Research Article
CODEN (USA): IJPLCP

[Sachan et al., 3(5): May, 2012]
ISSN: 0976-7126

INTERNATIONAL JOURNAL OF PHARMACY & LIFE SCIENCES

QSAR analysis of soil sorption coefficients for polar organic chemicals: substituted anilnes & phenols

Madhu Mishra[2], Shailja Sachan[1*], R.S. Nigam[3] and Vikash Pandey [2]

1, Department of Chemistry, Govt. New Science College, Rewa, (M.P.) - India
2, Department of Chemistry, A.P.S.University, Rewa, (M.P.) - India
3, Department of Chemistry, Rajiv Gandhi College, Sherganj, Satna, (M.P.) - India

Abstract

Based on various topological molecular descriptors, viz. Weiner index, various Randic indices, different molecular connectivity indices, Balaban and different balaban type parameters, several QSAR models were built to estimate the soil sorption coefficients (log K_{oc}) of substituted anilines and phenols. The regression analysis of the data employing the multiple linear analysis. Results showed that a tetra-parametric model was excellent for modeling of these compounds.

Key-Words: QSAR, Balaban, Weiner index

Introduction

The soil coefficient K_{oc}, which determines the partitioning of an organic chemical between the soil sediment and aqueous solution, is an important environmental parameter. K_{oc} is the ratio between the concentration of a chemical adsorbed by the soil normalized to soil organic carbon and those dissolved in the soil water. Thus, K_{oc} is a frequently used parameter to indicate the physical movements of pollutants, chemical degradation. and biodegradation activity of a given species in environment[1-5], and it is of great use for the environmental risk assessment of organic chemicals. For some chemical species, quantitative structure-activity relationship (QSAR) modeling is a useful technique to correlate their physical, chemical, biological or environmental activities to their physicochemical property descriptors. Because the experimental determination is time-consuming and expensive, estimated values based on QSAR models are now widely used.

Nowadays, many QSAR models have been developed to predict the soil sorption coefficients of organic chemicals[6-9]. In these works, satisfactory results were reported for non-polar chemicals, but for polar chemicals such as aniline-type chemicals, phenol-type chemicals, alcohols, organic acid and etc., results usually were poor.

*** Corresponding Author**
E.mail: sachanshailja@gmail.com

This can be explained that, general QSAR models are developed for non-polar chemicals, but for polar chemicals, specific interactions between polar chemicals and appropriate soil/sediment constituents (hydrogen bonding, dipole interactions, charge transfer, and etc.) may exist[10]. This implies that the soil/sediment sorption behaviors of polar and non-polar organic chemicals are different. Thus, in order to improve the estimate quality for these polar chemicals classes, molecular descriptors which reflect other specific interactions should be also included in addition to n-octanol/water partition coefficients. As we know, descriptors derived from quantum chemical computation can clearly describe molecular structure and electronic properties, and these descriptors can be easily obtained. Therefore, the experiences in which quantum chemical descriptors are included in general QSAR models are popular for development of reliable QSAR models[11-13].

The purpose of this study was to systematically investigate the QSAR models of soil sorption coefficients for substituted anilines and phenols based on various topological molecular descriptors viz. Weiner index, various Randic indices, different molecular connectivity indices, Balaban and different balaban type parameters. The other purpose of this work was to try to explain soil sorption mechanism so that it should be possible in the future to obtain more accurate estimates of soil sorption coefficients for polar organic chemicals.

Material and Methods

Calculation of Molecular Descriptors

Experimentally observed soil sorption coefficients log K_{oc} for 42 substituted anilines and phenols were collected by literature[14,15]. E-DRAGON software was used to calculate the molecular descriptors at the basis of a fully optimization of the molecular geometry. In this study, in order to understand the nature of the soil sorption coefficients coefficients for polar organic compounds, we calculate the various molecular descriptors such as Weiner index(W), various Balaban and Balaban type descriptors(J, Jhetz, Jhetp, Jhetv, Jhetm, Jhete), Randic indices($^0\chi$, $^1\chi$, $^2\chi$, $^3\chi$), different connectivity parameters($^0\chi^v$, $^1\chi^v$, $^2\chi^v$, $^3\chi^v$).

Regression Analysis

NCSS 7.0 software was used to perform regression analysis. For the results of regression analysis, model adequacy was measured as the square of correlation coefficient (R^2), the adjusted R^2 for degree of freedom (AR^2), mean square error (M.S.E.), the F-value for analysis of variance (F) and the significance Q-test (Q=R/MSE). Single linear regression analysis was performed based on various topological descriptors, respectively.

In order to improve the quality of QSAR model for polar organic chemicals, more than one descriptor should be used in obtained models, thus a multilinear regression analysis was performed. The correlation coefficients between variables in the model were calculated by also using NCSS 7.0 software.

Presentation of Data

In order to compare the prediction ability of this study, the values of soil sorption coefficient (log K_{oc}) for 42 substituted anilines and phenols collected from the literature[15]. The structure and toxicity which is represented in terms of log K_{oc} were listed in Table 1. The topological molecular parameters Winer index(W),various Balaban type parameters (Jhetv and Jhetp), Randic indices($^2\chi$),and different connectivity descriptors($^0\chi^v$,$^1\chi^v$,$^2\chi^v$) calculated using DRAGON software are listed in Table 2. Correlation matrix is given in Table 3. Various regression equations which are obtained by single and multiple linear regression are presented in Table 4. And after that, actual, predicted and residual values for best model are given in Table 5.

Results and Discussion

Single Linear Regression Analysis

Though single regression analysis, seven regression equations were obtained, but we can find that, for single regression analysis, three equations were satisfactory with R^2 larger than 0.8. These regression equations were listed in Table 4. In these three equations (eqn.1, 2 and 3) highest value of R^2 is obtained with second order connectivity (eqn.3), this means $^2\chi^v$ is the largest correlated descriptors with log K_{oc} than any other descriptors. This correlation is showed in Table 3. And eqn.1 and 2 are less significant because of low values of R^2, AR^2,F-ratio and Q-test.

Multiple Linear Regression Analysis

In order to improve the quality of QSAR model, multilinear regression analysis were performed. As we know, models with variables correlated with each other were of no significance. Successive regression analysis resulted into several binary combinations of $^2\chi^v$ with the Weiner index, Balaban, and connectivity indices used. The best bi-parametric model contained $^0\chi^v$ and $^2\chi^v$ (eqn.6). Here, In all these three models, second order connectivity have positive coefficient, and therefore, with increasing the value of $^2\chi^v$, toxicity also increases. The regression parameters and the quality of model expressed by eqn. 6 ,which indicate that addition of $^0\chi^v$, slightly improves the value of variance (R^2) increases from 0.87 to 0.88. In best tri-parametric equation contains the following independent variables: Jhetv, Jhetp and $^2\chi^v$. In this equation all regression coefficients (except the Jhetv) were positive sign which indicate that with increasing the value of coefficient of Jhetp and $^2\chi^v$, toxicity also increases. In our best tri-parametric model the regression parameters and the quality of model expressed by eqn.7, which indicates that addition of the Balaban type indices significantly improves the correlation coefficient and R^2 increases from 0.88 to 0.92. Also, the quality factor Q increases from 6.2048 to 9.2511.

When Balaban type indices, connectivity parameter and Randic parameter have been tried, a four parametric model is obtained. This model contain one Randic, one connectivity and two Balaban type indices. The adjusted R^2 and value of quality factor are in favour of this combination. A very significant improvement is observed in the variance.

Based on equation 8, we would attempt to explain mechanisms of soil sorption for polar compounds of substituted anilines and phenols. K_{oc} stands for the hydrophobic properties of organic chemicals, compounds with large K_{oc} values will tend to be adsorbed more easily by organic phase than by water phase.

References

1. Hodson J., Williams N.A., The estimation of the adsorption coefficients (K_{oc}) for soils, by high performance liquid chromatography, Chemosphere, 17, (1988), 66-77.

2. Meylan W., Howard P.H., Molecular topology/fragment contribution method for predicting soil sorption coefficients, Environ. Sci. Technol., 26, (1992), 1560-1567.
3. Muller M., Kordel W., Comparision of screening methods for the estimation of adsorption coefficients on soil, Chemosphere, 32, (1996), 2493-2504.
4. Kortvelyesi T., Gorgenyi M., Correlation between retention indices and quantum chemical descriptors of ketones and aldehydes on stationary phases of different polarity, Anal. Chim. Acta., 428, (2001), 1773-1782.
5. Moss G.P., Dearden J.C., Quantitative structure permeability relationship (QSPRs) for percutaneous absorption, Toxicol. in Vitro, 16(3), (2002), 299-317.
6. Sabljic A., Protic M., Relationship between molecular connectivity indices and soil sorption coefficients of polycyclic aromatic hydrocarbons, Bull. Environ. Contam. Toxicol., 28, (1982), 162-165.
7. Sabljic A., Prediction of the nature and strength of soil sorption of organic pollutants by molecular topology, J. Agric. Food. Chem., 32, (1984), 243-246.
8. Abdul A.S., Gibon T.L., Statistical correlations for predicting partition coefficient for nonpolar organic contaminants between aquifer organic carbon and water, Hazardous Waste Hazardous Mater, 4, (1987), 211-222.
9. Bakul H.R., Shyam R.A., QSAR models to predict effect of ionic strength on sorption of chlorinated benzenes and phenols at sediment-water interphase, Water Research, 35(14), (2001), 3391-3401`.
10. Oepen B.V., Kordel W., Sorption of polar and nonpolar compounds to soils. Process, measurements and experience with the applicability of the modified OECD-guideline.106, Chemosphere, 22, (1991), 285-304.
11. Chen J., Peijnenburg W.J.G.M., Quan W., The application of quantum chemical and statistical technique in developing QSPRs for the photo-hydrolysis quantum yields of substituted aromatic halides, Chemosphere, 37, (1998), 1169-1186.
12. Anna K., Mikael H., Multivariate characterization of polycyclic aromatic hydrocarbons using semi-emperical molecule orbital calculations and physical data, Chemosphere, 50(5), (2003), 627-637.
13. Fabiana A.L.R., Marcia M.C.F., QSPR models of boiling point, octanol-water partition coefficient and retention time index of polycyclic aromatic hydrocarbons, J. Mol. Struct. : THEOCHEM, 663(1-3), (2003), 109-126.
14. Sabljic A., Gusten H., QSAR modeling of soil sorption. Improvements and systematic of log Koc vs log Kow correlations, Chemosphere, 31, (1995), 4489-4514.
15. Liu G., YU J., QSAR analysis of soil sorption coefficients for polar organic chemicals : substituted anilines and phenols, Water Research, 39,(2005), 2048-2055.

Table 1: Structure and toxicity of 42 aniline and phenol derivative

Comp. No.	Chemical name	Log K_{oc}
1.	PHENOL	1.43
2.	2,3-DICHLOROPHENOL	2.65
3.	2,4-DICHLOROPHENOL	2.75
4.	2,4,6-TRICHLOROPHENOL	3.02
5.	2,4,5-TRICHLOROPHENOL	3.36
6.	3,4,5-TRICHLOROPHENOL	3.56
7.	2,3,4,6-TETRACHLOROPHENOL	3.35
8.	PENTACHLOROPHENOL	3.73
9.	4-BROMOPHENOL	2.41
10.	4-NITROPHENOL	2.37
11.	2-CHLOROPHENOL	2.60
12.	3-CHLOROPHENOL	2.54
13.	3,4-DICHLOROPHENOL	3.09
14.	3,5-DIMETHYLPHENOL	2.83
15.	2,3,5-TRIMETHYLPHENOL	3.61
16.	4-METHYLPHENOL	2.70

17.	2-METHOXYPHENOL	1.56
18.	3-METHOXYPHENOL	1.50
19.	3-HYDROXYPHENOL	0.98
20.	4,5,6-TRICHLOROGUAIACOL	2.80
21.	TETRACHLOROGUAIACOL	2.85
22.	CATECHOL	2.03
23.	ANILINE	1.41
24.	3-METHYLANILINE	1.65
25.	4-METHYLANILINE	1.90
26.	4-CHLOROANILINE	1.96
27.	4-BROMOANILINE	1.96
28.	3-TRIFLUOROMETHYLANILINE	2.36
29.	3-CHLORO-4-METHOXYANILINE	1.93
30.	3-METHYL-4-BROMOANILINE	2.26
31.	2.4-DICHLOROANILINE	2.72
32.	2.6-DICHLOROANILINE	3.25
33.	3,5-DICHLOROANILINE	2.11
34.	3,4-DICHLOROANILINE	2.29
35.	2,3,4-TRICHLOROANILINE	2.60
36.	2, 3,4,5- TETRACHLOROANILINE	3.03
37.	2, 3,5,6- TETRACHLOROANILINE	3.94
38.	PENTACHLOROANILINE	4.62
39.	3,5-DINITROANILINE	2.55
40.	N-METHYLANILINE	2.28
41.	N,N-DIMETHYLANILINE	2.26
42.	DIPHENYLAMINE	2.78

Table 2: Calculated parameter of 42 substituted aniline and phenol

Comp.No.	W	Jhetv	Jhetp	$^{2}\chi$	$^{0}\chi^{V}$	$^{1}\chi^{v}$	$^{2}\chi^{v}$
1.	42	2.651	2.571	2.743	3.834	2.134	1.336
2.	82	2.971	3.018	3.745	5.947	3.102	2.358
3.	84	2.909	2.952	3.873	5.947	3.096	2.439
4.	110	3.072	3.160	4.390	7.004	3.579	2.969
5.	111	3.05	3.136	4.381	7.004	3.579	2.939
6.	110	3.076	3.167	4.390	7.004	3.579	2.916
7.	140	3.238	3.366	4.768	8.060	4.069	3.387
8.	174	3.402	3.565	5.155	9.117	4.558	3.808
9.	62	2.822	2.799	3.365	5.721	3.027	2.392
10.	120	2.248	2.060	4.264	5.020	2.634	1.774
11.	60	2.816	2.806	3.239	4.891	2.618	1.859
12.	61	2.773	2.763	3.377	4.891	2.612	1.916
13.	84	2.911	2.956	3.873	5.947	3.096	2.413
14.	84	2.904	2.845	4.023	5.679	2.956	2.346
15.	110	3.072	3.018	4.39	6.602	3.378	2.726
16.	62	2.738	2.673	3.365	4.757	2.545	1.836
17.	86	2.223	2.086	3.43	5.165	2.663	1.672
18.	88	2.19	2.057	3.546	5.165	2.657	1.701
19.	61	2.519	2.408	3.377	4.204	2.269	1.520
20.	182	2.766	2.723	4.959	8.334	4.114	3.200
21.	220	2.963	2.948	5.368	9.391	4.604	3.650
22.	60	2.551	2.437	3.239	4.204	2.275	1.489

23.	42	2.836	2.775	2.743	3.964	2.199	1.411
24.	61	2.927	2.877	3.377	4.887	2.610	1.914
25.	62	2.888	2.839	3.365	4.887	2.610	1.911
26.	62	2.888	2.901	3.365	5.021	2.677	1.988
27,	62	2.981	2.98	3.365	5.851	3.092	2.467
28.	148	2.351	2.123	5.335	5.521	2.927	2.123
29.	116	2.456	2.361	4.064	6.352	3.206	2.301
30.	84	3.128	3.128	3.873	6.773	3.509	2.863
31.	84	3.043	3.112	3.873	6.077	3.161	2.509
32.	82	3.11	3.183	3.745	6.077	3.167	2.45o
33.	84	3.038	3.107	4.023	6.077	3.155	2.576
34.	84	3.045	3.115	3.873	6.077	3.161	2.488
35.	109	3.225	3.345	4.250	7.134	3.650	2.928
36.	140	3.351	3.507	4.768	8.190	4.134	3.431
37.	140	3.349	3.503	4.768	8.190	4.134	3.452
38.	174	3.506	3.696	5.155	9.247	4.623	3.873
39.	240	2.373	2.143	5.82	6.337	3.198	2.297
40.	64	2.478	2.371	2.912	4.887	2.661	1.616
41.	88	2.393	2.262	3.642	5.834	3.029	2.23
42.	264	1.815	1.714	5.244	7.274	4.321	2.857

Table 3: Correlation matrix

	$\log K_{oc}$	W	Jhetv	Jhetp	$^{2}\chi$	$^{0}\chi^{v}$	$^{1}\chi^{v}$	$^{2}\chi$
$\log K_{oc}$	1	0.5502	0.6361	0.6828	0.6781	0.8945	0.8984	0.9350
W		1	-0.0810	-0.0578	0.9277	0.7431	0.7752	0.6528
Jhetv			1	0.9941	0.1464	0.5122	0.4584	0.6275
Jhetp				1	0.1569	0.5424	0.4967	0.6570
$^{2}\chi$					1	0.8094	0.8119	0.7629
$^{0}\chi^{v}$						1	0.9886	0.9798
$^{1}\chi^{v}$							1	0.9724
$^{2}\chi^{v}$								1

Table 4: Regression equation (N=42)

Regression Equations	Equations	R^2	AR^2	MSE	F-RATIO	Q-value
$\log K_{oc}$ =-1.6463+0.7032(0.0556)$^{0}\chi^{v}$	1	0.8001	0.7952	0.2592	160.148	3.4509
$\log K_{oc}$ =-2.1222+1.5017(0.1160)$^{1}\chi^{v}$	2	0.8073	0.8025	0.2499	167.550	3.5954
$\log K_{oc}$ =-1.1276+1.5637(0.0938)$^{2}\chi^{v}$	3	0.8742	0.8711	0.1631	278.067	5.7325
$\log K_{oc}$ =-1.1710-0.0023(0.0016)W+0.6781(0.1222)$^{2}\chi^{v}$	4	0.8805	0.8744	0.1589	143.745	5.9049
$\log K_{oc}$ =-1.6296+0.2921(0.1766)Jhetp+1.4313(0.1218)$^{2}\chi^{v}$	5	0.8825	0.8765	0.1563	146.436	6.0100
$\log K_{oc}$ =-0.6691-0.4258(0.2127)$^{0}\chi^{v}$+2.4512(0.4525)$^{2}\chi^{v}$	6	0.8860	0.8801	0.1517	151.492	6.2048
$\log K_{oc}$=2.0683-6.1021(1.3431)Jhetv+5.2394(1.0984)Jhetp+1.2807(0.1047)$^{2}\chi^{v}$	7	0.9239	0.9178	0.1039	153.672	9.2511
$\log K_{oc}$ =1.6004-7.7176(1.3781)Jhetv+6.9497(1.1959)Jhetp+0.3995(0.1472)$^{2}\chi$+0.7115(0.2310)$^{2}\chi^{v}$	8	0.9365	0.9296	0.0890	136.412	10.8671

Table 5: Observed (Obs.), Predicted (Pre.) Residual Value Obtained using Equation 08

Row	Actual	Predicted	Residual
1	1.460	1.055	0.405
2	2.840	2.819	0.021
3	3.060	2.948	0.112
4	3.690	3.719	-0.029
5	3.720	3.697	0.023
6	4.140	3.699	0.441
7	4.450	4.318	0.132
8	5.120	4.890	0.230
9	2.590	2.320	0.270
10	1.910	1.533	0.377
11	2.150	1.985	0.165
12	2.500	2.114	0.386
13	3.330	2.942	0.388
14	2.350	2.237	0.113
15	2.920	2.560	0.360
16	1.940	1.697	0.243
17	1.320	1.501	-0.181
18	1.580	1.621	-0.041
19	0.800	1.325	-0.525
20	3.190	3.436	-0.246
21	3.630	3.962	-0.332
22	0.950	1.203	-0.253
23	0.940	1.098	-0.158
24	1.400	1.716	-0.316
25	1.390	1.746	-0.356
26	1.880	2.232	-0.352
27	2.260	2.404	-0.144
28	2.290	1.852	0.438
29	1.850	2.315	-0.465
30	2.530	2.783	-0.253
31	2.910	3.076	-0.166
32	2.820	2.959	-0.139
33	2.900	3.187	-0.287
34	2.690	3.066	-0.376
35	3.680	3.739	-0.059
36	4.570	4.457	0.113
37	4.460	4.460	0.000
38	5.080	5.044	0.036
39	1.890	2.139	-0.249
40	1.660	1.267	0.393
41	2.310	1.894	0.416
42	3.5	3.633	-0.133

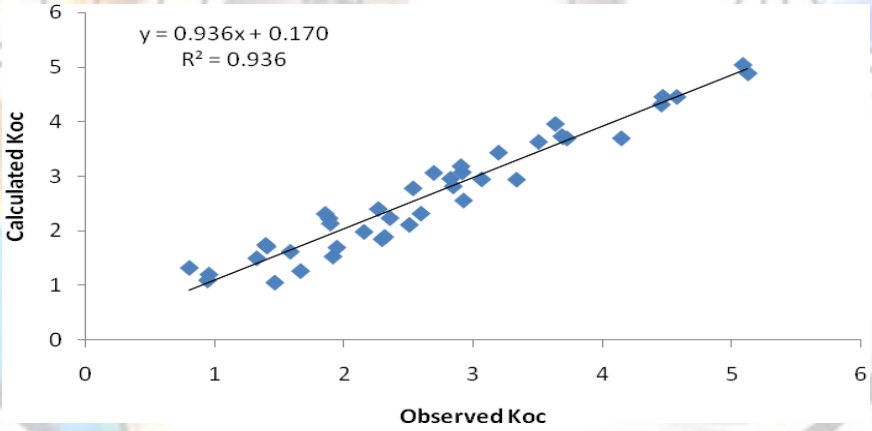

Fig 1: Graph plotted between observed and calculated activity