RELIABLE QSAR FOR ESTIMATING KOC FOR PERSISTENT ORGANIC POLLUTANTS: CORRELATION WITH MOLECULAR CONNECTIVITY INDICES

Manish Rao Ambedkar[1], Kushal Kumar[2], Madhu Gupta[3]

[1, 3] *Department Of Chemistry, M. M. H. College Ghaziabad, C.C.S.University Meerut U.P (India)*

[2] *IIMT College of Engineering Greater Noida , Uttar Pradesh Technical University Lucknow U.P. (India)*

ABSTRACT

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 the animal and human reproduction, development and immunological function. Several recent studies have shown that the logarithmic n-octanol/water partition coefficient (logKow) may not be a good predictor for estimating soil sorption coefficients of persistent organic pollutants (POPs), defined here as chemicals with LogKo/w greater then 5. Thus an alternative QSAR model was developed that seems to provide reliable estimates for the soil sorption coefficients of persistent organic pollutants (POPs). Quantitative structure–activity relationship (QSAR) model for soil absorption coefficients of 18 persistent organic pollutants (POPs) is analyzed using multiple linear regression analysis (MLRA) followed by statistical evaluation by SPSS software (IBM). In order to indicate the influence of different molecular descriptors on soil sorption coefficients values and well understand the important structural factors affecting the experimental values, a set of physiochemical and topological parameters were taken into consideration. The proposed models gave the following results: the square of correlation coefficient, R^2, for the models with one, two and three molecular descriptors are 0.5960, 0.6738 and 0.8948.

Keywords: Persistent Organic Pollutant (Pops) , Quantitative Structure Activity Relationship (QSAR), SPSS .

I. 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. The Stockholm Convention, held in May 2001, focuses on eliminating or reducing releases of 12 POPs, the so-called “Dirty Dozen”. These 12 chemicals include

aldrin, chlordane, DDT, dieldrin, endrin, heptachlor, mirex, and toxaphene used principally as pesticides, two industrial chemicals polychlorinated biphenyls and hexachlorobenzene used in industry but also produced unintentionally together with dioxins and furans. The chemicals known as persistent organic pollutants act as powerful pesticides and serve a range of industrial purposes. Some POPs are also released as unintended by-products of combustion and industrial processes.While the risk level varies from POP to POP, by definition all of these chemicals share four properties:

1) They are highly toxic;

2) they are persistent, lasting for years or even decades before degrading into less dangerous forms;

3) they evaporate and travel long distances through the air and through water; and

4) they accumulate in fatty tissue.

This is a dangerous combination. The persistence and mobility of POPs means that they are literally everywhere in the world, even in the Arctic, Antarctica, and remote Pacific islands. Their attraction to fatty tissue, known as "bioaccumulation", means that even though a poison is first dispersed widely and thinly it gradually starts to concentrate as organisms consume other organisms as they move up the food chain. The chemicals reach magnified levels – up to many thousands of times greater than background levels – in the fatty tissues of creatures at the top of the food chain, such as fish, predatory birds, and mammals, including human beings. Worse still, during pregnancy and breastfeeding these POPs are often passed on to the next generation. Human beings and other mammals are thus exposed to the highest levels of these contaminants when they are most vulnerable – in the womb and during infancy, when their bodies, brains, nervous systems, and immune systems are in the delicate process of construction.

Direct contact with POPs can cause acute effects – accidents with pesticides, for example, have killed and seriously sickened agricultural workers. But the kind of harm caused to human beings by low levels of POPs – cancer, immune-system disruption, nervous-system damage, liver damage, memory loss, endocrine disruption, birth defects, other reproductive problems – can be difficult to prove conclusively. It is hard to demonstrate that someone's immune system is weaker than it might have been, let alone that a particular chemical is the culprit. Nervous-system damage may result in something as basic and yet as nebulous as a lower level of intelligence. Once again, this can be hard to demonstrate beyond challenge. But unless precautionary action is taken to curtail exposure to these chemicals, millions of people – not to mention millions of other creatures ranging from lake trout to penguins – are likely to suffer terrible harm. They may also cause toxic effects on the animal and human reproduction, development and immunological function.[1]

Typical examples of these chemicals are Polychlorinated bi phenyls (PCBs), Polychlorinated dibenzo p-dioxims and furans, polycyclic aromatic hydrocarbons (PAHs) and pesticides such as DDT, Chlordane and Heptachlor. To deal with POPs, the nations of the world really will have to work together as a team. That will be good for eliminating the use of these dangerous chemicals and if such cooperation becomes a habit, it could be good for facing up to many other global problems as well.

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 normalizes to organic carbon (Koc). Since measured Koc data are

not available for majority of those chemicals, numerous correlations have been developed relating Koc to other physicochemical or structural descriptors, in particular with the n-octanal/water partition coefficient (Kow).[2-3] This has enabled simple and fast estimation of Koc for POPs.[4]

However, *Aleksandar Sabljic* et al have observed the breakdown in linear relationship between LogKoc and LogKow as a chemical's LogKow reached the range of 6 to 7. Furthermore, Baker et all [5] Have demonstrated using a high quality database of LogKoc test using a high quality database of Koc values for 18 POPs that there is only a week correlation between LogKoc and LogKo/w (R^2=0.2940) and its application will result in predicted values which may be off by a factor of 15.[6-9] Many coworkers has shown correlation with LogKoc.

II. MATERIALS AND METHODS

2.1. Data Set

All data of the present investigation were obtained from the reference *(Aleksandar Sabljic et al., 2000)*. The data set for this investigation consisted of 18 POPs is reported into (Table no. 1).

2.2. Molecular Descriptor Generation

To obtain a QSAR model, compounds are often represented by the molecular descriptors. The calculation process of the molecular descriptors was described as below: The two-dimensional molecular structures of 18 POPs were drawn by Chem Sketch 12.0 then calculated some parameters. Then this optimize structure files were exported into software Dragon 6.0 to calculate all kinds of descriptors. The software Dragon 6.0 can calculate Physicochemical parameters, constitutional, topological, geometrical, descriptors and has been successfully used in various QSAR researches. Then value of all parameters put into NCSS statistical and data

analysis software or SPSS (We can also use MSTAT & NCSS instead of SPSS) statistical and data analysis software to get data regression and correlation. Constitutional descriptors are related to the number of atoms and bonds in each molecule. Topological descriptors include valence and non-valence molecular connectivity indices calculated from the hydrogen-suppressed formula of the molecule, encoding information about the size, composition, and the degree of branching of a molecule. The topological descriptors describe the atomic connectivity in the molecule. The geometrical descriptors describe the size of the molecule and require 3Dcoordinates of the atoms in the given molecule. The electrostatic descriptors reflect characteristics of the charge distribution of the molecule. The quantum chemical descriptors offer information about binding and formation energies, partial atom charge, dipole moment, and molecular orbital energy levels.

III. RESULTS AND DISCUSSION

By using the multiple linear regression analysis (MLRA) method of QSAR, regression models were developed for 18 POPs. To select the sets of descriptors that are most relevant to logKow & logKoc values and effectively show the relation between descriptors , logKow & logKoc values of these compounds, three subsets with the descriptors from one to three were determined to establish the QSAR models. However, Baker et al have observed the

breakdown in linear relationship between LogKoc and LogKow as a chemical's LogKow reached the range of 6 to 7. Multi-linear regression method for descriptor selection proceeds with a reselections of descriptors by sequentially eliminating descriptors which do not match any of the following criteria: (i) the F-test greater than one unit; (ii) R^2 value less than a value defined at the start (default 0.01); (iii) the student's t-test less than that defined (default 0.1); and (iv) duplicate descriptors having a higher squared inter-correlation coefficient than a predetermined level (usually 0.8). The next step involves correlation of the given property with (i) the top descriptor in the above list with each of the remaining descriptors, and (ii) the next one with each of the remaining descriptors, etc. The goodness of the correlation is tested by the correlation coefficient (R^2) and The stability of the correlations was tested against the cross-validated coefficient (R^2 CV). Besides, it will demonstrate which descriptors have bad or missing values, which descriptors are insignificant, and which descriptors are highly intercorelated .This information will be helpful in reducing the number of descriptors involved in the search for the best QSAR/QSPR model. Comparison with *Aleksandar Sabljic* et al we have observed that in our case R^2 for models with one, two and three molecular descriptors are 0.5960, 0.6738 and 0.8948. Our results are much more superior then the result reported by *Aleksandar Sabljic* et al. Therefore simple 2D QSAR reported by us is much betters then the QSAR modeling of *Aleksandar Sabljic* et al.

Following topological indices have been calculated using Dragon software and they are reported in (Table no. 2). The calculating connectivity indices have been calculated is in (Table no. 3). Topological parameters used for modeling LogKoc value for set of 18 compound's used in present sturdy, when single parameter is used 17 mono-parametric model have been obtained. No significant mono-parametric model is obtained. When two parameter are taken together 10 bi-parametric model have been obtained. Out of 10 bi-parametric models the best model contains LogKow and J. The R^2 value of best model is 0.6377.The best model is as given below.

3.1 Best Bi-Parametric Model

LogKoc=0.3730(±0.1448)LogKow -1.0270(±0.2722)J +5.2507

N=18, Se=0.0805, R^2=0.6377, R^2A=0.5894, F-Ratio=13.2040, Q=9.9200

In the above model LogKow has positive coefficient and J has negative coefficient. Suggesting that the high value of LogKow and low value of J will favor the modeling of LogKoc.

When three parameters are taken together eight tri-parametric models have been obtained. These models are statistically are better then there bi-parametric models. The best model contains BAC, J and Jhetm having R^2 value 0.8277. The model is as given below.

3.2 Best Tri Parametric Model

LogKoc=0.039(±0.0082)BAC -6.3928(±1.0153)J 1.9091(±0.3849)Jhetm +11.4779

N=18, Se=0.0574, R^2=0.8277, R^2A=0.7908, F-Ratio=22.4230, Q=15.8498

In the above model BAC and Jhetm have positive coefficient and J has negative coefficient, suggesting that the high value of BAC and Jhetm and low value of J will favor the modeling of LogKoc.

The observed and estimated activity along with difference the LogKoc values for the compounds used in the present study using the best model are reported in (Table no. 5). Also the predictive power of the model comes out to be 0.8277 as demonstrated in (Figure no. 1). This is also confirm on the basis of cross-validated parameters as reported in (Table no. 7). A study of (Table no. 6) shows that compound no. 6 is serious outlier there for it is not considered in deriving new models. However we observed improvement in quality of statistics only in all the previously discussed models. They are reported in (Table no. 8). The new tri-parametric model has now R^2 value 0.8948 and in the best using model the LogKoc values calculated for different compounds and they are reported in (Table no. 9). A plot between observed and estimated LogKoc values has been drawn which shown in (Figure no. 2). The cross-validated parameters are for these models are reported in (Table no. 10). The best model after deleting of compound no. 6 is as given below.

3.3 Best Tri Parametric Model without considereing compound no. 6

LogKoc=0.0335(±0.0068)BAC -6.1480(±0.8094)J 1.8226(±0.3067)Jhetm +11.4419

N=18, Se=0.0453, R^2=0.8948, R^2A=0.8706, F-Ratio=36.8700, Q=20.8816

IV. CONCLUSION

A quantitative structure–activity relationship model was derived to study the logKow and LogKoc values of a diverse set of 17 POPs. Three QSAR models were developed with the squared correlation coefficient ($R^{2)}$ of one, two and three molecular descriptors are 0.5960, 0.6738 and 0.8948. These models showed strong predictive ability. The present work provides an effective method for the prediction of the logKow and LogKoc values for the POPs. This study also showed that the utility of the QSAR treatment involving descriptors derived solely from chemical structure and the correlation equation and descriptors can be used for the prediction of the logKow and LogKoc values for unknown structures. Following conclusion may be drawn on the basis of above discussion. To understand the correlated values among the parameters a correlation matrix has been obtained which is demonstrated in (Table no. 4). Close look at this matrix has been obtained which is demonstrated in (Table no. - 4).Close look at this table revels that…

CO-RELATION

1. The LogKoc has poor co-relation with the Topological parameters used.
2. LogKow has poor co-relation with all the parameter we have used.
3. W has moderate co-relation with Balaban and Balaban type indices, poor co-relation with BAC and good co-relation with 0X, 1χ, 2χ and 1χv.
4. J and all the Balaban type indices have good co-relation with among themselves and with 0χ, 1χ, 2χ and 3χ.
5. BAC has poor co-relation with all the parameters used.
6. All the Ranadic and Kiren Hall type of connectivity type indices has good co-relation among themselves except 3χv which has poor co-relation.

REFERENCES

1. Wania, F. ; Mackay, D. ; ***Environ. Sci. Technol.*** 30 (1996) 390A-396A.
2. Sabljic, A. ; Gusten, H. ; Verhaar, H. ; Hermens, J. ; ***Chemosphere*** 31 (1995) 4489-4514.
3. Baker, J. R. ; Mihelcic, J. R. ; Luehrs, D. C. ; Hickey, J. P. ; ***Water Environ. Res.*** 69 (1997) 136-145.
4. Baker , J. R. ; Mihelcic , J. R. ; Shea, E. ; ***Chemosphere*** 41 (2000) 813-817.
5. Sabljic , A. ; ***J. Chromatogr.*** 319 (1985) 1-8.
6. Sabljic , A. ; ***Environ. Sci. Technol.*** 21 (1987) 358-366.
7. Sabljic , A. ; ***Environ. Health Perspect.*** 83 (1989) 179-190.
8. Gusten, H. ; Sabljic, A. ; Horvatic , D. ; ***Chemosphere*** 23 (1991) 199-213.
9. Meyland , W. ; Howard , P. H. ; Boethling , R. S. ; ***Environ. Sci. Technol.*** 26 (1992) 1560-1567

Table- 1 Details of compounds with their activity LogKoc and logKow values used in the present study.

S.No.	Chemical name	logKoc	logKow
1	Benza[a]anthracene	5.6200	5.7900
2	Benzo[a]pyrene	6.6400	5.9700
3	Chlordane	5.3600	5.8000
4	4,4'-DDT	4.6700	6.9100
5	1,2:5,6dibenzaanthracene	5.9400	6.5000
6	Fluoranthene	4.8800	5.1600
7	Hexachlorobenzene	4.3100	5.7300
8	Methoxychlor	4.9000	5.0800
9	2,2',3,4,4',5'hexachlorobiphenyl	5.9300	7.2500
10	2,2',3,4',5',6hexachlorobiphenyl	5.7900	6.8600
11	2,2',4,4',5,5'hexachlorobiphenyl	5.8600	7.4400
12	2,2',5,5'tetrachlorobiphenyl	5.4100	6.0900
13	2,2',3,4,5'pentachlorobiphenyl	5.7300	6.8500
14	2,2',3,5',6 pentachlorobiphenyl	5.5500	6.5500
15	2,2',3,4',5' pentachlorobiphenyl	5.6900	6.6700
16	Pentachlorobenzene	4.5900	5.1800
17	pentachlorophenol	4.5200	5.1200
18	2,3,7,8-TCDD	6.6600	6.4200

Table - 2 Values of calculated topological parameters for the compounds used in the present study.

Comp.no.	W	J	JhetZ	Jhetm	Jhetv	Jhete	Jhetp	BAC
1	553.0000	1.5120	1.9630	1.9630	1.9630	1.9630	1.9630	0.0000
2	680.0000	1.4870	2.1490	2.1490	2.1490	2.1490	2.1490	0.0000
3	469.0000	2.1490	2.5950	2.6050	2.2070	2.3180	2.3080	65.0000
4	678.0000	2.0370	2.6890	2.6940	2.4740	2.5390	2.5330	27.0000
5	971.0000	1.3460	1.7390	1.7390	1.7390	1.7390	1.7390	0.0000
6	364.0000	1.6770	2.5160	2.5160	2.5160	2.5160	2.5160	0.0000
7	174.0000	2.7600	4.8310	4.8680	3.6040	3.9230	3.8940	37.0000
8	932.0000	1.9910	2.5950	2.5970	2.1810	2.5190	2.1530	35.0000
9	571.0000	2.0620	3.3050	3.3160	2.9000	3.0180	3.0080	37.0000
10	555.0000	2.1260	3.4170	3.4280	2.9830	3.1090	3.0980	37.0000

11	573.0000	2.0550	3.2930	3.3030	2.8910	3.0080	2.9980	37.0000
12	412.0000	2.0240	3.2050	3.2130	2.8800	2.9760	2.9680	17.0000
13	486.0000	2.0500	3.2690	3.2780	2.8960	3.0050	2.9960	26.0000
14	472.0000	2.1160	3.3830	3.3930	2.9820	3.0990	3.0890	26.0000
15	486.0000	2.0490	3.2670	3.2760	2.8940	3.0040	2.9940	26.0000
16	140.0000	2.6250	4.5490	4.5810	3.4560	3.7430	3.7180	26.0000
17	174.0000	2.7600	4.6700	4.6980	3.4020	3.9330	3.5650	37.0000
18	571.0000	1.6880	2.5360	2.5400	1.5460	2.3980	1.4670	17.0000

Table - 3 Values of calculated connectivity indices for the compounds used in the present study.

Comp.no.	**0χ**	**1χ**	**2χ**	**3χ**	**0χv**	**1χv**	**2χv**	**3χv**
1	11.9490	8.9160	7.9860	7.2780	10.1350	6.4820	5.0250	3.9710
2	13.1040	9.9160	9.1280	8.6240	11.1880	7.2500	5.7440	4.6730
3	13.6190	8.1690	8.8580	8.8770	14.5350	8.4150	9.1900	9.2800
4	14.0440	8.8760	8.9320	6.1370	13.3660	7.3430	7.3200	3.8930
5	14.5180	10.8990	9.8910	9.2130	12.5500	8.1720	6.4820	5.2780
6	10.5350	7.9490	7.1390	6.7330	8.7740	5.5650	4.2550	3.4120
7	9.4640	5.4640	5.1550	4.9760	9.8030	4.9020	4.1520	4.0040
8	15.4580	9.9520	9.2700	6.9530	13.9140	7.4340	6.8900	3.8800
9	13.4470	8.4140	8.0040	6.9970	13.1130	6.9670	5.8540	4.6640
10	13.4470	8.4140	8.0350	6.8230	13.1130	6.9670	5.8780	4.4770
11	13.4470	8.3970	8.1340	6.7300	13.1130	6.9610	5.9360	4.3800
12	11.7070	7.5750	7.1190	5.5900	11.0000	5.9940	4.9360	3.2760
13	12.5770	8.0030	7.4960	6.4280	12.0560	6.4830	5.3540	4.1120
14	12.5770	8.0030	7.5280	6.2540	12.0560	6.4830	5.3780	3.9260
15	12.5770	8.0030	7.4980	6.4100	12.0560	6.4830	5.3540	4.1050
16	8.5940	5.0370	4.7680	4.1970	8.7470	4.4120	3.7300	3.2040
17	9.4640	5.4640	5.1550	4.9760	9.1170	4.5580	3.8080	3.4430
18	12.8610	8.5420	8.3640	7.1850	11.6610	6.3940	5.2590	3.8570

Table- 4 Correlation matrix showing inter-correlation among all the parameters with the activity.

	logKoc	***log kow***	***W***	***J***	***JhetZ***	***Jhetm***	***Jhetv***	***Jhete***	***Jhetp***	***BAC***
logKoc	1.0000									
log Kow	0.5422	1.0000								
W	0.4987	0.3304	1.0000							
J	-0.6917	-0.2210	-0.7349	1.0000						
JhetZ	-0.5818	-0.1600	-0.7944	0.9534	1.0000					
Jhetm	-0.5825	-0.1613	-0.7947	0.9537	1.0000	1.0000				
Jhetv	-0.5820	-0.0450	-0.7402	0.8612	0.9211	0.9204	1.0000			
Jhete	-0.5595	-0.1243	-0.7796	0.9339	0.9918	0.9913	0.9374	1.0000		
Jhetp	-0.5894	-0.0489	-0.7481	0.8723	0.9270	0.9265	0.9982	0.9361	1.0000	
BAC	-0.3108	0.1137	-0.2498	0.6787	0.4967	0.4969	0.4179	0.4765	0.4339	1.0000
0χ	0.5177	0.4997	0.9206	-0.5903	-0.7016	-0.7025	-0.6349	-0.6760	-0.6433	0.0661
1χ	0.6224	0.3479	0.9568	-0.8733	-0.9101	-0.9106	-0.8119	-0.8881	-0.8225	-0.3651
2χ	0.6044	0.3997	0.9313	-0.7932	-0.8844	-0.8847	-0.8137	-0.8716	-0.8184	-0.1690

3χ	0.6555	0.2308	0.7347	-0.7815	-0.8445	-0.8438	-0.7999	-0.8646	-0.7916	-0.1866
0χv	0.3623	0.5755	0.6829	-0.2438	-0.4070	-0.4076	-0.3626	-0.3966	-0.3592	0.4686
1χv	0.5231	0.4528	0.8394	-0.6241	-0.7605	-0.7604	-0.6732	-0.7649	-0.6666	0.0855
2χv	0.2347	0.3069	0.6243	-0.3153	-0.5448	-0.5443	-0.5078	-0.5669	-0.4924	0.4074
3χv	0.2037	0.1110	0.2143	-0.1229	-0.3097	-0.3080	-0.3179	-0.3711	-0.2850	0.4927

	0χ	*1χ*	*2χ*	*3χ*	*0χv*	*1χv*	*2χv*	*3χv*
0χ	1.0000							
1χ	0.8887	1.0000						
2χ	0.9410	0.9676	1.0000					
3χ	0.7190	0.8427	0.8662	1.0000				
0χv	0.8981	0.6188	0.7553	0.5547	1.0000			
1χv	0.9291	0.8640	0.9461	0.8647	0.8700	1.0000		
2χv	0.8025	0.6114	0.7756	0.6846	0.8886	0.9059	1.0000	
3χv	0.3983	0.2719	0.4389	0.6621	0.5819	0.6611	0.7912	1.0000

Table- 5 Regression parameters and quality of correlation with Topological and connectivity indices.

Model No.	Parameter used	Ai=(1…3)	B	Se	R^2	R^2A	F –Ratio	Q=R/Se
1	**LogKow**	0.4927(±0.1909)	2.3988	0.1088	0.2940		6.6617	4.9836
2	**W**	0.0015(±0.0007)	4.6669	0.1122	0.2487		5.2961	4.4447
3	**J**	-1.1816(±0.3084)	7.8441	0.0935	0.4784		14.6744	7.3974
4	**Jhetz**	-0.4505(±0.1574)	6.8480	0.1053	0.3385		8.1870	5.5252
5	**Jhetm**	-0.4458(±0.1555)	6.8379	0.1052	0.3393		8.2165	5.5370
6	**Jhetv**	-0.6773(±0.2366)	7.2408	0.1053	0.3388		8.1968	5.5276
7	**Jhete**	-0.6107(±0.2262)	7.1763	0.1073	0.3130		7.2904	5.2140
8	**Jhetp**	-0.5978(±0.2048)	7.0796	0.1046	0.3474		8.5164	5.6348
9	**BAC**	-0.0123(±0.0094)	5.7555	0.1231	0.0966		1.7105	2.5248
10	**0χ**	0.1918(±0.0792)	3.0674	0.1108	0.2680		5.8573	4.6722
11	**1χ**	0.2779(±0.0874)	3.1933	0.1013	0.3873		10.1160	6.1434
12	**2χ**	0.2866(±0.0944)	3.2426	0.1032	0.3653		9.2077	5.8565
13	**3χ**	0.3399(±0.0979)	3.1743	0.0978	0.4297		12.0559	6.7026
14	**0χv**	0.1397(±0.0898)	3.8156	0.1207	0.1313		2.4173	3.0020
15	**1χv**	0.3214(±0.1309)	3.3533	0.1103	0.2736		6.0276	4.7422
16	**2χv**	0.1204(±0.1247)	4.7745	0.1259	0.0551		0.9330	1.8644
17	**3χv**	0.1036(±0.1245)	4.9994	0.1268	0.0415		0.6924	1.6065
18	**LogKow** **Jhete**	0.4362(±0.1604) -0.5456(±0.1927)	4.2928	0.0907	0.5399	0.4786	8.8010	8.1012
19	**J** **Jhetm**	-2.5690(±0.9898) 0.6518(±0.4434)	8.6253	0.0903	0.5441	0.4833	8.9500	8.1686
20	**J** **Jhetz**	-2.5699(±0.9865) 0.6601(±0.4471)	8.6080	0.0902	0.5446	0.4838	8.9680	8.1814
21	**LogKow** **Jhetz**	0.4188(±0.1602) -0.3934(±0.1365)	4.0791	0.0902	0.5455	0.4849	9.0010	8.1882
22	**LogKow**	0.4182(±0.1603)	4.0732	0.0901	0.5456	0.4850	9.0050	8.1980

	Jhetm	-0.3890(±0.1350)						
23	**3χ** **LogKow**	0.2905(±0.0880) 0.3752(±0.1542)	1.1832	0.0855	0.5911	0.5366	10.8420	8.9921
24	**1χv** **2χv**	1.0639(±0.2391) -0.6844(±0.1997)	2.3390	0.0853	0.5927	0.5384	10.9120	9.0254
25	**LogKow** **Jhetv**	0.4698(±0.1475) -0.6502(±0.1889)	4.2621	0.0840	0.6055	0.5529	11.5130	9.2635
26	**LogKow** **Jhetp**	0.4676(±0.1464) -0.5722(±0.1634)	4.1166	0.0833	0.6116	0.5598	11.8080	9.3883
27	**LogKow** **J**	0.3730(±0.1448) -1.0270(±0.2722)	5.2293	0.0805	0.6377	0.5894	13.2040	9.9200
28	**1χv** **J** **Jhetm**	0.4616(±0.1567) -4.0014(±0.9405) 1.7008(±0.5069)	5.2507	0.0734	0.7185	0.6582	11.9100	11.5482
29	**0χv** **J** **Jhetz**	0.2178(±0.0677) -4.0986(±0.9086) 1.4988(±0.4373)	6.5561	0.0708	0.7381	0.6819	13.1490	12.1345
30	**BAC** **J** **Jhete**	0.0304(±0.0092) -4.6496(±0.9736) 1.7656(±0.5197)	9.1212	0.0706	0.7397	0.6839	13.2590	12.1821
31	**3χ** **2χv** **LogKow**	0.4728(±0.0969) -0.2783(±0.0980) 0.4527(±0.1300)	1.0389	0.0705	0.7406	0.6850	13.3210	12.2068
32	**2χv** **J** **Jhetm**	0.4690(±0.1388) -6.0829(±1.2880) 2.5336(±0.6527)	7.2625	0.0694	0.7489	0.6951	13.9190	12.4695
33	**2χv** **J** **Jhetz**	0.4698(±0.1382) -6.0761(±1.2791) 2.5615(±0.6561)	7.1841	0.0692	0.7504	0.6969	14.0300	12.5181
34	**BAC** **J** **Jhetz**	0.0391(±0.0082) -6.3584(±1.0123) 1.9185(±0.3883)	11.4034	0.0576	0.8269	0.7898	22.2950	15.7871
35	**BAC** **J** **Jhetm**	0.0393(±0.0082) -6.3928(±1.0153) 1.9091(±0.3849)	11.4779	0.0574	0.8277	0.7908	22.4230	15.8498

Table- 6 Observed and calculated activity for the compounds using model- 35 (Table - 5)

Comp. No.	**Observed log koc**	**Calculated log koc**	**Residual**
1	5.6200	5.5600	0.0600
2	6.6400	6.0740	0.5660
3	5.3600	5.2650	0.0950
4	4.6700	4.6590	0.0110
5	5.9400	6.1930	-0.2530
6	4.8800	5.5600	-0.6800
7	4.3100	4.5800	-0.2700
8	4.9000	5.0820	-0.1820
9	5.9300	6.0790	-0.1490
10	5.7900	5.8840	-0.0940
11	5.8600	6.0990	-0.2390
12	5.4100	5.3400	0.0700

13	5.7300	5.6510	0.0790
14	5.5500	5.4490	0.1010
15	5.6900	5.6540	0.0360
16	4.5900	4.4630	0.1270
17	4.5200	4.2550	0.2650
18	6.6600	6.2030	0.4570

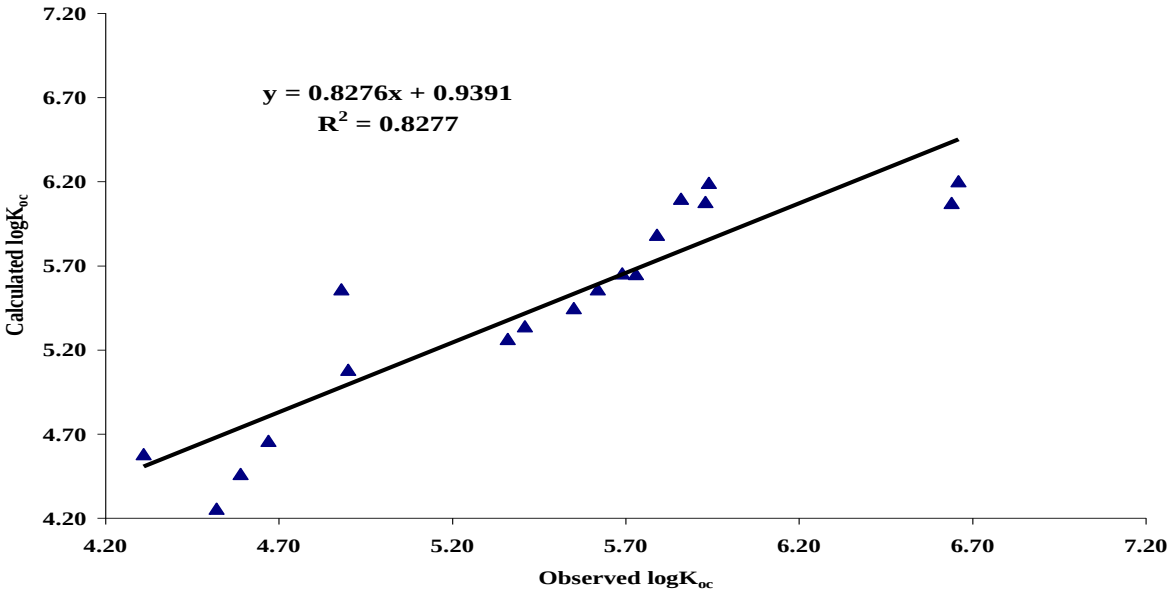

Fig- 1 Correlation between Observed and Calculated activity using model no-35 (Table- 5)

Table- 7 cross validated values for Topological and connectivity indices.

S.no.	Parameters used	PRESS	SSY	PRESS/SSY	R^2cv	PSE	S_{PRESS}
1.	**J**	4.1513	3.8074	1.0903	-0.0903	0.4802	0.5093
2.	**logKow,J**	2.8830	5.0756	0.5680	0.4320	0.4002	0.4384
3.	**J,Jhetm,BAC**	1.3710	6.5877	0.2081	0.7919	0.2759	0.3129

Table- 8 Regression parameters and quality of correlation with Topological and connectivity indices after deleting one compound (Compound no. 6).

Modeln o.	Parameter used	Ai=(1…3)	B	Se	R^2	R^2a	F Ratio	Q=R/Se
1	**J**	-1.3224(±0.2811)	8.1906	0.0826	0.5960	-	22.1311	9.3463
2	**log Kow** **J**	0.2786(±0.1525) -1.1598(±0.2762)	6.1167	0.0769	0.6738	0.6272	14.4620	10.6742
3	**J** **Jhetm** **BAC**	0.0335(±0.0068) -6.1480(±0.8094) 1.8226(±0.3067)	11.4419	0.0453	0.8948	0.8706	36.8700	20.8816

Table- 9 Observed and calculated activity for the compounds after deleting one compound (compound no. 6) using model no. 3 (Table - 8).

Comp. No.	Observed log koc	Calculated log koc	Residual
1	5.6200	5.7240	-0.1040
2	6.6400	6.2170	0.4230
3	5.3600	5.1540	0.2060
4	4.6700	4.7330	-0.0630
5	5.9400	6.3360	-0.3960
6	4.3100	4.5850	-0.2750
7	4.9000	5.1070	-0.2070
8	5.9300	6.0470	-0.1170
9	5.7900	5.8580	-0.0680
10	5.8600	6.0670	-0.2070
11	5.4100	5.4240	-0.0140
12	5.7300	5.6840	0.0460
13	5.5500	5.4870	0.0630
14	5.6900	5.6860	0.0040
15	4.5900	4.5230	0.0670
16	4.5200	4.2750	0.2450
17	6.6600	6.2630	0.3970

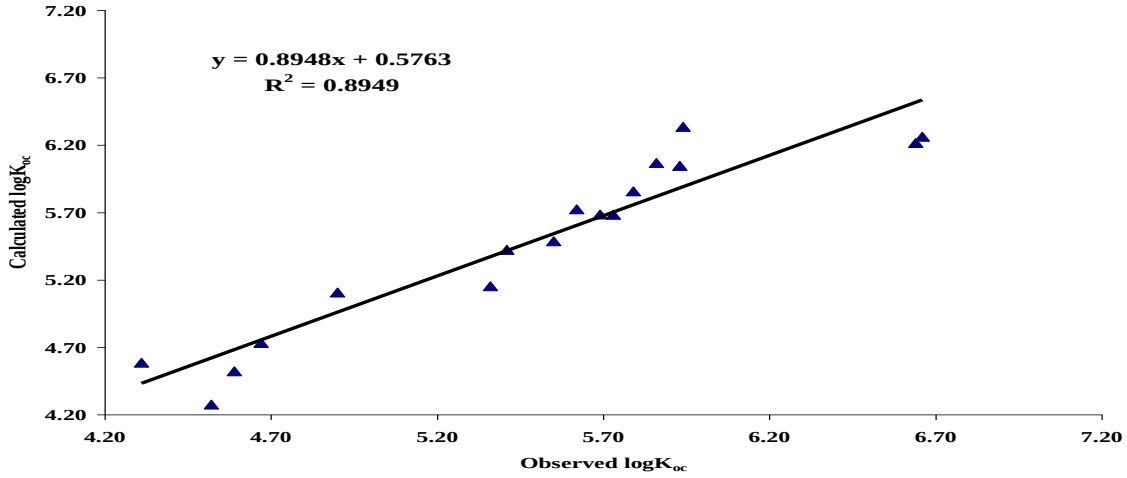

Fig- 2 Correlation between Observed and Calculated activity using model no. 3 (Table- 8)

Table- 10 cross validated values for Topological and connectivity indices after deleting of one compound (compound no. - 6).

S.no.	Parameters used	PRESS	SSY	PRESS/SSY	R^2cv	PSE	S_{PRESS}
1.	**J**	3.0775	4.5405	0.6777	0.3223	0.4254	0.4529
2.	**logKow,J**	2.4847	5.1333	0.4840	0.5160	0.3823	0.4212
3.	**J,Jhetm,BAC**	0.8011	6.8168	0.1175	0.8825	0.2170	0.2482