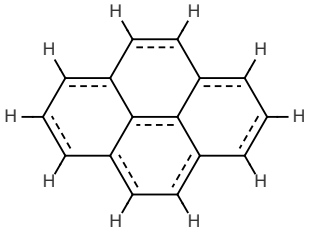
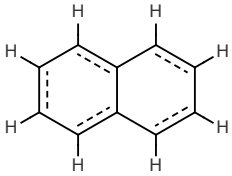
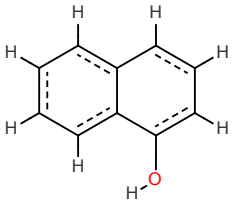
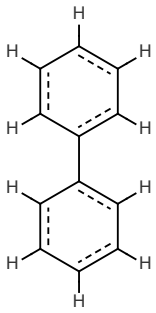
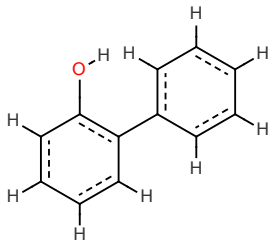
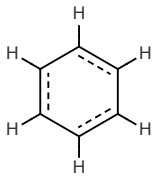
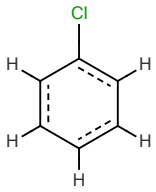
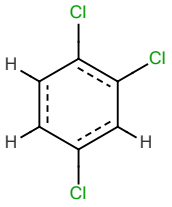
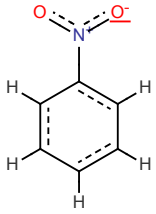
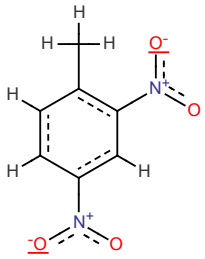
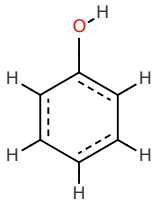
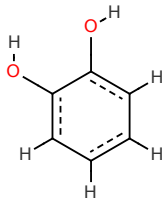
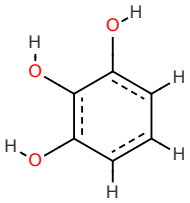
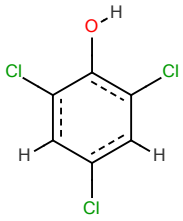
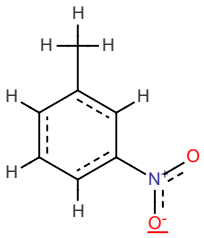
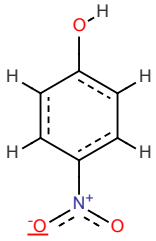
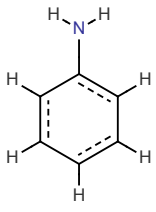
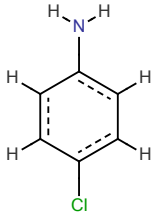
Supplementary materials

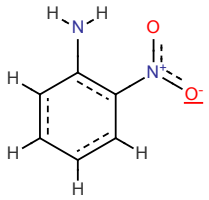
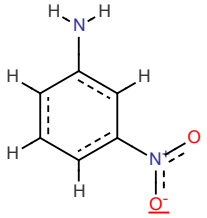
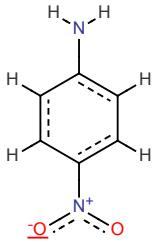
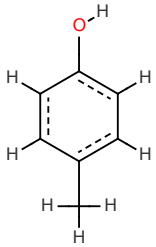
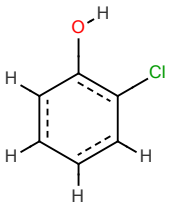
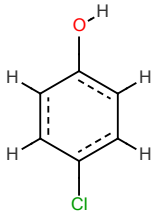
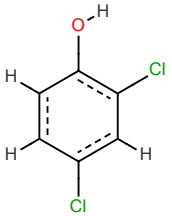
Methods and materials

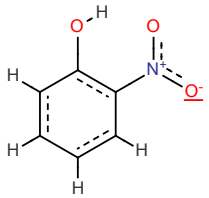
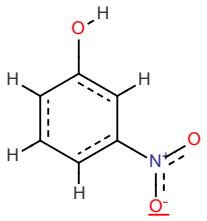
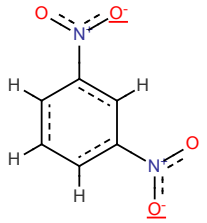
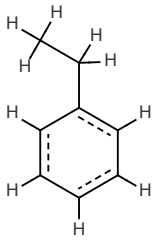
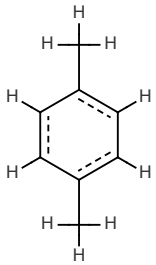
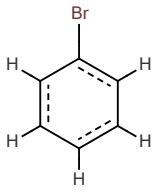
Table S1 The chemical name, experimental $logK_{\infty}$ and calculated $logK_{\infty}$ values of the MLR models obtained from dataset 1

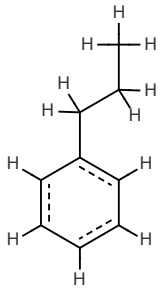
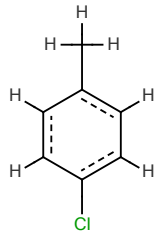
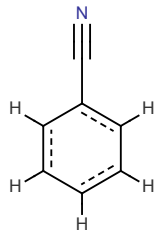
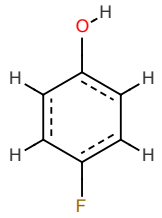
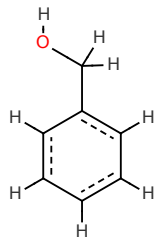
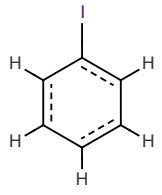
No.	Chemical name	Chemical structure	$logK_{\infty}$					
			Expt.	Calc.				
				Model M1	Model M2	Model M3	Model M4	Model M5
1	pyrene		4.01	3.83	3.75	3.76	3.76	3.71
2*	naphthalene		1.63	1.23	1.07	1.08	1.13	1.09
3*	1-naphthol		0.76	1.46	1.28	1.29	1.33	1.29
4	biphenyl		2.05	2.06	1.97	1.98	2.01	1.97

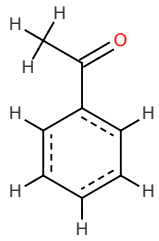
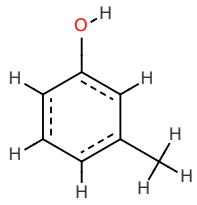
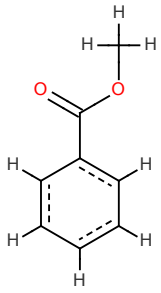
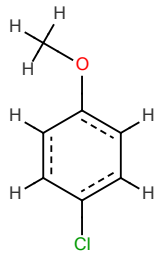
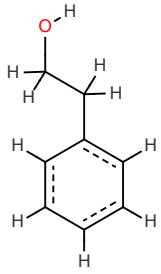
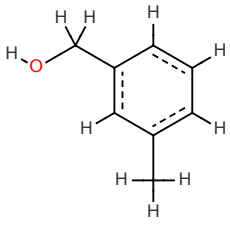
5	2-phenylphenol		1.63	2.29	2.19	2.20	2.22	2.19
6	benzene		-0.45	-0.73	-0.82	-0.82	-0.74	-0.77
7*	chlorobenzene		-0.33	-0.05	-0.09	-0.09	-0.03	-0.05
8*	1,2,4-trichlorobenzene		1.17	1.30	1.45	1.44	1.46	1.45
9	nitrobenzene		0.33	0.53	0.42	0.44	0.51	0.46
10	2,4-dinitrotoluene		2.38	2.24	2.30	2.32	2.33	2.29
11	phenol		-0.54	-0.51	-0.64	-0.64	-0.57	-0.59

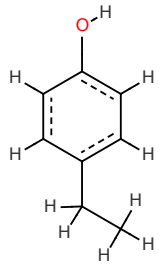
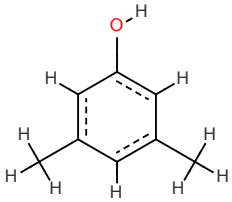
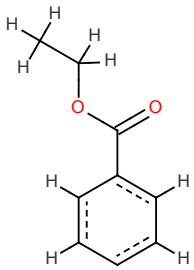
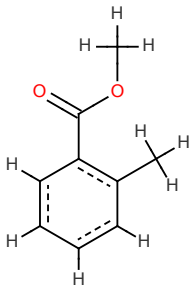
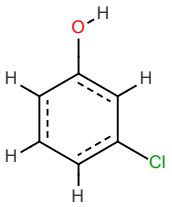
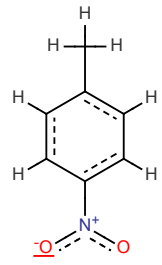
12	catechol		0.21	0.33	0.25	0.17	0.28	-0.39
13	pyrogallol		1.18	1.17	1.16	1.22	1.14	0.96
14	2,4,6-trichlorophenol		1.43	1.53	1.69	1.68	1.69	1.68
15	3-nitrotoluene		1.03	1.12	1.18	1.20	1.22	1.19
16	4-nitrophenol		0.77	0.76	0.61	0.62	0.68	0.64
17*	aniline		-0.77	-0.41	-0.54	-0.54	-0.47	-0.49
18	4-chloroaniline		-0.66	0.27	0.22	0.22	0.27	0.25

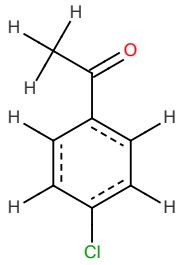
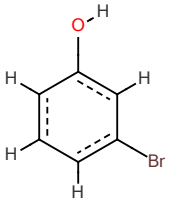
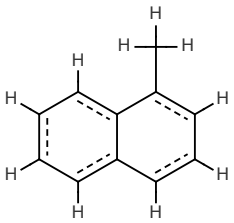
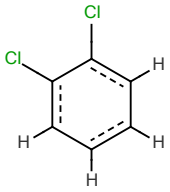
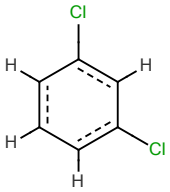
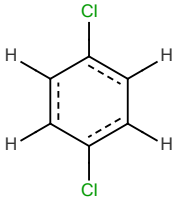
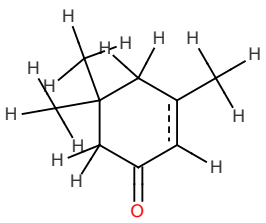
19	2-nitroaniline		1.6	0.85	1.54	1.14	0.78	1.31
20*	3-nitroaniline		0.72	0.85	0.71	0.72	0.78	0.74
21*	4-nitroaniline		0.95	0.85	0.71	0.72	0.78	0.74
22	4-methylphenol		0.06	0.09	0.11	0.11	0.13	0.13
23	2-chlorophenol		0.08	0.17	0.12	0.12	0.17	0.15
24	4-chlorophenol		0.74	0.17	0.12	0.12	0.17	0.15
25	2,4-dichlorophenol		0.96	0.85	0.90	0.89	0.92	0.91

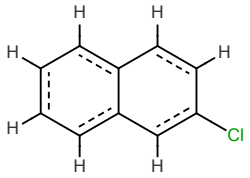
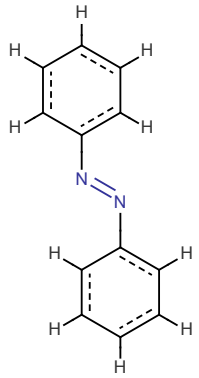
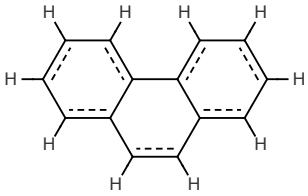
26	2-nitrophenol		0.56	0.76	0.61	1.03	0.68	1.20
27	3-nitrophenol		0.92	0.76	1.02	0.62	0.68	0.64
28	1,3-dinitrobenzene		1.46	1.65	1.53	1.55	1.60	1.54
29	ethylbenzene		0.19	0.31	0.11	0.10	0.10	0.12
30*	4-xylene		0.26	0.45	0.27	0.26	0.26	0.28
31	bromobenzene		0.5	0.48	0.46	0.45	0.49	0.48

32	propylbenzene		0.76	0.77	0.41	0.40	0.37	0.41
33	4-chlorotoluene		0.82	0.54	0.69	0.69	0.70	0.69
34	benzonitrile		0.04	0.31	0.39	0.41	0.50	0.43
35*	4-fluorophenol		-0.32	-0.31	-0.49	-0.49	-0.43	-0.44
36	benzyl alcohol		-0.9	-0.65	-0.67	-0.69	-0.66	-0.64
37	iodobenzene		0.88	0.84	0.92	0.91	0.94	0.93

38	acetophenone		0.26	0.10	0.51	0.51	0.53	0.53
39	3-methylphenol		0.08	0.09	0.10	0.10	0.12	0.12
40*	methyl benzoate		0.7	0.36	0.68	0.68	0.68	0.69
41*	4-chloroanisole		1.07	0.85	0.98	0.99	0.99	0.99
42	phenethyl alcohol		-0.46	-0.20	-0.39	-0.42	-0.41	-0.37
43	3-methylbenzl alcohol		-0.15	-0.06	0.19	0.18	0.17	0.20

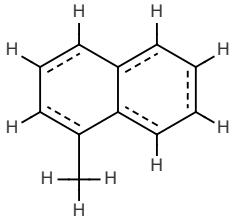
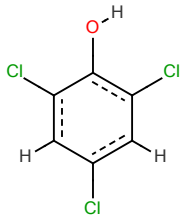
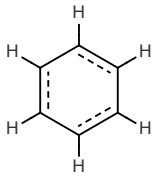
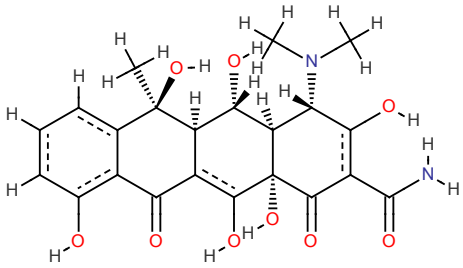
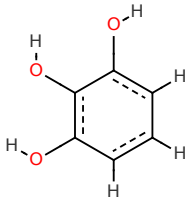
44	4-ethylphenol		0.62	0.54	0.39	0.38	0.38	0.40
45	3,5-dimethylphenol		0.49	0.68	0.54	0.53	0.52	0.54
46	ethyl benzoate		1.14	0.81	1.09	1.08	1.06	1.08
47	methyl 2-methylbenzoate		1.12	0.95	1.18	1.17	1.15	1.17
48	3-chlorophenol		0.62	0.17	0.12	0.53	0.17	0.71
49*	4-nitrotoluene		1.44	1.12	1.19	1.21	1.23	1.20

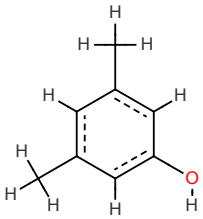
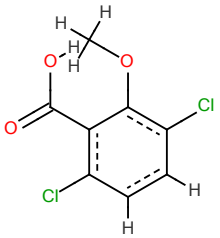
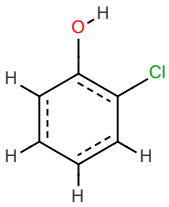
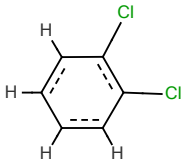
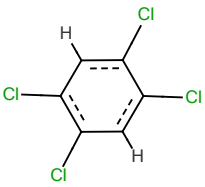
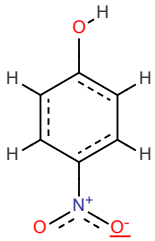
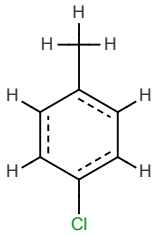
50	4-chloroacetophenone		1.28	1.44	1.33	1.33	1.33	1.32
51	3-bromophenol		0.79	0.70	0.68	0.68	0.71	0.70
52	1-methylnaphtalene		1.89	1.82	1.94	1.95	1.95	1.93
53	2-dichlorobenzene		0.56	0.62	0.67	0.67	0.70	0.69
54	3-dichlorobenzene		0.65	0.62	0.67	0.67	0.70	0.69
55*	4-dichlorobenzene		0.51	0.62	0.67	0.67	0.70	0.69
56*	isophorone		0.01	0.51	0.14	0.08	0.00	0.11

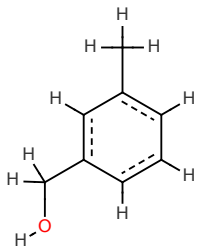
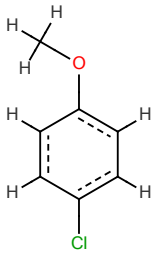
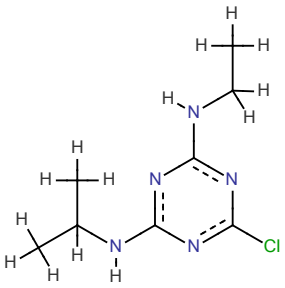
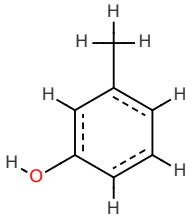
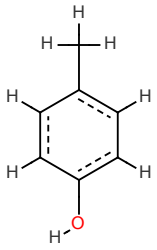
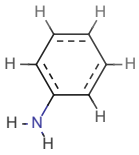
57	2-chloronaphthalene		2.73	2.57	1.83	1.84	1.86	1.83
58	azobenzene		2.72	2.80	2.90	2.92	2.94	2.89
59*	phenanthrene		3.29	3.01	2.87	2.88	2.90	2.85

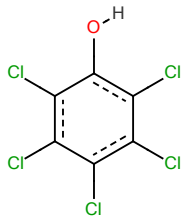
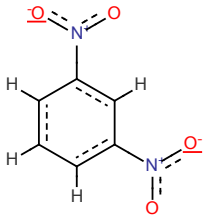
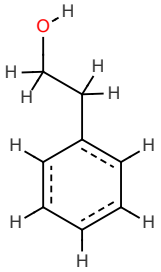
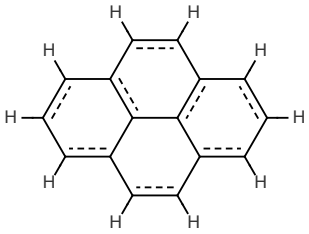
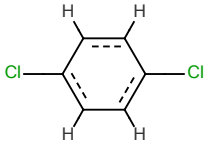
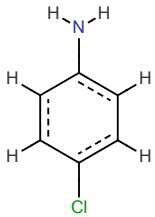
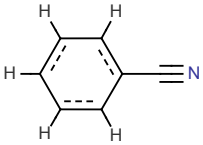
* denotes the test set compounds

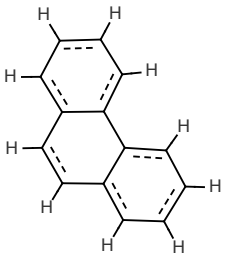
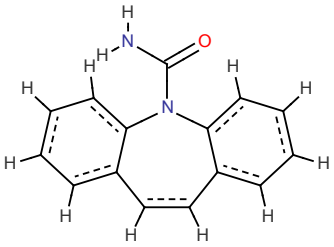
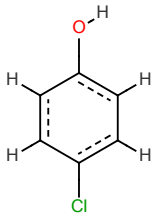
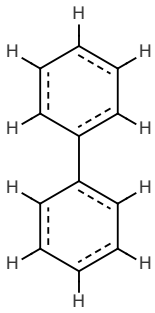
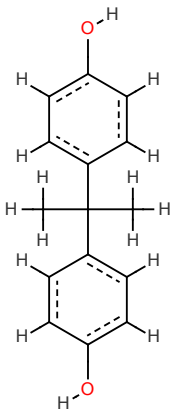
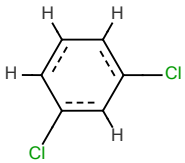
Table S2 The chemical name, experimental $logK_{SA}$ and calculated $logK_{SA}$ values of the MLR models developed from dataset 2

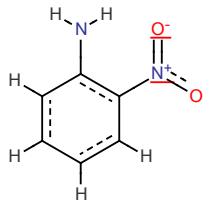
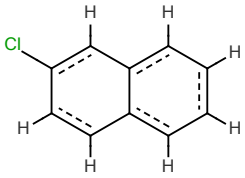
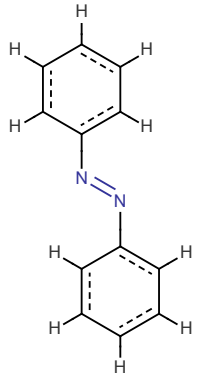
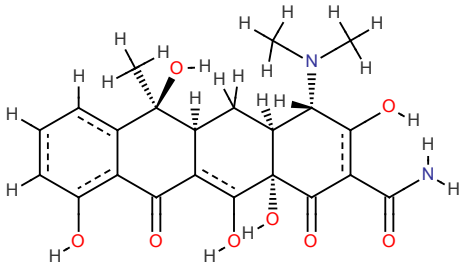
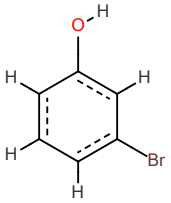
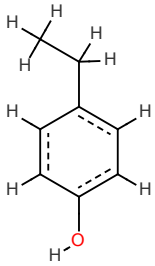
No.	Chemical name	Chemical Structure	$logK_{SA}$				
				Calc.			
			Expt.	Model N1	Model N2	Model N3	Model N4	Model N5
1*	1-Methylnaphthalene		-0.48	-0.33	-0.41	-0.40	-0.40	-0.36
2	2,4,6-trichlorophenol		-0.81	-0.62	-0.79	-1.28	-1.27	-0.61
3	benzene		-2.47	-2.90	-2.81	-2.52	-2.58	-2.82
4	oxytetracycline		-0.23	0.12	0.34	0.15	0.15	0.29
5	pyrogallol		-0.98	-1.45	-1.57	-1.37	-1.40	-1.36

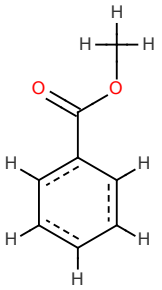
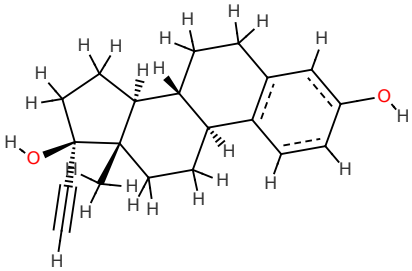
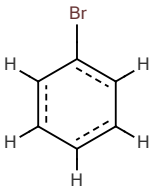
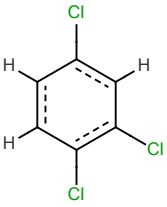
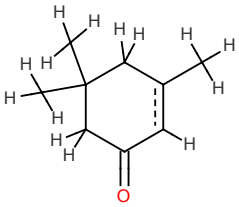
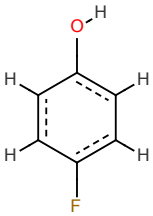
6	3,5-dimethylphenol		-1.88	-1.87	-1.88	-2.05	-2.09	-2.13
7	dicamba		-2.64	-2.17	-1.98	-1.72	-1.72	-2.59
8	2-chlorophenol		-2.16	-1.96	-2.01	-1.99	-2.02	-1.80
9	ortho-dichlorobenzene		-1.81	-1.08	-1.12	-1.40	-1.41	-1.19
10*	1,2,4,5-tetrachlorobenzene		0.2	0.07	-0.02	0.10	0.16	0.48
11	p-nitrophenol		-1.45	-1.14	-1.10	-1.10	-1.11	-1.37
12	4-chlorotoluene		-1.55	-1.39	-1.45	-1.69	-1.71	-1.70

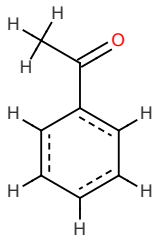
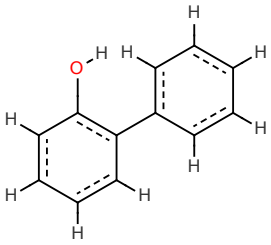
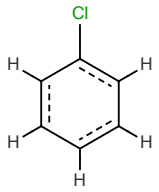
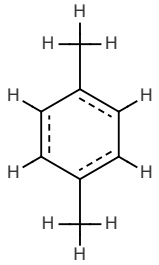
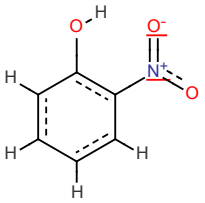
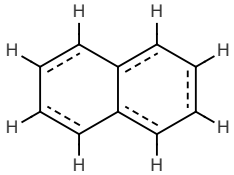
13	3-methylbenzyl alcohol		-2.52	-2.35	-2.75	-2.85	-2.28	-2.08
14	4-chloroanisole		-1.3	-1.92	-1.75	-1.96	-1.99	-1.62
15	atrazine		-0.55	-0.36	-1.34	-0.42	-0.40	-0.42
16	3-methylphenol		-2.29	-2.18	-2.21	-2.17	-2.22	-2.32
17*	4-methylphenol		-2.18	-2.18	-2.21	-2.17	-2.22	-2.30
18	aniline		-3.01	-3.22	-3.12	-2.97	-3.05	-3.07

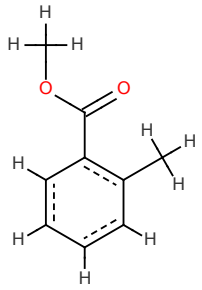
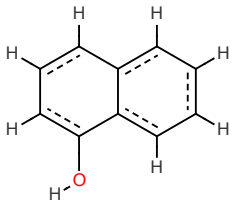
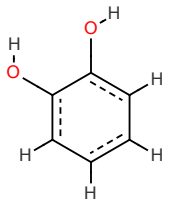
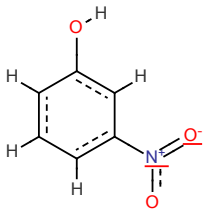
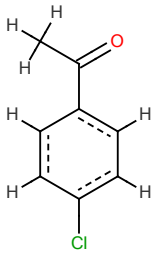
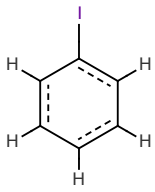
19	pentachlorophenol		-0.7	-0.88	-0.97	-0.92	-0.88	-0.89
20*	1,3-dinitrobenzene		-0.75	0.33	-0.13	0.28	0.32	-0.05
21*	phenylethylalcohol		-2.83	-2.53	-2.75	-2.85	-2.28	-2.20
22	pyrene		1.77	2.02	1.80	2.00	2.06	1.61
23	p-dichlorobenzene		-1.86	-2.04	-2.06	-1.40	-1.41	-1.42
24	4-chloroaniline		-2.9	-2.57	-2.52	-2.61	-2.68	-2.61
25*	benzonitrile		-2.33	-3.15	-2.78	-2.84	-2.91	-2.61

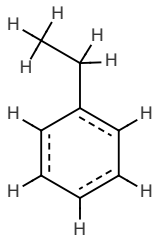
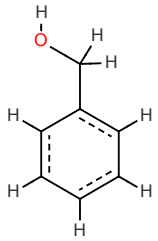
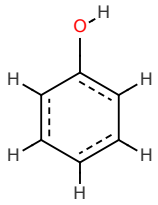
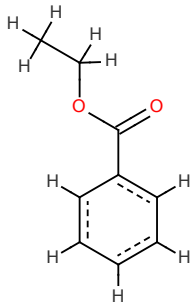
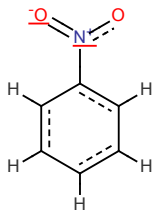
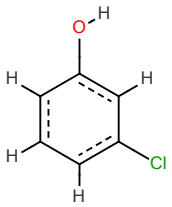
26	phenanthrene		1.13	0.96	0.89	0.92	0.96	0.76
27	carbamazepine		-1.47	-1.25	-0.73	-0.99	-1.03	-1.02
28	4-chlorophenol		-1.5	-1.96	-2.08	-2.07	-2.11	-2.03
29	biphenyl		-0.17	-0.20	-0.02	-0.23	-0.22	-0.32
30*	bisphenol A		-0.73	-0.20	0.20	-0.51	-0.50	-0.45
31*	m-dichlorobenzene		-1.72	-1.03	-1.12	-1.40	-1.41	-1.43

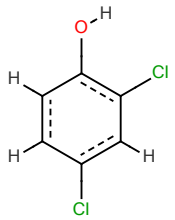
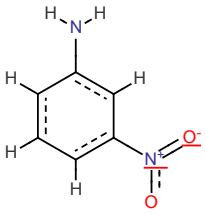
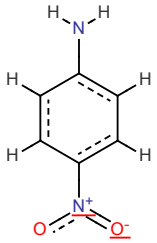
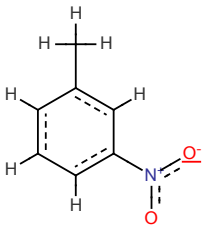
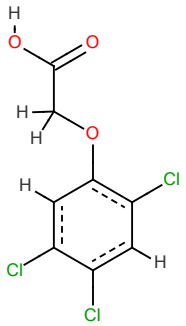
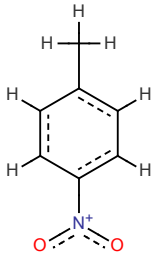
32	2-nitroaniline		-0.64	-1.46	-1.71	-1.37	-1.39	-1.38
33*	2-chloronaphthalene		0.36	0.10	-0.04	-0.07	-0.06	-0.14
34	azobenzene		0.35	-1.18	-0.63	-1.05	-1.08	-1.20
35	tetracycline		0.21	-0.38	-0.15	-0.25	-0.28	-0.12
36	3-bromophenol		-1.58	-1.67	-1.76	-1.75	-1.78	-1.93
37	4-ethylphenol		-1.75	-2.13	-1.88	-2.05	-2.09	-1.89

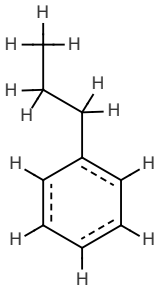
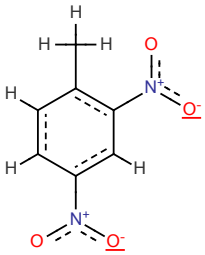
38*	methylbenzoate		-1.67	-1.75	-1.39	-1.60	-1.62	-1.45
39*	17α-ethynylestradiol		0.99	1.07	1.36	0.84	0.87	0.91
40	bromobenzene		-1.87	-1.80	-1.83	-1.85	-1.88	-1.96
41	1,2,4-trichlorobenzene		-1	-1.02	-1.08	-0.70	-0.67	-0.48
42	isophorone		-2.36	-1.92	-2.01	-2.32	-2.38	-2.37
43	4-fluorophenol		-2.69	-1.97	-2.03	-1.98	-2.02	-2.14

44	acetophenone		-2.11	-2.10	-1.90	-2.06	-2.10	-2.08
45	2-phenylphenol		-1.16	-0.81	-0.53	-0.71	-0.72	-0.75
46	chlorobenzene		-2.35	-2.00	-2.01	-2.01	-2.05	-2.12
47	p-xylene		-2.11	-1.73	-1.76	-1.95	-1.98	-1.96
48*	2-nitrophenol		-1.69	-1.17	-1.10	-1.10	-1.11	-1.25
49	naphthalene		-0.45	-0.80	-0.93	-0.65	-0.67	-0.74

50	methyl 2-methylbenzoate		-1.25	-1.53	-1.13	-1.53	-1.55	-1.23
51	1-naphthol		-1.24	-1.15	-1.20	-0.94	-0.97	-1.00
52*	catechol		-1.95	-1.99	-2.06	-1.82	-1.86	-1.89
53	3-nitrophenol		-1.32	-1.14	-1.10	-1.10	-1.11	-1.39
54*	4-chloroacetophenone		-1.09	-1.57	-1.45	-1.85	-1.88	-1.58
55	iodobenzene		-1.49	-1.59	-1.64	-1.69	-1.71	-1.80

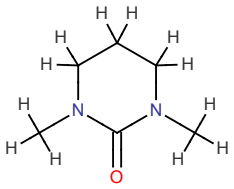
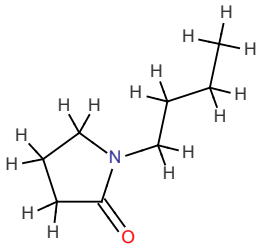
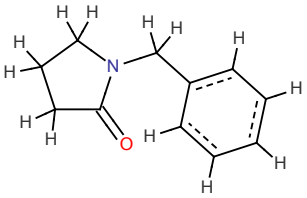
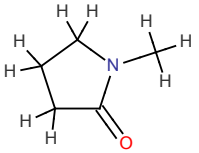
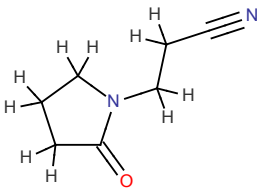
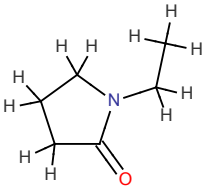
56	ethylbenzene		-2.18	-2.02	-1.76	-1.95	-1.98	-1.73
57	benzylalcohol		-3.27	-2.66	-3.04	-2.94	-2.38	-2.22
58	phenol		-2.73	-2.48	-2.50	-2.23	-2.28	-2.49
59	ethyl benzoate		-1.23	-1.66	-1.13	-1.53	-1.55	-1.44
60	nitrobenzene		-1.86	-1.46	-1.39	-1.38	-1.40	-1.58
61	3-chlorophenol		-1.75	-1.93	-2.04	-2.03	-2.06	-2.77

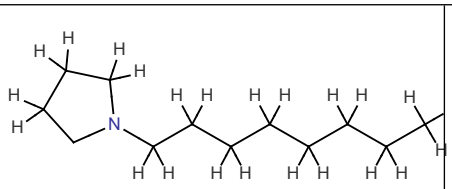
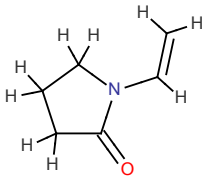
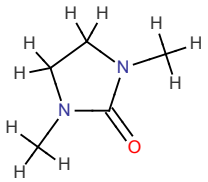
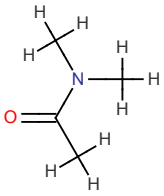
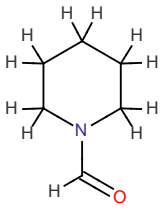
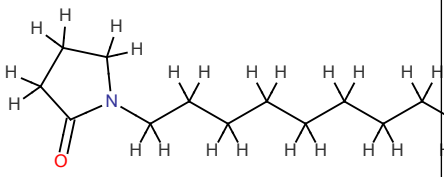
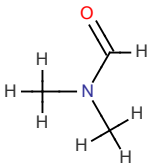
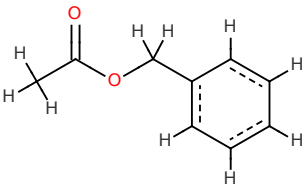
62*	2,4-dichlorophenol		-1.28	-1.34	-1.48	-1.72	-1.74	-1.33
63	3-nitroaniline		-1.53	-1.28	-1.71	-1.21	-1.23	-1.37
64*	4-nitroaniline		-1.29	-1.13	-1.71	-1.06	-1.07	-1.21
65	*m*-nitrotoluene		-1.17	-1.28	-1.17	-1.39	-1.40	-1.50
66	2,4,5-trichlorophenoxyacetic acid		-2.51	-2.60	-2.62	-2.64	-2.64	-1.27
67	4-nitrotoluene		-0.93	-1.28	-1.17	-1.39	-1.40	-1.48

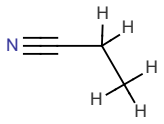
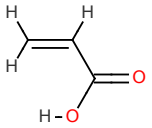
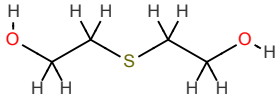
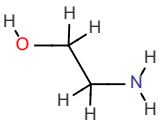
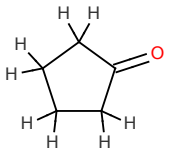
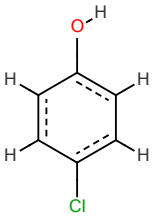
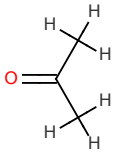
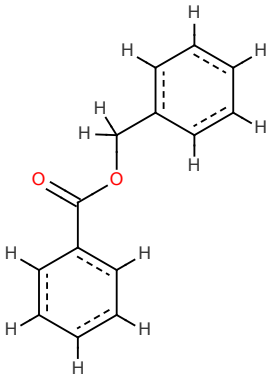
68*	propylbenzene		-1.61	-1.64	-1.23	-1.64	-1.65	-1.51
69	2,4-dinitrotoluene		0.21	0.32	-0.06	0.13	0.17	-0.01

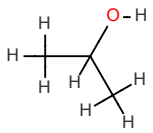
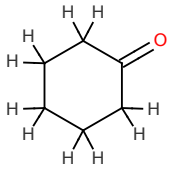
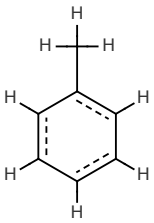
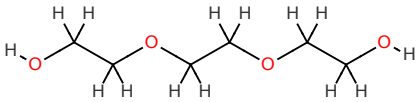
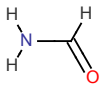
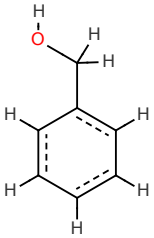
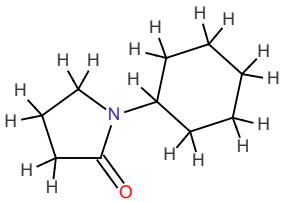
* denotes the test set compounds

Table S3 The chemical name, experimental $\log C_{max}$ and calculated $\log C_{max}$ values of model P1 developed from dataset 3

No.	Chemical name	Chemical structure	$\log C_{max}$	
			Expt.	Calc.
1	1,3-Dimethyltetrahydro-2(1H)-pyrimidinone		-0.19	-0.32
2	1-Butylpyrrolidin-2-one		-0.55	-0.78
3	1-Benzylpyrrolidin-2-one		-0.74	-0.64
4*	1-Methylpyrrolidin-2-one		-0.94	-1.21
5	3-(2-Oxo-1-pyrrolidinyl)propanenitrile		-0.94	-0.97
6	*N*-Ethyl-pyrrolidinone		-1.00	-0.86

7	*N*-Octyl-pyrrolidone		-1.04	-1.12
8	*N*-Vinyl-pyrrolidinone		-1.08	-1.16
9	Dimethyl-imidazolidinone		-1.08	-0.86
10	Dimethylacetamide		-1.39	-1.55
11*	*N*-Formyl-piperidine		-1.41	-0.93
12	*N*-Dodecyl-pyrrolidone		-1.52	-1.29
13*	Dimethylformamide		-1.64	-1.58
14	Benzyl acetate		-1.72	-2.09

15*	Propionitrile		-1.82	-1.86
16	Acrylic acid		-1.86	-2.25
17	2,2'-thiodiethanol		-1.87	-2.01
18	Ethanolamine		-1.88	-1.85
19*	Cyclopentanone		-1.89	-2.00
20	Chlorophenol		-1.92	-2.02
21	Acetone		-1.96	-1.88
22	Benzyl benzoate		-1.96	-2.21

23	Isopropyl alcohol		-1.98	-2.00
24	Cyclohexanone		-2.17	-2.30
25*	Toluene		-2.30	-2.23
26	Triethyleneglycol		-2.43	-2.47
27	Formamide		-3.52	-3.54
28	Benzyl alcohol		-3.55	-2.19
29*	*N*-Cyclohexyl-pyrrolidinone		0.54	0.58

* denotes the test set compounds

Data pretreatment: It was performed to exclude constant and highly inter-correlated descriptors based on variance and correlation cut-off value using the software

developed in our laboratory.[1]

Stepwise regression: Stepwise regression is a method for developing of multiple linear regression equations step by step which are altered by adding or removing a predictor variable. Forward selection and backward elimination are two parts of stepwise regression method. Forward selection starts with no variable and then 'statistically significant' variables are included one by one. In case of backward elimination, initially, all the candidate variables are selected and then deleting statistically insignificant variables one by one. The objective function of the selection in stepwise regression may be "F-for-inclusion", also known as "steeping criteria". The F-value is square of t-value of incoming variable which signifies the regression coefficient. In stepwise regression process, a multiple term linear regression equation is built up using a "stepping criteria" also known as "Fisher criteria". In this present study, we have fixed the "stepping criteria" or "Fisher criteria" F=4 to enter and F=3.9 to remove[1] because at this value of the F-criterion, the descriptors are considered to be significant at the 95% confidence level. In this study, the stepwise regression has been performed using initial pool of descriptors and selected the model descriptors and kept aside. After removing the selected descriptors from the initial pool, stepwise regression was done again and the model descriptors were selected and so on. The variable selection approach using stepwise regression technique was applied for the datasets containing 59 and 69 organic pollutants only. In this way, we have selected 43 descriptors (reduced pool) for the dataset containing

59 organic pollutants and 47 descriptors (reduced pool) for the dataset containing 69 organic pollutants.

***Genetic Method*:** Genetic algorithm (GA) is taken from an analogy with the evolution of DNA. This algorithm was developed from the concept of Holland's genetic algorithm (1975)[2] and Friedman's (1990)[3] multivariate adaptive regression splines (MARS) algorithm.[4] Roger and Hopfinger first introduced this technique in QSAR studies[5] GA is an iterative heuristic approach that mimics the natural selection process. This heuristic method is used to speed up the process and find out a satisfactory solution where exhaustive search is impractical. GA generates solution to optimize problems using techniques inspired by natural evaluation. Different steps involved in GA for selection of significant variables during QSAR model development are inheritance, crossover, mutation and selection. Here, the significant variables are selected based on the MAE-based criteria. In case of dataset containing 29 organic contaminants, we have run GA using logarithmic value of defined dispersibility index ($\log C_{max}$) of single wall carbon nanotubes (SWCNTs) for development of some equations. Finally, we have made a reduced pool using the descriptors obtained from GA equations for further study.

2.4.4. Best subset selection: The best descriptor combination out of total descriptor sets is selected by the best subset selection software developed in our laboratory[1] by evaluating the possible combinations of descriptors. Different statistical parameters

such as R^2, Q^2, Q^2_{F1}, Q^2_{F2} are also calculated by this method. In this work, we have run best subset selection using the reduced pool of descriptors obtained after applying different variable selection approaches and selected best five multiple linear regression models based on MAE based criteria.[6]

Results and discussion

Table S4 Definition, frequency and contribution of all the descriptors obtained from the MLR models (models developed by using 59 organic pollutants)

Sl. no.	Name of descriptors	Contribution	Discussion	Mechanism	Frequency of descriptors
1	X0v	+ve	Valence connectivity index order 0.	Hydrophobic interaction	5
2	nArOH	+ve	Number of aromatic hydroxyls.	Hydrogen bonding interactions.	1
3	B01[C-O]	-ve	Presence/absence of C-O fragment at topological distance 1.		1
4	B06[C-Cl]	+ve	Presence/absence of C-Cl fragment at topological distance 6.	Here, size plays an important role for adsorption affinity. The size enhances the surface area of molecules which	1

				regulate the hydrophobic interactions between organic pollutants and MWCNTs.	
5	Ui	+ve	Unsaturation index.	π-π interactions.	1
6	F03[O-O]	+ve	Frequency of O-O at topological distance 3.	Enhanced electron density, electrostatic interactions.	2
7	F04[N-O]	+ve	Frequency of N-O at topological distance 4.	Hydrogen bonding and electrostatic interactions.	1
8	ETA_BetaP	+ve	A measure of electronic features of the molecules relative to molecular size.	Electrostatic interactions.	4
9	minsCH$_3$	+ve	Minimum atom E-state –CH$_3$.	Hydrophobic interactions.	4
10	B03[O-O]	+ve	Presence/absence of O-O at topological distance 3.	Electrostatic interactions, hydrogen bonding.	1

11	nHBint4	+ve	Count of E-state descriptors of strength for potential hydrogen bond of path length 4.	Hydrogen bonding interaction.	2

Table S5 Definition and contribution of all the descriptors obtained from the MLR models (models developed by using 69 organic pollutants)

Sl. No.	Name of descriptor	Contribution	Discussion	Mechanism	Frequency of descriptors
1	ETA_Epsilon_3	+ve	Summation of electronegativity or ϵ value relative to the total no. of atoms including hydrogen in the connected molecular graph of the reference alkane	Electrostatic interaction	2
2	X1A	-ve	Average connectivity index of order 1	Hydrophobic interaction	1

3	X2A	-ve	Average connectivity index of order 2	Hydrophobic interaction	1
4	nOHp	-ve	Number of primary alcohols	π-π interaction	2
5	VAdjMat	+ve	Vertex adjacency information (magnitude)	Hydrophobic interaction	1
6	F04(O-Cl)	-ve	Number of (O..Cl) fragment at topological distance 4	π-π interaction	1
7	B05[Cl-Cl]	-ve	Presence/absence of Cl - Cl at topological distance 5	π-π interaction	2
8	MLOGP2	+ve	Squared Moriguchi octanol-water partition coefficient (logP^2)	Hydrophobic interaction	5
9	T(N..N)	+ve	Sum of topological distances between N..N	π-π interaction	4
10	O%	+ve	Percentage of O atoms	Hydrogen bonding, π-π interaction, electrostatic interaction	5
11	T(O..Cl)	-ve	Sum of topological distances between O..Cl	π-π interaction	4

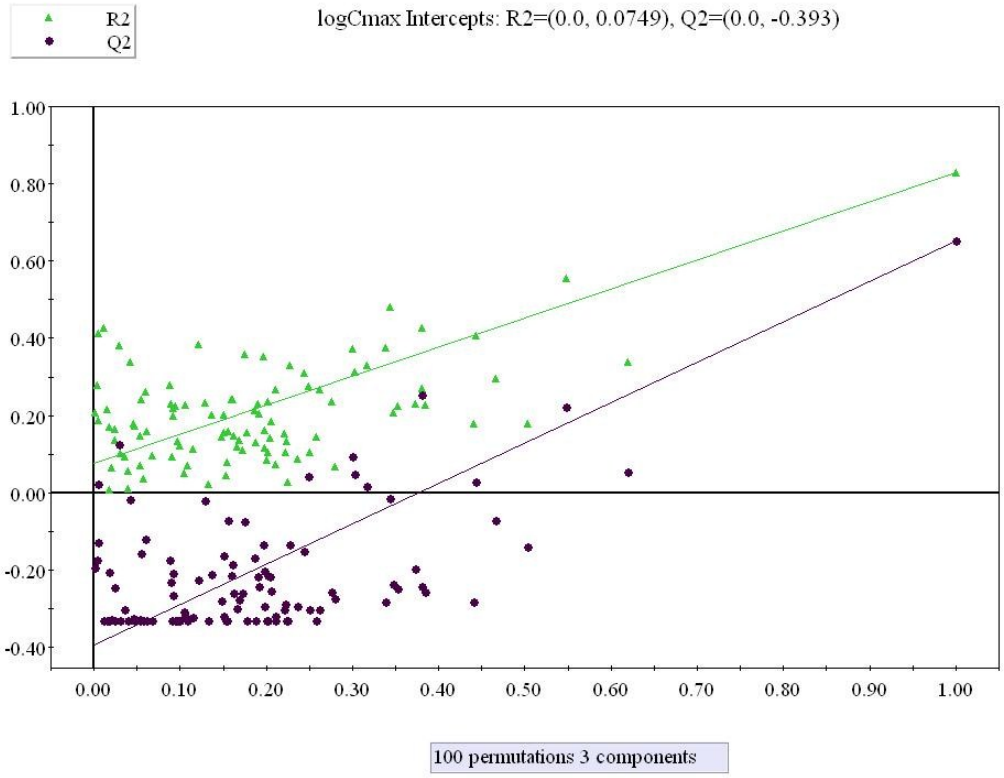


Fig. S1. Randomization plot of PLS model (model P1).

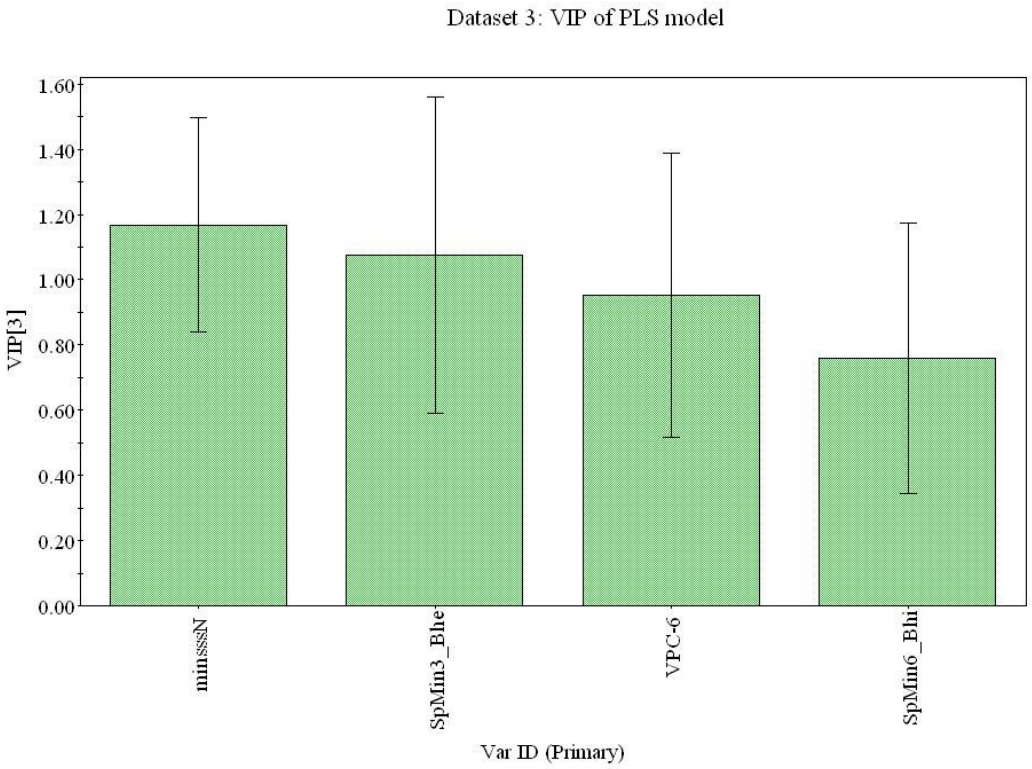


Fig. S2. VIP plot for PLS model (model P1)

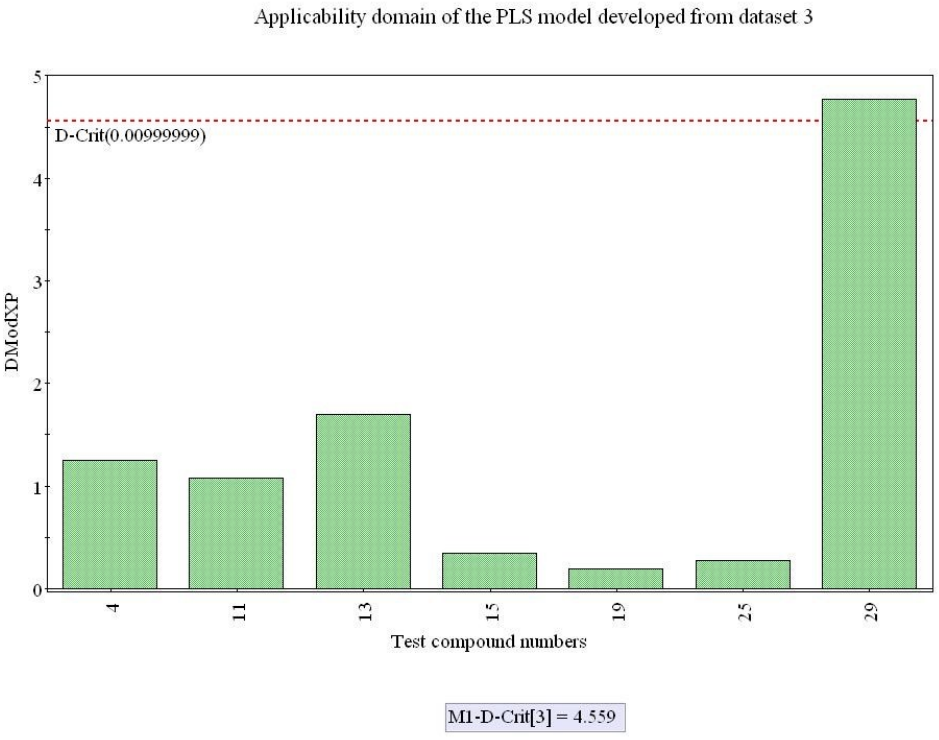


Fig. S3. DModX plot for PLS model (model P1).

References

1 (http://teqip.jdvu.ac.in/QSAR_Tools/DTCLab)

2 J. H. Holland, Adaptation in natural and artificial systems.An introductory analysis with application to biology, control, and artificial intelligence, University of Michigan Press, Ann Arbor, MI, 1975.

3 J. Friedman, Multivariate Adaptive Regression Splines, Technical Report No. 102, Laboratory for Computational Statistics, Department of Statistics, Stanford University, Stanford, CA, Nov 1988, (revised Aug 1990)

4 A. J. Hopfinger, M. G. Koehler and D. Rogers, Molecular modeling of polymers, quantitative structure–property relationship analyses of multicomponent systems containing polymers, *In Macromolecular Symposia.H¨uthig&WepfVerlag*., 1995, **98**, 1087–1100.

5 D. Rogers and A. J. Hopfinger, Application of genetic function approximation to quantitative structure-activity relationships and quantitative structure-property relationships, *J. Chem. Inf. Comput. Sci.,* 1994, **34**, 854–866.

6 K. Roy, R. N. Das, P. Ambure and R. B. Aher, Be aware of error measures. Further studies on validation of predictive QSAR models, *Chemom.Intell. Lab. Syst*., 2016,**152**, 18-33.