195	1.704	−0.1354
18	3-$CONH_2$	−0.193	−0.278	0.0853	−0.344	0.1510
19	3-CHO	0.140	0.150	−0.0099	0.245	−0.1049
20	3-NO_2	0.762	0.973	−0.2106	0.739	0.0234
21	3-OCH_3	0.670	0.390	0.2802	0.425	0.2431
22	3-$COCH_3$	0.317	0.529	−0.2118	0.478	−0.1611
23	4-CH_3	0.796	0.417	0.3787	0.511	0.2854
24	4-C_2H_5	0.804	0.628	0.1761	0.680	0.1242
25	4-OCH_3	0.544	0.541	0.0030	0.597	−0.0533
26	4-OC_2H_5	0.829	1.180	−0.3511	1.036	−0.2074
27	4-OC_4H_9	1.420	1.352	0.0685	1.110	0.3103
28	4-F	0.253	0.441	−0.1881	0.493	−0.2401
29	4-Cl	0.559	0.721	−0.1616	0.722	−0.1635
30	4-Br	0.461	0.817	−0.3559	0.801	−0.3405
31	4–1N	0.569	0.424	0.1447	0.499	0.0700
32	4-CH_2CN	0.132	0.577	−0.4451	0.604	−0.4719
33	4-CH_2Cl	1.180	0.662	0.5183	0.604	0.4734
34	4-CH=NOH	0.678	0.434	0.2440	0.430	0.2484
35	4-NHC_6H_5	1.886	1.420	0.4656	1.679	0.2068
36	4-CH_2OH	0.101	0.321	−0.2201	0.354	−0.2929
37	4-$COOCH_3$	0.398	0.872	−0.4736	0.889	−0.4911
38	4-$COOC_2H_5$	0.710	0.754	−0.0492	0.632	0.0777
39	4-$CONH_2$	−0.179	0.015	−0.1938	0.018	−0.1972
40	4-CHO	0.203	0.301	−0.0981	0.415	−0.2123
41	4-NO_2	1.301	0.895	0.4058	0.766	0.5352

Obs, observed; est., estimated; Res., residue (difference between observed and estimated toxicity); compound **1** is omitted as being an outlier.

on eq 3 is the most appropriate model for modeling $_P(IGC_{50})$ of the set of nitrobenzenes under present study.

In order to obtain a general approximation of the predictive performances of the studied models, we have estimated the root-mean-square error (RMS) for the studied models employing the following relationship:

$$RMS = \left[\left(Obs \quad P^{(IGC50)} - Est \quad P^{(IGC_{50})}\right)^2\right]^{0.5}$$

where *N* is the number of the compounds involved in the model.

The RMS so calculated for the models eqs 2 and 3 are found as 0.2458 and 0.2356, respectively. This indicates that both the models, more or less, have similar predictive potential, the model represented by eq 3 is slightly better than the model represented by eq 2. This further confirms our aforementioned findings.

Conclusion

The primary objective of this paper was to investigate how far the PI Index can be useful in modeling toxicity of nitrobenzene derivatives under present investigation. Another important objective was to examine how the combination(s) of the PI Index with other distance-based topological indices will be useful for this purpose.

Results presented indicated that the PI Index alongwith other topological indices used alone are incapable of predicting toxicity of nitrobenzenes. However, multiparameteric regressions, that is, tetra- and penta-parametric gave statistically significant models. The most significant model found was the pentaparametric model

consisting of PI, Sz, J, MRI and V_w. The results also show that Sz is a better index than W in developing a multiparametric model for predicting toxicity of the nitrobenzene derivatives under present investigation.

Experimental

Toxicity of nitrobenzenes derivatives

Nitrobenzenes (Fig. 1) are widely used in industrial chemicals, and consequently have high potential for environmental pollution. They have been reported[1] to be present in surface water. They are reactive chemicals, being reported to be uncouplers of oxidative phosphorylation[2] and may be regard as pro-electrophites yielding the corresponding potentially high toxic C-nitroso compounds.

Because of their widespread use, the toxicity of a nitrobenzene derivatives has been quite extensively examined and they have been the subject of a number of QSAR/QSTR studies.[29–32]

In Table 1, we have listed 41 nitrobenzene derivatives and their toxicities $_P(IGC_{50})$. The corresponding toxicity values are those reported by Deander et al.[29] and are used in the present study after converting to log units.

PI (Padmakar–Ivan) Index

The PI (Padmakar–Ivan) Index was introduced[23–28] very recently as a modification of the Sz Index, as the latter index was considered as a modification of the Wiener Index (W) for cyclic compounds and that, for trees (acyclic compound), Sz coincides with W. The potential of the PI Index in QSAR/QSTR/QSPR has yet to be evaluated. The present paper is a continuation of this effort that resulted in half a dozen recent studies. We decided, therefore, to see how well the PI Index performs to the nitrobenzene toxicity data presented in Table 1. We are motivated to explore the potential of the PI Index under different circumstances. The challenge is apparent: the PI Index is considered to be a modification of the Sz Index and the set of nitrobenzene derivatives consisted of only one cycle with one, two or more tree-like (acyclic) side chains.

The details of the calculations of the PI Index are given in our earlier communications.[23,24] We therefore give only a brief account for the calculation of the PI Index.

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, connecting vertices u and v then we write $e = uv$.

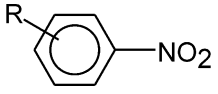

Figure 1. Nitrogenzene derivatives used in the present study.

We define the $e = uv$ 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; $n_{ev}(e|G)$ is the number or edges lying closer to the vertex v than the vertex u. Edges equidistant from both ends of the edge uv are not counted (taken into account), then the PI Index is defined as:

$$PI = \sum [n_{eu}(e|G) + n_{ev}(e|G)] \quad (5)$$

The summation goes over all edges of G.

PI values calculated for the set of 41 nitrobenzene derivative are recorded in Table 2.

Supporting topological indices

In addition to the PI Index, we have have also used, the oldest and widely used Wiener Index (W), the well studied Szeged Index (Sz),[21,22] the highly discriminately used Balaban Index (J),[33,36] successfully used molecular redudancy index (MRI),[37–40] and the widely used van der Waals volume (V_w).[43–45] Sz and J are produced from the path lengths. MRI is an information theoretic topological index, while V_w is a property related to steric parameter. These indices account for molecular hydrophobicity, an important parameter in many toxicological models. All these indices are well defined in the literature and there is no need to give any detail information for their calculations. However, a brief account of their calculations is given below.

Wiener Index (W)[19,20]

The Wiener Index (W),[19] the first topological index reported in the chemical literature, may be calculated from the given distance matrix $\boldsymbol{D}(\boldsymbol{G})$ of a hydrogen-suppressed chemical graph $\boldsymbol{G}$ as the sum of the entries of the upper triangular distance submatrix. The distance matrix $\boldsymbol{D}(\boldsymbol{G})$ of a nondirected graph $\boldsymbol{G}$ with W indices is a real symmetric $n \times n$ matrix with elements d_{ij} equal to the distance before vertices v_i and v_j in $\mathbf{G}$. Each diagonal element d_{ii} of $\boldsymbol{D}(\boldsymbol{G})$ is zero. We then calculate W using following expression:

$$W = W(G) = \frac{1}{2}\sum_{ij} d_{ij} = \sum_{h} h \quad gh \quad (6)$$

where, gh is the number of unordered pairs of vertices whose distance is h.

Szeged Index (Sz)[21,22]

Let e be an edge of the molecular graph G. Let $n_1(e|G)$ be the number of vertices of G lying close to one end of e; let $n_2(e|G)$ be the number of vertices of G lying closer to the other end of e. Then the Szeged index (Sz) is depicted as:

$$Sz = Sz(G) = \sum_{e} n_1(e|G) n_2(e|G) \quad (7)$$

with the summation going over all edges of G.

In cyclic graphs, there are edges equidistant from both the ends of edge e; by definition of Sz, such edges are not taken into account for the calculation of Sz.

Balaban Index (J)[33–35]

The Balaban Index, $J = J(G)$, was introduced by Balaban in 1982 as the average distance shows connectivity index. It is defined as below:

$$J = J(G) = \frac{M}{\mu + 1}\left(d_i.d_j\right)^{0.5} \quad (8)$$

where, M is the number of edges of G; μ is the cyclomatic number of G; and d_i, d_j are the distance sums, where $i, j = 1,2,3,\ldots,N$. The cyclomatic number is defined as the minimum number of edges that must be removed from G in order to transform it to the related acyclic graph.

Molecular redudancy index (MRI)[37–42]

The molecular redudance index (MRI) is derived from information theory and molecular graph theory[37–42] and is defined as under:

$$MRI = MRI(G) = \frac{\sum n_i \log n_i}{N \log N} \quad (9)$$

where, ni is the number of atoms of the same kind in the ith set, i is the number of different atoms in the molecule.

The topological indices PI, W, J and Sz were calculated for carbon–hydrogen suppressed graphs of the nitrobenzene derivatives under present investigation. Furthermore, W, Sz and J indices were calculated using the program provided by Professor Istavan Lukovits, Hungarian Academy of Sciences, Budapest, Hungary.

van der Waal's volume (V_w)[43–45]

The van der Waal's volume (V_w) is the empirical substituent parameter accounting for the size or bulk of a molecule or a substituent. It has also been found to be one of the most fundamental characteristics of the drug structure controlling the biological activity. In the present study, V_w for the compound was calculated as described in the literature.[43–45]

All the aforementioned molecular descriptors as calculated using the corresponding expressions are presented in Table 2.

Materials and methods

We carried out QSAR/QSTR analysis for 41 substituted nitrobenzene for which toxicity data in *T. pyriformis* have been reported. The *T. pyriformis* data have been the subject of several QSAR/QSTR analysis, where traditional physicochemical indices were used in modeling the data using multiple linear regression[46,47] and it was necessary to remove some outliers to obtain good models. However, no attempt is made to use the PI Index and its combination with other topological index to arrive at an excellent model of modeling toxicity of nitrobenzene derivatives (Table 1).

Statistical analysis. Statistical analyses were carried out based on the methodology reported in the literature.[46–50]

Acknowledgements

One of the authors (P.V.K.) is thankful to Prof. Ivan Gutman for introducing him (P.V.K.) to this functioning field, namely Chemical Graph Theory and Topology. The authors are thankful to Prof. Istavan Lukovits for providing software and to Prof. Subhash Basak for his personal interest.

References and Notes

1. Randic, M.; Basak, S. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 614.
2. Seydel, J. K.; Burger, H.; Saxena, A. K.; Coleman, M. D.; Smith, S. N.; Perris, A. D. *Quant. Struct. Act. Relat.* **1999**, *18*, 43.
3. Garcia-Domenech, R.; deGregorio Alapont, C.; deJulian-Ortiz, J. V.; Galvez, J. *Bioorg. Med. Chem.* **1997**, *7*, 567.
4. Basak, S. C. *Med. Sci. Res.* **1987**, *15*, 605.
5. Singh, A. *Comp. Biochem. Physiol.* **1999**, *123*, 241.
6. Gute, B. D.; Basak, S. C. *SAR and QSAR Environ. Res.* **1997**, *7*, 117.
7. Burden, F. A.; Winkler, D. A. *Chem. Res. Toxicol.* In press.
8. Barratt, M. D.; Castell, J. V.; Chamberlain, M.; Courhes, R. D.; Dearden, J. C.; Fewtem, J. H., et al. *ATLA* **1995**, *23*, 410.
9. Hansch, C.; Fujita, J. *J. Am. Chem. Soc.* **1969**, *86*, 1616.
10. Devisiers, S.; Domine, D. *SAR QSAR Environ. Res.* **1999**, *10*, 61.
11. Basak, S.; Grunwald, G. D.; Guts, B. D.; Balasubramanian, K.; Optiz, D. *J. Chem. Int. Comput. Sci.* **2000**, *40*, 885.
12. Schultz, T. W.; Sinks, G. D.; Hunter, R. B. *SAR QSAR Environ Res.* **1995**, *3*, 27.
13. Bradbury, S. P. *SAR QSAR Environ. Res.* **1994**, *2*, 89.
14. Karcher, W., Devillers, J., Eds. *Practical Application of Quantitative Structure Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*. Kluwer Academic: Dordrecht, 1990.
15. Devillers, J., Ed. *Comparative QSAR*. Taylor and Francis: London, 1998.
16. Devillers, J.; Balaban, A. T. *Topological Indices and Related Descriptors in QSAR and QSPR*; Gordon and Breach: Williston, USA, 2000.
17. Todeschini, R.; Consonni, V. *Hand Book of Molecular Descriptors*; Wiley-VCH: Weinheim, 2000.
18. Diuden, M. V., Ed. *QSPR/QSAR Studies by Molecular Descriptors*; Nova Science. Calcutta, India, 2000.
19. Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
20. Karmakar, S.; Khadikar, P. V.; Mandloi, M.; Joshi, S.; Agrawal, V. K. *Indian. J. Chem.* **2001**, *40A*, 12.
21. Gutman, I., *Graph Theory Notes New York* **1994**; *27*, 9.
22. Khadikar, P. V.; Deshpande, N. V.; Kale, P. P.; Dobrynin, A.; Gutman, I.; Domotor, G. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 547.
23. Khadikar, P. V. On a Novel Structural Descriptor PI. *Nat. Acad. Sci. Lett.* **2000**, *23*, 113.

24. Khadikar, P. V.; Karmarkar, S.; Agrawal, V. K. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 934.
25. Khadikar, P. V.; Kale, P. P.; Deshpande, N. V.; Karmarkar, S.; Agrawal, V. K. *J. Math. Chem.* **2001**, *29*, 143.
26. Agrawal, V. K.; Khadikar, P. V. *Bioorg. Med. Chem.* **2001**, *9*, 3035.
27. Khadikar, P. V.; Karmarkar, S.; Agrawal, V. K. *Nat. Acad. Sci. Lett.* **2000**, *23*, 124.
28. Khadikar, P. V.; Karmarkar, S.; Agrawal, V. K. *Nat. Acad. Sci. Lett.* **2000**, *23*, 165.
29. Dearden, J. C.; Cronin, M. T. D.; Schaltz, T. W.; Lin, D. T. *Quant. Struct Act. Relat.* **1995**, *14*, 427.
30. Pronin, M. T. D.; Bryant, S. E.; Deavien, J. C.; Schulter, J. W. *SAR. QSAR Environ. Res.* **1995**, *3*, 1.
31. Cronin, M. T. D.; Gregory, B. W.; Schultz, J. W. *Chem. Res. Toxicol.* **1998**, *11*, 902.
32. Cronin, M. T. D.; Schultz, T. W. *Chemosphere* **1996**, *32*, 1453.
33. Balaban, A. T. *Chem. Phys. Lett.* **1982**, *89*, 399.
34. Balaban, A. T. *Pure. Appl. Chem.* **1983**, *55*, 199.
35. Khadikar, P. V.; Sharma, S.; Sharma, V.; Joshi, S.; Lukovits, I.; Kaveeshwar, M. *Bull. Soc. Chem. Belg.* **1997**, *106*, 767.
36. Sapre, N. S.; Sikarwar, A.; Khadikar, P. V. *Oxid. Commun.* **2001**, *24*, 28.
37. Kier, L. B. *J. Pharm. Sci.* **1980**, *69*, 801.
38. Agrawal, V. K.; Khadikar, P. V. *Bioorg. Med. Chem.* **2001**, *9*, 2787.
39. Agrawal, V. K.; Sachan, S.; Khadikar, P. V. *Acta Pharm.* **2000**, *50*, 281.
40. Agrawal, V. K.; Joseph, S.; Khadikar, P. V. *Nat. Acad. Sci. Lett.* **2000**, *23*, 57.
41. Agrawal, V. K.; Joseph, S.; Khadikar, P. V.; Karmarkar, S. *Acta Pharm.* **2000**, *50*, 329.
42. Agrawal, V. K.; Sachan, S.; Khadikar, P. V. *Polish J. Pharm.* **2000**, *59*, 39.
43. Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.
44. Moriguchi, I.; Kanada, Y. *Chem. Pharm. Bull.* **1977**, *25*, 926.
45. Moriguchi, I.; Kanada, Y.; Komatsu, K. *Chem. Pharm. Bull.* **1976**, *24*, 1799.
46. Pogliani, L. *Amino Acids* **1994**, *6*, 141.
47. Pogliani, L. *J. Phys. Chem.* **1996**, *100*, 18065.
48. Chaterjee, S.; Hadi, A. S.; Price, B. *Regression Analysis in Examples*, 3rd ed.; Wiley: New York, 2000.
49. Lucic, B.; Trinajstic, N. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 121.
50. Lucic, B.; Amic, D.; Trinajstic, N. *J. Chem. Inf. Comput. Sci.* **2001**, *40*, 403.