Pergamon

Bioorganic & Medicinal Chemistry 10 (2002) 1761–1766

BIOORGANIC & MEDICINAL CHEMISTRY

Study on Quantitative Structure–Toxicity Relationships of Benzene Derivatives Acting by Narcosis

Padmakar V. Khadikar,[a,*] Keshav C. Mather,[b] Shalini Singh,[b] Anjani Phadnis,[c] Anjali Shrivastava[d] and Manorama Mandaloi[d]

[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 486 003, India*
[c]*Department of Chemistry, R.S. College, Indore 452 001, India*
[d]*Department of Chemistry, Holkar Model and Autonomous College, Indore 452 001, India*

Received 17 September 2001; accepted 12 January 2002

Abstract—Hydrophobicity (log*P*) as well as quantiative structure–toxicity relationships (QSTRs) of some benzene derivatives acting by narcosis have been established based on narcotic mechanisms of action and toxicity data to the fathead minnow, *Daphnia magna* and *Vibrio fischeri* using information-theoretic topological index (Id). Excellent results are obtained in multiparametric regression upon introduction of dummy parameters (indicator variables). Consistent increase in R^2_A values indicated that inspite of collinarity between Id and one of the indicator variables (I_{3-6}) the proposed models are statistically significant. © 2002 Published by Elsevier Science Ltd.

Introduction

One of the current interests in medicinal chemistry, environmental science and toxicology is the ranking of chemical substances with respect to their potential hazardous effects on humans, wild life, aquatic flora and fauna.[1,2] Quantitative structure–activity (QSARs) and quantitative structure–toxicity (QSTRs) relationships have provided a valuable approach in research into the toxicity of organic chemicals in such studies.

Zhao et al.[3] have reported toxicities of chemicals (benzene derivatives) acting by narcosis while, Russom et al.[4] have reported acute toxicity data on such chemicals. They have discussed structure–toxicity relationships based on some physicochemical parameters. However, no quantitative structure–toxicity relationship (QSTR) for the set of benzene derivatives (Table 1) using topological indices has been reported so far.

Currently, there is much interest in using QSTR studies in the predictivity of toxicity of organic compounds, particularly those present in the various industrial effluents, in view of their hazardous effects on environmental ecology and aquatic life. Research carried out using model systems indicate narcosis—a reversible state of arrested activity of various protoplastoxic action of many industrially important chemicals.

At this stage it is interesting to record that physicochemical parameters are not readily available for any arbitrary (real or hypothetical) chemical structures so as to use them in developing QSTRs. Consequently, much effort has gone into searching for parameters which give acceptable QSTRs, and can be readily derived from chemical structure only. Graph theoretical variables, namely topological indices, fall into such category. In particular, there is a recent upsurge of interest in the use of topological indices in QSTR studies. A topological index is a numerical descriptor of molecular structure and is sensitive to size, shape, symmetry, and heterogeneity of atomic environments in the molecule.[5,6]

In recent years, information-theoretic formalism has been applied to molecular graphs to derive topological indices which are useful in the QSTR studies.[7–9] Information-theoretic topological indices are capable of predicting the pharmacological, as well as toxic properties of a variety of bioactive molecules. Use of these indices in risk assessment of chemicals and toxicology is excellently described by Basak.[2,8]

*Corresponding author. Tel.: +91-731-531906; fax: +91-931-531906; e-mail: anjalee_shirvastava@rediffmail.com

0968-0896/02/$ - see front matter © 2002 Published by Elsevier Science Ltd.
PII: S0968-0896(02)00055-X

In view of the above, we have undertaken the present investigation, in that we have modelled hydrophobicity as well as toxicity of benzene derivatives (Table 1) using information-theoretic topological index, Id. Our aim is not to substitute the Id part with log*P* in the analysis of toxicity of benzene derivatives. Otherwise importance of log*P* for toxicity analysis is well known.[1–3] The set of benzene derivatives consists of halo- and methyl-substituted derivatives. In addition, we have also included benzenes so as to investigate the effect on substitution of the parent benzene moiety. The results, as discussed below, indicate that Id is an excellent topological index for modeling *Vibrio fischeri* toxicity (pEC_{50}), *Daphnia magna* toxicity (pIC_{50}), fathead minnow toxicity (pLC_{50}), as well as log*P* of the set of compounds under present investigation. As will be seen below even monoparametric models based on Id alone gave excellent results. These results were improved in multiparametric correlations upon introduction of dummy parameters (indicator parameter) related to the number of halo-substitutions in the aromatic skeleton. The toxicity data as well as log*P* were adopted from the literature.[3,4]

Results and Discussion

Table 1 lists toxicity data, log*P*, Id, and indicator parameters (I_2, I_{3-6}, I_2) for the set of 20 benzene derivatives used in the present study. The data show that, like halo-substitution, the methyl substitution also enhances toxicity of benzene. Also, that the toxicity changes with the position of the aforementioned substituents in the aromatic nucleus. The toxicity increases with substitution of 2-positions that is why 2-xylene is more toxic than 3-xylene. A perusal of Table 1 shows that degeneracy exists in log*P*, toxicities, as well as molecular descriptor, Id. The degeneracy in Id is obvious as it belongs to first-generation topological index as described by Balaban.[10] Balaban[10] has shown that in spite of observed degeneracy, the first-generation topological indices are quite successful in developing QSPR/QSAR, and thus QSTR. This is found to be the case in the present investigation also.

Initial study has shown that simple regression (monoparametric) using the Id index alone gave excellent results. However, with a hope of obtaining still better results, we have also carried out multiparametric correlations[11–13] by using indicator parameters.

Before a multivariate analysis is undertaken, it is convenient to tailor the data in certain ways to make the calculations easier. Normally, it is sufficient to pre-process the data by means of auto-scaling and mean-centering the variables. Auto-scaling gives each variable unit variance and hence the same chance to contribute to a calculated model, whereas, mean-centering facilitates interpretation. The best way to do so is to obtain a correlation matrix which accounts for the relationship between each independent variable and dependent variable. The correlation matrix is very useful for determining which independent variables are likely to help explain variation in the dependent variable. Such a matrix obtained in the present study is given in Table 2.

In a correlation matrix, we look for correlations close to ± 1.0 since that indicates changes in the independent variables are linearly related to changes in the dependent variable. We can also use a correlation matrix to determine the extent to which independent variables are correlated with one another, that is to estimate autocorrelation/collinearity. This can be useful in determining if certain independent variables are redundant and not needed in the model.

A perusal of the correlation matrix (Table 2) indicates that Id and I_{3-6} are significantly correlated. This indicates that multivariate models containing these two parameters may suffer from the defect of collinearity.

Table 1. Chemicals tested, toxicological data, Id-index, and indicator parameters (I_2, I_{3-6}, I_1)

Compd	Name of the compound	Id	log*P*	pEC_{50}	pIC_{50}	pLC_{50}	I_2	I_{3-6}	I_1
1	Hexachlorobenzene	2.35	5.70	6.32	—	6.38	0	1	0
2	1,2,4,5-Tetra chloro-benzene	2.19	4.75	5.51	—	5.80	0	1	0
3	1,2,4-Trichlorobenzene	2.10	4.16	4.50	4.70	4.78	0	1	0
4	1,2,3-Trichlorobenzene	2.10	4.04	4.53	5.13	4.89	0	1	0
5	1,4-Dichlorobenzene	2.01	3.57	4.39	4.58	4.56	1	0	0
6	1,3-Dichlorobenzene	2.01	3.57	4.24	4.41	4.26	1	0	0
7	1,2-Dichlorobenzene	2.01	3.45	4.38	4.83	4.19	1	0	0
8	Chlorobenzene	1.90	2.56	3.86	4.26	3.82	0	0	1
9	1,4-Dibromobenzene	2.01	3.87	4.54	5.00	—	1	0	0
10	1,3-Dibromobenzene	2.01	3.87	4.99	4.94	—	1	0	0
11	Bromobenzene	1.90	3.01	3.78	4.40	3.94	0	0	1
12	1-Chloro-4-Bromo-benzene	2.01	3.72	4.50	4.86	—	1	0	0
13	2,4,5-Trichlorotoluene	2.19	4.66	4.86	4.99	—	0	1	0
14	2,5-Dichlorotoluene	2.10	4.07	4.38	4.76	—	1	0	0
15	4-Chlorotoluene	2.01	3.35	3.88	4.70	4.33	0	0	1
16	4-Xylene	2.01	3.14	3.68	3.79	4.08	0	0	0
17	3-Xylene	2.01	3.14	3.65	3.75	3.82	0	0	0
18	2-Xylene	2.01	3.09	4.08	4.52	3.81	0	0	0
19	Toluene	1.90	2.64	3.08	3.57	3.43	0	0	0
20	Benzene	1.79	2.14	3.34	3.54	3.62	0	0	0

For log*P*, see refs 3 and 4.

However, as will be discussed latter, the problems of collinearity can be resolved on the basis of adjustable R^2 (R^2_A). Furthermore, in spite of collinearity the information content of the parameters might be quite different. Neglecting collinearily related parameters will deprive us of such useful information which might be significant for the model.

The data presented in Table 2 also show high collinearity among log$P$, Id, and pEC_{50}. This indicates that single regressions based on log$P$ or Id will give an excellent model for modeling, monitoring and estimating pEC_{50}. Also, I_{3-6} is significantly correlated with log$P$ and pEC_{50}. It means that I_{3-6} will be another useful parameter for modeling both log$P$ and pEC_{50}. Other parameters are not that correlated with log$P$ and pEC_{50}. Compared to pEC_{50}, the correlations of parameters with pIC_{50} and pLC_{50}, are less significant.

It is worth mentioning that partition coefficient (P) between 1-octonol/water, expressed as log$P$, is useful for predicting toxicity as well as pharmacological activity of organic compounds acting as a drug. This is made clear by the correlation matrix presented in Table 2. However, as stated earlier our aim is not to use log$P$ for the analysis of toxicity of the benzene derivatives used. Instead, our main objective is to propose a topological method for the estimation of hydrophobicity and toxicity of the compounds used.

Consequent to the above, we have attempted several multiparametric correlations for modeling log$P$, pEC_{50}, pIC_{50} and pLC_{50} of the compounds under the present investigation. The statistically significant correlations are presented in Tables 3–5 and are discussed below. The regression parameters and quality of these correlations are given in Table 6.

Table 3 records statistically significant regressions for modeling log$P$. The multiparametric model based on Id and some combinations of indicator parameters gave excellent results. However, the quality of the model goes on increasing with the addition of indicator parameters so that the tetraparametric model (Table 3) gave the most significant model:

$$\begin{aligned}\log P = {} & 5.5150\ (\pm 0.4450)\ \mathrm{Id} \\ & + 0.4862\ (\pm 0.0878) I_2 \\ & + 0.5660 (+0.1383)\ I_{3-6} \\ & + 0.2776\ (\pm 0.1005) I_1 - 7.3039\end{aligned} \quad (1)$$

The model (eq 1) indicates that in addition to Id the indicator parameters I_2, I_{3-6}, I_1 also play dominant role in modeling log$P$. These three indicator parameters are related to the number of halo-substitutions in the aromatic nucleus. It is worth mentioning that a biparametric model containing Id and I_2 is found to be better than the monoparametric model based on Id (Table 6). No other combination of I_{3-6} or I_1 gave better results than this biparametric model. Recall that I_2 is used for dihalogenated benzene derivatives. This means that dihalogenation is a primary requirement for the exhibition of log$P$, that is for the exhibition of toxicity of benzene derivatives used.

The aforementioned tetraparametric model is used for the set of 20 benzene derivatives for modeling their

Table 2. Correlation matrix for the correlation of toxicity data and molecular descriptors of the tested chemicals

	log$P$	pEC_{50}	pIC_{50}	pLC_{50}	Id	I_2	I_{3-6}	I_1
log$P$	1.0000							
pEC_{50}	0.9191	1.0000						
pIC_{50}	−0.4389	−0.4634	1.0000					
pLC_{50}	0.0313	0.0485	−0.5530	1.0000				
Id	0.9637	0.8342	−0.4763	0.1417	1.0000			
I_2	0.0844	0.1870	0.3772	−0.5197	−0.0505	1.0000		
I_{3-6}	0.7995	0.6442	−0.4343	0.3107	0.7554	−0.4237	1.0000	
I_1	−0.3032	−0.2508	0.1230	0.1552	−0.3346	−0.3083	−0.2425	1.0000

Table 3. Regression equations, parameters and quality of correlation for modeling log$P$

$\log P = 6.3995\ (\pm 0.4176)\ \mathrm{Id} - 9.3564$	(1)
$\log P = 6.4442\ (\pm 0.3729)\ \mathrm{Id} + 0.2199\ (\pm 0.0926)\ I_2 - 9.5241$	(2)
$\log P = 6.6307\ (\pm 0.3936)\ \mathrm{Id} + 0.2629\ (\pm 0.0960)\ I_2 + 0.1763\ (\pm 0.1371)\ I_1 - 9.9445$	(3)
$\log P = 5.1750\ (\pm 0.5290)\ \mathrm{Id} + 0.3854\ (\pm 0.0950)\ I_2 + 0.4717\ (\pm 0.1595)\ I_{3-6} - 7.1222$	(4)
$\log P = 5.2150\ (\pm 0.4450)\ \mathrm{Id} + 0.4862\ (\pm 0.0878)\ I_2 + 0.566\ (\pm 0.1384)\ I_{3-6} + 0.2776\ (\pm 0.1005)\ I_1 - 7.3039$	(5)

Table 4. Regression equations, parameters and quality of correlation for modeling pEC_{50}

$pEC_{50} = 5.3625\ (\pm 0.8355)\ \mathrm{Id} - 6.5967$	(6)
$pEC_{50} = 5.4370\ (\pm 0.7828)\ \mathrm{Id} + 0.3667\ (\pm 0.1944)\ I_2 - 6.8764$	(7)
$pEC_{50} = 3.7066\ (\pm 1.2740)\ \mathrm{Id} + 0.5923\ (\pm 0.2288)\ I_2 + 0.6430\ (\pm 0.3842)\ I_{3-6} - 3.6018$	(8)
$pEC_{50} = 3.7673\ (\pm 1.2249)\ \mathrm{Id} + 0.7455\ (\pm 0.2417)\ I_2 + 0.7863\ (\pm 0.3810)\ I_{3-6} + 0.4216\ (\pm 0.2765)\ I_1 - 3.8777$	(9)

log$P$. 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, the most significant model based on four parameters given above fall well within the thumb rule. The proposed model, therefore, can be used for modeling, monitoring, and estimating log$P$ of benzene derivatives used.

The above model contains the indicator parameter I_{3-6} which is linearly correlated with Id. This is also the case with triparametric model 4 given in Table 3. Table 6 shows that R^2_A goes on increasing as we pass from bi- to tetra-parametric models. This consistent increase in R^2_A indicates that in spite of observed collinearity between Id and I_{3-6}, the models containing both of these parameters are statistically significant. Note that if a variable is added that does not contribute its fair shair, the R^2_A will actually decline. R^2_A is a measure of the % explained variation in the dependent variable that takes into account the relationship between the number of cases and number of independent variables in the regression model. Whereas R^2 will always increase when an independent variable is added. R^2_A will decrease if the added variable does reduce the unexplained variation enough to offset the loss of degrees of freedom.

Selection of descriptors to be used in structure–property–activity–toxicity 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. However, what is important is not whether two descriptors parallel to one another, that is, duplicate much of the same structural information but whether they are in those parts that are important for structure–property–activity–toxicity correlations. If they differ in the domain which is important for the property/activity/toxicity considered both descriptors should be retained. If they differ in parts that are not relevant for the correlation of considered property/activity/toxicity then one of them can be discarded. In our case, Id and I_{3-6} give different types of information and are, therefore, important for toxicity consideration of benzene derivatives used. Hence, in spite of their observed collinearity they are to be retained in the proposed models.

Table 5. Regression equations, parameters and quality of correlation for modeling pIC_{50} and pLC_{50}

$$pIC_{50} = -5.7307\ (\pm 2.4933)\ \mathrm{Id} + 15.6751 \quad (10)$$
$$pIC_{50} = -5.5158\ (\pm 2.3519)\ \mathrm{Id} + 1.0579\ (\pm 0.5840)\ I_2 + 14.8683 \quad (11)$$
$$pLC_{50} = 5.6617\ (\pm 0.2102)\ \mathrm{Id} - 7.0519 \quad (12)$$

We now discuss the statistically significant expressions for modeling pEC_{50}. Four such regressions are found to be statistically important and are recorded in Table 4. Here also a statistically significant monoparametric model containing Id is obtained whose quality is improved in biparametric model upon the addition of I_2, Hence, here also for enhancing pEC_{50} (toxicity), dihalogenation is the basic requirement.

Like the earlier case of modeling log$P$, in the modeling of pEC_{50}, the quality goes on increasing as we pass from mono- to tetraparametric models so that the tetraparametric model is again found to be the best (Table 6). Hence, based on earlier reasoning this model is the most suitable model for modeling pEC_{50} and is found as under:

$$\begin{aligned} pEC_{50} = {} & 3.7673\ (\pm 1.2249)\ \mathrm{Id} \\ & + 0.7455\ (\pm 0.2417) I_2 \\ & + 0.7863 (\pm 0.3810)\ I_{3-6} \\ & + 0.4216\ (\pm 0.2765) I_1 - 3.8777) \end{aligned} \quad (2)$$

The regression parameters as well as quality of different models (Table 6) are in favor of the formulationed model.

We have also attempted regression analysis in modeling pIC_{50}. However, the results are not encouraging. Only a fairly significant biparametric model (Tables 5 and 6) was obtainable. Higher order multiparametric correla-

Table 6. Regression parameters and quality correlations

Model no.	Parameters used	*Se*	R^2_A	*R*	*F*	*P*	*Q*
Modeling logP							
1.	Id	0.2212	—	0.9637	234.797	9.060×10^{-12}	4.3360
2.	Id, I_2	0.1973	0.9402	0.9729	150.400	1.546×10^{-11}	4.9310
3.	Id, I_2, I_1	0.1936	0.9424	0.9755	104.715	9.935×10^{-11}	5.0340
4.	Id, I_2, I_{3-6}	0.1936	0.9589	0.9826	148.888	6.520×10^{-12}	6.034
5.	Id, I_2, I_{3-6}, I_1	0.1375	0.9710	0.9885	159.879	4.190×10^{-12}	7.1330
Modeling of pEC_{50}							
6.	Id	0.4426	—	0.8342	41.191	4.843×10^{-6}	1.8348
7.	Id, I_2	0.4141	0.7189	0.8652	25.301	8.019×10^{-6}	2.0334
8.	Id, I_2, I_{3-6}	0.3938	0.7459	0.8866	19.589	1.321×10^{-5}	2.2514
9.	Id, I_2, I_{3-6}, I_1	0.3784	0.7653	0.9028	16.489	2.299×10^{-5}	2.3358
Modeling pIC_{50}							
10.	Id	1.3207	—	−0.4763	5.283	0.337	0.3536
11.	Id, I_2	1.2442	0.2757	0.5933	4.014	0.0250	0.4759
Modeling pLC_{50}							
12.	Id	0.1892	—	0.9326	208.721	1×10^{-12}	—

tions does give improved values for the correlation coefficient (R). However, all of them suffered from the defect that there exist one or more correlating parameters in that their standard deviations were much higher than the corresponding coefficients respective in the regression model. Such models are not statically allowed and, therefore, are discarded and not discussed.

In the case of modeling pLC_{50} the situation is highly favourable. A statistically significant monoparametric model is obtained for modeling pLC_{50} (Tables 5 and 6). This shows that the Id is an appropriate parameter in modeling pLC_{50} of the benzene derivatives used in the present investigation. Higher parametric models gave improved R-values but all these models suffer from the defect that coefficients of the parameters involved are much smaller than their standard deviation. Such models are not statistically allowed. Therefore, statistically the most significant model for modeling pLC_{50} is as under:

$$pLC_{50} = 5.6617\ (\pm 0.2102)\ \mathrm{Id} - 7.0519 \tag{3}$$

In order to confirm our findings, we have calculated quality factor Q[15] ($= R/Se$). This quality factor Q takes into account the simultaneous contributions R/S in developing quality of regression.

The data presented in Table 6 show that based on Q-values tetraparametric model-5 (Table 6) is the most appropriate model for modeling log$P$ of benzene derivatives. Similarly, Q-values suggest that models 9 and 12 are the most appropriate models for modeling pEC_{50} and pLC_{50}, respectively. Thus, Q-values confirms our results.

Conclusion

The aforementioned results and discussion lead us to the following conclusions:

1. If log$P$ is considered as a measure of toxicity of benzene derivatives used in the present investigation then the same can be modeled very successfully by the information theoretic topological index used, namely the Id index.
2. Dihalogenation is a basic requirement for the exhibition of toxicity. Slight improvement resulted in higher order halogen substitution.
3. Like log$P$, pEC_{50} and pLC_{50} can also be modeled successfully using the Id index.
4. Id is not a good parameter for modeling pIC_{50} of the benzene derivatives under the present study.
5. In all cases quality of the model goes on increasing as we pass from mono- to tetraparametric model.
6. In spite of collinearity between Id and I_{3-6} the proposed models are free from collinearity defect as evident from consistent increase in R^2_A values and also because Id and I_{3-6} provide quite different information useful and important for the proposed models.

Experimental

Information-theoretic-topological index (Id)

A non-hydrogen suppressed molecular graph is considered for obtaining information-theoretic topological index, namely, Id, used in the present investigation.

If G is a non-hydrogen suppressed molecular graph with vertex set $\chi(G)$ and $A_i(i-1,2,\ldots,k)$ is a partition π of $\chi(G)$, then a probability scheme is given by the following expression:

$$\begin{pmatrix} A_1 & A_2 & \ldots\ldots & A_k \\ P_1 & P_2 & \ldots\ldots & P_k \end{pmatrix} \tag{4}$$

where $P_i = n_i/n$, n_i and n and n are the cardinalities of A_i and $\chi(G)$, respectively. Then, the degree of complexion, Id(G) is given by the equation given below:

$$\mathrm{Id}(G) = \sum_{i=1}^{k} n_i \frac{d_i}{d} \log_2 \frac{d}{d_i} \tag{5}$$

Note that Id is measured in bits.

Eq (5) is used for the calculation of Id for the set of benzene derivatives used in the present study and are recorded in Table 1.

Source of log$P$

The hydrophobic parameter log$P$ (octanol–water) was taken from the literature.[3] Usually, log$P$ values are determined experimentally by analyzing only one of the two phases. Data used here are obtained by two-phase analysis reported by Hansch and Leo.[14]

Source of toxicity data

V. fischeri as well as *D. magna* toxicity data were adopted from the work of Zhad et al.,[3] while fathead minnow toxites were those reported by Russom et al.[4]

D. magna were cultured parthenogenetically in an environment chamber at 22±°C with photoperiod of 14 h daylight/10 h darkness. *Daphnia* were fed a diet of green algae, and 6–24 h old *Daphnia* were used in toxicity tests. Acute toxicity tests on organic chemicals were conducted for 24 h using a static method, with 10 *Daphnia* in 25 mL of test water. Each test used 60 *Daphnia* and the number of immobilizations was determined regularly. The immobilization concentrations in 5% of *Daphnia* at 24 h was determined statistically as the IC_{50} (mol/l) and converted into pIC_{50} units.

V. fischeri testing made by Model Toxicity Analyser Dxy-2. Ecotoxicological descriptors were the concentration values causing a 50% inhibition of bioluminesecence after 15 min exposure (15 min–EC_{50},

mol/L). These EC_{50} values were also converted into pEC_{50} units.

Acute toxicity data (LC_{50}; mol/L) for the fathead minnow (*Pimephales promelas*) were taken from the work of Russom et al.[4]

Indicator parameters

Indicator variables (parameters) are dummy parameters sometimes used in multiple regression analysis. These variables take on two values, usually zero and one. The two values signify that the observation belongs in one of the two possible categories. The numerical value of dummy variables are not intended to reflect a quantitative ordering of categories but only serve to identify category or class membership. In the present study, we have used three dummy variables namely I_2, I_{3-6} and I_1. These variables assume the value unity in dihalo-, tri- or more halo-, and mono-halo-substitution in the aromatic nucleus respectively. In the absence of such halo-substitution the values of I_2, I_{3-6} and I_1 were taken as zero.

Statistical analysis

All statistical analyses were performed by least squares linear regression using the software supplied by I. Lukovits. Hurgarian Academy of Science, Budapest, Hungary. For each, regression, the following descriptive information is provided: number of observations used in the analysis (n), standard error of estimation (Se), multiple correlation coefficient (R), adjustable R^2 (R^2_A), square of the correlation coefficient adjusted for degrees of freedom (R^2), Fisher's criterion (F), and probability (P).

Acknowledgements

The authors are thankful to Professor Ivan Gutman for introducing them to this facinating field of Chemical Graph Theory and Topology. They are also thankful to Subhash Basak for his kind interest and supports in undergoing research related to information theory. Thanks are also due to the referees for their excellent suggestions.

References and Notes

1. Schultz, J. W.; Kier, L. B.; Hall, L. H. *Bull. Environ. Contam. Toxicol.* **1982**, *28*, 379.
2. Gute, B. D.; Basak, S. C. *SAR and QSAR in Environ. Res.* **1997**, *7*, 117.
3. Zhao, Y. H.; Cronin, M. J. D.; Dearden, J. C. *Quant. Struct. Act. Relat.* **1998**, *17*, 131.
4. Russom, C. L.; Bradbury, S. P.; Broderius, S. J.; Hammertheister, D. E.; Drummod, R. A. *Environ. Toxicol. Chem.* **1997**, *16*, 948–967.
5. Randic, M.; Basak, S. K. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 614.
6. Devillers, J., Balaban, A.J., Eds. *Topological Indices and Related Descriptors in QSAR and QSPR*; Gordon and Breach Science: Amsterdam, 1999.
7. Basak, S. C.; Grunwald, G. D.; Gute, B. D.; Balasubraniam, K.; Opitz, D. *J. Chem. Int. Comput. Sci.* **2000**, *40*, 885.
8. Basak, S. C.; Bertelsen, S.; Grunwald, G. D. *Toxicology Letters* **1995**, *79*, 239.
9. Bondev, D. *Information—Theoretic Indices for Characterization of Chemical Structures*; Research Studies: Chichester, 1983.
10. Balaban, A. J. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 23.
11. Chaterjee, S.; Hadi, A. S.; Pric, B. *Regression Analysis by Examples*, 3rd ed.; Wiley: New York, 2000.
12. Lucic, B.; Trinajstic, N. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 121.
13. Lucic, B.; Trinajstic, N. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 610.
14. Hansch, C.; Leo, A. *Substituted Constants for Correlation Analysis in Chemistry and Biology*; John Wiley and Sons: Brisbane, Toronio, New York, Chichester, 1979.
15. Pogliani, L. *Amino Acids* **1994**, *6*, 141.