Available online at www.sciencedirect.com

ELSEVIER

SCIENCE DIRECT®

Bioorganic & Medicinal Chemistry 13 (2005) 1185–1194

Bioorganic &
Medicinal
Chemistry

QSTR with extended topochemical atom indices. Part 5: Modeling of the acute toxicity of phenylsulfonyl carboxylates to *Vibrio fischeri* using genetic function approximation☆

Kunal Roy* and Gopinath Ghosh

Drug Theoretics and Cheminformatics Lab, Division of Medicinal and Pharmaceutical Chemistry,
Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700 032, India

Received 11 October 2004; revised 8 November 2004; accepted 9 November 2004
Available online 26 November 2004

Abstract—In continuation of our recent efforts to model the acute toxicity of 56 phenylsulfonyl carboxylates to *Vibrio fischeri* using principal component factor analysis, the present paper deals with modeling of the same data set with extended topochemical atom (ETA) indices using genetic function approximation (GFA) as the statistical tool. The statistical quality of the best model using ETA descriptors is as follows: $n = 56$, $Q^2 = 0.771$, $R_a^2 = 0.861$, $R = 0.935$, $F = 69.0$ (df5,50), $s = 0.172$, AVRES = 0.128. This equation was better in statistical quality than that obtained previously using principal component factor analysis as the data-preprocessing step. An attempt was also made to model the data set with different nonETA parameters (topological indices including Wiener, Hosoya *Z*, molecular connectivity, kappa shape, Balaban *J* and E-State parameters apart from physicochemical parameters like Alog*P*98, MolRef and H_bond_acceptor) and the best model showed the following quality: $n = 56$, $Q^2 = 0.805$, $R_a^2 = 0.820$, $R = 0.913$, $F = 53.5$ (df4,51), $s = 0.196$, AVRES = 0.136. An attempt to use both ETA and nonETA parameters lead to an equation which is marginally better than the best ETA model: $n = 56$, $Q^2 = 0.779$, $R_a^2 = 0.865$, $R = 0.937$, $F = 71.7$ (df5,50), $s = 0.169$, AVRES = 0.127. Use of the ETA indices suggested negative contributions of steric bulk, functionality of C_{10}, volume of substituents on C_{10} and functionalities of chloro and nitro substituents at X_1 position and positive contributions of functionalities of C_4 and C_1 and presence of substituents with electronegative atoms, especially at R_2 and R_3 positions. Furthermore, absence of substituents at R_2 and R_3 positions decreases toxicity. This study corroborates the results obtained from principal component factor analysis and suggests that ETA parameters are sufficiently rich in chemical information to encode the structural features contributing significantly to the acute toxicity of phenylsulfonyl carboxylates to *V. fischeri*.
© 2004 Elsevier Ltd. All rights reserved.

1. Introduction

Accumulating evidences suggest that humans and domestic and wildlife species have suffered adverse health consequences from exposure to environmental chemicals.[1] Animal testing is still considered essential to the support of risk assessment, but is often too costly and time consuming to be applied to the full range of chemicals for which some level of toxicological screening is necessary and desired.[2] Faced with the task of screening a large number of chemicals, for an increasing array of toxicity endpoints, using limited resources, quantitative structure–activity relationships (QSARs) have been used in many diverse problem settings as a complement to experimental data.[2] QSARs have emerged as an indispensable tool for predicting ecotoxicological hazard of new chemicals. The regulatory agencies have to rely on QSAR techniques as these can rapidly predict potential ecotoxicological hazard at minimum cost.[3] The US Environmental Protection Agency (EPA) designed ASTER (Assessment Tools for Evaluation of Risk), which is an integration of AQUIRE (database of aquatic toxicity) and QSAR (database of physicochemical properties and QSAR models) to assist regulators in producing assessments.[4] Apart from prediction of ecological and human health effects, QSARs are being used to help industry design safer chemicals for commercial use.[2] QSARs have also been used in exploring the mechanism of toxic actions of chemicals.[5]

Keywords: QSAR; QSTR; Extended topochemical atom index; ETA; TAU; VEM.

☆ Presented in APTI International Convention 2004, Andhra University, Visakhaptanam, India (2–3 October 2004).

* Corresponding author. Tel.: +91 33 2414 6676; fax: +91 33 2414 6677; e-mail: kunalroy_in@yahoo.com

0968-0896/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2004.11.014

Different QSAR methods have been used for ecotoxicological risk assessment. A QSAR model based on atom type E-state index has been reported for the prediction of toxicity of fathead minnow for a diverse set of 140 organic chemicals.[6] The toxicity of organic chemicals to *Chlorella vulgaris* was modeled using multiple linear regression and *k*-nearest neighbours methods.[7] Different classification techniques like discriminant analysis, quadratic discriminant analysis, regularized discriminant analysis, etc., have been used for toxicity prediction.[8] The toxicity of diverse aromatic compounds to *Tetrahymena pyriformis* was modeled using partial least squares and Bayesian regularized neural network.[9,10] 3-D QSAR studies on toxicity of chlorophenols have been reported using the CATALYST software.[11]

We have recently introduced[12–15] extended topochemical atom (ETA) indices as an extension of the TAU concept in the valence electron mobile (VEM) environment,[16–26] and modeled different toxicity data (phenol toxicity,[12] fish toxicity[13] and nitrobenzene toxicity[14]) to establish the utility of ETA indices in modeling studies. Very recently, we have reported[15] modeling of the acute toxicity of 56 phenylsulfonyl carboxylates to *V. fischeri* using principal component factor analysis and principal component regression analysis. The present paper deals with modeling of the same toxicity data of phenylsulfonyl carboxylates using ETA indices applying genetic function approximation (GFA).

GFA,[27,28] a powerful new technique, automates the search for QSAR models by combining genetic algorithm with statistical modeling tools. It rapidly produces a population of statistically valid structure–activity models, rather than a single model. Thousands of candidate models are created and tested during evolution, where only the superior models survive, which are then used as 'parents' for the creation of the next generation of the candidate models. The predictiveness of GFA candidate models is estimated using Friedman's lack-of-fit (LOF) error measures, which uses least square error, number of samples in the data set and number of terms in the model. The frequency of use of a particular descriptor in the population of equations may indicate the relevance of the descriptor in the prediction of the response variable.[29]

The statistical qualities of the best regression equations obtained from GFA in the present study were compared to those of the equations obtained previously using principal component factor analysis as the data pre-processing step.[15] Attempt has also been made to develop models using selected nonETA indices and compare these to the ETA models. Finally, models were developed using both ETA and nonETA descriptors.

2. Materials and methods

Definitions of some of the basic parameters used in the ETA scheme are given below.

2.1. The core count (α)

The core count [α] for a nonhydrogen vertex was defined as[12]

$$\alpha = \frac{Z - Z^v}{Z^v} \cdot \frac{1}{\mathrm{PN} - 1}. \tag{1}$$

In Eq. 1, PN stands for period number. Hydrogen atom being considered as reference, α for hydrogen is taken to be zero. The α values of different atoms (which are commonly found in organic compounds) have high correlation ($r = 0.946$) with (uncorrected) van der Waals volume.[12] Thus, $\sum\alpha$ values of all nonhydrogen atoms of a molecule (instead of vertex count N_{V}) may be taken as a gross measurement of molecular bulk.

2.2. The electronegativity measure (ε)

We defined a term ε as a measure of electronegativity in the following manner:[12]

$$\varepsilon = -\alpha + 0.3Z^V. \tag{2}$$

It was found[12] that ε has good correlation ($r = 0.937$) with Pauling's electronegativity scale.

2.3. The VEM count β

The VEM count β of the ETA scheme was defined as[12]

$$\beta = \sum xv + \sum y\pi + \delta. \tag{3}$$

In the above equation, x is the contribution of a sigma bond having values of 0.5 for two bonded atoms of similar electronegativity ($\Delta\varepsilon \leqslant 0.3$) and 0.75 for two bonded atoms of different electronegativity ($\Delta\varepsilon > 0.3$). Again, in the case of pi bonds, contributions (y) are considered depending on the type of the double bond: (i) for pi bond between two atoms of similar electronegativity ($\Delta\varepsilon \leqslant 0.3$), y is taken to be 1; (ii) for pi bond between two atoms of different electronegativity ($\Delta\varepsilon > 0.3$) or for conjugated (nonaromatic) pi system, y is considered to be 1.5; (iii) for aromatic system, y is taken as 2. δ is a correction factor of value 0.5 per atom with lone pair of electrons capable of resonance with aromatic ring (e.g., nitrogen of aniline, oxygen of phenol, etc.).

For a given part (substructure) of a molecular graph, $\sum\beta_s$ and $\sum\beta_{\mathrm{ns}}$ may be calculated considering all bonds (sigma bonds for the former and pi bonds and lone pair of electrons for the latter) in the substructure. $\sum\beta'_{\mathrm{s}}$ (defined as $[\sum\beta_s]/N_{\mathrm{V}}$, N_{V} being the vertex count) may be taken as a relative measure of number of electronegative atoms in the substructure while $\sum\beta'_{\mathrm{ns}}$ (defined as $[\sum\beta_{\mathrm{ns}}]/N_{\mathrm{V}}$) may be taken as a relative measure of electron-richness (unsaturation) of the substructure.

2.4. The VEM vertex count γ

The VEM vertex count γ_i of the ith vertex in a molecular graph was defined as[12]

$$\gamma_i = \frac{\alpha_i}{\beta_i}. \tag{4}$$

In the equation above, α_i stands for α value for the ith vertex and β_i stands for VEM count considering all bonds connected to the atom and lone pair of electrons (if any).

2.5. The composite index η

The composite index η was defined in the following manner:[12]

$$\eta = \sum_{i<j} \left[\frac{\gamma_i \gamma_j}{r_{ij}^2}\right]^{0.5}. \tag{5}$$

In Eq. 5, r_{ij} stands for the topological distance between ith atom and jth atom. Again, when all heteroatoms in the molecular graph are replaced by carbon and multiple bonds are replaced by single bond, corresponding molecular graph is considered as the reference alkane and the corresponding composite index value is designated as η_R.[12]

2.6. The functionality index η_F

Considering functionality as the presence of heteroatoms (atoms other than carbon or hydrogen) and multiple bonds, functionality index η_F was calculated[12] as $\eta_R - \eta$. To avoid dependence of functionality on vertex count or bulk, another term η'_F was defined as η_F/N_V.

2.7. The atom level index

The contribution of a particular position or vertex to functionality can be determined in the following manner:[12]

$$[\eta]_i = \sum_{j \neq i} \left[\frac{\gamma_i \gamma_j}{r_{ij}^2}\right]^{0.5}. \tag{6}$$

In Eq. (6), $[\eta]_i$ stands for contribution of the ith vertex to η. Similarly, contribution of the ith vertex $[\eta_R]_i$ to η_R can be computed. Contribution of the ith vertex $[\eta_F]_i$ to functionality was defined[12] as $[\eta_R]_i - [\eta]_i$. To avoid dependence of this value on N_V, a related term $[\eta'_F]_i$ was defined[12] as $[\eta_F]_i/N_V$.

2.8. The local index η^{local}

When only bonded interactions are considered ($r_{ij} = 1$), the corresponding composite index may be written as η^{local}.

$$\eta^{\text{local}} = \sum_{i<j, r_{ij}=1} (\gamma_i \gamma_j)^{0.5}. \tag{7}$$

In the similar way, η_R^{local} for the corresponding reference alkane may also be calculated. Local functionality contribution (without considering global topology), η_F^{local}, may be calculated as $\eta_R^{\text{local}} - \eta^{\text{local}}$.

2.9. The branching index η_B

Branching is calculated with respect to η value of the corresponding normal alkane (straight chain compound of same vertex count obtained from the reference alkane), η_N^{local}, which may be conveniently calculated as (for compounds having $N_V \geqslant 3$):[12]

$$\eta_N^{\text{local}} = 1.414 + (N_V - 3)0.5. \tag{8}$$

Branching index η_B can be calculated as $\eta_N^{\text{local}} - \eta_R^{\text{local}} + 0.086 N_R$, where N_R stands for the number of rings in the molecular graph of the reference alkane. The N_R term in the branching index expression represents a correction factor for cyclicity. To calculate branching contribution in comparison to the molecular size, another term η'_B was defined[12] as η_B/N_V.

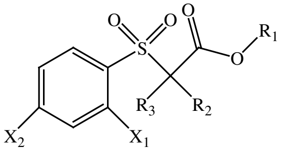
2.10. The shape indices

The terms like $(\sum\alpha)_p/\sum\alpha$, $(\sum\alpha)_Y/\sum\alpha$ and $(\sum\alpha)_X/\sum\alpha$ can be used as the shape parameters. $(\sum\alpha)_p$, $(\sum\alpha)_Y$ and $(\sum\alpha)_X$ stand for summation of α values of the vertices that are joined to one, three and four other nonhydrogen vertices respectively in the molecular graph.

In the present communication, utility of ETA parameters is demonstrated through a Quantitative Structure–Toxicity Relationship (QSTR) study taking acute toxicity of phenylsulfonyl carboxylates to *V. fischeri*[30] as the model data set (Table 1). The definitions of important ETA parameters are given in Table 2. Genetic Function Approximation (GFA) was used as the statistical tool for generating the multiple regression equations. The study was done using the GFA module under QSAR + environment of CERIUS2 software[31] All default settings were used for the analysis (number of crossovers = 5000, linear terms, smoothness factor = 1, mutation probability for adding new term = 50%).

The nonhydrogen common atoms of the compounds were so numbered that these maintain same serials in all the models (Fig. 1). The calculations of η, η_R, η_F, η_B, η_{local} and contributions of different vertices to η_F were done, using distance matrix and VEM vertex counts as inputs, by the GW-BASIC programs KRETA1 and KRETA2 developed by one of the authors.[32] We have also modeled the toxicity data using other selected topological and physicochemical variables and compared the ETA models with nonETA ones. The values for the nonETA topological descriptors and physicochemical variables for the compounds have been generated by QSAR + and Descriptor + modules of the Cerius2 version 4.8 software.[31] The various topological indices calculated are Wiener,[33] Hosoya Z[34] Balaban J[35] connectivity indices ($^0\chi, {}^1\chi, {}^2\chi, {}^3\chi_P, {}^3\chi_C, {}^3\chi_{CH}, {}^0\chi^v, {}^1\chi^v, {}^2\chi^v, {}^3\chi_P^v, {}^3\chi_C^v, {}^3\chi_{CH}^v$),[36–39] kappa shape indices ($^1\kappa$, $^2\kappa$, $^3\kappa$, $^1\kappa_\alpha$, $^2\kappa_\alpha$, $^3\kappa_\alpha$),[40] E-state parameters.[41,42] Among the physicochemical variables, molar refractivity (MolRef), hydrophobicity (Alog*P*98) and H_bond_acceptor[43] were considered. Finally

Table 1. Observed, calculated and predicted toxicity of phenylsulfonyl carboxylates to *V. fischeri*



1-56

Sl.	Structural features[a]					Toxicity to *V. fischeri* (pC)						
	R_1	*R_2*	*R_3*	*X_1*	*X_2*	Obsd[b]	Calcd[c]	Pred[c]	Calcd[d]	Pred[d]	Calcd[e]	Pred[e]
1	CH_3	$–[CH_2]_2–$		H	H	2.28	2.13	2.12	2.20	2.19	2.21	2.20
2	CH_3	$–[CH_2]_3–$		H	H	2.12	1.92	1.90	1.86	1.83	1.83	1.80
3	CH_3	$–[CH_2]_4–$		H	H	1.91	1.73	1.72	1.75	1.73	1.75	1.74
4	CH_3	$–[CH_2]_5–$		H	H	1.81	1.56	1.55	1.64	1.63	1.67	1.66
5	CH_3	$–[CH_2]_2–$		H	NO_2	2.12	2.03	2.01	1.97	1.96	2.00	1.98
6	$CH(CH_3)_2$	$–[CH_2]_2–$		H	NO_2	1.78	1.84	1.84	1.83	1.83	1.84	1.85
7	$CH(CH_3)_2$	$–[CH_2]_3–$		H	NO_2	1.81	1.66	1.65	1.48	1.47	1.46	1.44
8	$CH(CH_3)_2$	$–[CH_2]_5–$		H	NO_2	1.45	1.37	1.37	1.27	1.26	1.29	1.28
9	$CH(CH_3)_2$	$–[CH_2]_6–$		H	NO_2	1.05	1.24	1.24	1.15	1.16	1.21	1.21
10	CH_3	$–[CH_2]_2–$		H	Br	1.89	1.87	1.87	2.03	2.05	2.00	2.01
11	CH_3	$–[CH_2]_3–$		H	Br	1.76	1.69	1.69	1.69	1.69	1.63	1.61
12	CH_3	$–[CH_2]_4–$		H	Br	1.60	1.53	1.52	1.58	1.58	1.55	1.54
13	CH_3	$–[CH_2]_5–$		H	Br	1.31	1.38	1.38	1.47	1.48	1.47	1.49
14	CH_3	$–[CH_2]_2–$		H	Cl	1.96	2.03	2.04	2.06	2.07	2.07	2.08
15	CH_3	$–[CH_2]_3–$		H	Cl	1.92	1.83	1.82	1.72	1.70	1.70	1.68
16	$CH(CH_3)_2$	$–[CH_2]_2–$		H	Cl	1.86	1.83	1.83	1.90	1.91	1.91	1.92
17	$CH_2(CH_2)_2CH_3$	$–[CH_2]_2–$		H	Cl	1.70	1.74	1.75	1.72	1.72	1.83	1.84
18	$CH(CH_3)_2$	$–[CH_2]_4–$		H	Cl	1.51	1.49	1.49	1.45	1.45	1.46	1.46
19	$CH(CH_3)_2$	$–[CH_2]_5–$		H	Cl	1.32	1.35	1.35	1.34	1.34	1.38	1.38
20	$CH(CH_3)_2$	$–[CH_2]_6–$		H	Cl	0.90	1.22	1.23	1.23	1.24	1.29	1.32
21	$CH(CH_3)_2$	$–[CH_2]_2–$		H	CH_3	1.96	1.81	1.80	1.92	1.92	1.87	1.86
22	$CH(CH_3)_2$	$–[CH_2]_3–$		H	CH_3	1.46	1.63	1.64	1.58	1.58	1.49	1.50
23	CH_3	$–[CH_2]_2–$		H	CH_3	2.22	2.01	2.00	2.08	2.07	2.02	2.00
24	CH_2CH_3	$–[CH_2]_2–$		H	CH_3	1.92	1.90	1.90	1.97	1.97	1.95	1.95
25	CH_2CH_3	$–[CH_2]_3–$		H	CH_3	1.68	1.71	1.72	1.63	1.62	1.57	1.56
26	$CH(CH_3)_2$	$–[CH_2]_4–$		H	CH_3	1.22	1.47	1.49	1.47	1.48	1.41	1.43
27	$CH(CH_3)_2$	$–[CH_2]_5–$		H	CH_3	1.09	1.33	1.35	1.36	1.37	1.34	1.36
28	CH_3	$–[CH_2]_5–$		H	CH_3	1.40	1.48	1.48	1.52	1.53	1.49	1.50
29	CH_3	H	H	H	NO_2	1.29	1.42	1.45	1.37	1.37	1.52	1.56

30	$CH(CH_3)_2$	H	H	H	NO_2	1.28	1.26	1.26	1.21	1.21	1.38	1.39
31	CH_3	H	H	Cl	NO_2	0.44	0.81	1.28	1.22	1.28	0.81	1.26
32	$CH(CH_3)_2$	H	H	Cl	NO_2	1.13	0.72	0.37	1.07	1.07	0.72	0.38
33	CH_3	H	H	NO_2	H	1.49	1.42	1.41	1.37	1.36	1.50	1.51
34	$CH(CH_3)_2$	H	H	NO_2	H	1.34	1.26	1.25	1.22	1.22	1.37	1.37
35	CH_3	H	H	NO_2	Cl	1.33	1.37	1.38	1.22	1.21	1.35	1.36
36	$CH(CH_3)_2$	H	H	NO_2	Cl	1.45	1.23	1.18	1.08	1.05	1.22	1.21
37	CH_3	H	CH_3	H	NO_2	1.48	1.46	1.45	1.52	1.53	1.59	1.60
38	CH_3	CH_3	CH_3	H	NO_2	1.42	1.83	1.86	1.61	1.63	1.58	1.60
39	CH_3	CH_2CH_3	CH_2CH_3	H	NO_2	1.36	1.53	1.53	1.36	1.36	1.41	1.41
40	CH_3	$CH_2(CH_2)_2CH_3$	$CH_2(CH_2)_2\ CH_3$	H	NO_2	1.10	1.00	0.99	0.88	0.86	1.06	1.06
41	CH_3	CH_2Ph	CH_2Ph	H	NO_2	0.60	0.51	0.47	0.83	0.86	0.72	0.73
42	CH_2CH_3	$CH_2(CH_2)_2CH_3$	$CH_2(CH_2)_2CH_3$	H	NO_2	1.08	0.95	0.94	0.76	0.73	0.98	0.97
43	CH_2CH_3	CH_3	CH_2Ph	H	NO_2	0.98	1.05	1.05	1.11	1.12	1.05	1.06
44	CH_2CH_3	CH_3	$CH_2CH{=}CH_2$	H	NO_2	1.12	1.45	1.46	1.31	1.32	1.29	1.30
45	CH_2CH_3	CH_3	CH_2-1-Naph	H	NO_2	0.83	0.68	0.64	0.91	0.91	0.83	0.84
46	$CH(CH_3)_2$	$CH_2(CH_2)_2CH_3$	$CH_2(CH_2)_2CH_3$	H	NO_2	1.05	0.90	0.88	0.74	0.70	0.90	0.88
47	Cyclohexyl	H	CH_3	H	NO_2	1.19	1.23	1.24	1.17	1.17	1.30	1.31
48	CH_3	H	$CH_2CO_2CH_2CH_3$	H	NO_2	1.00	1.07	1.08	1.06	1.08	1.10	1.10
49	$CH(CH_3)_2$	H	$CH_2CO_2CH(CH_3)_2$	H	NO_2	0.92	0.86	0.85	0.93	0.93	0.87	0.86
50	$CH(CH_3)_2$	$CH_2CO_2CH_2CH_3$	$CH_2CO_2CH_2CH_3$	H	NO_2	0.66	0.84	0.88	0.69	0.70	0.48	0.44
51	CH_3	$=CHPh$		H	NO_2	0.82	0.83	0.83	0.87	0.87	0.84	0.84
52	CH_2CH_3	$=CHPh$		H	NO_2	0.75	0.79	0.79	0.75	0.75	0.77	0.77
53	$CH(CH_3)_2$	$=CHPh$		H	NO_2	0.64	0.75	0.76	0.74	0.74	0.69	0.70
54	$CH_2CH(CH_3)_2$	$=CHPh$		H	NO_2	0.66	0.71	0.71	0.51	0.48	0.62	0.61
55	$CH(CH_3)_2$	$=CHPh$		H	CH_3	0.89	0.68	0.64	0.88	0.88	0.72	0.67
56	$CH(CH_3)_2$	$=CHPh$		H	Cl	0.80	0.71	0.70	0.85	0.85	0.76	0.75

---

[a] In cases of compounds **1**–**28**, R_2 and R_3 substituents correspond to alicyclic rings of different size.

[b] Obsd = observed (Ref. 30); Calcd = calculated, Pred = predicted (leave-one-out).

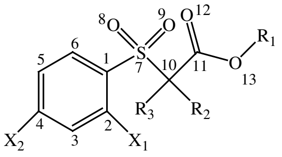
[c] From Eq. 9.

[d] From Eq. 13.

[e] From Eq. 15.

Table 2. Definitions of important ETA parameters used in exploring QSAR of toxicity of phenylsulfonyl carboxylates

Parameter	Definition
$\sum \alpha$	Sum of α values of all nonhydrogen vertices of a molecule
$[\sum \alpha]_X$	Sum of α values of all nonhydrogen vertices each of which is joined to four other vertices of the molecule
$[\sum \alpha]_P$	Sum of α values of all nonhydrogen vertices each of which is joined to only one other vertex of the molecule
$[\sum \alpha]_{R_1}$	Sum of α values of all nonhydrogen vertices of R_1 position
$[\eta'_F]_1$	Functionality for carbon atom at 1 position
$[\eta'_F]_4$	Functionality for carbon atom at 4 position
$[\eta'_F]_{10}$	Functionality for carbon atom at 10 position
$[\eta'_F]_{13}$	Functionality for oxygen atom at 13 position
$[\eta'_F]_{X_1-Cl}$	Functionality for chlorine substituent at X_1 position in the phenyl ring
$[\eta'_F]_{X_1-NO_2}$	Functionality for nitro substituent at X_1 position in the phenyl ring
$\sum \beta'_s$	$=[\sum \beta_s]/N_v$
$[\sum \beta'_s]_{Sub}$	Sum of $\sum \beta'_s$ values of substituents (R_2 and R_3) on C_{10}; $\sum \beta'_s$ is defined as $[\sum \beta_s]/N_v$ for nonhydrogen substituent(s); in case, hydrogen is present in the substituent position, the value for that position is taken as zero
$[\sum \beta'_{ns}]_{Sub}$	Sum of $\sum \beta'_{ns}$ values of substituents on C_{10}; $\sum \beta'_{ns}$ is defined as $[\sum \beta_{ns}]/N_v$ for nonhydrogen substituent(s); in case, hydrogen is present in the substituent position, the value for that position is taken as zero



1-56

Figure 1. General structure of phenylsulfonyl carboxylates: the common atoms have been numbered 1–13.

attempt was made to use both ETA and nonETA descriptors in modeling of the toxicity data.

The statistical quality of the equations[44] was judged by the parameters like explained variance (R_a^2, i.e., adjusted R^2), correlation coefficient (r or R), standard error of estimate (s) and variance ratio (F) at specified degrees of freedom (df). PRESS (leave-one-out) statistics[45,46] were calculated and leave-one-out cross-validation R^2 (Q^2), predicted residual sum of squares (PRESS), standard deviation based on PRESS (S_{PRESS}), standard deviation of error of prediction (SDEP) and average absolute predicted residual ($Pres_{av}$) were reported. Finally, 'leave-many-out' cross-validation was applied on the final equations. All the accepted equations have regression constants and F ratios significant at 95% and 99% levels, respectively, if not stated otherwise.

3. Results and discussion

3.1. Results with ETA indices

Three best equations for the ETA descriptors were selected from the population of equations generated based on the values of LOF as well as adjusted R^2 values (R_a^2) (shown in Table 3). From these equations the following was chosen as the best one based on statistical

Table 3. List of selected GFA-derived equations

Type of variables	Equation No.	Equations	Statistics			
			LOF	R^2	R_a^2	R
ETA	E1	$pC = 9.001[\sum \alpha]_X/\sum \alpha + 3.873\sum \beta'_s - 4.726[\eta'_F]_{X_1\_Cl} -0.126[\sum \alpha]_{Sub} + 0.362[\sum \beta'_s]_{Sub} - 1.943$	0.039	0.873	0.861	0.935
	E2	$pC = 8.938[\sum \alpha]_X/\sum \alpha - 3.371[\eta'_F]_{X_1\_Cl} + 0.511[\sum \beta'_s]_{Sub} +27.046[\eta'_F]_1 + 0.194[\sum \alpha]_{R_1} - 2.867$	0.039	0.873	0.860	0.934
	E3	$pC = 6.503[\sum \alpha]_X/\sum \alpha + 0.819[\sum \beta'_s]_{Sub} - 0.136\sum \alpha +0.966[\eta'_F]_{X_1\_NO_2} + 1.512$	0.039	0.861	0.850	0.928
NonETA	NE1	$pC = 0.942\,{}^3\chi_{CH} - 0.118\,{}^1\kappa_\alpha - 0.200S\_sssCH -0.272\ S\_ssssC + 2.930$	0.048	0.833	0.820	0.913
	NE2	$pC = 1.513\,{}^3\chi_{CH} + 0.029S\_ssO + 0.042S\_sCH_3 - 0.229\,{}^0\chi^V +0.267\,{}^3\chi_P^V + 2.500$	0.051	0.834	0.817	0.913
	NE3	$pC = 1.559\,{}^3\chi_{CH} + 0.044S\_sCH_3 - 0.207\,{}^0\chi^V + 0.240\,{}^3\chi_P^V -0.064S\_dsCH + 2.524$	0.053	0.829	0.812	0.911
ETA + NonETA	ENE1	$pC = 1.179\,{}^3\chi_{CH} - 5.502[\eta'_F]_{X_1\_Cl} + 7.937[\eta'_F]_4 - 8.861[\eta'_F]_{10} -0.079\,{}^1\kappa + 2.604$	0.035	0.878	0.865	0.937
	ENE2	$pC = 0.944\,{}^3\chi_{CH} + 7.401[\sum \alpha]_X/\sum \alpha - 4.850[\eta'_F]_{X_1\_Cl} -0.013MolRef + 1.483$	0.035	0.873	0.863	0.935
	ENE3	$pC = 0.824\,{}^3\chi_{CH} + 9.245[\sum \alpha]_X/\sum \alpha - 5.065[\eta'_F]_{X_1\_Cl} -0.107[\sum \alpha]_{Sub} + 0.469$	0.036	0.871	0.860	0.933

significance of the regression coefficients and intercorrelation among predictor variables.

$$
\begin{aligned}
pC = {} & 9.001(\pm 2.302)\left[\sum \alpha\right]_X \Big/ \sum \alpha \\
& + 3.873(\pm 4.154)\sum \beta'_{\mathrm{s}} - 4.726(\pm 2.509)[\eta'_{\mathrm{F}}]_{X_1-\mathrm{Cl}} \\
& - 0.126(\pm 0.045)\left[\sum \alpha\right]_{\mathrm{Sub}} \\
& + 0.362(\pm 0.305)\left[\sum \beta'_{\mathrm{s}}\right]_{\mathrm{Sub}} - 1.943,
\end{aligned}
$$

$$
\begin{aligned}
& n = 56,\ Q^2 = 0.771,\ R_{\mathrm{a}}^2 = 0.861, \\
& R^2 = 0.873,\ R = 0.935,\ s = 0.172, \\
& F = 69.0(\mathrm{df}5, 50),\ \mathrm{AVRES} = 0.128, \\
& \mathrm{PRESS} = 2.674,\ \mathrm{SDEP} = 0.219, \\
& S_{\mathrm{PRESS}} = 0.231,\ Pres_{\mathrm{av}} = 0.153.
\end{aligned} \tag{9}
$$

The 95% confidence intervals are shown within parentheses. The regression coefficient of $\sum \beta'_{\mathrm{s}}$ is significant at 90% level. The intercorrelation ($|r|$) matrix among the predictor variables is given in Table 4. The positive coefficient of the shape parameter $[\sum \alpha]_X / \sum \alpha$ indicates that X type branching (e.g., atom 10) is a conducive factor for the toxicity. This implies that presence of hydrogen at R_2 and R_3 positions will reduce toxicity. We may recall here an equation obtained in the previous communication[15], which is given below:

$$
\begin{aligned}
pC = {} & -1.232(\pm 1.273)\left[\sum \alpha\right]_P \Big/ \sum \alpha \\
& + 40.358(\pm 8.588)[\eta'_{\mathrm{F}}]_{13} - 3.266(\pm 2.799)[\eta'_{\mathrm{F}}]_{X_1-\mathrm{Cl}} \\
& + 0.776(\pm 0.279)\left[\sum \beta'_{\mathrm{s}}\right]_{\mathrm{Sub}} \\
& - 0.635(\pm 0.168)\left[\sum \beta'_{\mathrm{ns}}\right]_{\mathrm{Sub}} - 2.212,
\end{aligned}
$$

$$
\begin{aligned}
& n = 56,\ Q^2 = 0.704,\ R_{\mathrm{a}}^2 = 0.826, \\
& R^2 = 0.842,\ R = 0.918,\ s = 0.192, \\
& F = 53.3(\mathrm{df}5, 50),\ \mathrm{AVRES} = 0.139, \\
& \mathrm{PRESS} = 3.464,\ \mathrm{SDEP} = 0.249, \\
& S_{\mathrm{PRESS}} = 0.263,\ Pres_{\mathrm{av}} = 0.168.
\end{aligned} \tag{10}
$$

As is seen in Eq. 10 that $[\sum \alpha]_P / \sum \alpha$ bears a negative coefficient, presence of primary (pendant) type vertex will reduce the toxicity. This means that presence of branching as well as absence of primary vertex increases the toxicity. This is clearly indicative of favoring contribution of an alicyclic ring at R_2 and R_3 positions to the toxicity. Again the negative coefficient of $[\sum \alpha]_{\mathrm{sub}}$ in Eq. 9 clearly indicates that size of the alicyclic ring or substituents at R_2/R_3 positions contributes negatively to the toxicity. The negative coefficients of $\sum \beta'_{\mathrm{s}}$ and $[\sum \beta'_{\mathrm{s}}]_{\mathrm{sub}}$ in Eq. 9 suggest that presence of electronegative atoms, especially at R_2 and R_3 positions increases the toxicity. The negative coefficient of $[\eta'_{\mathrm{F}}]_{X_1-\mathrm{Cl}}$ in Eq. 9 indicates that the toxicity decreases as functionality value of Cl substituent at X_1 position increases.

Other two equations obtained from ETA descriptors and listed in Table 3 show positive coefficients of $[\eta'_{\mathrm{F}}]_1$, $[\sum \alpha]_{R_1}$ and $[\eta'_{\mathrm{F}}]_{X_1-\mathrm{NO}_2}$ and negative coefficients of $\sum \alpha$. These indicate positive contributions of the functionality at position 1, size of R_1 substituent, functionality contribution of nitro substituent at X_1 position while negative contribution of the overall molecular bulk. As the volume of R_2/R_3 positions shows negative contribution while the size of R_1 substituent shows positive contribution to the toxicity, opposite contributions of these volume terms explain the negative parabolic relation with respect to overall volume obtained in the previous communication:[15]

$$
\begin{aligned}
pC = {} & -3.761(\pm 2.689)[\eta'_{\mathrm{F}}]_{X_1-\mathrm{Cl}} - 6.511(\pm 3.469)[\eta'_{\mathrm{F}}]_{10} \\
& + 0.656(\pm 0.289)\left[\sum \beta'_{\mathrm{s}}\right]_{\mathrm{Sub}} - 0.588(\pm 0.327)\sum \alpha \\
& + 0.017(\pm 0.015)\left[\sum \alpha\right]^2 + 5.630,
\end{aligned}
$$

$$
\begin{aligned}
& n = 56,\ Q^2 = 0.726,\ R_{\mathrm{a}}^2 = 0.837, \\
& R^2 = 0.852,\ R = 0.923,\ s = 0.186, \\
& F = 57.4(\mathrm{df}5, 50),\ \mathrm{AVRES} = 0.136, \\
& \mathrm{PRESS} = 3.207,\ \mathrm{SDEP} = 0.239, \\
& S_{\mathrm{PRESS}} = 0.253,\ Pres_{\mathrm{av}} = 0.163.
\end{aligned} \tag{11}
$$

Such negative parabolic relations with respect to volume and lipophilicity have also been reported by other authors.[47,48]

Eq. 11 was the best equation obtained in the previous communcation[15] based on factor analysis as the pre-processing step and considering R_{a}^2 as the statistical measure

Table 4. Intercorrelation ($|r|$) matrix

	$[\sum \alpha]_p / \sum \alpha$	$[\sum \alpha]_{\mathrm{Sub}}$	$[\sum \beta'_{\mathrm{s}}]$	$[\sum \beta'_{\mathrm{s}}]_{\mathrm{Sub}}$	$[\eta'_{\mathrm{F}}]_4$	$[\eta'_{\mathrm{F}}]_{10}$	$[\eta'_{\mathrm{F}}]_{X_1\_\mathrm{Cl}}$	${}^3\chi_{\mathrm{CH}}$	${}^1\kappa_\alpha$	S_ssssC	S_sssCH
$[\sum \alpha]_X / \sum \alpha$	1.000										
$[\sum \alpha]_{\mathrm{Sub}}$	0.355	1.000									
$[\sum \beta'_{\mathrm{s}}]$	0.033	0.502	1.000								
$[\sum \beta'_{\mathrm{s}}]_{\mathrm{Sub}}$	0.451	0.371	0.301	1.000							
$[\eta'_{\mathrm{F}}]_4$	0.092	0.640	0.719	0.560	1.000						
$[\eta'_{\mathrm{F}}]_{10}$	0.569	0.221	0.309	0.556	0.438	1.000					
$[\eta'_{\mathrm{F}}]_{X_1\_\mathrm{Cl}}$	0.159	0.254	0.272	0.444	0.306	0.229	1.000				
${}^3\chi_{\mathrm{CH}}$	0.533	0.316	0.262	0.443	0.066	0.165	0.090	1.000			
${}^1\kappa_\alpha$	0.644	0.794	0.342	0.008	0.470	0.130	0.066	0.446	1.000		
S_ssssC	0.633	0.373	0.282	0.583	0.398	0.845	0.268	0.215	0.145	1.000	
S_sssCH	0.449	0.022	0.151	0.006	0.038	0.012	0.029	0.167	0.329	0.359	1.000

of comparison. When leave-one-out crossvalidation R^2 was considered, the following was the best equation:[15]

$$\begin{aligned} pC = {} & -3.831(\pm 2.758)[\eta'_{\mathrm{F}}]_{X_1-\mathrm{Cl}} + 0.493(\pm 0.299)\left[\sum \beta'_{\mathrm{s}}\right]_{\mathrm{Sub}} \\ & - 9.457(\pm 3.907)[\eta'_{\mathrm{F}}]_{10} + 26.236(\pm 3.646)[\eta'_{\mathrm{F}}]_{13} \\ & - 0.006(\pm 0.002)\left[\sum \alpha\right]^2 \\ & n = 56,\ Q^2 = 0.732,\ R_{\mathrm{a}}^2 = 0.826, \\ & R^2 = 0.839,\ R = 0.916,\ s = 0.192, \\ & F = 613.6(\mathrm{df}5, 51),\ \mathrm{AVRES} = 0.142, \\ & \mathrm{PRESS} = 3.135,\ \mathrm{SDEP} = 0.237, \\ & S_{\mathrm{PRESS}} = 0.250,\ Pres_{\mathrm{av}} = 0.166. \end{aligned} \tag{12}$$

It is clear that the best equation obtained in this communication based on genetic function algorithm is better than equations 11 and 12 obtained previously based on factor analysis considering both R_{a}^2 and Q^2 values.

3.2. Results with nonETA indices

While working with nonETA descriptors, three best equations were selected from the population of generated models (Table 3). From these equations the following was chosen as the best one based on statistical significance of the regression coefficients and intercorrelation among predictor variables.

$$\begin{aligned} pC = {} & 0.942(\pm 0.652)^3\chi_{\mathrm{CH}} - 0.118(\pm 0.021)^1\kappa_{\alpha} \\ & - 0.200(\pm 0.121)S\_\mathrm{sssCH} \\ & - 0.272(\pm 0.084)S\_\mathrm{ssssC} + 2.930, \\ & n = 56,\ Q^2 = 0.805,\ R_{\mathrm{a}}^2 = 0.820,\ R^2 = 0.833, \\ & R = 0.913,\ s = 0.196,\ F = 63.5(\mathrm{df}4, 51), \\ & \mathrm{AVRES} = 0.136,\ \mathrm{PRESS} = 2.276,\ \mathrm{SDEP} = 0.202, \\ & S_{\mathrm{PRESS}} = 0.211,\ Pres_{\mathrm{av}} = 0.147. \end{aligned} \tag{13}$$

The intercorrelation ($|r|$) matrix among the predictor variables is given in Table 4. Eq. 13 shows importance of connectivity, kappa shape and E-state parameters. When compared to Eq. (9), it is observed that adjusted R^2 value of Eq. 13 is lower in comparison to that of Eq. 9. However, Q^2 value of the Eq. 9 is somewhat lower than that of Eq. 13. The best equation obtained from nonETA indices in the previous communication[15] based on factor analysis was the following:

$$\begin{aligned} pC = {} & -0.021(\pm 0.004)\mathrm{MolRef} + 1.240(\pm 0.675)^3\chi^v_{\mathrm{CH}} \\ & - 0.179(\pm 0.088)S\_\mathrm{ssssC} + 0.257(\pm 0.194)S\_\mathrm{ddsN} \\ & - 0.022(\pm 0.025)S\_\mathrm{sCl} + 2.885, \\ & n = 56,\ Q^2 = 0.763,\ R_{\mathrm{a}}^2 = 0.798,\ R^2 = 0.816, \\ & R = 0.903,\ s = 0.207,\ F = 44.3(\mathrm{df}5, 50), \\ & \mathrm{AVRES} = 0.139,\ \mathrm{PRESS} = 2.769, \\ & \mathrm{SDEP} = 0.222,\ S_{\mathrm{PRESS}} = 0.235,\ Pres_{\mathrm{av}} = 0.156. \end{aligned} \tag{14}$$

Eq. 13 is superior to Eq. 14 with respect to both equation statistics and crossvalidation statistics.

3.3. Results with ETA and nonETA indices

While working with both ETA and nonETA descriptors, three best equations were selected from the population of models generated (Table 3). From these equations the following was chosen as the best one based on statistical of the regression coefficients and intercorrelation among predictor variables.

$$\begin{aligned} pC = {} & 1.179(\pm 0.545)^3\chi_{\mathrm{CH}} - 5.502(\pm 2.314)[\eta'_{\mathrm{F}}]_{X_1\_\mathrm{Cl}} \\ & + 7.937(\pm 4.931)[\eta'_{\mathrm{F}}]_4 - 8.861(\pm 2.993)[\eta'_{\mathrm{F}}]_{10} \\ & - 0.079(\pm 0.015)^1\kappa + 2.604, \\ & n = 56,\ Q^2 = 0.779,\ R_{\mathrm{a}}^2 = 0.865,\ R^2 = 0.878, \\ & R = 0.937,\ s = 0.169,\ F = 71.7(\mathrm{df}5, 50), \\ & \mathrm{AVRES} = 0.127, \mathrm{PRESS} = 2.585, \\ & \mathrm{SDEP} = 0.215,\ S_{\mathrm{PRESS}} = 0.227,\ Pres_{\mathrm{av}} = 0.153. \end{aligned} \tag{15}$$

The intercorrelation ($|r|$) matrix among the predictor variables is given in Table 4. When Eq. 15 is compared to Eq. 9, it is observed that there is marginal improvement in equation statistics and cross-validation statistics on using nonETA parameters with ETA ones. This shows that the ETA parameters are sufficiently rich in chemical information to encode the structural features contributing significantly to the acute toxicity of phenylsulfonyl carboxylates to *V. fischeri*. When Eq. (15) is compared to Eq. 13 involving different nonETA parameters, the adjusted R^2 value in the former case is found to be higher though Q^2 value in the latter case is somewhat larger.

The best equation obtained in the previous communication[15] involving both ETA and nonETA descriptors using factor analysis as the data-preprocessing step was the following:

$$\begin{aligned} pC = {} & -0.151(\pm 0.026)^0\chi^v + 1.202(\pm 0.546)^3\chi^v_{\mathrm{CH}} \\ & - 10.872(\pm 2.963)[\eta'_{\mathrm{F}}]_{10} - 4.532(\pm 2.382)[\eta'_{\mathrm{F}}]_{X_1-\mathrm{Cl}} \\ & - 0.296(\pm 0.297)[\eta'_{\mathrm{F}}]_{X_2-\mathrm{NO}_2} + 4.000, \\ & n = 56,\ Q^2 = 0.779,\ R_{\mathrm{a}}^2 = 0.861,\ R^2 = 0.874, \\ & R = 0.935, s = 0.172,\ F = 69.4(\mathrm{df}5, 50), \\ & \mathrm{AVRES} = 0.127,\ \mathrm{PRESS} = 2.589,\ \mathrm{SDEP} = 0.215, \\ & S_{\mathrm{PRESS}} = 0.228,\ Pres_{\mathrm{av}} = 0.152. \end{aligned} \tag{16}$$

On comparison, it is found that adjusted R^2 value of Eq. 15 is slightly higher than that of Eq. 16 while Q^2 values of both are the same.

Leave-20%-out crossvalidation was applied on Eqs. 9, 13 and 15 and the results are shown in Table 5. Crossvalidation statistics indicate robustness of the formulated models.

Table 5. Results of leave-20%-out cross-validation applied on Eqs. 9, 13 and 15 *Model equation*, $pC = \sum \beta_i x_i + \alpha$

Key Eq. no.	Number of cycles	Average regression coefficients (standard deviations)	Statistics Q^2 (Average *Pres*)
9	5[a]	$9.073\ (0.268)\ [\sum \alpha]_X / \sum \alpha + 4.116(0.471) \sum \beta'_s - 4.604(2.622)[\eta'_F]_{X_1\_Cl}$ $-0.125(0.004)[\sum \alpha]_{Sub} + 0.358(0.046)[\sum \beta'_s]_{Sub} - 2.098$	0.784 (0.146)
13	5[a]	$0.939(0.091)^3\chi_{CH} - 0.118(0.002)^1\kappa_\alpha - 0.200(0.024)S\_sssCH$ $-0.272(0.018)S\_ssssC + 2.927(0.060)$	0.817 (0.143)
15	5[a]	$1.174(0.088)^3\chi_{CH} - 5.403(2.446)[\eta'_F]_{X_1\_Cl} + 8.079(1.081)[\eta'_F]_4$ $-8.844(0.522)[\eta'_F]_{10} - 0.079(0.002)^1\kappa + 2.595(0.081)$	0.788 (0.147)

Q^2 denotes cross-validated R^2. Average *Pres* means average of absolute values of *predicted residuals*.

[a] Compounds were deleted in five cycles in the following manner: (**1**, **6**, **11**, **16**, ... ,**51**, **56**), (**2**, **7**, **12**, **17**, ... ,**47**, **52**), ... ,(**5**, **10**, **15**, **20**, ... ,**50**, **55**).

4. Conclusion

This study corroborates the results obtained from principal component factor analysis[15] and suggests that ETA parameters are sufficiently rich in chemical information to encode the structural features contributing significantly to the acute toxicity of phenylsulfonyl carboxylates to *V. fischeri*. This study also indicates the potential of genetic function approximation technique in developing statistically valid models when a large number of descriptors are available.

Acknowledgements

One of the authors (K.R.) thanks the All India Council for Technical Education (AICTE), New Delhi for financial grant under the Career Award for Young Teachers (CAYT) scheme.

References and notes

1. Kavlock, R. J.; Daston, G. P.; DeRosa, C.; Fenner-Crisp, P.; Gray, L. E.; Kaattari, S.; Lucier, G.; Luster, M.; Mac, M. J.; Maczka, C.; Miller, R.; Moore, J.; Rolland, R.; Scott, G.; Sheehan, D. M.; Sinks, T.; Tilson, H. A. *Environ. Health Perspect.* **1996**, *104*(Suppl 4), 715.
2. McKinney, J. D.; Richard, A.; Waller, C.; Newman, M. C.; Gerberick, F. *Toxicol. Sci.* **2000**, *56*, 8.
3. Comber, M. H. I.; Walker, J. D.; Watts, C.; Hermens, J. *Environ. Toxicol. Chem.* **2003**, *22*, 1822.
4. Russom, C. L.; Anderson, E. B.; Greenwood, B. E.; Pilli, A. *Sci. Total Environ.* **1991**, *109–110*, 667.
5. Ren, S. *Environ. Toxicol.* **2002**, *17*, 119.
6. Huuskonen, J. *Chemosphere* **2003**, *50*, 949.
7. Cronin, M. T.; Netzeva, T. I.; Dearden, J. C.; Edwards, R.; Worgan, A. D. *Chem. Res. Toxicol.* **2004**, *17*, 545.
8. Mazzatorta, P.; Benfenati, E.; Lorenzini, P.; Vighi, M. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 105.
9. Cronin, M. T.; Schultz, T. W. *Chem. Res. Toxicol.* **2001**, *14*, 1284.
10. Burden, F. R.; Winkler, D. A. *Chem. Res. Toxicol.* **2000**, *13*, 436.
11. Briens, F.; Bureau, R.; Rault, S. *Ecotoxicol. Environ. Saf.* **1999**, *43*, 241.
12. Roy, K.; Ghosh, G. *Internet Electron. J. Mol. Des.* **2003**, *2*, 599, http://www.biochempress.com.
13. Roy, K.; Ghosh, G. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 559.
14. Roy, K.; Ghosh, G. *QSAR Comb. Sci.* **2004**, *23*, 99.
15. Roy, K.; Ghosh, G. *QSAR Comb. Sci.* **2004**, *23*, 526.
16. Pal, D. K.; Sengupta, C.; De, A. U. *Indian J. Chem. Sect. B.* **1988**, *27*, 734.
17. Pal, D. K.; Sengupta, C.; De, A. U. *Indian J. Chem. Sect. B* **1989**, *28*, 261.
18. Pal, D. K.; Sengupta, M.; Sengupta, C.; De, A. U. *Indian J. Chem. Sect. B* **1990**, *29*, 451.
19. Pal, D. K.; Purkayastha, S. K.; Sengupta, C.; De, A. U. *Indian J. Chem. Sect. B.* **1992**, *31*, 109.
20. Roy, K.; Pal, D. K.; De, A. U.; Sengupta, C. *Indian J. Chem. Sect. B.* **1999**, *38*, 664.
21. Roy, K.; Pal, D. K.; De, A. U.; Sengupta, C. *Indian J. Chem. Sect. B.* **2001**, *40*, 129.
22. Roy, K.; Saha, A. *J. Mol. Model.* **2003**, *9*, 259.
23. Roy, K.; Saha, A. *Internet Electron. J. Mol. Des.* **2003**, *2*, 288, http://www.biochempress.com.
24. Roy, K.; Saha, A. *Internet Electron. J. Mol. Des.* **2003**, *2*, 475, http://www.biochempress.com.
25. Roy, K.; Chakroborty, S.; Ghosh, C. C.; Saha, A. *J. Indian Chem. Soc.* **2004**, *81*, 115.
26. Roy, K.; Saha, A. *Indian J. Chem. Sect. A.* **2004**, *43*, 1369.
27. Rogers, D.; Hopfinger, A. J. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854.
28. Fan, Y.; Shi, L. M.; Kohn, K. W.; Pommier, Y.; Weinstein, J. N. *J. Med. Chem.* **2001**, *44*, 3254.
29. Makhija, M. T.; Kulkarni, V. M. *Bioorg. Med. Chem.* **2002**, *10*, 1483.
30. Liu, X.; Yang, Z.; Wang, L. *SAR QSAR Environ. Res.* **2003**, *14*, 183.
31. CERIUS2 version 4.8 is a product of Accelrys Inc., San Diego, CA.
32. The GW-BASIC programs KRETA1 and KRETA2 were developed by Kunal Roy and standardized using known data sets.
33. Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
34. Hosoya, H. *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332.
35. Balaban, A. T.; Catana, C.; Dawson, M.; Niculescu-Duvaz, I. *Rev. Roum. Chim.* **1990**, *35*, 997.
36. Randic, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609.
37. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic: New York, 1976.
38. Kier, L. B.; Hall, L. H. *J. Pharm. Sci.* **1983**, *72*, 1170.
39. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Analysis*; Research Studies: Letchworth, England, 1986.
40. Kier, L. B. *Prog. Clin. Biol. Res.* **1989**, *291*, 105.
41. Kier, L. B.; Hall, L. H. *Pharm. Res.* **1990**, *7*, 801.
42. Hall, L. H.; Mohney, B. K.; Kier, L. B. *Quant. Struct–Act. Relat.* **1993**, *12*, 44.
43. *Comprehensive Medicinal Chemistry*; Hansch, C., Sammes, P. G., Taylor, J. B., Eds.; Pergamon: Oxford, 1990; Vol. 4.
44. Snedecor, G. W.; Cochran, W. G. *Statistical Methods*; Oxford and IBH: New Delhi, 1967; pp 381–418.

45. Wold, S.; Eriksson, L.. In *Chemometric Methods in Molecular Design*; Waterbeemd, H., Van de, Eds.; VCH: Weinheim, 1995; pp 312–317.
46. Debnath, A. K. In *Combinatorial Library Design and Evaluation*; Ghose, A. K., Viswanadhan, V. N., Eds.; Marcel Dekker: New York, 2001; pp 73–129.
47. Chand, P.; Babu, Y. S.; Bantia, S.; Rowland, S.; Dehghani, A.; Kotian, P. L.; Hutchison, T. L.; Ali, S.; Brouillette, W.; El-Kattan, Y.; Lin, T. H. *J. Med. Chem.* **2004**, *47*, 1919.
48. Goslinski, T.; Golankiewicz, B.; De Clercq, E.; Balzarini, J. *J. Med. Chem.* **2002**, *45*, 5052.