Research Article [Sachan *et al.*, 3(5): May, 2012]
CODEN (USA): IJPLCP **ISSN: 0976-7126**

INTERNATIONAL JOURNAL OF PHARMACY & LIFE SCIENCES

Reliable QSAR for estimating K_{oc} for persistent organic pollutants

Madhu Mishra[1], Shailja Sachan[2], Vikash Pandey[1] and Santosh Tiwari[1]
1, Department of Chemistry, A.P.S.University, Rewa, (M.P.) - India
2, Department of Chemistry, Govt. New Science College, Rewa, (M.P.) - India

Abstract

Several recent studies have shown that n-octanol/water partition coefficients may not be a good predictor for estimating soil sorption coefficients of persistent organic pollutants (POPs), defined here as chemicals with log K_{ow} greater than 5. Thus, an alternative QSAR model was developed that seems to provide reliable estimates for the soil sorption coefficients of persistent organic pollutants (POPs). This model is based on a set of calculated topological molecular descriptors viz. W, J, $^1\chi$, $^2\chi$, Jhetz, Jhetm, Jhete, Jhetv, and Jhetp and evaluated soil sorption data for 32 POPs. The regression analysis gave a penta-parametric correlation involving W, J, Jhetm, Jhetv and $^1\chi$ as the most significant QSAR model for modeling of POPs. The developed QSAR model was rationalized in terms of potential hydrophobic interactions between persistent organic pollutants and soil organic matrix.

Key-Words: Soil Sorption, Persistent, Bioaccumulation, Hydrophobic interaction, QSAR

Introduction

Persistent organic pollutants (POPs) are chemicals that persist for a very long time in the environment and consequently may concentrate to a high level (10^6) in the food chain. They may also cause toxic effects on animal and human reproductions, development and immunological function. Several years ago under the auspices of the UN Environment Programme (UNEP) stated that POPs were "one of the great environmental challenges the world faces[1-3]". Subsequently, in July 1998, officials from 92 countries agreed to take action on 12 POPs by the year 2000 and acknowledged that the number of chemicals that could be described as POPs may be very large. Typical examples of these chemicals are polychlorinated biphenyls (PCBs), polychlorinated dibenzo-p-dioxins and furans, polycyclic aromatic hydrocarbons (PAHs), and pesticides such as DDT, chlordane, and heptachlor.

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

POPs have a wide range of Industrial, anthropogenic and agricultural applications. They include pesticides such as DDT (dichlorodiphenyltrichloroethane) and lindane, in addition to petroleum hydrocarbons which are organic chemicals composed of fused benzene rings formed during incomplete combustion of coal, oil, petrol and wood[4]. POPs have the ability to volatilize and travel great distances through the atmosphere to become deposited in remote regions.

The chemicals also have the ability to bioaccumulate and biomagnify, and can bioconcentrate (i.e. become more concentrated) up to 70,000 times their original concentrations5. POPs may continue to poison non-target organisms in the environment and increase risk to humans by disruption in the endocrine, reproductive, and immune systems, cancer, neurobehavioral disorders, infertility and mutagenic effects, although very little is currently known about these chronic effects. Some POPs have been banned, while others continue to be used[6].

Polychlorinated biphenyls (PCB) formerly widely used industrially as dielectric fluids in transformers and capacitors, as hydraulic and heat transfer fluids, and as plasticizers, are now of concern as prevalent, persistent, and toxic pollutants[7-9]. PCB produced widespread global contamination of water and soil and bioaccumulated in food chains due to their high hydrophobicity and chemical stability[10,11].

One chemical property that is of particular importance in evaluating the fate and persistence of POPs in the environment is the soil/water partition coefficient normalized to organic carbon (K_{oc}), since measured K_{oc} data are not available for majority of those chemicals, numerous correlations have been developed relating K_{oc} to other structural descriptors. This has enabled simple and fast estimation of K_{oc} for POPs.

The aim of this study is to develop a reliable QSAR model for estimating K_{oc} of POPs. Several molecular descriptors will be used in correlations with log K_{oc} data because they have been proven very successful in correlations with various classes of POPs, i.e. chlorinated benzenes, PCBs, polycyclic aromatic hydrocarbons (PAHs), chlorinated dibenzodioxins[12-23]. For the training set we have used 32 chemicals (pollutants) with the evaluated K_{oc} values from a high quality database described earlier[24-26].

Material and Methods

The structure of 32 persistent organic pollutants together with their log K_{oc} data are given in table 1.

Table 1: The list of compound studied and their K_{oc} values

Comp. No.	Compound's name	Log Koc
01.	BENZANTHRACENE	5.62
02.	BENZOPYRENRE	6.64
03.	CHLORDANE	5.36
04.	4,4' - DDT	4.67
05.	1,2;5,6-DIBENZANRHRACENE	5.94
06.	FLUORANTHENE	4.88
07.	HEXACHLOROBENZENE	4.31
08.	METHOXYCHLOR	4.90
09.	2,2',3,4,4',5'-HEXACHLORO BIPHENYL	5.93
10.	2,2',3,4',5',6- HEXACHLORO BIPHENYL	5.79
11.	2,2',4,4',5,5'-- HEXACHLORO BIPHENYL	5.86
12.	2,2',5,5'- TETRACHLORO BIPHENYL	5.41
13.	2,2',3,4,5'- PENTACHLORO BIPHENYL	5.73
14.	2,2',3,5',6- PENTACHLORO BIPHENYL	5.55
15.	2,2',3,4',5'- ENTACHLOROBIPHENYL	5.69
16.	PENTACHLOROBENZENE	5.59
17.	PENTACHLOROPHENOL	4.52
18.	2,3,7,8-TCDD	6.66
19.	9-METHYLANTHRACENE	4.81
20.	TETRACENE	5.81
21.	7,12-DIMETHYLBENZANTHRACENE	5.37
22.	9-METHYLCHOLANTHRENE	6.10
23.	2,2'-DICHLOROBIPHENYL	3.92
24.	2,4'-DICHLOROBIPHENYL	4.14
25.	2,2',4-TRICHLOROBIPHENYL	4.84
26.	2,2',5-TRICHLOROBIPHENYL	4.23
27.	2,4,4'-TRICHLOROBIPHENYL	4.62
28.	2,2',6,6'-TETRACHLOROBIPHENYL	5.01
29.	2,3',4',5-TETRACHLOROBIPHENYL	4.84
30.	2,2',4,5,5'-PENTACHLOROBIPHENYL	6.00
31.	ALDRIN	4.69
32.	MIREX	4.88

Results and Discussion

QSAR analysis is one of the most effective approaches for optimizing lead compounds and designing new drugs. Excellent QSAR model can aid in understanding the mechanism of the action of drugs and may save the cost and time in the course of developing a new drug when compared with empirical procedure.[27-29]

In the present study, we consider the different topological molecular descriptors. The Randic indices $^{1}\chi$ will be used primarily in correlations with log K_{oc} data because they have been proven very successful in correlations with various classes of POPs, i.e. chlorinated benzene, PCBs, polycyclic aromatic hydrocarbons (PAHs), chlorinated dibenzodioxins. In addition several calculated topological descriptors (i.e. Winner, Randic, Balaban and Balaban type indices) will also be tested in correlation with log K_{oc} data used to validate the predictive potential of developed QSAR model.

A preliminary analysis revealed that out of the original set of 32 persistent organic pollutants (POPs) two compounds (nos.7 and 32) are serious outliers. QSAR studies on the persistent organic pollutants were done by correlated toxicity against with other descriptors. Among several models one "best"

model from POPs was chosen for further analysis based on statistical parameters.

To obtain models one has to calculate molecular descriptors which can be used in modeling the biological toxicity of the compounds. For this purpose various topological parameters W, J, $^1\chi$, $^2\chi$, Jhetz, Jhetm, Jhete, Jhetv, and Jhetp are listed in Table 2. In fact topological indices have been found to model the activity of such compounds effectively. The intercorrelation of the aforementioned topological indices and their relationship with toxicity (K_{oc}) is presented in Table 3. The statistical quality of the regression equations were justified by parameters like percent of explained variance (regression coefficient representd by R^2), adjusted R^2 represented by AR^2, probability factor related to F-ratio (Fisher's constant), Quality factor (Q=R/MSE), mean square error (M.S.E).

A perusal of Table 3 shows that $^1\chi$ is the most suitable molecular descriptor for modeling toxicity of the compounds studied and that other molecular descriptor i.e. topological indices ($^2\chi$ and W) have similar potential. On the other hand all Balaban and Balaban type indices (J, Jhetz, Jhetm, Jhete, Jhetv, and Jhetp) shows negative correlation with toxicity.

The data presented in Table 3 shows that Jhetz and Jhetm highly correlated with each other. A comparatively less colinearity is observed in between J and Jhetz, Jhetv and Jhete, Jhetp and Jhete.

Several multiple regressions where attempt using correlation matrix from the programme and the best results are considered and discussed in developing QSAR modeling of POPs. Any significant mono-parametric correlation is not obtained in this study. Very small values of variance suggest that instead of mono-parametric correlation one should go for multi-parametric modeling. So, multiple regression analysis has been performed on the data set and the obtained significant models and their statistical characteristics are given below in Table 4.

In the above model the change in statistics is observed. The variance of 79% suggests that as compared to the 4-parametric model the present model is much better. These model shows that topological descriptors W, J, Jhetm, Jhetv, and $^1\chi$ are responsible for the modeling of these compounds. However, the rule of thum prevents us in further addition of parameter, and hence it is final model which is obtained deleting compound no. 7 and 32 as outliers.

In order to obtain further support in favour of our result we have also estimated the toxicity of POPs and found that the estimated values are very close to experimental toxicity. The residual, i.e., the difference between the experimental and estimated values confirm these feelings.

The final support in our favour is obtained by plotting the estimated activity against the experimental activity and such a plot showed is the best suitable model for modelling of POPs in the present study.

References

1. Wania F., Mackay D., Tracking the distribution of persistent organic pollutants, *Environ. Sci. Technol. 30*, **(1996)**, 390A-396A.
2. Gramatica P., Consolaro F., QSAR approach to POPs screening for atmospheric persistence, *Chemosphere, 43*, **(2001)**, 655-664.
3. Baker J.R., Mihelcic J. R., *Chemosphere, 41*, **(2000)**, 813-817.
4. Said T.O., Hamed M.A.F., *Egyptian Journal Of Aquatic Research, 32(1)*, **(2006),** 235-245.
5. Meylan W. M., Howard P. H., Improved method for estimating bioconcentration/bioaccumulation factor from octanol/water partition coefficient, *Environ. Toxicol. Chem., 18*, **(1999),** 664–672.
6. Uddameri V., Kuchanur M., *Chemosphere, 54*, **(2004),** 771–776.
7. Baker J.R., Mihelcic J. R.,*Chemosphere, 45*, **(2001)**, 213-221.
8. Chiou C.T., McGroddy S. E., Partition characteristics of polycyclic aromatic hydrocarbons on soils and sediments, *Environ. Sci. Technol., 32*, **(1998),** 264-269.
9. Karickhoff S.W., Brown D.S., Sorption of hydrophobic pollutants on natural sediments, *Water Res., 13*, **(1979)**, 241-248.
10. Sabljic A., Gusten H., Predicting Henry's law constants for the polychlorinated biphenyls, *Chemosphere, 19*, **(1989),** 1503-1511.
11. Sabljic A., Gusten H., Modelling n-octanol/water partition coefficients by molecular topology : chlorinated benzenes and biphenyls, *Environ. Sci. Technol., 27*, **(1993)**, 1394-1402.
12. Sabljic A., Protic M., Molecular connectivity: a novel method for prediction of bioconcentration factors of hazardous chemicals, *Chem.-Biol. Interact, 42*, **(1982)**, 301-310.
13. Sabljic A., Calculation of retention indices by molecular topology: chlorinated benzenes, *J. Chromatogr. 319*, **(1985),** 1-8.
14. Sabljic A., On the prediction of soil sorption coefficients of organic pollutants from

molecular structure, *Environ. Sci. Technol., 21*, **(1987),** 358-366.
15. Bahnick D.A., Doucette W.J., Use of molecular connectivity indices to estimate soil sorption coefficients for organic chemicals, *Chemosphere, 17*, **(1988)**, 1703-1715.
16. Sabljic A., Quantitative modeling of soil sorption for xenobiotic chemicals, *Environ. Health Perspect, 83*, **(1989),** 179-190.
17. Sabljic A., Gusten H., Modelling n-octanol/water partition coefficients by molecular topology : chlorinated benzenes and biphenyls, *Environ. Sci. Technol., 27*, **(1993)**, 1394-1402.
18. Sabljic A., Gusten H., QSAR modeling of soil sorption. Improvements and systematic of log Koc vs log Kow correlations, *Chemosphere, 31*, **(1995)**, 4489-1514.
19. Sabljic A., Gusten H., Predicting Henry's law constants for the polychlorinated biphenyls, *Chemosphere, 19*, **(1989),** 1503-1511.
20. Gerstl Z., Estimation of organic chemical sorption by soils, *J. Con. Hydrology, 6*, **(1990),** 357-375.
21. Gusten H., Sabljic A., Modelling n-octanol/water partition coefficient by molecular topology: polycyclic aromatic hydrocarbons and their alkyl derivatives, *Chemosphere, 23*, **(1991),** 199-213.
22. Meyland W., Howard P.H., Molecular topology/fragment contribution method for predicting soil sorption coefficients, *Environ. Sci. Technol., 26*, **(1992),** 1560-1567.
23. Sekusak S., Sabljic A., Calculation of retention indices by molecular topology: part 3. Chlorinated dibenzodioxins, *J. Chromatogr., 628*, **(1993),** 69-79.
24. Baker J.R., Mihelcic J.R., Evaluation of estimation methods for organic carbon normalized sorption coefficients, *Water Environ. Res., 69*, **(1997),** 136-145.
25. Baker J.R., Mihelcic J. R., *Chemosphere, 41*, **(2000)**, 813-817.
26. Sabljic A., Gusten H., QSAR modeling of soil sorption. Improvements and systematic of log Koc vs log Kow correlations, Chemosphere, 31, **(1995),** 4489-1514.
27. Santos-Filho O.A., Hopfinger A.J., A search for sources of drug residence by the 4D-QSAR analysis of a set of antimalarial dihydrofolate reductase inhibitors, *J. Comput. Aided Mol. Des., 15*, **(2001),** 1-12.
28. Nanbo A., Nanbo T., Mechanistic study on n-demethylation catalysed with P450 by QSAR using electronic properties of 4-substituted N,N-dimethylaniline, *Quant. Struct. Act. Relat., 21*, **(2002),** 613-616.
29. Wang X.D., Lin Z.F., 2D/3D-QSAR comparative study on mutagenicity of nitroaromatics, *Sci. China Ser B-Chem., 48(3)*, **(2005),** 246-252.

Table 2: Calculated Parameters of 32 Persistent Organic Pollutants

Comp.No.	W	J	Jhetz	Jhetm	Jhetv	Jhete	Jhetp	^1X	^2X
1.	553	1.512	2.268	2.268	2.268	2.268	2.268	8.916	7.986
2.	680	1.487	2.231	2.231	2.231	2.231	2.231	9.916	9.128
3.	459	2.203	2.68	2.691	2.264	2.382	2.372	8.114	8.972
4.	484	2.057	2.638	2.642	2.476	2.525	2.521	8.088	7.688
5.	971	1.346	2.019	2.019	2.019	2.019	2.019	10.899	9.891
6.	364	1.677	2.516	2.516	2.516	2.516	2.516	7.949	7.139
7.	174	2.760	4.831	4.868	3.604	3.923	3.894	5.464	5.155
8.	932	1.991	2.595	2.597	2.181	2.519	2.153	9.952	9.270
9.	571	2.062	3.305	3.316	2.900	3.018	3.008	8.414	8.004
10.	555	2.126	3.417	3.428	2.983	3.109	3.098	8.414	8.035
11.	573	2.055	3.293	3.303	2.891	3.008	2.998	8.397	8.134
12.	412	2.024	3.205	3.213	2.880	2.976	2.968	7.575	7.119

13.	486	2.050	3.269	3.278	2.896	3.005	2.996	8.003	7.496
14.	472	2.116	3.383	3.393	2.982	3.099	3.089	8.003	7.528
15.	486	2.049	3.267	3.276	2.894	3.004	2.994	8.003	7.498
16.	140	2.625	4.549	4.581	3.456	3.743	3.718	5.037	4.768
17.	174	2.760	4.608	4.636	3.368	3.888	3.528	5.464	5.155
18.	571	1.688	2.536	2.540	1.546	2.398	1.467	8.542	8.364
19.	326	1.760	2.598	2.598	2.598	2.598	2.598	7.360	6.512
20.	569	1.465	2.197	2.197	2.197	2.197	2.197	8.899	8.072
21.	696	1.621	2.383	2.383	2.383	2.383	2.383	9.771	8.857
22.	804	1.437	2.077	2.077	2.077	2.077	2.077	10.310	9.753
23.	287	1.963	3.047	3.052	2.847	2.907	2.902	6.788	5.851
24.	301	1.867	2.889	2.893	2.715	2.768	2.764	6.771	5.946
25.	352	1.960	3.074	3.080	2.815	2.892	2.886	7.182	6.485
26.	346	1.994	3.130	3.137	2.861	2.942	2.935	7.182	6.485
27.	368	1.872	2.925	2.931	2.696	2.765	2.759	7.165	6.579
28.	394	2.125	3.379	3.388	3.011	3.119	3.110	7.609	6.926
29.	426	1.952	3.081	3.089	2.784	2.872	2.865	7.575	7.087
30.	488	2.041	3.254	3.263	2.885	2.994	2.984	7.986	7.626
31.	448	1.954	2.310	2.317	2.032	2.113	2.106	8.276	9.153
32.	680	2.078	2.553	2.566	2.078	2.210	2.198	9.500	12.250

Table 3: Correlation Matrix

	log K_{oc}	**W**	**J**	**Jhetz**	**Jhetm**	**Jhetv**	**Jhete**	**Jhetp**	**^1x**	**^2x**
log K_{oc}	1	0.6162	-0.4739	-0.3697	-0.3697	-0.4277	-0.3724	-0.4303	0.6414	0.5264
W		1	-0.6412	-0.6892	-0.6886	-0.6962	-0.7029	-0.6978	0.9625	0.8517
J			1	0.9151	0.916	0.7614	0.8629	0.7955	-0.765	-0.5231
Jhetz				1	0.9999	0.8998	0.9843	0.9167	-0.8181	-0.7042
Jhetm					1	0.8984	0.9834	0.9157	-0.8177	-0.7026
Jhetv						1	0.9355	0.9963	-0.7759	-0.7651
Jhete							1	0.9379	-0.8246	-0.7723
Jhetp								1	-0.7809	-0.7381
^1x									1	0.8901
^2x										1

Table 4: Obtained QSAR Models For 32 POPs

No.	Equation	statistical characteristics
1.	K_{oc} =-3.3249-0.0034(0.0021)W-1.4496(0.6340)J+ 1.3679(0.3706)Jhetm+1.1100(0.3822)$^1\chi$	N=32, R^2=0.6096, AR^2=0.5518, MSE=0.2274, F=10.540, Q=3.4258
2.	K_{oc} =2.4597-2.0757(0.6559)J+1.8559(0.4709)Jhetm-0.9597(0.4468)Jhetv+0.4733(0.1090)$^1\chi$	N=32, R^2=0.6345, AR^2=0.5804, MSE=0.2129, F=11.718, Q=3.7414
3.	K_{oc} =1.9705-1.9310(0.6394)J+1.8305(0.4730)Jhetm-0.8606(0.4166)Jhetp+0.4835(0.1088)$^1\chi$	N=32, R^2=0.6304, AR^2=0.5757, MSE=0.2153, F=11.515, Q=3.6877
4.	K_{oc}=-2.9602-0.0050(0.0019)W-2.0862(0.5940)J+2.4442(0.4815)Jhetm-1.2790(0.4224)Jhetv+1.3384(0.3433)$^1\chi$	N=32, R^2=0.7114, AR^2=0.6559, MSE=0.1746, F=12.816, Q=4.8307
5	K_{oc}=-5.5226-0.0064(0.0017)W-1.6656(0.5512)J+2.7450(0.4329)Jhetm-1.5979(0.3718)Jhetv+1.6398(0.3065)$^1\chi$	N=30, R^2=0.7931, AR^2=0.7500, MSE=0.1268, F=18.399, Q=7.0228
	In this model compound no. 7 and 32 are outliers.	

Table 5: Observed (Obs.), Predicted (Pre.) Residual Value Obtained using Eq. 12

Row	Observed	Predicted	Residual
1	5.620	5.627	-0.007
2	6.640	6.449	0.191
3	5.360	4.932	0.428
4	4.670	4.499	0.171
5	5.940	6.182	-0.242
6	4.880	5.265	-0.385
7	4.900	5.134	-0.234
8	5.930	5.638	0.292
9	5.790	5.810	-0.020
10	5.860	5.588	0.272
11	5.410	5.097	0.313
12	5.730	5.433	0.297
13	5.550	5.591	-0.041
14	5.690	5.432	0.258
15	4.590	4.518	0.072
16	4.520	5.066	-0.546
17	6.660	6.505	0.155
18	4.810	4.500	0.310
19	5.810	5.493	0.317
20	5.370	6.060	-0.690
21	6.100	6.205	-0.105
22	3.922	4.323	-0.403
23	4.140	4.139	0.001
24	4.840	4.684	0.156
25	4.230	4.749	-0.519
26	4.620	4.481	0.139
27	5.010	5.371	-0.361
28	4.840	4.940	-0.100
29	6.000	5.384	0.616
30	4.690	5.027	-0.337

Res. = Residual = difference between observed and Predicted %Inhibition of various POPs

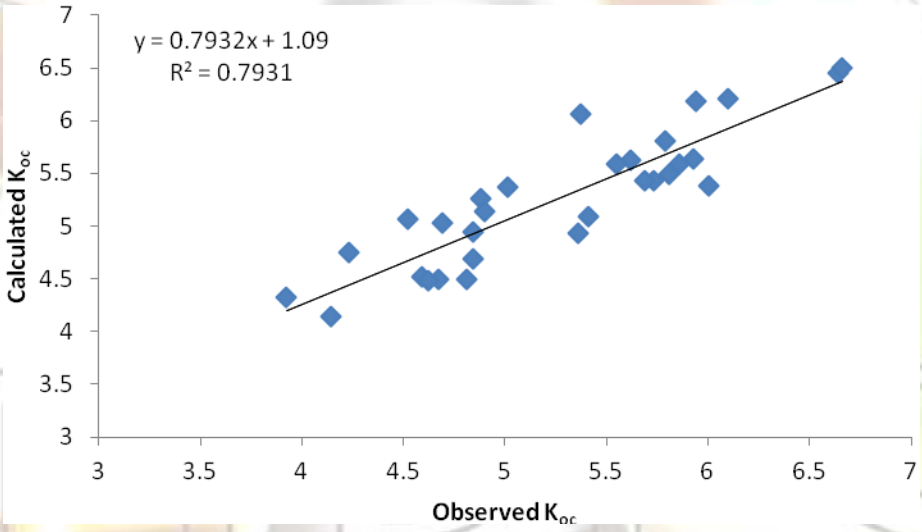

Fig 1 : Graph plotted between observed and calculated activity