Journal of Physical and Theoretical Chemistry
of Islamic Azad University of Iran, 13 (2) 209-215: Summer 2016
(J. Phys. Theor. Chem. IAU Iran)
ISSN 1735-2126

Application of topological and physicochemical descriptors: QSTR analysis of the toxicity of benzene derivatives

H. Hosseini and F.Shafiei*

Department of Chemistry, Faculty of Science, Arak Branch, Islamic Azad University, Arak, Iran

Received February 2016; Accepted August 2016

ABSTRACT

Quantitative Structure-Activity Relationship (QSAR) models are useful in understanding how chemical structure relates to the biological activity and the toxicity of natural and synthetic chemicals. In the present investigation the applicability of various topological indices and physicochemical descriptor are tested for the QSAR study on benzene derivatives. The topological indices used for the QSAR analysis were Szeged (Sz), Randic (^{1}X) (the first order molecular connectivity), Balaban (J), HyperWiener (HW), Wiener Polarity (WP) and Harary (H) indices. The physicochemical descriptor is also used in the study (n-octanol/water partition coefficient (logP). For obtaining appropriate QSTR model we have used multiple linear regression (MLR) techniques and followed back ward regression analysis. The results have shown that best models are obtained by multi parametric analysis. The toxicities of 45 benzene derivatives are well predicted by a tri parametric model consisting of HyperWiener (HW), Wiener Polarity (WP) and partition coefficient (logP) as the correlating parameters. The predictive ability of the model is discussed on the basis of predictive correlation coefficient.

Keywords: QSAR; Topological indices; benzene derivatives; physicochemicald escriptors

1. INTRODUCTION

Benzene derivatives compounds have been used for many years in the chemical and pharmaceutical industry as food additives, gasoline additives, solvent, medicine and so on.

Benzene, one of the industrial effluents that have great hazard and extensive pollution scope in the world, has brought about serious pollution against water bodies in the environment and caused great hazard against human.

The health risks associated with exposure to benzene have been known for many years. The compound has both chronic and acute effects whether ingested by mouth, taken in through the respiratory system, or absorbed through the skin. Acute effects resulting from inhalation include irritation of the mucous membrane, headache, instability, euphoria, convulsions, excitement or depression, and unconsciousness.

The ingestion of benzene has been associated with the development of bronchitis and pneumonia, while exposure through the skin can cause drying,

*Corresponding author: f-shafiei@ia-arak.ac.ir

blistering, and erythema (redness). Death can result from exposure to high concentrations of benzene.

Quantitative structure – activity relationships (QSAR) have provided a valuable approach in research into the toxicityof organic chemicals[1, 2]. Many investigators have used the 1-octanol/water partition coefficient (logP) dependent QSAR as a basis for predicting the toxicity.Recently, some QSAR studies applying theoretical approaches to predict the toxicity have been reported [3-7].

A QSAR analysis has been carried out on the toxicities of 40 mono-substituted nitro benzene [8].

Topological sub-structural molecular was used to assess acute aquatic toxicity of a series of 69 benzene derivatives[9].

A quantitative structure-toxicity model correlating toxic potency [1/log(IGC(50)] with hydrophobicity quantified by the 1-octanol/water partition coefficient (logP) and electrophilic reactivity quantified by the molecular orbital parameter, maximum superdelocalizability (S(max)), was developed[10].

In the present study, we have made QSAR models on benzene derivative compounds were based on topological indices(TI) and physicochemical descriptor (logP).The multiple linear regression (MLR) techniques and back ward methods are used for modeling the toxicity of 45 benzene derivatives.

The topological indices used for the QSAR analysis were WienerPolarity [11], Szeged [12], first order molecular connectivity [13], Balaban [14], HyperWiener [15]and Harary [16] indices.

The physicochemical descriptor used for the QSAR analysis was the partition coefficients (logP).

The main objective of this work is to obtain QSAR models that could be used to predict toxicity of benzene derivative compounds.

MATERIALS AND METHODS

A data set of 45 benzene derivatives was selected, due to the toxicological action reported in the literature consulted [17, 18].The partition coefficients (logP)of these benzene derivatives are taken fromchemicalize software [19].

A set of benzene derivatives used in the present investigation are recorded in Table 1, along with their toxicities. Toxicities are expressed as the negative logarithm of the lethal concentration of a benzene derivative and noted by $-\log (LC_{50})$.

Topological indices

All the used topological indices were calculated using all hydrogen suppressed graph by deleting all the carbon hydrogen as well as heteroatomic hydrogen bonds from the structure of the benzene derivatives. The descriptors were calculated with chemicalize software [19]. Six topological indices tested in the present study are recorded in Table 2.

Statisticalanalysis

Structure- Property models (MLR models) aregenerated using the multilinear regression procedure of SPSS version 16. The toxicity is used as the dependent variable and topological indices and logPas the independent variables. The statistical significance of the screened models was judged by the multiple correlation coefficients (R), standard error of estimation (Se), adjusted R-squared (R^2_{adj}), the F value (Fischer statistic) and the sig (significant).

Table 1. Comparison between predictedand observed values of toxicity (−log [LC_{50}]) of respect benzene derivatives

compounds	Comp. No.	−log [LC_{50}]	compounds	Comp. No.	−log [LC_{50}]
Bromobenzene	1	3.86	2,4-Dichlorotoluene	24	4.54
Phenol	2	3.51	Chlorobenzene	25	3.77
1,2-Dichlorobenzene	3	4.40	1,3,5-Trinitrobenzene	26	5.29
3-Chlorotoluene	4	4.30	1,2,3,4-Tetrachlorobenzene	27	5.43
1,3-Dihydroxybenzene	5	3.04	2,3,4,5,6-Pentachlorophenol	28	6.06
3-Hydroxyanisole	6	3.21	1,3-Dichlorobenzene	29	4.30
4-Methyl-3-nitroaniline	7	3.77	2-Chlorophenol	30	4.02
2,4-Dimethylphenol	8	3.86	3-Methylphenol	31	3.29
2,6-Dimethylphenol	9	3.75	2,3-Dinitrotoluene	32	5.01
3-Nitrotoluene	10	3.63	1,4-Dimethylbenzene	33	4.21
2,6-Dinitrotoluene	11	3.99	2,3,4,5-Tetrachlorophenol	34	5.72
4-Methyl-2,6-dinitroaniline	12	4.21	2,3,6-Trinitrotoluene	35	6.37
5-Methyl-2,6-dinitroaniline	13	4.18	4-Methylphenol	36	3.58
5-Methyl-2,4-dinitroaniline	14	4.92	4-Methyl-3,5-dinitroaniline	37	4.46
2,4-Dinitrotoluene	15	3.75	1,3,5-Trichlorobenzene	38	4.74
4-Nitrophenol	16	3.36	Benzene	39	3.40
4-Chlorotoluene	17	4.33	2-Nitrotoluene	40	3.57
2,4,6-Trichlorophenol	18	4.33	1,4-Dinitrobenzene	41	5.22
Toluene	19	3.32	2-Methyl-3,6-dinitroaniline	42	5.34
3-Methyl-6-nitroaniline	20	3.80	2-Methyl-4,6-dinitrophenol	43	5.00
4-Methyl-2-nitroaniline	21	3.79	2,5-Dinitrotolueno	44	5.15
1,2,4-Trichlorobenzene	22	5.00	1,2-Dinitrobenzene	45	5.45
3,4-Dichlorotoluene	23	4.74			

Table2: Benzene derivatives and their topological indices,partition coefficientsused in present study

Comp. No.	$^1\chi$	J	H	HW	WP	Sz	logP
1	3.39	1.82	12.92	71	5	78	2.74
2	3.39	1.82	12.92	71	5	78	1.67
3	3.8	2.28	16.17	106	8	106	3.18
4	3.79	2.23	16.08	110	7	108	3.09
5	3.79	2.23	16.08	110	7	108	1.37
6	4.33	1.98	19.15	176	9	146	1.51
7	5.11	2.25	26.67	315	14	232	1.60
8	4.2	2.09	19.53	160	10	144	2.70
9	4.22	2.15	19.67	151	11	140	2.70
10	4.7	2.32	22.72	245	11	186	2.43
11	6.04	2.4	34.6	545	19	348	2.37
12	6.43	2.7	39.2	669	31	420	2.84
13	6.45	2.72	39.13	667	22	418	2.84
14	6.43	2.65	38.83	698	21	430	2.19
15	6.02	2.33	34.3	576	18	360	2.37
16	4.7	2.26	22.6	262	11	192	1.61
17	3.79	2.19	16.03	115	7	110	3.09
18	4.61	2.49	23.28	215	13	184	3.48
19	3.39	1.82	12.92	71	5	78	2.49
20	5.11	2.22	26.6	327	14	236	2.25
21	5.11	2.27	26.67	315	14	232	2.25

Continued Table 2

22	4.2	2.09	19.53	160	10	144	3.79
23	4.2	2.09	19.53	160	10	144	3.69
24	4.2	2.09	19.53	160	10	144	3.69
25	3.39	1.82	12.92	71	5	78	2.58
26	6.91	2.46	42.6	906	21	516	1.79
27	4.63	2.52	23.37	211	14	182	4.39
28	5.46	2.76	31.6	357	21	282	4.69
29	3.79	2.23	16.08	110	7	108	3.18
30	3.8	2.28	6.17	106	8	106	2.27
31	3.79	2.23	16.08	110	7	108	2.18
32	6.04	2.47	34.83	511	19	336	2.37
33	3.79	2.19	16.03	115	7	110	3.00
34	5.04	2.39	27.32	281	17	230	4.09
35	7.36	2.83	47.97	1036	26	588	2.31
36	2.18	2.19	16.3	115	7	110	2.18
37	6.43	2.7	39.02	669	21	420	1.54
38	4.18	2.08	19.5	159	9	144	3.79
39	3	2	10	42	3	54	1.97
40	4.72	2.4	22.9	231	12	180	2.43
41	5.61	2.3	29.74	521	15	314	1.85
42	6.45	2.64	38.87	717	22	434	2.19
43	6.43	2.66	38.85	691	21	428	2.06
44	6.02	2.28	34.14	616	18	372	2.37
45	5.63	2.54	30.43	416	16	278	1.85

RESULTS

Several linear QSAR models involving one-seven descriptors are established and strongest multivariable correlations are identified by the back ward method are significant at the 0.05 level and regression analysis of the SPSS program.

In the first of this study we drown scattering plots of toxicity versus the six topological indices, and logP. Some of these plots are given in Fig. (1-4), respectively.

Distribution of the dependent variable against the independent variable for 45 chemicals employed in developing quantitative structure- toxicity relationship.

QSTR models for toxicity

Model 1

$-\log [LC_{50}] = 0.413 + 0.147^{1}\chi + 0.413\ J - 0.019\ H + 0.004\ HW - 0.082\ WP + 0.0001\ Sz + 0.724\ logP$ (1)

N= 45 R= 0.847 R^2 =0.7174

$R^2_{adj} = 0.663$ Se=0.4694 F= 13.372

sig = 0.000

Model 2

$-\log [LC_{50}] = 0.402 + 0.150^{1}\chi + 0.724\ J - 0.019\ H + 0.004\ HW - 0.082WP + 0.724logP$ (2)

N= 45 R= 0.847 R^2 =0.7174

$R^2_{adj} = 0.672$ Se=0.4632 F= 16.022

sig = 0.000

Model 3

$-\log [LC_{50}] = 0.383 + 0.11^{1}\chi + 0.716\ J + 0.003\ HW - 0.089WP + 0.717logP$ (3)

N= 45 R= 0.846 R^2 =0.7157

$R^2_{adj} = 0.678$ Se=0.4586 F= 19.561

sig = 0.000

Model 4

$-\log [LC_{50}] = 0.709+0.715\ J +0.004\ HW - 0.081WP+0.717logP$ (4)

N= 45 R= 0.845 R^2 =0.7140
$R^2_{adj} = 0.685$ Se=0.4542 F= 24.870
sig = 0.000

Model 5

$-\log [LC_{50}] = 1.984+0.004\ HW - 0.056WP+0.734logP$ (5)

N= 45 R= 0.837 R^2 =0.7006
$R^2_{adj} = 0.678$ Se=0.4590 F= 31.861
sig = 0.000

It turns out that the toxicity has a good correlation with all six topological indices and logPas well as with HW, WP and logP (Eq. (5)).

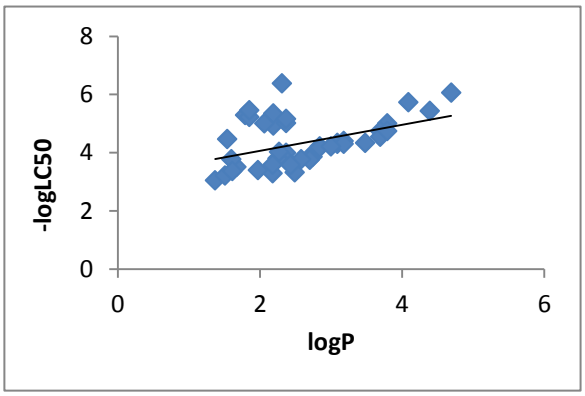

Fig. 1. Plot of the relationship between logP and the observed toxicity.

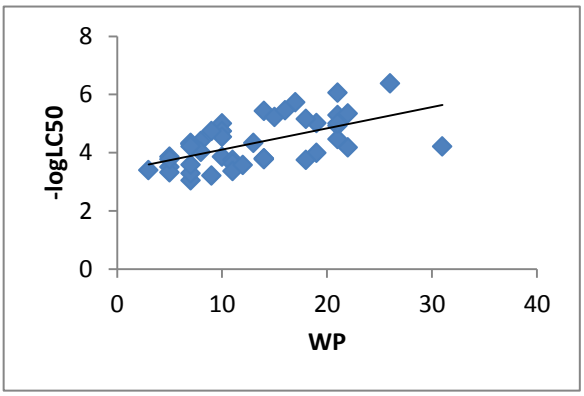

Fig. 2. Plot of the relationship between Wiener Polarity and the observed toxicity.

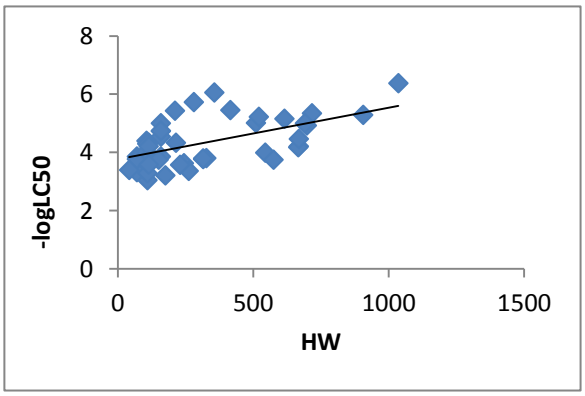

Fig. 3. Plot of the relationship between Hyper Wiener and the observed toxicity.

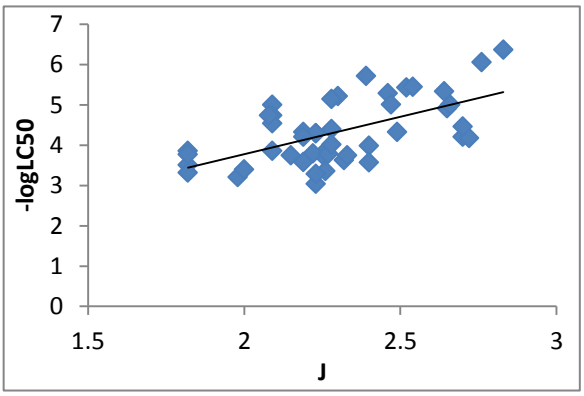

Fig. 4. Plot of the relationship between Balaban index and the observed toxicity.

DISCUSSION

We studied the relationship between topological indices ($^1\chi$,HW,WP,Sz,H and J),the partition coefficient (logP)and the toxicity.

The elaborated QSAR models (Eqs 1 – 5) reveal that thetoxicity of the benzene derivatives could be explained by tri, tetra, penta, hexa and hepta parameters. All of models can explain about 70% of the experimental variance of the dependent variable (−log [LC_{50}]). The R^2_{adj} value of the models is about 0.7, indicating that our modelsare stable and can be efficiently used for estimating the toxicity of other benzene derivatives for which no experimental data are available. The combination of the tri parameters (HW, WP and logP) recorded in Eq (5) has the highest F of Fischer (F = 31.861) andhas the lowest number of parameters which explain that the model (5) for predict toxicity is better than another models.($R^2$and R^2_{adj} all of models is same)

The comparison between the observed data and predicted values using Eq (5) of toxicity is presented in Table 3. The linear relations between the observed and predicted values of the toxicity of 45 benzene derivatives show inFig. (5).

CONCLUSION

The present study shows that physicochemical descriptors expressing the partition coefficient (logP) in combination with the topological indices are useful for the prediction of the toxicity ofbenzene derivatives. The best QSAR model (Eq(5)) is able to describe about 70% of the variance in the experimental toxicity and could be efficiently usedfor estimating the toxicity of other benzene derivatives for which no experimental data are available.

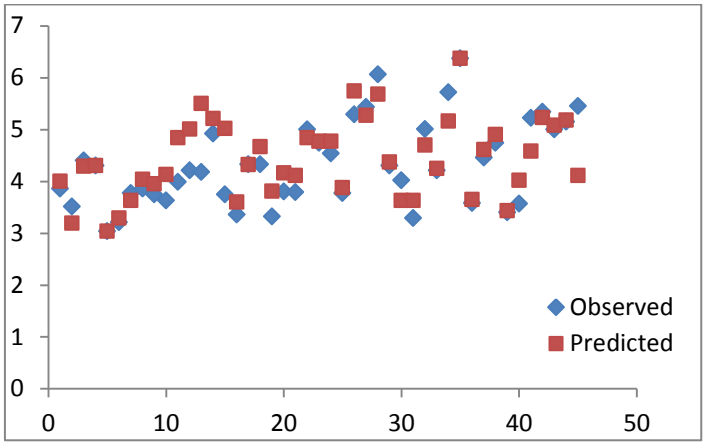

Fig. 5. Comparison between the predicted and observed values of toxicity by MLR.

Table 3. Comparison between predictedand observed values of toxicity (−log [LC_{50}]) of respect benzene derivatives

Comp. No.	Observed −log [LC50]	Predicted −log [LC_{50}]	Residual	Comp. No.	Observed −log [LC_{50}]	Predicted −log [LC_{50}]	Residual
1	3.86	4.00	-0.14	24	4.54	4.77	-0.23
2	3.51	3.19	0.32	25	3.77	3.88	-0.11
3	4.40	4.29	0.11	26	5.29	5.74	-0.45
4	4.30	4.30	0.00	27	5.43	5.27	0.16
5	3.04	3.04	0.00	28	6.06	5.68	0.38

Continued Table 3

6	3.21	3.29	-0.08	29	4.30	4.37	-0.07
7	3.77	3.63	0.14	30	4.02	3.63	0.39
8	3.86	4.04	-0.18	31	3.29	3.63	-0.34
9	3.75	3.95	-0.20	32	5.01	4.70	0.31
10	3.63	4.13	-0.50	33	4.21	4.25	-0.04
11	3.99	4.84	-0.85	34	5.72	5.16	0.56
12	4.21	5.01	-0.80	35	6.37	6.37	0.00
13	4.18	5.50	-1.32	36	3.58	3.65	-0.86
14	4.92	5.21	-0.29	37	4.46	4.61	-0.15
15	3.75	5.02	-1.27	38	4.74	4.90	-0.16
16	3.36	3.60	-0.24	39	3.40	3.43	-0.03
17	4.33	4.32	-0.01	40	3.57	4.02	-0.45
18	4.33	4.67	-0.34	41	5.22	4.58	0.64
19	3.32	3.81	-0.49	42	5.34	5.23	0.11
20	3.80	4.16	-0.36	43	5.00	5.08	-0.08
21	3.79	4.11	-0.32	44	5.15	5.18	-0.03
22	5.00	4.84	0.16	45	5.45	4.11	1.34
23	4.74	4.77	-0.03				

REFERENCES

[1] S. A. Senior, M. D. Madbouly and A.M. EIMassry, Chemosphere. 85(1) (2011) 7.

[2] A. Can, Toxicol. Lett. 230 (2014) 434.

[3] J. Q. Shi, J. Cheng and F. Y. Wang, Ecotoxicol. Environ. Saf. 78 (2012) 134.

[4] G. Jing, Z. Zhou and J. Zhuo, Chemosphere. 86 (2012) 76.

[5] AM. Richard, Chem. Res. Toxicol. 19 (10) (2006) 1257.

[6] DV. Eldred, PC. Jurs, SAR & QSAR Environ. Res. 10 (1999) 75.

[7] A. A. Toropov, B. F. Rasulev and J. Leszczynski, QSAR Comb. Sci. 26 (5) (2007) 686.

[8] V. K. Agrawal, P.V. Khadikar, Bioorg. Med. Chem. 9 (2001) 3035.

[9] M. P. Gonzales, A. M. Helguera and M. A. Cabrera, Bioorg. Med. Chem. 13 (2005) 1775.

[10] T. W. Schultz, Chem. Res. Toxicol. 12 (1999) 1262.

[11] A. Behmaram, H. Yousefi- Azari, A. R. Ashrafi, Appl. Math. Lett. 25 (2012) 1510.

[12] P. V. Kadikar, N. V. Deshpande and G. Domotor, J. Chem. Inf. Compt. Sci, 35 (1995) 547.

[13] M. Randic', J. Am. Chem. Soc. 97 (1975) 6609.

[14] A. T. Balaban, Chem. Phys. Lett., 89 (1982) 399.

[15] I. Gutman, Croat. Chem. Acta, 77 (1-2) (2004) 61.

[16] C. K. Das, B. Zhou and N. Trinajstic, J. Math.Chem. (2009) 1369.

[17] A. A. Toropov, A.P. Toropova, J. Mol. Struct. (Theochem) 578(2002) 129.

[18] C. Bertinetto, C. Duce and M. R. Tine, Math. Comput.Chem. 70 (2013) 1005.

[19] Web search engine developed by ChemAxon; software available at http:// WWW. Chemicalize. Org.