Prediction of Log$(IGC_{50})^{-1}$ for Benzene Derivatives to Ciliate *Tetrahymena pyriformis* from Their Molecular Descriptors

Mohammad Hossein Fatemi* and Hanieh Malekzadeh

Chemometrics Laboratory, Department of Chemistry, Mazandaran University, Babolsar, Iran

Received August 17, 2009; E-mail: mhfatemi@umz.ac.ir

The purpose of this study was to develop the structure–toxicity relationships for a large group of organic compounds including 392 substituted benzenes to the ciliate *Tetrahymena pyriformis* (Log$(IGC_{50})^{-1}$) using interpretable molecular descriptors. These descriptors were calculated using DRAGON and CODESSA software. Multiple linear regression and artificial neural network methods were used as linear and nonlinear feature-mapping techniques. The best obtained model was derived by MLR with seven descriptors which are: the molecular weight, the radial distribution function, the Kier shape index, the 26th component of atom-centered descriptors type of R–CX–R, the topographic electronic index, the H atoms attached to CO groups, the 24th component of atom-centered descriptors of R–CH–R. These descriptors can encode different features of molecules which are responsible for their steric, electronic, and lipophilicity interactions. The best obtained model had statistics of $R^2 = 0.822$, $F = 1386.806$, and $SE = 0.312$ for training and $R^2 = 0.815$, $F = 361.384$, and $SE = 0.337$ for prediction. The presented model shows better statistical parameters in comparison with a previous model. The reliability of the model was evaluated by using the leave-many-out cross-validation method ($Q^2 = 0.819$ and $SPRESS = 0.32$) as well as by *y*-scrambling.

Benzene derivatives have been used for many years in the chemical industry.[1,2] They are used as solvents, additives, cooling agents, insecticides, herbicides, and also organic synthesis. Many of these chemicals were released into the environment and accumulated in nearly all environmental compartments, especially in aquatic ecosystems. It is therefore beneficial to carry out a detailed study of their potential hazard to aquatic ecosystems. Aquatic toxicology is the science that deals with the effects of substances and physicochemical conditions on plants and animals that live in water. This includes the study of mechanisms by which changes in the quality of water or food supply affect growth, reproduction, behavior or survival of organisms. As a branch of ecotoxicology, aquatic toxicology has made an immense contribution to our understanding of how natural and man-made substances affect the living environment.[3] An important aspect of modern toxicology research is the prediction of toxicity of chemicals from their molecular structure. Since experimental measurements of aquatic toxicity are extremely time-consuming and expensive, it is imperative to develop robust quantitative structure–activity relationships (QSARs) that can predict toxicity of organic pollutants. QSARs establish mathematical relationships between physical, chemical, biological, or environmental activities of interest and measurable or computable parameters such as topological, physicochemical, stereochemical, or electronic indices.[4–7] QSARs have provided a valuable approach in research on the toxicity of organic chemicals in such studies. There are many papers about the QSAR predictions of toxicity of organic compounds.[8] For example, Ramos et al.[9] proposed a model for fish toxicity estimation based on three principal components derived from 11 physicochemical and quantum chemical descriptors. Hermens et al.[10] discussed QSAR models for the inhibition of bioluminescence of *P. phosphoreum* by 22 non-reactive organic compounds (aliphatic and aromatic). They found linear relationships between *n*-octanol/water partition coefficient (Log P) and toxicity. Improvements of their QSAR models occurred when molar refractivity (MR) was added as an independent variable. Blum and Speece[11] developed QSAR models for toxicity of a large set of organic compounds (including substituted phenols, benzene, and aliphatic compounds) to environmental bacteria. They successfully used Log P, solvation energy, and molecular connectivity as independent variables in QSAR investigations. Sixt et al.[12] investigated the acute toxicity of organic compounds to *P. phosphoreum* using "theoretical linear solvation energy relationship" (TLSER) parameters. Similarly, Kamlet et al.[13] found molar volume and solvatochromic parameters to be the most important descriptors for toxicity prediction of substituted aliphatic and aromatic compounds. In the same way, Castillo-Garit et al. developed a QSAR model for prediction of aquatic toxicity of benzene derivatives against *Tetrahymena pyriformis*.[14] They used MLR techniques for model development and obtained two stochastic and non-stochastic models. The coefficients of multiple determination (R^2) of the obtained stochastic model are: $R^2_{train} = 0.799$ and $R^2_{prediction} = 0.797$, while these values for the non-stochastic model are: $R^2_{train} = 0.791$ and $R^2_{prediction} = 0.762$.

In order to improve these models we decided to calculate a large number of descriptors and use multiple linear regression (MLR) and artificial neural network (ANN) as linear and nonlinear feature-mapping techniques to develop some QSAR models for prediction of benzene derivatives toxicities.

Published on the web February 26, 2010; doi:10.1246/bcsj.20090213

Experimental

Data Set. The compounds involved in this paper and their values of toxicity against the aquatic ciliate *T. pyriformis* were found in a paper published by Castillo-Garit et al.[14] The toxicities of these compounds are expressed as Log(IGC$_{50}$)$^{-1}$ where IGC$_{50}$ means 2-day (i.e., eight generations) static 50% inhibitory growth concentration. The *T. pyriformis* can reproduce itself in 6 h, so that in 48 h there are $48/6 = 8$ generations. The chemical names and the corresponding toxicities of these 392 benzene derivatives are listed in Table 1. These compounds were sorted according to their Log(IGC$_{50}$)$^{-1}$ value, then the training, internal and external test sets were chosen in desired distances from each others, consisting of 308, 42, and 42 members respectively. In development of the ANN model, the training set was used to train the network, the internal test set was used to prevent overfitting during the training of the network, and the external test set was used to evaluate the prediction power of the obtained model. In the case of MLR models, the training set was used to model development, and the internal and external test sets were considered as a test set which was used in evaluation of MLR models.

Descriptors Calculation. To obtain a QSAR model, compounds are often represented by molecular descriptors. The molecular descriptors used to search for the best model were calculated by the DRAGON 3.0 program[15] on the basis of the minimum energy molecular geometries optimized by the HyperChem package 7.0,[16] based on the AM1 semiempirical method. In addition, electronic descriptors were calculated by the MOPAC 6.0 package.[17] Also, the HyperChem and MOPAC output files were transferred into CODESSA software.[18–20] This software can calculate more than 400 constitutional, topological, geometrical, electrostatic, and quantum chemical descriptors and has been successfully used in various QSPR research.[21–23]

After calculation of the molecular descriptors, those that stayed constant for all molecules were eliminated and pairs of variables with a correlation coefficient greater than 0.9 were classified as inter-correlated and one in each correlated pair was deleted. In order to find the best QSAR model, three separate MLR models were developed using DRAGON calculated descriptors, CODESSA calculated descriptors and combined DRAGON and CODESSA calculated descriptors. The most statistically significant descriptors in each case were selected using the elimination selection stepwise regression variable selection method. The quality of each model was determined by examining the regression's statistical parameters. The following parameters were verified in each case: Fisher-ratio p-level [p(F)], standard deviation of the regression (*SE*) and R^2 value. Descriptors which appeared in the best MLR model were used as inputs for development of the ANN model.

Neural Networks. A detailed description of the theory behind an artificial neural network has been adequately described elsewhere.[24–31] In this work the ANN program was written in MATLAB 7.3.0 in our laboratory. This network was feed-forward fully connected with three layers which are input, hidden, and output layers. The number of nodes in the input and output layers are defined by the complexity of the problem being solved. The input layer receives the experimental information and the output layer contains the response sought. Descriptors appearing in the best MLR model were used as inputs for generation of the ANN. The hidden layer codes the information obtained from the input layer, and delivers it to the output layer. The number of nodes in the hidden layer may be considered as an adjustable parameter.[32] The number of nodes in the hidden layer was optimized. The initial weights ranged between -0.3 and $+0.3$. The initial bias values were set to be one. These values were optimized during the network training. It is known that neural networks can become over-trained. There are several methods for overcoming this problem. One of the superior methods is to use an external test set to validate the generalization of the network and monitoring the extent of over-training during training of the network.[29] Then that optimized and trained network is used to calculate the Log(1/IGC$_{50}$) of the external test set.

Results

In order to build a QSAR model with good generalized performance, a preliminary analysis for the quality of the data set (mainly the detection of outliers) was performed by modeling the complete set of 392 chemicals. The results revealed that there did not exist any significant outlier in the data set; then some QSAR models for the whole data set were developed by using DRAGON and CODESSA calculated descriptors, and multiple linear regression.

Linear Model with CODESSA Calculated Descriptors. To select the sets of descriptors that were most relevant to Log(IGC$_{50}$)$^{-1}$ value, the heuristic correlations performed for the whole set provided the optimal equations for different groups of descriptors (constitutional, geometrical, topological, and electrostatic) separately. Then, the selected descriptors of each group were considered in one spreadsheet for the selection of the most important descriptors using stepwise multiple regression variable selection. The numbers of descriptors in the final model were determined on the basis of the data set size and the values of R^2, the F and the standard error (*SE*). A major decision in developing successive QSAR is when to stop adding descriptors to the model during the stepwise regression procedure. A simple technique to control the model expansion is the "break point" procedure. In this method, the improvement of the statistical quality of the models is analyzed by plotting the square of correlation coefficient and standard error values of the obtained models versus the number of descriptors involved in each model. Consequently, the model corresponding to the break point is considered as the best/optimum model. Thus, the stepwise-multiple linear regression was applied to the training set and statistics of obtained MLR equations up to 21 descriptors were studied. The application of the "break point" algorithm led to the conclusion that the best model had seven parameters. This linear model has the following specifications:

$$\begin{aligned}\mathrm{Log(IGC_{50})^{-1}} &= 0.072(\pm 0.246) + 0.006(\pm 0.002)\mathrm{WNSA1} \\ &\quad - 0.189(\pm 0.045)\mathrm{NNA} + 0.440(\pm 0.051)^2\chi^{\nu} \\ &\quad - 2.441(\pm 0.458)\mathrm{RNHA} - 0.011(\pm 0.008)\mathrm{S_{YZ}} \\ &\quad + 0.018(\pm 0.006)\mathrm{S_{ZX}} - 0.129(\pm 0.040)\mathrm{WPSA3} \quad (1)\end{aligned}$$

where WNSA1 is the weighted partial negatively charged surface area (PNSA1*TMSA/1000), NNA is the number of N atoms, $^2\chi^{\nu}$ is the Kier and Hall index (order 2), RNHA is the relative number of H atoms, S_{YZ} is the shadow plane YZ, S_{ZX} is the shadow plane ZX, WPSA$^{(3)}$ is the weighted partial

Table 1. Experimental and ANN and MLR Predicted Values of $[\mathrm{Log}(1/\mathrm{IGC}_{50})]$[a)]

No.	Compounds	CAS	EXP	ANN		MLR(3)		$\mathrm{Resid}_{\mathrm{ANN}}$	$\mathrm{Resid}_{\mathrm{MLR(3)}}$
001	Benzene	71-43-2	−0.12	1.97		−0.45		−0.34	2.09
002	*p*-Xylene	106-42-3	0.25	1.72	e	0.15	t	−0.10	1.47
003	1-Phenyl-2-butanol	120055-09-6	−0.16	1.60		−0.01		0.15	1.76
004	Toluene	108-88-3	0.25	0.91		−0.21		−0.46	0.66
005	Butylbenzene	104-51-8	1.25	1.60	e	0.77	t	−0.48	0.35
006	Pentylbenzene	538-68-1	1.79	1.72		0.93		−0.85	−0.07
007	Benzylamine	100-46-9	−0.24	0.01		−0.68		−0.45	0.25
008	Isopropylbenzene	98-82-8	0.69	0.34	e	0.48	t	−0.21	−0.36
009	6-Phenyl-1-hexanol	2430-16-2	0.87	2.01	i	0.82	t	−0.05	1.14
010	5-Phenyl-1-pentanol	10521-91-2	0.42	0.38		0.37		−0.05	−0.04
011	α,α-Dimethylbenzenepropanol	103-05-9	−0.07	1.58		0.19		0.26	1.65
012	4-Phenyl-1-butanol	3360-41-6	0.12	0.46	e	0.22	t	0.10	0.34
013	3-Phenyl-1-propanol	122-97-4	−0.21	1.19		−0.19		0.01	1.40
014	Benzyl alcohol	100-51-6	−0.83	1.42		−0.64		0.19	2.25
015	1-Phenethylethanol	98-85-1	−0.66	0.29		−0.44		0.22	0.95
016	4-Ethylbenzyl alcohol	768-59-2	0.07	1.06		−0.12		−0.19	0.99
017	3-Phenyl-1-butanol	2722-36-3	0.01	1.72		0.09		0.08	1.71
018	(*R*)-1-Phenyl-1-butanol	22144-60-1	−0.01	0.14		0.04		0.05	0.15
019	4-Biphenylmethanol	3597-91-9	0.92	0.83		0.47		−0.45	−0.09
020	4-Ethylbiphenyl	5707-44-8	1.97	0.77		1.45		−0.52	−1.20
021	Biphenyl	92-52-4	1.05	1.09		0.65		−0.39	0.04
022	(±)-2-Phenyl-2-butanol	1565-75-9	0.06	0.84		−0.08		−0.14	0.78
023	(±)-1,2-Diphenyl-2-propanol	5342-87-0	0.80	1.96		0.85		0.05	1.15
024	1,1-Diphenyl-2-propanol	29338-49-6	0.75	0.02	e	1.03	t	0.28	−0.73
025	3,4-Dimethylaniline	95-64-7	−0.16	0.07	i	0.17	t	0.34	0.23
026	3-Aminobenzyl alcohol	1877-77-6	−1.13	1.97		−0.62		0.51	3.10
027	4-Butoxyaniline	4344-55-2	0.61	0.00	i	0.72	t	0.12	−0.61
028	4-Pentyloxyaniline	39905-50-5	0.97	0.00		0.83		−0.14	−0.97
029	4-Hexyloxyaniline	39905-57-2	1.38	0.60		1.11		−0.27	−0.78
030	4-Methylaniline	106-49-0	−0.05	1.65	i	−0.11	t	−0.06	1.70
031	4-Isopropylaniline	99-88-7	0.22	0.15	i	0.44	t	0.23	−0.07
032	3-Ethylaniline	587-02-0	−0.03	0.82		0.06		0.10	0.85
033	4-Ethylaniline	589-16-2	0.03	0.00		0.07		0.04	−0.03
034	3-Methylaniline	108-44-1	0.28	1.80		−0.19		−0.47	1.52
035	4-Butylaniline	104-13-2	1.07	0.41		0.65		−0.42	−0.66
036	(2-Bromoethyl)benzene	103-63-9	0.42	1.73		0.88		−0.24	1.31
037	2-Methylaniline	95-53-4	−0.16	0.64	e	−0.12	t	0.04	0.80
038	2,6-Diisopropylaniline	24544-04-5	0.76	1.14		0.84		0.08	0.37
039	Aniline	62-53-3	−0.23	1.73	i	−0.37	t	−0.14	1.96
040	2-Ethylaniline	578-54-1	−0.22	0.14		0.04		0.26	0.36
041	2,6-Diethylaniline	579-66-8	0.31	1.12		0.28		−0.03	0.82
042	Thioanisole	100-68-5	0.18	1.06		0.26		0.08	0.88
043	4-Methoxyphenol	150-76-5	−0.14	0.65		−0.20		−0.06	0.79
044	3,4,5-Trimethylphenol	527-54-8	0.93	0.53		0.41		−0.52	−0.4
045	Benzyl chloride	100-44-7	0.06	0.00	i	0.04	t	−0.02	−0.06
046	4-Methylanisole	104-93-8	0.25	1.50		−0.08		−0.33	1.25
047	2,3,5-Trimethylphenol	697-82-5	0.36	0.00		0.34		−0.02	−0.36
048	2,4,6-Trimethylphenol	527-60-6	0.42	0.10	i	0.46	t	0.04	−0.32
049	4-*tert*-Butylphenol	98-54-4	0.91	1.11		0.79		−0.12	0.20
050	4-*tert*-Pentylphenol	80-46-6	1.23	0.81		0.95		−0.28	−0.42
051	2,3,6-Trimethylphenol	2416-94-6	0.28	0.54	e	0.46	t	0.18	0.26
052	Phenetole	103-73-1	−0.14	0.92		−0.07		0.07	1.06
053	Anisole	100-66-3	−0.10	0.13		−0.33		−0.23	0.23
054	2,4-Dimethylphenol	105-67-9	0.14	0.10	i	0.23	t	0.09	−0.04
055	2-Phenyl-3-butyn-2-ol	127-66-2	−0.18	0.47	i	0.03	t	0.21	0.65
056	*p*-Cresol	106-44-5	−0.16	0.07	i	−0.01	t	0.15	0.24
057	4-Ethylphenol	123-07-9	0.21	2.11		0.15		−0.06	1.90
058	4-Propylphenol	645-56-7	0.64	1.15		0.44		−0.19	0.51

Continued on next page.

Continued.

No.	Compounds	CAS	EXP	ANN		MLR(3)		$Resid_{ANN}$	$Resid_{MLR(3)}$
059	3-Ethylphenol	620-17-7	0.29	2.16		0.15		−0.14	1.87
060	4-Nonylphenol	104-40-5	2.47	1.86	e	2.38	t	−0.09	−0.61
061	*m*-Cresol	108-39-4	−0.08	0.04	e	−0.01	t	0.07	0.12
062	*o*-Cresol	95-48-7	−0.29	0.64		−0.15		0.14	0.93
063	2-Ethylphenol	90-00-6	0.16	0.09		0.12		−0.04	−0.07
064	Phenol	108-95-2	−0.35	1.37	i	−0.27	t	0.07	1.72
065	2-Allylphenol	1745-81-9	0.33	1.46		−0.02		−0.35	1.13
066	Iodobenzene	591-50-4	0.36	1.03		0.28		−0.07	0.67
067	4-Chloroaniline	106-47-8	0.05	2.32		0.07		0.03	2.27
068	2-Tolunitrile	529-19-1	−0.24	0.25		0.11		0.35	0.49
069	4-Hydroxyphenethyl alcohol	501-94-0	−0.83	0.74		−0.37		0.46	1.57
070	2-Chloro-4-methylaniline	615-65-6	0.18	0.59		0.31		0.12	0.41
071	2-Chloroaniline	95-51-2	−0.17	1.76		0.05		0.22	1.93
072	5-Pentylresorcinol	500-66-3	1.31	1.39		1.11		−0.20	0.08
073	3-Methoxyphenol	150-19-6	−0.33	0.61		−0.22		0.11	0.94
074	4-Hexylresorcinol	136-77-6	1.80	0.02		1.32		−0.48	−1.78
075	4-Chloro-3,5-dimethylphenol	88-04-0	1.20	0.02		0.66		−0.54	−1.18
076	4-Bromotoluene	106-38-7	0.47	0.08		0.45		−0.02	−0.39
077	1-Bromo-4-ethylbenzene	1585-07-5	0.67	0.04		0.71		0.04	−0.63
078	4-Chloroanisole	623-12-1	0.60	1.03		0.21		−0.39	0.43
079	4-Chloro-3-methylphenol	59-50-7	0.80	1.67		0.42		−0.38	0.87
080	1,3-Dihydroxybenzene	108-46-3	−0.65	0.57		−0.24		0.41	1.23
081	Bromobenzene	108-86-1	0.08	0.82	e	0.33	t	0.24	0.74
082	4-Chlorophenol	106-48-9	0.54	1.48		0.17		−0.37	0.94
083	4-Iodophenol	540-38-5	0.85	2.34		0.41		−0.45	1.49
084	2-(4-Chlorophenyl)ethylamine	156-41-2	0.14	0.25		−0.03		−0.17	0.11
085	4-Chlorobenzylamine	104-86-9	0.16	1.29		−0.16		−0.33	1.13
086	2,4-Dichloroaniline	554-00-7	0.56	0.41		0.59		0.03	−0.15
087	Chlorobenzene	108-90-7	−0.13	2.12	e	0.14	t	0.27	2.26
088	3-Chloroaniline	108-42-9	0.22	0.33		0.07		−0.15	0.11
089	1,2-Dimethyl-4-nitrobenzene	99-51-4	0.59	0.42		0.71		0.12	−0.17
090	4-(Pentyloxy)benzaldehyde	5736-91-4	1.18	0.89	e	1.13	t	−0.05	−0.29
091	4-Nitrotoluene	99-99-0	0.65	0.02	i	0.57	t	−0.08	−0.63
092	4-Isopropylbenzaldehyde	122-03-2	0.67	0.03	e	0.61	t	−0.06	−0.64
093	1,2-Dimethyl-3-nitrobenzene	83-41-0	0.56	0.55		0.70		0.14	−0.01
094	3-Chlorophenol	108-43-0	0.87	0.77		0.17		−0.70	−0.10
095	3-Nitrotoluene	99-08-1	0.42	1.33		0.45		0.03	0.91
096	2-Nitrotoluene	88-72-2	0.26	0.00	i	0.55	t	0.28	−0.26
097	1,4-Dibromobenzene	106-37-6	0.68	0.32		0.84		0.16	−0.36
098	Benzaldehyde	100-52-7	−0.20	1.86		−0.36		−0.16	2.06
099	3-Ethoxy-4-hydroxybenzaldehyde	121-32-4	0.02	0.16		0.10		0.08	0.14
100	3-Methoxy-4-hydroxybenzaldehyde	121-33-5	−0.03	0.12	e	−0.02	t	0.01	0.15
101	4-Hydroxypropiophenone	70-70-2	0.12	0.13		0.34		0.22	0.01
102	2,4-Dichlorophenol	120-83-2	1.04	1.33		0.68		−0.36	0.29
103	Valerophenone	1009-14-9	0.56	0.48	i	0.94	t	0.38	−0.08
104	Propiophenone	93-55-0	−0.07	0.51		0.23		0.30	0.58
105	Butyrophenone	495-40-9	0.21	0.18		0.52		0.31	−0.03
106	2-Hydroxybenzaldehyde	90-02-8	0.42	2.50		−0.27		−0.70	2.08
107	Heptanophenone	1671-75-6	1.56	1.85		1.37		−0.18	0.29
108	Acetophenone	98-86-2	−0.46	0.00		−0.09		0.37	0.46
109	Nitrobenzene	98-95-3	0.14	0.62		0.19		0.05	0.48
110	Octanophenone	1674-37-9	1.89	0.17		1.67		−0.21	−1.72
111	2,5-Dichloroaniline	95-82-9	0.58	0.45		0.60		0.02	−0.14
112	3,4-Dichlorotoluene	95-75-0	1.07	0.17	i	0.95	t	−0.12	−0.90
113	3-Nitroaniline	99-09-2	0.03	1.27		0.20		0.17	1.24
114	3,5-Dichloroaniline	626-43-7	0.71	1.96		0.61		−0.11	1.25
115	4-Bromo-6-chloro-*o*-cresol	7530-27-0	1.28	1.09		1.04		−0.24	−0.18
116	1,2-Dichlorobenzene	95-50-1	0.53	0.96		0.56		0.03	0.43

Continued on next page.

Continued.

No.	Compounds	CAS	EXP	ANN		MLR(3)		$Resid_{ANN}$	$Resid_{MLR(3)}$
117	3-Nitroanisole	555-03-3	0.72	1.42	i	0.46	t	−0.25	0.70
118	Benzophenone	119-61-9	0.87	0.08		0.96		0.09	−0.79
119	3-Chloro-5-methoxyphenol	65262-96-6	0.76	1.69		0.32		−0.44	0.93
120	4-Nitrobenzyl chloride	100-14-1	1.18	1.89		0.61		−0.57	0.71
121	2,4-Dibromophenol	615-58-7	1.40	1.66		0.96		−0.44	0.26
122	2-Amino-5-chlorobenzonitrile	5922-60-1	0.44	0.59	e	0.50	t	0.06	0.15
123	2-Hydroxy-4-methoxyacetophenone	552-41-0	0.55	0.01		0.09		−0.45	−0.55
124	3,5-Dichlorophenol	591-35-5	1.56	1.25		0.71		−0.85	−0.31
125	4-Chlorobenzaldehyde	104-88-1	0.40	0.19		0.17		−0.23	−0.21
126	4-Chlorobenzophenone	134-85-0	1.50	0.42	i	1.65	t	0.15	−1.08
127	1,3,5-Trichlorobenzene	108-70-3	0.87	0.88		1.12		0.25	0.01
128	2,4,5-Trichloroaniline	636-30-6	1.30	0.05		1.13		−0.17	−1.25
129	4-Bromobenzophenone	90-90-4	1.26	0.36		1.61		0.36	−0.90
130	1,2,4-Trichlorobenzene	120-82-1	1.08	0.11		1.09		0.02	−0.97
131	2,4,6-Trichlorophenol	88-06-2	1.41	0.17		1.23		−0.18	−1.24
132	4-Ethoxy-2-nitroaniline	616-86-4	0.76	1.01		0.49		−0.27	0.25
133	5-Bromovanillin	2973-76-4	0.62	0.41		0.50		−0.12	−0.21
134	4-Nitrophenetole	100-29-8	0.83	0.35	e	0.75	t	−0.08	−0.48
135	4-Chloro-2-nitrotoluene	89-59-8	0.82	0.46		0.94		0.12	−0.36
136	1-Bromo-3-nitrobenzene	585-79-5	1.03	1.57		0.86		−0.17	0.54
137	4-Bromo-2,6-dichlorophenol	3217-15-0	1.78	0.18		1.36		−0.42	−1.60
138	2-Chloro-6-nitrotoluene	83-42-1	0.68	0.00		0.95		0.27	−0.68
139	2,3,5,6-Tetrachloroaniline	3481-20-7	1.76	1.55	i	1.85	t	0.09	−0.21
140	3-Nitrobenzonitrile	619-24-9	0.45	1.13		0.49		0.04	0.68
141	2,4,5-Trichlorophenol	95-95-4	2.10	0.26		1.22		−0.88	−1.84
142	1,2,4,5-Tetrachlorobenzene	95-94-3	2.00	0.61		1.64		−0.36	−1.39
143	4-Methyl-2-nitroaniline	89-62-3	0.37	0.16		0.33		−0.03	−0.21
144	1-Chloro-3-nitrobenzene	121-73-3	0.73	0.32		0.71		−0.02	−0.41
145	2-Nitroaniline	88-74-4	0.08	2.17		0.08		0.01	2.10
146	2,3,4,5-Tetrachloroaniline	634-83-3	1.96	0.48		1.65		−0.31	−1.48
147	2,4,6-Tribromophenol	118-79-6	1.91	0.68		1.62		−0.29	−1.23
148	2-Bromo-5-nitrotoluene	7149-70-4	1.16	0.44		1.11		−0.05	−0.72
149	1-Fluoro-3-iodo-5-nitrobenzene	3819-88-3	1.09	0.42		1.28		0.20	−0.67
150	2-Nitrophenol	88-75-5	0.67	0.89		0.26		−0.40	0.22
151	2-Chloro-4-nitroaniline	121-87-9	0.75	0.23		0.72		−0.02	−0.51
152	5-Hydroxy-2-nitrobenzaldehyde	42454-06-8	0.33	0.52		0.31		−0.02	0.19
153	3,4,5,6-Tetrabromo-*o*-cresol	576-55-6	2.57	0.61		2.52		−0.05	−1.96
154	2,3,4,6-Tetrachlorophenol	58-90-2	2.18	0.01	i	1.95	t	−0.23	−2.18
155	1-Fluoro-4-nitrobenzene	350-46-9	0.10	0.41		0.64		0.54	0.31
156	Pentafluoroaniline	771-60-8	0.26	0.70		1.16		0.90	0.44
157	1-Bromo-2-nitrobenzene	577-19-5	0.75	0.61		0.82		0.07	−0.07
158	3,5-Dibromosalicylaldehyde	90-59-5	1.65	0.66		1.08		−0.57	−0.99
159	3,5-Dichloronitrobenzene	618-62-2	1.13	0.16	i	1.42	t	0.29	−0.97
160	4-Chloro-3-nitrophenol	610-78-6	1.27	0.05	i	0.95	t	−0.32	−1.22
161	2,3,4,5-Tetrachlorophenol	4901-51-3	2.72	0.53		1.75		−0.97	−2.18
162	Thiobenzamide	2227-79-4	0.09	0.23		−0.27		−0.37	0.14
163	1-Chloro-4-nitrobenzene	100-00-5	0.43	0.15		0.73		0.30	−0.27
164	α,α,α,4-Tetrafluoro-*m*-toluidine	2357-47-3	0.77	0.18	e	0.48	t	−0.29	−0.59
165	1-Chloro-2-nitrobenzene	88-73-3	0.68	0.05	i	0.80	t	0.12	−0.63
166	4-Chloro-6-nitro-*m*-cresol	7147-89-9	1.63	0.35	i	1.19	t	−0.43	−1.27
167	Pentachlorophenol	87-86-5	2.07	1.18		2.31		0.24	−0.89
168	1,3-Dinitrobenzene	99-65-0	0.76	0.43		0.89		0.13	−0.33
169	2,4-Dinitrotoluene	121-14-2	0.87	0.05		1.02		0.15	−0.82
170	4,5-Dichloro-2-nitroaniline	6641-64-1	1.66	0.01	e	1.32	t	−0.34	−1.65
171	Pentafluorophenol	771-61-9	1.63	1.18		1.33		−0.30	−0.45
172	Pentabromophenol	608-71-9	2.66	0.24		2.99		0.33	−2.42
173	3-Chloro-4-fluoronitrobenzene	350-30-1	0.80	0.04		1.02		0.22	−0.75
174	1,4-Dinitrobenzene	100-25-4	1.30	0.34		0.91		−0.39	−0.96

Continued on next page.

Continued.

No.	Compounds	CAS	EXP	ANN		MLR(3)		$Resid_{ANN}$	$Resid_{MLR(3)}$
175	3,4-Dichloronitrobenzene	99-54-7	1.16	0.26		1.24		0.08	−0.90
176	2,5-Dichloronitrobenzene	89-61-2	1.13	0.46		1.22		0.09	−0.67
177	2,4-Dichloro-6-nitroaniline	2683-43-4	1.26	0.04		1.13		−0.13	−1.22
178	3,4-Dinitrobenzyl alcohol	79544-31-3	1.09	0.02		1.16		0.08	−1.06
179	2,4-Dichloronitrobenzene	611-06-3	0.99	0.01	i	1.30	t	0.40	−0.98
180	2,3-Dichloronitrobenzene	3209-22-1	1.07	0.01		1.19		0.12	−1.06
181	1,2-Dinitrobenzene	528-29-0	1.25	0.52		0.85		−0.40	−0.73
182	Phenyl isothiocyanate	103-72-0	1.41	0.47	i	0.53	t	−0.89	−0.94
183	3-Trifluoromethyl-4-nitrophenol	88-30-2	1.65	1.18		0.74		−0.91	−0.47
184	2,6-Diiodo-4-nitrophenol	305-85-1	1.81	0.35		1.85		0.04	−1.46
185	2,4-Dichloro-6-nitrophenol	609-89-2	1.75	0.87		1.34		−0.41	−0.88
186	1,3,5-Trichloro-2-nitrobenzene	18708-70-8	1.43	0.56		1.71		0.28	−0.86
187	1,2,4-Trichloro-5-nitrobenzene	89-69-0	1.53	0.31		1.75		0.22	−1.22
188	1,2,3-Trichloro-4-nitrobenzene	17700-09-3	1.51	0.04		1.73		0.22	−1.46
189	2-Chloro-5-nitrobenzaldehyde	6361-21-3	0.53	0.02		0.84		0.31	−0.51
190	Pentafluorobenzaldehyde	653-37-2	0.82	0.91		1.27		0.45	0.09
191	2,4-Dinitroiodobenzene	709-49-9	2.12	0.87		1.63		−0.49	−1.25
192	2,3,5,6-Tetrachloronitrobenzene	117-18-0	1.82	0.83		2.24		0.42	−0.99
193	2,5-Dinitrophenol	329-71-5	1.04	0.83	e	0.93	t	−0.11	−0.21
194	2,4-Dinitroaniline	97-02-9	0.72	0.00		0.76		0.04	−0.72
195	2,3,4,5-Tetrachloronitrobenzene	879-39-0	1.78	0.01		2.28		0.50	−1.77
196	1,2,3-Trifluoro-4-nitrobenzene	771-69-7	1.89	0.81	i	1.26	t	−0.63	−1.08
197	1,2-Dichloro-4,5-dinitrobenzene	6306-39-4	2.21	0.02		2.11		−0.10	−2.19
198	2,6-Dinitroaniline	606-22-4	0.84	0.15		0.52		−0.32	−0.69
199	4,6-Dinitro-2-methylphenol	534-52-1	1.73	0.68		0.99		−0.73	−1.04
200	4-*tert*-Butyl-2,6-dinitrophenol	4097-49-8	1.80	1.20	e	2.04	t	0.24	−0.60
201	1-Bromo-2,4-dinitrobenzene	584-48-5	2.31	0.25		1.52		−0.79	−2.05
202	2,4-Dinitrophenol	51-28-5	1.06	0.01		0.77		−0.29	−1.05
203	1,5-Dichloro-2,3-dinitrobenzene	28689-08-9	2.42	0.74		2.09		−0.33	−1.68
204	6-Chloro-2,4-dinitroaniline	3531-19-9	1.12	1.25		1.28		0.16	0.13
205	2-Bromo-4,6-dinitroaniline	1817-73-8	1.24	1.18		1.42		0.18	−0.06
206	2,3,4,6-Tetrafluoronitrobenzene	314-41-0	1.87	0.15		1.45		−0.42	−1.72
207	2,6-Dinitrophenol	573-56-8	0.83	0.64		0.73		−0.09	−0.19
208	1-Chloro-2,4-dinitrobenzene	97-00-7	2.16	1.17		1.37		−0.78	−0.98
209	2,4-Dinitro-1-fluorobenzene	70-34-8	1.71	0.45		1.16		−0.54	−1.26
210	Pentafluoronitrobenzene	880-78-4	2.43	0.14	e	1.98	t	−0.45	−2.29
211	Tetrachloro-1,4-dinitrobenzene	20098-38-8	2.82	0.67	e	3.12	t	0.30	−2.15
212	1,5-Difluoro-2,4-dinitrobenzene	327-92-4	2.08	1.3		1.45		−0.63	−0.70
213	1,3-Dinitro-2,4,5-trichlorobenzene	2678-21-9	2.60	0.575		2.33		−0.27	−2.03
214	1,2,4-Trichloro-3,5-dinitrobenzene	6284-83-9	2.19	0.00		2.31		0.12	−2.19
215	4-Chloro-3,5-dinitrobenzaldehyde	1930-72-9	2.66	1.05	i	1.64	t	−1.02	−1.61
216	1-Phenyl-2-propanol	14898-87-4	−0.62	0.00		−0.32		0.30	0.62
217	4-Methylbenzyl alcohol	589-18-4	−0.49	0.19		−0.38		0.11	0.68
218	(±)-1-Phenyl-2-pentanol	705-73-7	0.16	1.13		0.25		0.09	0.97
219	4-Isopropylbenzyl alcohol	536-60-7	0.18	0.02		0.19		0.01	−0.16
220	2-(*p*-Tolyl)ethylamine	3261-62-9	−0.04	0.88		−0.28		−0.25	0.92
221	4-Methylbenzylamine	104-84-7	−0.01	0.04		−0.43		−0.42	0.05
222	3-Methylbenzyl alcohol	587-03-1	−0.24	1.35		−0.38		−0.14	1.59
223	3-Phenyl-2-propen-1-ol	104-54-1	−0.08	0.00		−0.33		−0.25	0.08
224	4-*tert*-Butylbenzyl alcohol	877-65-6	0.48	0.05		0.52		0.05	−0.42
225	4-Methylphenethyl alcohol	699-02-5	−0.26	0.21		−0.23		0.03	0.47
226	1-Phenylethylamine	618-36-0	−0.18	0.16		−0.53		−0.35	0.34
227	2-Methylbenzyl alcohol	89-95-2	−0.43	0.02		−0.44		−0.01	0.45
228	2-Methyl-1-phenyl-2-propanol	100-86-7	−0.41	1.08		−0.09		0.32	1.49
229	*N*-Methylphenethylamine	589-08-2	−0.41	1.23		−0.47		−0.06	1.64
230	*β*-Methylphenethylamine	582-22-9	−0.28	0.04	e	−0.08	t	0.20	0.32
231	(±)-1-Phenyl-1-butanol	22135-49-5	−0.09	2.51		0.04		0.13	2.60
232	(±)-1-Phenyl-1-propanol	93-54-9	−0.43	0.11		−0.17		0.25	0.54

Continued on next page.

Continued.

No.	Compounds	CAS	EXP	ANN		MLR(3)		$Resid_{ANN}$	$Resid_{MLR(3)}$
233	Phenethyl alcohol	60-12-8	−0.59	0.27	i	−0.37	t	0.22	0.87
234	2-Phenyl-1-propanol	1123-85-9	−0.4	0.00		−0.15		0.25	0.40
235	2-Phenyl-2-propanol	617-94-7	−0.57	0.14		−0.21		0.36	0.71
236	2-Phenyl-1-butanol	89104-46-1	−0.11	0.33	i	0.23	t	0.34	0.44
237	Benzhydrol	91-01-0	0.50	2.32		0.52		0.02	1.82
238	Benzaldoxime	622-32-2	−0.11	0.99		−0.20		−0.09	1.11
239	3,5-Dimethylaniline	108-69-0	−0.36	0.02		0.07		0.43	0.38
240	4-*tert*-Butylaniline	769-92-6	0.36	0.00	i	0.85	t	0.49	−0.36
241	2,4-Dimethylaniline	95-68-1	−0.29	0.22		−0.06		0.23	0.51
242	4-Phenylbutyronitrile	2046-18-6	0.15	0.89		0.64		0.49	0.74
243	2,4,6-Trimethylaniline	88-05-1	−0.05	0.08		0.18		0.23	0.13
244	3-Phenylpropionitrile	645-59-0	−0.16	0.06		0.18		0.34	0.22
245	4-*sec*-Butylaniline	30273-11-1	0.61	0.88		0.61		0.00	0.27
246	2,3-Dimethylaniline	87-59-2	−0.43	0.27		−0.01		0.42	0.70
247	Benzyl cyanide	140-29-4	−0.36	0.03		−0.11		0.25	0.39
248	2,5-Dimethylaniline	95-78-3	−0.33	0.050		0.05		0.38	0.38
249	α-Methylbenzyl cyanide	1823-91-2	0.01	0.02		0.28		0.27	0.01
250	2-Isopropylaniline	643-28-7	0.12	1.20		0.14		0.02	1.08
251	2,6-Dimethylaniline	87-62-7	−0.43	0.63	i	0.07	t	0.51	1.06
252	*N*-Ethylaniline	103-69-5	0.07	0.00		−0.17		−0.24	−0.07
253	2-Propylaniline	1821-39-2	0.08	0.06		0.32		0.24	−0.02
254	*N*-Methylaniline	100-61-8	0.06	1.03		−0.38		−0.44	0.97
255	2-Amino-4-*tert*-butylaniline	1199-46-8	0.37	1.58		0.76		0.4	1.21
256	2-Methoxyaniline	90-04-0	−0.69	0.81		−0.35		0.34	1.50
257	3-Phenylpyridine	1008-88-4	0.47	1.43		0.12		−0.35	0.96
258	2-Aminobenzyl alcohol	5344-90-1	−1.07	0.00	i	−0.68	t	0.39	1.07
259	2-Benzylpyridine	101-82-6	0.38	0.35		0.52		0.14	−0.03
260	3,5-Di-*tert*-butylphenol	1138-52-9	1.64	0.19		1.91		0.27	−1.45
261	Phenyl propargyl sulfide	5651-88-7	0.54	0.00	e	0.71	t	0.17	−0.54
262	4-Ethoxyphenol	622-62-8	0.01	0.00		0.06		0.05	−0.01
263	4-Butoxyphenol	122-94-1	0.70	0.06	e	0.82	t	0.12	−0.64
264	4-Benzylpyridine	2116-65-6	0.63	0.06		0.51		0.44	0.00
265	2-Phenylpyridine	1008-89-5	0.27	1.48		0.35		0.08	1.21
266	3,4-Dimethylphenol	95-65-8	0.12	0.06		0.15		0.03	−0.06
267	3-*tert*-Butylphenol	585-34-2	0.74	0.55		0.78		0.05	−0.19
268	3,5-Dimethylphenol	108-68-9	0.11	0.00		0.14		0.03	−0.11
269	6-*tert*-Butyl-2,4-dimethylphenol	1879-09-0	1.16	1.59		1.14		−0.02	0.44
270	4-Isopropylphenol	99-89-8	0.47	0.19		0.46		−0.01	−0.27
271	3-Isopropylphenol	618-45-1	0.61	0.18		0.45		−0.16	−0.43
272	2,3-Dimethylphenol	526-75-0	0.12	1.43		0.11		−0.01	1.31
273	2,5-Dimethylphenol	95-87-4	0.14	0.03		0.10		−0.04	−0.11
274	4-Hydroxy-3-methoxybenzyl alcohol	498-00-0	−0.70	0.00		−0.43		0.27	0.70
275	2-Isopropylphenol	88-69-7	0.61	0.00		0.46		−0.15	−0.61
276	3-Amino-*m*-cresol	53222-92-7	−0.55	0.04		−0.23		0.32	0.59
277	4-Chloro-2-methylaniline	95-69-2	0.35	0.67		0.22		−0.13	0.32
278	2-Methoxy-4-propenylphenol	97-54-1	0.75	1.42		0.22		−0.53	0.67
279	2,4,6-Tris(dimethylaminomethyl)phenol	90-72-2	−0.52	0.29		−0.23		0.29	0.81
280	2-Fluoroaniline	348-54-9	−0.37	0.95		−0.17		0.2	1.32
281	4-Aminobenzyl cyanide	3544-25-0	−0.76	0.08		−0.12		0.64	0.85
282	3-Iodoaniline	626-01-7	0.65	0.00		0.30		−0.35	−0.65
283	3-Cinnamonitrile	4360-47-8	0.16	0.18	i	0.18	t	0.02	0.02
284	3-Fluorobenzyl alcohol	456-47-3	−0.39	0.00		−0.32		0.07	0.39
285	3-Cyanoaniline	2237-30-1	−0.47	0.81		−0.18		0.29	1.28
286	4-Fluorophenol	371-41-5	0.02	1.91	e	0.06	t	0.05	1.89
287	2-Iodoaniline	615-43-0	0.35	1.85		0.29		−0.05	1.50
288	3-Fluoroaniline	372-19-0	−0.10	0.02		−0.13		−0.04	0.12
289	4-Chloro-2-methylphenol	1570-64-5	0.70	2.4		0.39		−0.31	1.72
290	4-Chloro-3-ethylphenol	14143-32-9	1.08	1.44		0.64		−0.44	0.36

Continued on next page.

Continued.

No.	Compounds	CAS	EXP	ANN		MLR(3)		$Resid_{ANN}$	$Resid_{MLR(3)}$
291	2-Chloro-4,5-dimethylphenol	1124-04-5	0.69	0.01		0.66		−0.03	−0.68
292	3,5-Dimethoxyphenol	500-99-2	−0.09	0.96		−0.06		0.03	1.05
293	4-Hydroxybenzyl cyanide	14191-95-8	−0.38	1.09		−0.02		0.37	1.48
294	4-Bromo-2,6-dimethylphenol	2374-05-2	1.16	2.14	e	0.93	t	−0.23	0.98
295	2-Bromobenzyl alcohol	18982-54-2	0.10	0.31	e	0.17	t	0.07	0.21
296	2-Chloro-5-methylphenol	615-74-7	0.54	0.03		0.39		−0.14	−0.51
297	2-Fluorophenol	367-12-4	0.19	0.00		−0.06		−0.25	−0.19
298	4-(Dimethylamino)benzaldehyde	100-10-7	0.23	1.73		−0.07		−0.30	1.50
299	4-Bromophenol	106-41-2	0.68	0.46		0.31		−0.37	−0.22
300	3-Chloro-2-methylaniline	95-79-4	0.50	0.67	i	0.44	t	−0.06	0.18
301	3-Chloro-4-methylaniline	95-74-9	0.39	0.00	e	0.45	t	0.06	−0.39
302	3-Chloro-2-methylaniline	87-60-5	0.38	0.59		0.24		−0.14	0.22
303	4-Chlorophenethyl alcohol	1875-88-3	0.32	0.46	i	0.18	t	−0.14	0.14
304	4-Chlorobenzyl alcohol	873-76-7	0.25	0.00		−0.11		−0.36	−0.25
305	2-Bromo-4-methylphenol	6627-55-0	0.60	0.21		0.54		−0.05	−0.39
306	1,3,5-Trimethyl-2-nitrobenzene	603-71-4	0.86	0.02		0.92		0.06	−0.84
307	3-Chlorobenzyl alcohol	873-63-2	0.15	0.10	e	0.02	t	−0.13	−0.05
308	2-Bromophenol	95-56-7	0.33	0.26		0.29		−0.04	−0.06
309	4-Hydroxy-3-methoxybenzonitrile	4421-08-3	−0.03	0.00	e	0.15	t	0.18	0.03
310	3-Nitrobenzyl alcohol	619-25-0	−0.22	0.00		0.04		0.26	0.22
311	4-Bromophenylacetonitrile	16532-79-9	0.60	0.01		0.56		−0.04	−0.59
312	4-Methoxybenzonitrile	874-90-8	0.10	0.00		−0.04		−0.14	−0.10
313	2-Hydroxy-4,5-dimethylacetophenone	36436-65-4	0.71	0.00		0.49		−0.21	−0.71
314	2-Anisaldehyde	135-02-4	0.15	0.03		−0.24		−0.40	−0.12
315	4-Chlororesorcinol	95-88-5	0.13	0.01		0.28		0.15	−0.12
316	Methyl 4-methylaminobenzoate	18358-63-9	0.31	0.01		−0.25		−0.56	−0.3
317	4-Phenoxybenzaldehyde	67-36-7	1.26	0.01		1.35		0.09	−1.26
318	3-Hydroxy-4-methoxybenzaldehyde	621-59-0	−0.14	0.09		−0.12		0.02	0.23
319	4-Biphenylcarbaldehyde	3218-36-8	1.12	0.03		0.76		−0.36	−1.09
320	2,4,5-Trimethoxybenzaldehyde	4460-86-0	−0.10	0.32		0.01		0.11	0.42
321	4-Benzoylaniline	1137-41-3	0.68	0.07		0.94		0.26	−0.60
322	3-Anisaldehyde	5991-31-1	0.23	0.07		−0.21		−0.44	−0.15
323	*n*-Propyl cinnamate	7778-83-8	1.23	0.09		0.62		−0.61	−1.14
324	*trans*-Ethyl cinnamate	103-36-6	0.99	0.23		0.28		−0.71	−0.76
325	Hexanophenone	942-92-7	1.19	0.03	i	1.25	t	0.06	−1.16
326	*n*-Butyl cinnamate	538-65-8	1.53	0.85		0.90		−0.63	−0.68
327	4-Chlorobenzyl cyanide	140-53-4	0.66	0.28		0.41		−0.25	−0.38
328	*trans*-Methyl cinnamate	103-26-4	0.58	0.08		−0.01		−0.59	−0.5
329	Ethyl 4-methoxybenzoate	94-30-4	0.77	0.92		0.10		−0.67	0.15
330	Phenylacetohydrazide	937-39-3	−0.48	0.22	e	−0.28	t	0.20	0.70
331	3-Hydroxybenzaldehyde	100-83-4	0.08	0.24		−0.24		−0.32	0.16
332	2,6-Dichlorophenol	87-65-0	0.73	0.29		0.68		−0.05	−0.44
333	Benzyl methacrylate	2495-37-6	0.65	0.16		0.42		−0.23	−0.49
334	Isoamyl 4-hydroxybenzoate	6521-30-8	1.48	0.17		0.96		−0.52	−1.31
335	Benzyl 4-hydroxyphenyl ketone	2491-32-9	1.07	0.86		1.11		0.04	−0.21
336	Benzyl benzoate	120-51-4	1.45	1.14		0.99		−0.45	−0.31
337	4-Benzoylphenol	1137-42-4	1.02	0.39	e	1.22	t	0.2	−0.63
338	2-Methyl-5-nitrophenol	5428-54-6	0.66	0.82		0.51		−0.15	0.16
339	3-Acetoamidophenol	621-42-1	−0.16	0.74		0.10		0.26	0.90
340	4-Cyanobenzamide	3034-34-2	−0.38	0.82		−0.20		0.18	1.21
341	2-Nitrobiphenyl	86-00-0	1.3	1.35		1.11		−0.19	0.06
342	5-Chloro-2-hydroxybenzamide	7120-43-6	0.59	0.43	e	0.17	t	−0.42	−0.15
343	3-Nitrophenol	554-84-7	0.51	1.22		0.30		−0.21	0.71
344	Phenyl-1,3-dialdehyde	626-19-7	0.18	0.23	e	−0.13	t	−0.31	0.05
345	Ethyl 4-bromobenzoate	5798-75-4	1.33	1.01		0.63		−0.70	−0.32
346	2,4-Dihydroxyacetophenone	89-84-9	0.25	0.39		0.07		−0.18	0.14
347	3-Chlorobenzophenone	1016-78-0	1.55	1.29		1.47		−0.08	−0.25
348	Phenyl 4-hydroxybenzoate	17696-62-7	1.37	0.87		1.29		−0.08	−0.5

Continued on next page.

Continued.

No.	Compounds	CAS	EXP	ANN		MLR(3)		$Resid_{ANN}$	$Resid_{MLR(3)}$
349	Phenyl benzoate	93-99-2	1.35	1.97		1.20		−0.15	0.62
350	2-Hydroxy-4-methoxybenzophenone	131-57-7	1.42	1.47		1.13		−0.29	0.07
351	Benzylidenemalononitrile	2700-22-3	0.64	0.46	e	0.63	t	−0.01	−0.18
352	4-Nitrophenyl phenyl ether	620-88-2	1.58	0.99		1.90		0.32	−0.59
353	Resorcinol monobenzoate	136-36-7	1.11	1.28		1.30		0.19	0.18
354	4-Bromophenyl 3-pyridyl ketone	14548-45-9	0.82	0.04		1.17		0.35	−0.78
355	3-Nitroacetophenone	121-89-1	0.32	0.58		0.57		0.25	0.26
356	3-Nitrobenzaldehyde	99-61-6	0.11	0.02		0.32		0.2	−0.09
357	Ethyl phenylcyanoacetate	4553-07-5	−0.02	0.51		0.39		0.42	0.53
358	2-Nitroanisole	91-23-6	−0.07	0.48		0.29		0.36	0.55
359	3-Methyl-2-nitrophenol	4920-77-8	0.61	0.45	e	0.63	t	0.02	−0.15
360	2,5-Diphenyl-1,4-benzoquinone	844-51-9	1.48	0.29		1.56		0.08	−1.18
361	2-Nitrobenzamide	610-15-1	−0.72	0.02	e	0.21	t	0.93	0.74
362	Methyl 2,5-dichlorobenzoate	2905-69-3	0.81	0.35	i	0.89	t	0.08	−0.46
363	2-Nitrobenzaldehyde	552-89-6	0.17	0.08		0.18		0.01	−0.09
364	4-Methyl-2-nitrophenol	119-33-5	0.57	1.51		0.52		−0.05	0.94
365	2,2′,4,4′-Tetrahydroxybenzophenone	131-55-5	0.96	0.47		1.13		0.17	−0.49
366	4-Nitrobenzaldehyde	555-16-8	0.2	0.27		0.32		0.12	0.07
367	5-Methyl-2-nitrophenol	700-38-9	0.59	0.70		0.51		−0.08	0.11
368	3,5-Dichlorosalicylaldehyde	90-60-8	1.55	0.04	e	0.98	t	−0.57	−1.51
369	2-(Benzylthio)-3-nitropyridine	69212-31-3	1.72	0.08		1.54		−0.18	−1.63
370	Ethyl 4-nitrobenzoate	99-77-4	0.71	0.24		0.63		−0.08	−0.48
371	2,4-Dichlorobenzaldehyde	874-42-0	1.04	0.01		0.69		−0.35	−1.03
372	2,3,4-Trichloroacetophenone	13608-87-2	1.34	0.02		1.45		0.11	−1.32
373	2,2′-Dihydroxybenzophenone	835-11-0	1.16	0.53		0.83		−0.33	−0.63
374	Methyl 4-nitrobenzoate	619-50-1	0.39	0.26		0.34		−0.05	−0.13
375	2-Chloromethyl-4-nitrophenol	2973-19-5	0.75	0.04		0.70		−0.05	−0.71
376	α,α,α-Trifluoro-*p*-cresol	402-45-9	0.62	0.00		0.10		−0.49	−0.62
377	Dimethyl nitroterephthalate	5292-45-5	0.43	0.27		0.42		−0.01	−0.16
378	Thioacetanilide	637-53-6	−0.01	0.01	i	0.38	t	0.39	0.01
379	2-Nitroresorcinol	601-89-8	0.66	0.16	e	0.49	t	−0.17	−0.5
380	3,5-Dibromo-4-hydroxybenzonitrile	1689-84-5	1.16	0.92		1.25		0.09	−0.24
381	Pentafluorobenzyl alcohol	440-60-8	−0.2	0.63	e	1.13	t	1.33	0.83
382	Methyl 4-chloro-2-nitrobenzoate	42087-80-9	0.82	2.24	i	1.00	t	0.18	1.42
383	1-Fluoro-2-nitrobenzene	1493-27-2	0.23	0.85		0.46		0.23	0.62
384	α,α,α-Trifluoro-*o*-toluidine	393-39-5	−0.02	0.08		0.20		0.22	0.1
385	3-Hydroxy-4-nitrobenzaldehyde	704-13-2	0.27	0.58		0.20		−0.07	0.31
386	2,5-Dibromonitrobenzene	3460-18-2	1.37	0.00	i	1.74	t	0.37	−1.37
387	Benzoyl cyanide	613-90-1	0.31	0.15		0.10		−0.21	−0.16
388	4,5-Difluoro-2-nitroaniline	78056-39-0	0.75	0.02	i	0.86	t	0.11	−0.73
389	2,5-Difluoronitrobenzene	364-74-9	0.33	0.33		0.81		0.48	0.00
390	2,4-Dibromo-6-nitroaniline	827-23-6	1.62	0.1		1.39		−0.22	−1.44
391	4-Hydroxy-3-nitrobenzaldehyde	3011-34-5	0.61	0.01	e	0.54	t	−0.07	−0.6
392	Benzoyl isothiocyanate	532-55-8	0.10	0.06		0.45		0.35	−0.04

a) In the above table "e" represents the external test set, "i" represents the internal test set in the case of ANN model, and "t" represents the test set in the case of MLR model.

positively charged surface area (PPSA3*TMSA/1000). The model gave the statistics of $R^2 = 0.713$, $F = 741.699$, and $SE = 0.397$ for the training set and $R^2 = 0.691$, $F = 183.281$, and $SE = 0.437$ for the prediction set.

Linear Model with DRAGON Calculated Descriptors. The best MLR equation with DRAGON calculated descriptor using the toxicity as the dependent variable was derived in the same way as mentioned in the previous section. The best obtained model with seven descriptors has the following specification:

$$\begin{aligned}\mathrm{Log}(\mathrm{IGC}_{50})^{-1} = {} & -1.443(\pm 0.097) + 0.006(\pm 0.001)\mathrm{MW} \\ & - 0.173(\pm 0.018)\mathrm{RDF020e} + 0.074(\pm 0.011)\mathrm{Ts} \\ & + 0.115(\pm 0.022)\mathrm{nCaR} + 0.034(\pm 0.005)\mathrm{RDF040m} \\ & + 0.048(\pm 0.008)\mathrm{PCWTe} - 0.076(\pm 0.024)\mathrm{nHDon} \end{aligned} \quad (2)$$

where MW is the molecular weight, RDF020e is the radial distribution function −2.0/weighted by atomic sanderson electronegativies, Ts is the T total size index/weighted by atomic electrotopological states, nCaR is the number of substituted aromatic carbon (sp^2), RDF040m is the radial distribution

function −4.0/weighted by atomic masses, PCWTe is the partial charge weighted topological electronic charge, and nHDon is the number of acceptor atoms for H-bonds (N O F). This model has the statistical parameters of $R^2 = 0.80$, $F = 1175.194$, and $SE = 0.334$ for the training set and $R^2 = 0.81$, $F = 407.169$, and $SE = 0.331$ for the prediction set. The equation based on DRAGON calculated descriptors had better statistical parameters in prediction than the equation obtained from CODESSA calculated descriptors.

Combined Linear Model. We have also considered a combination of CODESSA and DRAGON calculated descriptors to explore better MLR models. The objective was to find out how much improvement in statistical quality of models is possible by combination of DRAGON and CODESSA calculated descriptors. After the combination of these descriptors in a new spreadsheet, the stepwise-multiple linear regression was applied to the training set and statistics of obtained MLR equations up to 21 descriptors were studied. The variations of R^2 and SE against the number of descriptors in the models were recorded and shown in Figure 1. As can be seen from this figure, seven descriptors model can be selected as the best model. This model has the following specifications:

$$\begin{aligned}\mathrm{Log(IGC_{50})^{-1}} &= -1.785(\pm 0.102) + 0.003(\pm 0.001)\mathrm{MW} \\ &- 0.117(\pm 0.019)\mathrm{RDF020e} + 0.241(\pm 0.023)^{1}\kappa \\ &+ 0.267(\pm 0.025)\mathrm{C026} - 0.484(\pm 0.062)\mathrm{T^E} \\ &+ 0.066(\pm 0.007)\mathrm{H046} + 0.097(\pm 0.015)\mathrm{C024} \qquad (3)\end{aligned}$$

where MW is the molecular weight, RDF020e is the radial distribution function −2.0/weighted by atomic Sanderson electronegativities, $^{1}\kappa$ is the Kier shape index (order 1), C026 is the 26th component of atom-centered descriptors of type R–CX–R, T^E is the topographic electronic index, H046 shows the H atoms attached to CO groups and C024 is the 24th component of atom-centered descriptors of type R–CH–R. The chemical meaning of these descriptors are described in the next section. This MLR model has the statistical parameters of $R^2 = 0.822$, $F = 1386.806$, and $SE = 0.312$ for training and $R^2 = 0.815$, $F = 361.384$, and $SE = 0.337$ for the prediction set. Comparison between these results and those obtained for previous models reveals the superiority of this model over other models. The calculated Log(IGC$_{50}$)$^{-1}$ for the training and test set using this model are shown in Table 1. The predicted Log(IGC$_{50}$)$^{-1}$ for the training and test set using this model were plotted against their experimental values in Figure 2. The linear correlation between response and selected variables is reliable.

Nonlinear Modeling. In order to investigate any nonlinear relationships between selected descriptors and Log(IGC$_{50}$)$^{-1}$, an artificial neural network was used. Descriptors that appeared in the eq 3 were used as inputs of developed ANN. In the first step the ANN parameters, including number of nodes in the input layer, number of nodes in the hidden layer, number of nodes in output layer, weights learning rate, biases learning rate and momentum were optimized. In order to determine the optimum number of hidden layer nodes several training sessions were conducted with different numbers of hidden nodes. The values of standard error for training set (SET) and internal test set (SEP) were calculated after each 1000 iterations

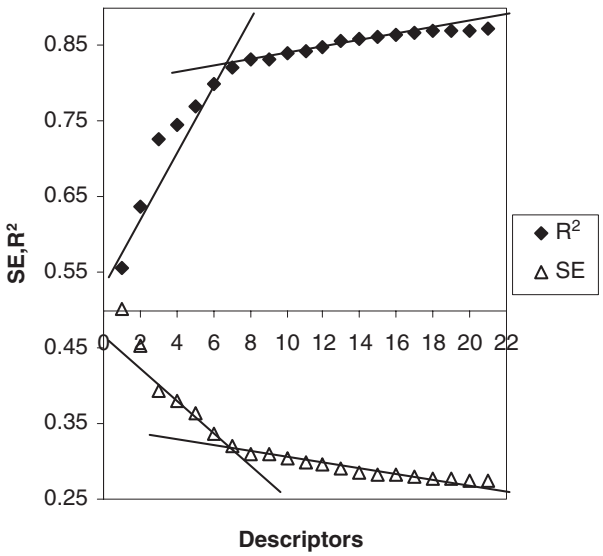

Figure 1. The variations of R^2 and SE against the number of descriptors for MLR model (3).

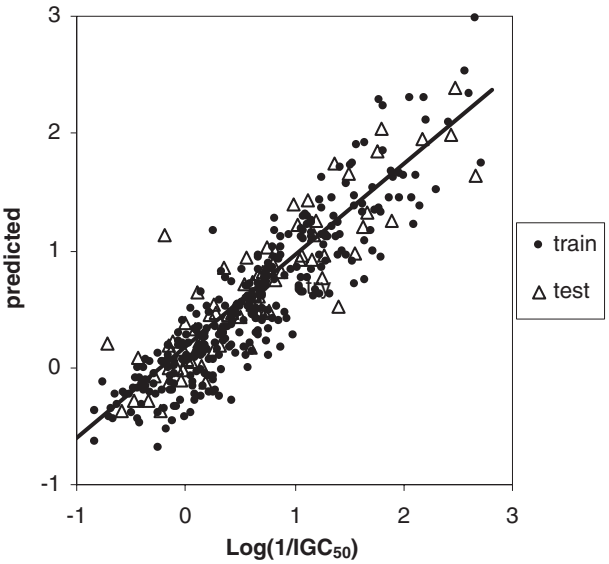

Figure 2. Plot of predicted toxicity by MLR model 3 versus experimental values.

and training epoch was stopped when overtraining began. Then the recorded values of SET and SEP were plotted against the number of hidden layer nodes, and the number of hidden nodes with minimum values of SET and SEP was chosen as optimum. Learning rates of weights and biases and also momentum values were optimized in a similar way. Table 2 shows the architecture and specifications of the optimized ANN's parameters. Then the optimized network was trained by using the training set for the adjustment of weights and bias values. To control the overfitting of the network during the training procedure, standard error in calculation of Log(IGC$_{50}$)$^{-1}$ for training and internal test sets were calculated and recorded to monitor the extent of the learning after each 1000 iterations. To maintain the predictive power of the network at a desirable level, training was stopped when standard error for prediction of Log(IGC$_{50}$)$^{-1}$ for the internal test set started to increase.

Then the trained ANN was used to predict the toxicity of the external test set in order to evaluate the predictive power of the network. The predicted values of $Log(IGC_{50})^{-1}$ for training, internal and external test sets are shown in Table 1. The statistical parameters of ANN model in calculation of $Log(IGC_{50})^{-1}$ are; $R^2 = 0.812$, $F = 308$, and $SE = 0.321$ for training set, $R^2 = 0.697$, $F = 92$, and $SE = 0.437$ for internal test set and $R^2 = 0.788$, $F = 148$, and $SE = 0.366$ for external test set.

Discussion

The comparison between statistical results of ANN model and those obtained by linear models 1–3 indicate that a linear model based on descriptors which were developed from the DRAGON and CODESSA calculated descriptors (model (3)) had the best statistical parameters in prediction of $Log(IGC_{50})^{-1}$. In the same QSAR work on this data set, Castillo-Garit et al. reported the coefficient of multiple determination of $R^2 = 0.733$ and $R^2 = 0.721$ and standard error of $SE = 0.394$ and $SE = 0.403$ for the stochastic and non-stochastic models respectively.[14] Then, they considered 6 compounds as outliers and deleted their information in further investigations. The obtained models had the statistical parameters of $R^2 = 0.799$, $SE = 0.343$ and $R^2 = 0.791$, $SE = 0.344$ for the training set and $R^2 = 0.797$, $SE = 0.364$ and $R^2 = 0.762$, $SE = 0.347$ for the prediction set for the stochastic and non-stochastic models respectively. Our best model (model (3)) gives the statistical parameters of $R^2 = 0.823$, $SE = 0.312$ for training and $R^2 = 0.815$, $SE = 0.337$ for test sets without deleting any data. Comparison between these values reveals the significant improvement of our models.

Descriptor Interpretation. A prerequisite of a good predictive model for any biological activity is that it should be transparent and mechanistically interpretable. To achieve transparency and mechanistic interpretability, the physicochemical meaning of parameters utilized for the modeling needs to be elucidated. Mechanistic interpretation of QSAR descriptors is seldom easy, especially for heterogeneous data sets. In this study our best model contained descriptors which were calculated by DRAGON and CODESSA software. These descriptors are: MW, RDF020e, C024, C026, H046, KSI, and T^E. Table 3 shows the correlation matrix between these descriptors. As can be seen from this table there are not any significant correlated descriptors in our model. Examination of descriptors included in the selected model reveals that they encode different aspects of the molecular structures. According to the Hansch model, steric, electronic and lipophilicity interactions eclipse molecule activity.[33] Descriptors which appeared in this model can encode different aspects of molecules which are responsible for these interactions. In order to evaluate the relative importance of each descriptor, their mean effects were calculated. Figure 3 represents the mean effects for each descriptor. As can be seen from this figure the order of importance of descriptors is: KSI > T^E > C026 > C024 > MW > RDF020e > H046. These descriptors can encode the strength of intramolecular bonding interactions and characterize the stability of the molecules, their conformational flexibility and other valency-related properties. The most important descriptor according to its mean effect is KSI. This descriptor is Kier-shape index which encodes the size of the molecule and has a direct relationship with the number of non-hydrogen atoms in the molecule. Other topological descriptors, including C024, C026, and H046, are atom-centered descriptors, which appeared in the model. These descriptors describe the size, branching, and composition of a molecule and relate to the dispersion interaction among molecules. The size of a molecule affects its penetration through cell membranes and therefore influences/effective concentration of chemicals in the cell body, the existence of geometrical and topological

Table 2. Architecture and Specifications of Optimized ANN Model

Parameter	Value
Number of nodes in the input layer	7
Number of nodes in the hidden layer	8
Number of nodes in output layer	1
Weights learning rate	0.6
Biases learning rate	0.6
Momentum	0.5

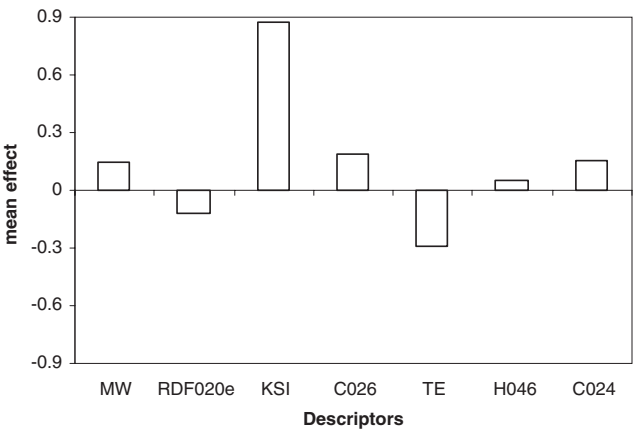

Figure 3. The calculated mean effects for descriptors in MLR model (3).

Table 3. Correlation Matrix between Selected Descriptors

	RDF020e	C024	C026	H046	MW	KSI	T^E
RDF020e	1						
C024	0.112	1					
C026	−0.233	−0.644	1				
H046	0.130	−0.079	−0.295	1			
MW	0.011	−0.188	0.582	−0.152	1		
KSI	0.314	−0.043	0.339	0.130	0.745	1	
T^E	0.619	0.007	0.036	0.276	0.162	0.603	1

descriptors in the model can encode these features of chemicals of interest. By increasing the size of an atom, the steric interaction increases, therefore their biological activities decrease, consequently the value of Log(IGC$_{50}$)$^{-1}$ increases. This observation is in agreement with positive signs of these descriptors (KSI, MW, C024, C026, and H046). Another descriptor in the model was the topographic electronic index of T^E which indicates the electronic distribution of the molecule and can encode some aspects of the molecule which are responsible for electronic interactions. This descriptor encodes the electron density over the molecule. The larger the T^E is, the lower the density of the electron cloud of the molecule. The density of the electron cloud on the molecule is a main factor that influences the polarity of the molecular, which may be lead to various polar interactions between the chemical, cell membrane and the receptor. Since the mean effect of T^E had negative sign by increasing of T^E values the Log(IGC$_{50}$)$^{-1}$ values were decreased.

The last descriptor is radial distribution function that encodes the probability distribution of finding an atom in a spherical volume of radius R. By increasing the distance between an atom and the center of the molecule, the value of this descriptor increases, therefore decreases the molecular activities due to steric interactions. All of these descriptors can encode different aspects of molecules which affect electronic, hydrophobic, and steric interactions and therefore affect molecular activities.

Diversity Test. One of the most critical aspects of constructing the training set is to warrant enough molecular diversity for it. In this study, diversity analysis was performed on the data set to make sure the structures of the training or test sets can represent those of the whole.[34] A MATLAB program was written to combine maximum dissimilarity search algorithms and a general multidimensional measurement of chemical similarity based on different molecular descriptors. The closer to one the distance is, the more diverse to each other the compound is. We considered a database of n compounds generated from m highly correlated chemical descriptors $\{X_j\}_{j=1}^{m}$. Each compound X_i is represented as the following vector:

$$X_i = (x_{i1}, x_{i2}, x_{i3}, \ldots, x_{im}) \quad \text{for} \quad i = 1, 2, \ldots, n \tag{4a}$$

where x_{ij} denotes the value of descriptor j of compound X_i. The collective database $X = \{X_i\}_{i=1}^{N}$ is represented a $n \times m$ matrix of X:

$$X = (X_1, X_2, \ldots, X_N)^T = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \tag{4b}$$

Here the superscript T denotes the vector/matrix transpose. A distance score for two different compounds X_i and X_j can be measured by the Euclidean distance norm based on the compound descriptors:

$$d_{ij} = \|X_i - X_j\| = \sqrt{\sum_{k=1}^{m} (x_{ik} - x_{jk})^2} \tag{4c}$$

The mean distances of one sample to the remaining ones were computed as follows:

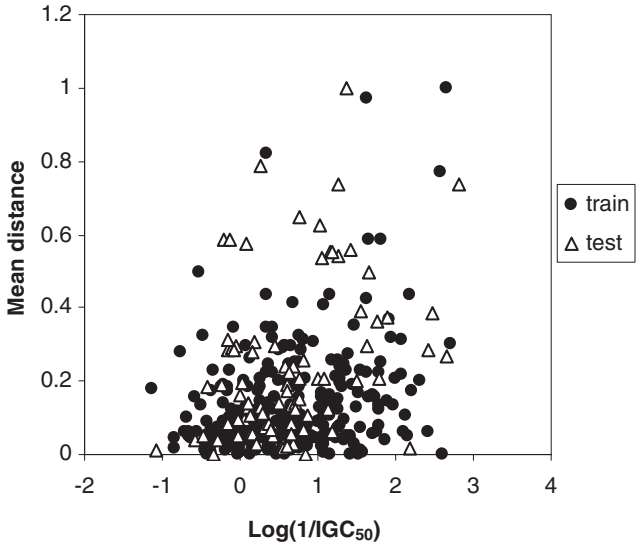

Figure 4. Diversity plot of samples for training, internal and external test set.

$$\overline{d_i} = \frac{\sum_{j=1}^{n} d_{ij}}{n-1} \quad i = 1, 2, \ldots, n \tag{4d}$$

Then the mean distances were normalized within the interval of zero to one. The mean distances of samples were plotted vs. Log(1/IGC$_{50}$) in Figure 4. As can be seen from this figure, the structures of the compounds are diverse in both training and test sets. The training set with a broad representation of the chemistry space was adequate to ensure the model stability and the diversity of the test set can prove the predictive capability of the model.

Validation of the Models. In order to ensure that the results were not conditioned by the data distribution in the descriptor space, the built model was also validated by evaluating the model's external predictive power on the selected prediction set by the leave-many-out (LMO) scheme. In this method a model is built with N–m compounds where m represents a group of randomly selected data points. Each m group is left out of the model derivation and predicted in turn. The outcomes from this procedure are a cross-validated correlation coefficient and standardized predicted error sum of squares (*SPRESS*), which are calculated according to the following formulas:

$$Q^2 = 1 - \frac{\sum (y_O - y_i)^2}{\sum (y_i - y_m)^2} \tag{5}$$

$$\text{SPRESS} = \sqrt{\frac{\sum (y_O - y_i)^2}{n - k - 1}} \tag{6}$$

In the above expressions, y_m is the mean of dependent variable (experimental values), n is the number of observations, and k is the number of independent variables in the regression equation. Q^2 is the proportion of variability in a data set that is accounted by a statistical model, and *SPRESS* is criteria of deviation from observed data. The obtained statistical results of leave-twenty-out cross-validation test on model (3) were: $Q^2 = 0.819$ and $SPRESS = 0.320$, which revealed the reliability of

Table 4. R^2 Values after Several Y-Randomization Tests

No.	R^2	No.	R^2
1	0.200	16	0.124
2	0.197	17	0.081
3	0.133	18	0.088
4	0.081	19	0.076
5	0.063	20	0.051
6	0.032	21	0.074
7	0.030	22	0.009
8	0.093	23	0.115
9	0.089	24	0.028
10	0.070	25	0.024
11	0.049	26	0.030
12	0.038	27	0.086
13	0.146	28	0.060
14	0.085	29	0.025
15	0.044	30	0.047

the obtained model. The model was further validated by applying the Y-randomization test. This technique was used to ensure the robustness of a QSAR model.[35–38] In this test, the dependent variable vector $[\mathrm{Log}(\mathrm{IGC}_{50})^{-1}]$ is randomly scrambled and a new QSAR model is developed using the original independent variable matrix. As can be seen from Table 4, after 30 repetitions the random models were found to have significantly lower R^2 values than the original model, which indicates that the good results in our original model (3) are not due to chance correlation or structural dependency of the training set.

Conclusion

Results of this study indicate that the multiple linear regression model can be used as a feature-mapping technique to quantitatively relate molecular structural descriptors of the largest benzene derivatives to their toxicities. Descriptors that appeared in this models are topological and electronic types which can encode features of molecules that are responsible for their steric, electronic and lipophilicity interactions. The obtained model was evaluated by using splitting of data set, leave-many-out cross-validation and *y*-scrambling methods. The result of these tests reveals the reliability of the obtained model.

References

1 S. Ikeno, C. Ogino, T. Ito, N. Shimizu, *Biochem. Eng. J.* **2003**, *15*, 193.

2 O. Pfohl, P. Avramova, G. Brunner, *Fluid Phase Equilib.* **1997**, *141*, 179.

3 R. G. V. Boelens, *Guidelines for the Use of Laboratory Tests and Aquati Organisms in the Control of Liquid Waste Discharges*, IIRS (Dublin) Occasional Report Series, **1980**, No. 4, p. 12.

4 G. Melagraki, A. Afantitis, K. Makridima, H. Sarimveis, O. Igglessi-Markopoulou, *J. Mol. Model.* **2006**, *12*, 297.

5 G. Melagraki, A. Afantitis, H. Sarimveis, O. Igglessi-Markopoulou, C. T. Supuran, *Bioorg. Med. Chem.* **2006**, *14*, 1108.

6 J. T. Leonard, K. Roy, *QSAR Comb. Sci.* **2004**, *23*, 387.

7 T. I. Netzeva, A. O. Aptula, S. H. Chaudary, J. C. Duffy, T. W. Schultz, G. Schüürmann, M. T. D. Cronin, *QSAR Comb. Sci.* **2003**, *22*, 575.

8 M. T. D. Cronin, J. C. Dearden, *Quant. Struct.-Act. Relat.* **1995**, *14*, 1.

9 E. U. Ramos, W. H. J. Vaes, H. J. M. Verhaar, J. L. M. Hermens, *Environ. Sci. Pollut. Res.* **1997**, *4*, 83.

10 J. Hermens, F. Busser, P. Leeuwangh, A. Musch, *Ecotoxicol. Environ. Saf.* **1985**, *9*, 17.

11 D. J. W. Blum, R. E. Speece, *Ecotoxicol. Environ. Saf.* **1991**, *22*, 198.

12 S. Sixt, J. Altschuh, R. Brüggemann, *Chemosphere* **1995**, *30*, 2397.

13 M. J. Kamlet, R. M. Doherty, G. D. Veith, R. W. Taft, M. H. Abraham, *Environ. Sci. Technol.* **1986**, *20*, 690.

14 J. A. Castillo-Garit, Y. Marrero-Ponce, J. Escobar, F. Torrens, R. Rotondo, *Chemosphere* **2008**, *73*, 415.

15 R. Todeschini, V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, Germany, **2000**.

16 *HyperChem*, Release 7.0 for Windows, Hypercube, Inc., **2002**.

17 J. J. P. Stewart, *J. Comput.-Aided Mater. Des.* **1990**, *4*, 1.

18 A. R. Katritzky, V. S. Lobanov, M. Karelson, *Comprehensive Descriptors for Structural and Statistical Analysis*, Reference Manual, Version 2, **1994**.

19 A. R. Katritzky, V. S. Lobanov, M. Karelson, *Chem. Soc. Rev.* **1995**, *24*, 279.

20 A. R. Katritzky, M. Karelson, V. S. Lobanov, *Pure Appl. Chem.* **1997**, *69*, 245.

21 J. Li, H. Liu, X. Yao, M. Liu, Z. Hu, B. Fan, *Chemom. Intell. Lab. Syst.* **2007**, *87*, 139.

22 A. R. Katritzky, O. V. Kulshyn, I. Stoyanova-Slavova, D. A. Dobchev, M. Kuanar, D. C. Faraa, M. Karelson, *Bioorg. Med. Chem.* **2006**, *14*, 2333.

23 M. H. Fatemi, F. Karimian, *J. Colloid Interface Sci.* **2007**, *314*, 665.

24 S. Haykin, *Neural Network*, Prentice-Hall, Englewood Cliffs NJ, **1994**, pp. 145–187.

25 M. T. Beal, H. B. Hagan, M. Demuth, *Neural Network Design*, PWS, Boston MA, **1996**, pp. 75–92.

26 N. K. Bose, P. Liang, *Neural Network Fundamentals*, McGraw-Hill, New York, **1996**, pp. 241–250.

27 M. Jalali-Heravi, M. H. Fatemi, *Anal. Chim. Acta* **2000**, *415*, 95.

28 M. Jalali-Heravi, M. H. Fatemi, *J. Chromatogr., A* **2001**, *915*, 177.

29 M. Jalali-Heravi, M. H. Fatemi, *J. Chromatogr., A* **1998**, *825*, 161.

30 M. Jalali-Heravi, M. H. Fatemi, *J. Chromatogr., A* **2000**, *897*, 227.

31 M. H. Fatemi, *J. Chromatogr., A* **2002**, *955*, 273.

32 J. Havel, M. Breadmore, M. Macka, P. R. Haddad, *J. Chromatogr., A* **1999**, *850*, 345.

33 L. M. A. Monzón, L. M. Yudi, *J. Electroanal. Chem.* **2006**, *591*, 46.

34 A. G. Maldonado, J. P. Doucet, M. Petitjean, B. T. Fan, *Mol. Diversity* **2006**, *10*, 39.

35 W. Wu, B. Walczak, D. L. Massart, S. Heuerding, F. Erni, I. R. Last, K. A. Prebble, *Chemom. Intell. Lab. Syst.* **1996**, *33*, 35.

36 B. Efron, *J. Am. Stat. Assoc.* **1983**, *78*, 316.

37 D. W. Osten, *J. Chemometr.* **1988**, *2*, 39.

38 S. Wold, L. Eriksson, *Statistical Validation of QSAR Results in Chemometrics Methods in Molecular Design*, ed. by H. Van de Waterbeemd, Wiley, VCH, Weinheim, **1995**.