QSPR Modeling of Lipid-Water Partition Coefficient by Optimization of Correlation Weights of Local Graph Invariants

Andrey A. Toropov† and Kunal Roy*,‡

Uzbekistan Academy of Sciences Research Institute 'Algorithm-Engineering',
700125 F. Khodjaeva Street 25, Tashkent, Uzbekistan, and Drug Theoretics and Cheminformatics Lab
Division of Medicinal and Pharmaceutical Chemistry, Department of Pharmaceutical Technology,
Jadavpur University, Kolkata 700 032, India

Received September 9, 2003

The optimization of correlation weights scheme was applied to model lipid-water partition coefficient (log *P*) of two sets of diverse functional aliphatic and aromatic compounds. In both cases, the optimized descriptors formulated based on the data of training sets generated statistically acceptable relations for the corresponding training sets, test sets, and combined sets. When the relations of log *P* values with the optimized molecular descriptors formulated based on the data of the training sets were used for calculation of log *P* values of the corresponding training sets, r_{pred}^2 values were found to be satisfactory (above 0.99) in both cases, which is indicative of the predictive potential of the scheme. The results indicate promising potential of the optimization of a correlation weights scheme in modeling studies.

INTRODUCTION

Hydrophobic interactions are of critical importance in many areas of chemistry, including enzyme−ligand interactions, drug−receptor interactions, transport of drug to the active site, the assembly of lipids in biomembranes, aggregation of surfactants, coagulation, and detergency, etc.[1,2] Hydrophobic "bonding" is actually not bond formation at all, but rather the tendency of hydrophobic molecules or hydrophobic parts of molecules to avoid water because they are not readily accommodated in the highly ordered hydrogen-bonded structure of water.[3] Hydrophobic interaction is favored thermodynamically because of increased entropy of the water molecules that accompanies the association of nonpolar molecules, which squeeze out water. The hydrophobic "bonding" resulting from an unwelcome reception of nonpolar molecules in water involves van der Waals forces, hydrogen bonding of water molecules in 3*D* structure, and other interactions.[4]

Hydrophobicities of solutes are readily determined by measuring partition coefficients (log *P*) using the shake-flask method (which involves distribution of a compound between an aqueous phase and an organic phase) and reversed phase high performance liquid chromatography.[5,6] Since the experimental determination of the partition coefficient of a large set of compounds is a very tedious job, several methods of calculations of log *P* values have been proposed by different groups of authors, e.g., Rekker et al.,[7] Leo et al.,[8,9] Crippen et al.,[10,11] Bodor et al.,[12] Klopman et al.,[13] Moriguchi et al.,[14] and Suzuki et al.[15] Many studies on the modeling of log *P* values using topological, topographic, quantum chemical, and other descriptors have been reported[16−22] where log *P* values have been the response variable to explore suitability of the descriptors/schemes in QSPR studies.

Selection of appropriate descriptors from the plethora of available descriptors is a real problem in modeling studies. One has to take care that such descriptors are chosen which extract a maximum amount of chemical information and, at the same time, the descriptors used in a multiple regression equation are not intercorrelated among themselves. The concept of flexible topological descriptors, originally introduced by Randic,[23−25] is a major breakthrough in this regard. The difficulties of multiple regression are not present in such an approach, which is based on regression with a single descriptor. Unlike usual topological descriptors, flexible topological descriptors do not have a definite predetermined formalism, which can be applied to any sets of compounds for the modeling of biological activity or physicochemical property. The formalism of such descriptors is defined based on an optimization procedure to obtain the best relation for a particular data set. Thus, the definition of the descriptors will vary from a data set to the other set, and the ultimate objective of the iterative optimization procedure is to get the best predictive model. Several promising descriptors have been proposed in this line, and their utilities have also been explored.[26−33] Among these descriptors, an interesting sort of flexible descriptors is based on the optimized correlation weights of the local graph invariants which was originally introduced by one of us (AAT).[31−33] This scheme has been successfully attempted to model different sets of biological activity and physicochemical property data.[34−41] In the present communication, we have applied the optimization of correlation weights scheme for modeling the lipid-water partition coefficient to show usefulness of the scheme. Two sets of data have been used for the present analysis. Though mostly straight chain aliphatic compounds have been con-

* Corresponding author phone: +91-33-2414 6676; e-mail: kunalroy_in@yahoo.com.
† Uzbekistan Academy of Sciences Research Institute 'Algorithm-Engineering'.
‡ Jadavpur University.

10.1021/ci034200g CCC: $27.50
Published on Web 12/12/2003

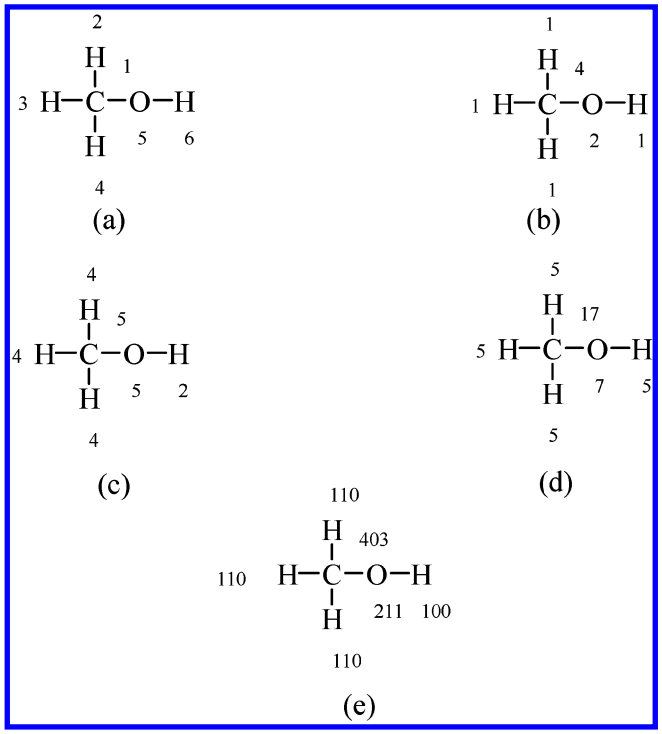

Figure 1. Local invariants of the LHFG of methanol: (a) arbitrary numbering of the vertices; (b) ^{0}EC; (c) ^{1}EC; (d) ^{2}EC; (e) NNC.

sidered in the first set, it contains a few alicyclic compounds also. The second set contains polynuclear aromatic compounds and aliphatic compounds of diverse functionality.

MATERIALS AND METHODS

The molecular descriptor was calculated based on labeled hydrogen filled graph (LHFG) in the following manner:

$$\mathrm{DCW}(a_k,\mathrm{LI}_k) = \sum_{k=1}^{n} \mathrm{CW}(a_k) + \sum_{k=1}^{n} \mathrm{CW}(\mathrm{LI}_k) \qquad (1)$$

In the above equation, the DCW term represents the molecular descriptor, CW terms represent the correlation weights, a_k is the chemical element of the kth vertex of the LHFG, and LI_k is the numerical value of some local invariants of the LHFG. As local invariants, we have used the Morgan extended connectivity of zero, first, and second orders (denoted by $^0\mathrm{EC}_k$, $^1\mathrm{EC}_k$, and $^2\mathrm{EC}_k$, respectively) and nearest neighboring codes (NNC).

Zero-order Morgan connectivity of an atom k is the adjacency count of that atom. Again, the first-order Morgan connectivity value of atom k is the sum of the zero-order Morgan connectivity values of the atoms that are connected to atom k. Similarly, the second-order Morgan connectivity value of atom k is the sum of the first-order Morgan connectivity values of the atoms that are connected to atom k. An example of calculations of Morgan connectivity indices for methanol is given in Figure 1.

The NNC of the kth vertex of the LHFG is calculated as

$$\mathrm{NNC}_k = 100N_\mathrm{T} + 10N_\mathrm{C} + N_\mathrm{H} \qquad (2)$$

In the above equation, N_T, N_C, and N_H represent the total number of vertices and the number of carbons and number of hydrogens, respectively, connected to the kth vertex. An example of the calculation of NNC for methanol is shown in Figure 1. Morgan connectivity values are local topological descriptors, while NNCs are mathematical functions of both number and kind of neighbors.

The partition coefficient (log P) values of set 1 (n = 168)[42,43] and set 2 (n = 139)[44] compounds were taken from the literature. In each case, the data set was divided into a training set and a test set, as listed in Tables 1 and 2. The starting value of each correlation weight was 1 and using the Monte Carlo iterative optimization procedure,[31−33] the best values of correlation weights [CW(a_k) and CW(LI_k)] (which give largest possible correlation coefficient between the log P values of the training set and the molecular descriptor [DCW]) were found out. Based on the optimized correlation weights, the molecular descriptor was finally defined, and this was then used to derive all the relations with log P values of both the training and test sets using the least-squares method of regression.

$$\log P = \alpha + \beta^*\mathrm{DCW}(a_k, \mathrm{LI}_k) \qquad (3)$$

The optimization of correlation weights was done using a program developed by one of the authors (AAT).[45] Least-squares linear regression analyses were done using a software program *RRR98* developed by the other author (KR).[46] Statistical quality of the equations[47] was judged by examining the parameters such as r_a^2 (adjusted r^2, i.e., explained variance), r (correlation coefficient), F (variance ratio) with df (degree of freedom), s (standard error of estimate), and *AVRES* (average of absolute values of residuals). Significance of the regression coefficients was judged by a 't' test. A compound was considered as an outlier for a particular equation when the residual exceeded twice the standard error of estimate of the equation. PRESS statistics were calculated for the training set by the "leave-one-out" (LOO) technique[48,49] using programs *KRPRES1* and *KRPRES2*,[35] and q^2 (cross-validation r^2 or predicted variance) along with SDEP (standard deviation of error of predictions) values were reported. The predictive capacity of the model was found out by its application on the test set and the value of r_pred^2.

RESULTS AND DISCUSSION

Among the molecular descriptors (DCW) defined in different ways [DCW(a_k, $^0\mathrm{EC}_k$), DCW(a_k, $^1\mathrm{EC}_k$), DCW(a_k, $^2\mathrm{EC}_k$), DCW(a_k, NNC_k)], DCW(a_k, NNC_k) gave the best correlation between the descriptor and the property (log P). The values of the optimized correlation weights of different local invariants (a_k and NNC_k) are shown in Tables 3 (set 1) and 4 (set 2). Based on the correlation weights as listed in Tables 3 and 4, the molecular descriptors (DCW) were calculated for all the compounds of sets 1 and 2 as listed in Tables 1 and 2. The calculation of the descriptor for methanol has been illustrated in Table 5.

QSPR for Set 1 Compounds. The results of the relations of log P values of different subsets of the training set with the molecular descriptor (DCW) are given in Table 6. It is observed that the descriptor could explain the variance of log P values to the extent of 98.0% for alcohols (n = 28), 98.7% for amines (n = 12), 99.9% for acids (n = 4), 98.9% for esters (n = 6), 91.9% for hydrocarbons (n = 15), and 98.3% for ketones (n = 8). However, in the cases of ethers (n = 5) and halocarbons (n = 6), poor relations were

Table 1. Optimized Molecular Descriptor and Observed and Calculated log *P* Values of Diverse Functional Compounds (Set 1)

Sl. no.	compound name	molecular descriptor (DCW)	log *P* obs.[a]	log *P* calc.[b]	log *P* res.[b]	Sl. no.	compound name	molecular descriptor (DCW)	log *P* obs.[a]	log *P* calc.[b]	log *P* res.[b]
				Training Set							
1	methanol	−0.635	−0.66	−0.55	−0.11	43	hexanoic acid	7.815	1.88	1.74	0.14
2	*n*-propanol	2.540	0.34	0.31	0.03	44	decanoic acid	15.195	4.09	3.73	0.36
3	*n*-pentanol	6.230	1.40	1.31	0.09	45	ethyl formate	2.247	0.23	0.23	0.00
4	*n*-hepatnol	9.920	2.34	2.31	0.03	46	ethyl acetate	4.441	0.73	0.82	−0.09
5	2-propanol	1.862	0.14	0.13	0.01	47	*n*-butyl acetate	8.131	1.73	1.82	−0.09
6	tert-butanol	2.678	0.37	0.35	0.02	48	methylpropionate	4.956	0.73	0.96	−0.23
7	2-methylbutanol	5.422	1.14	1.09	0.05	49	ethylbutyrate	8.131	1.73	1.82	−0.09
8	3-pentanol	5.552	1.14	1.12	0.02	50	*n*-pentyl acetate	9.976	2.23	2.32	−0.09
9	2-methyl-2-butanol	4.523	0.89	0.85	0.04	51	methyl *n*-butyl ether	7.061	1.53	1.53	0.00
10	2,2-dimethyl-1-propanol	5.600	1.36	1.14	0.22	52	methyl *tert*-butyl ether	5.354	1.06	1.07	−0.01
11	2-hexanol	7.397	1.61	1.62	−0.01	53	ethyl propyl ether	6.546	1.53	1.39	0.14
12	2-methyl-2-pentanol	6.368	1.39	1.34	0.05	54	ethylisopropyl ether	5.868	1.33	1.21	0.12
13	3-methyl-2-pentanol	6.589	1.41	1.4	0.01	55	ethyl cyclopropyl ether	4.332	1.24	0.79	0.45
14	3,3-dimethyl-1-butanol	7.445	1.86	1.64	0.22	56	chloroform	6.974	1.97	1.51	0.46
15	2-methyl-2-hexanol	8.213	1.87	1.84	0.03	57	methyl iodide	7.610	1.69	1.68	0.01
16	3-ethyl-3-pentanol	8.213	1.87	1.84	0.03	58	ethyl iodide	8.940	2.00	2.04	−0.04
17	2,4-dimethyl-2-pentanol	7.405	1.67	1.63	0.04	59	1-bromopropane	9.170	2.10	2.10	0.00
18	2,2-dimethyl-3-pentanol	8.612	1.69	1.95	−0.26	60	1-chlorobutane	11.994	2.39	2.87	−0.48
19	2,2,3-trimethyl-3-pentanol	9.428	1.99	2.17	−0.18	61	1-fluoropentane	10.073	2.33	2.35	−0.02
20	cyclohexanol	5.861	1.23	1.21	0.02	62	*n*-heptane	14.451	3.50	3.53	−0.03
21	3-penten-2-ol	4.132	0.81	0.74	0.07	63	cyclohexane	11.070	2.46	2.62	−0.16
22	1-hexen-3-ol	6.293	1.31	1.32	−0.01	64	cycloheptane	12.915	3.50	3.12	0.38
23	2-methyl-4-penten-3-ol	5.485	1.11	1.11	0.00	65	methylcyclohexane	12.107	2.76	2.90	−0.14
24	2,2,2-trifluoroethanol	2.898	0.41	0.41	0.00	66	cyclooctane	14.760	3.28	3.61	−0.33
25	ethylene glycol	−3.836	−1.93	−1.41	−0.52	67	1,2-dimethylcyclohexane	13.144	3.06	3.18	−0.12
26	allyl alcohol	1.436	0.17	0.01	0.16	68	2-heptene	13.031	3.20	3.15	0.05
27	*sec*-butanol	3.707	0.61	0.63	−0.02	69	4-methyl-1-pentene	10.694	2.50	2.51	−0.01
28	2,3-butanediol	−1.502	−0.92	−0.78	−0.14	70	1,6-heptadiene	12.243	2.90	2.93	−0.03
29	methylamine	−0.325	−0.57	−0.47	−0.10	71	1,5-hexadiene	10.398	2.40	2.43	−0.03
30	*n*-propylamine	2.850	0.48	0.39	0.09	72	cyclopentene	7.805	1.75	1.73	0.02
31	*n*-butylamine	4.695	0.75	0.89	−0.14	73	cycloheptene	11.495	2.57	2.73	−0.16
32	*n*-hexylamine	8.385	1.98	1.89	0.09	74	1-heptyne	12.360	2.98	2.96	0.02
33	isobutylamine	3.887	0.73	0.67	0.06	75	1-nonyne	16.050	3.98	3.96	0.02
34	methylethylamine	2.635	0.15	0.34	−0.19	76	1,8-nonadiyne	13.959	3.46	3.40	0.06
35	triethylamine	6.295	1.44	1.33	0.12	77	acetone	0.842	0.21	−0.15	0.36
36	di-*n*-butylamine	11.345	2.68	2.69	−0.01	78	2-pentanone	4.532	0.79	0.85	−0.06
37	*n*-propyl-*n*-butylamine	9.500	2.12	2.19	−0.07	79	3-pentanone	4.532	0.79	0.85	−0.06
38	*n*-propyl-*sec*-butylamine	8.822	1.91	2.01	−0.10	80	3-methyl-2-butanone	3.724	0.59	0.63	−0.04
39	*n*-propyl-isobutylamine	8.692	2.07	1.97	0.10	81	4-methyl-2-pentanone	5.569	1.09	1.13	−0.04
40	dimethyl-*n*-butylamine	7.325	1.70	1.60	0.10	82	2-heptanone	8.222	1.79	1.85	−0.06
41	acetic acid	0.435	−0.17	−0.26	0.09	83	5-nonanone	11.912	2.79	2.84	−0.05
42	propionic acid	2.280	0.25	0.24	0.01	84	2-nonanone	11.912	2.79	2.84	−0.05
				Test Set							
85	ethanol	0.695	−0.32	−0.19	−0.13	107	1-ethynylcyclohexanol	6.431	1.73	1.36	0.37
86	*n*-butanol	4.385	0.88	0.81	0.07	108	ethylamine	1.005	−0.13	−0.11	−0.02
87	n-hexanol	8.075	1.84	1.81	0.03	109	*n*-pentylamine	6.540	1.49	1.39	0.10
88	*n*-octanol	11.765	2.84	2.80	0.04	110	*n*-heptylamine	10.230	2.57	2.39	0.18
89	*n*-nonanol	13.610	3.15	3.30	−0.15	111	*sec*-butylamine	4.017	0.74	0.71	0.03
90	isobutanol	3.577	0.61	0.59	0.02	112	2-aminooctane	11.397	2.82	2.70	0.12
91	isopentanol	5.422	1.14	1.09	0.05	113	cyclohexylamine	6.171	1.49	1.29	0.20
92	1-methylbutanol	5.552	1.14	1.12	0.02	114	isopropylamine	2.172	0.26	0.21	0.05
93	3-methyl-2-butanol	4.744	0.91	0.91	0.00	115	di-*n*-propylamine	7.655	1.67	1.69	−0.02
94	3-hexanol	7.397	1.61	1.62	−0.01	116	diethylamine	3.965	0.57	0.69	−0.12
95	3-methyl-3-pentanol	6.368	1.39	1.34	0.05	117	methyl-*n*-butylamine	6.325	1.33	1.33	0.00
96	2-methyl-3-pentanol	6.589	1.41	1.40	0.01	118	piperidine	3.923	0.85	0.68	0.17
97	2-methyl-4-pentanol	6.589	1.41	1.40	0.01	119	ethylisopropylamine	5.132	0.93	1.01	−0.08
98	2,3-dimethyl-2-butanol	5.560	1.17	1.13	0.04	120	trimethylamine	2.305	0.27	0.25	0.02
99	3,3-dimethyl-2-butanol	6.767	1.19	1.45	−0.26	121	butyric acid	4.125	0.79	0.74	0.05
100	3-methyl-3-hexanol	8.213	1.87	1.84	0.03	122	*n*-propyl formate	4.092	0.73	0.73	0.00
101	2,3-dimethyl-2-pentanol	7.405	1.67	1.63	0.04	123	methyl acetate	3.111	0.23	0.46	−0.23
102	2,3-dimethyl-3-pentanol	7.405	1.67	1.63	0.04	124	*n*-propyl acetate	6.286	1.23	1.32	−0.09
103	2,4-dimethyl-3-pentanol	7.626	1.71	1.68	0.03	125	isopropyl acetate	5.608	1.03	1.14	−0.11
104	4-penten-1-ol	5.126	1.04	1.01	0.03	126	*sec*-butyl acetate	7.453	1.53	1.64	−0.11
105	1-penten-3-ol	4.448	0.81	0.83	−0.02	127	methyl butyrate	6.801	1.23	1.46	−0.23
106	2-hexen-4-ol	5.977	1.31	1.24	0.07	128	ethyl hexanoate	11.821	2.73	2.82	−0.09

Table 1 (Continued)

Sl. no.	compound name	molecular descriptor (DCW)	log *P* obs.[a]	log *P* calc.[b]	log *P* res.[b]	Sl. no.	compound name	molecular descriptor (DCW)	log *P* obs.[a]	log *P* calc.[b]	log *P* res.[b]
				Test Set							
129	ethyl heptanoate	13.666	3.23	3.32	−0.09	149	*n*-octane	16.296	4.00	4.03	−0.03
130	ethyl octanoate	15.511	3.73	3.82	−0.09	150	cyclopentane	9.225	2.05	2.12	−0.07
131	ethyl nonanoate	17.356	4.23	4.32	−0.09	151	methylcyclopentane	10.262	2.35	2.40	−0.05
132	ethyl decanoate	19.201	4.73	4.81	−0.08	152	1-pentene	9.657	2.20	2.23	−0.03
133	ethyl propionate	6.286	1.23	1.32	−0.09	153	2-pentene	9.341	2.20	2.15	0.05
134	ethyl isobutyrate	7.323	1.53	1.60	−0.07	154	1-hexene	11.502	2.70	2.73	−0.03
135	*n*-butyl pentanoate	13.666	3.23	3.32	−0.09	155	1-octene	15.192	3.70	3.73	−0.03
136	diethyl ether	4.701	1.03	0.89	0.14	156	1,4-pentadiene	8.553	1.90	1.94	−0.04
137	methyl *sec*-butyl ether	6.383	1.33	1.35	−0.02	157	cyclohexene	9.650	2.16	2.23	−0.07
138	methyl isobutyl ether	6.253	1.33	1.31	0.02	158	1-pentyne	8.670	1.98	1.97	0.01
139	di-*n*-propyl ether	8.391	2.03	1.89	0.14	159	1-hexyne	10.515	2.48	2.47	0.01
140	*n*-propyl isopropyl ether	7.713	1.83	1.71	0.12	160	1-octyne	14.205	3.48	3.46	0.02
141	methyl *n*-propyl ether	5.216	1.03	1.03	0.00	161	1,6-heptadiyne	10.269	2.46	2.40	0.06
142	methyl isopropyl ether	4.538	0.83	0.85	−0.02	162	2-butanone	2.687	0.29	0.35	−0.06
143	*n*-pentane	10.761	2.50	2.53	−0.03	163	2-hexanone	6.377	1.29	1.35	−0.06
144	2-methylbutane	9.953	2.30	2.31	−0.01	164	3-hexanone	6.377	1.29	1.35	−0.06
145	2-methylpentane	11.798	2.80	2.81	−0.01	165	3-methyl-2-pentanone	5.569	1.09	1.13	−0.04
146	3-methylpentane	11.798	2.80	2.81	−0.01	166	2-methyl-3-pentanone	5.569	1.09	1.13	−0.04
147	*n*-hexane	12.606	3.00	3.03	−0.03	167	3-heptanone	8.222	1.79	1.85	−0.06
148	2,4-dimethylpentane	12.835	3.10	3.09	0.01	168	2,4-dimethyl-3-pentanone	6.606	1.39	1.41	−0.02

[a] From refs 42 and 43. [b] From eq 4.

obtained, which requires further introspection. The average of the absolute values of the residuals was less than or equal to 0.10 in all cases. When all compounds of the training set ($n = 84$) were considered (Table 6), the following relation was obtained:

$$\log P = -0.377 + 0.270\text{DCW} \quad (4)$$

From Table 6, it can be observed that the above equation could predict 97.8% and explain 98.0% of the variance of the log *P* values of the training set. Out of 84 compounds, ethylene glycol, decanoic acid, ethyl cyclopropyl ether, chloroform, 1-chlorobutane, cycloheptane, cyclooctane, and acetone acted as outliers in the case of the modeling of all compounds (training set) with the molecular descriptor. Equation 4 was applied on the compounds of the training set and test set to calculate the log *P* values as shown in Table 1.

The results of relations of log *P* values of different subsets of the test set with the molecular descriptor (DCW) are given in Table 7. It is observed that the descriptor could explain the variance of log *P* values to the extent of 97.4% for alcohols ($n = 23$), 98.8% for amines ($n = 13$), 99.8% for esters ($n = 14$), 97.4% for ethers ($n = 7$), 99.6% for hydrocarbons ($n = 19$), and 99.9% for ketones ($n = 7$). The average of the absolute values of the residuals was less than 0.10 in all cases. When all compounds of the test set ($n = 84$) were considered (Table 7), the molecular descriptor could explain 99.2% of the variance. Out of 84 compounds, 3,3-dimethyl-2-butanol, 1-ethynylcyclohexanol, *n*-heptylamine, cyclohexylamine, methyl acetate, and methyl butyrate acted as outliers while modeling all compounds (test set) with the molecular descriptor. When eq 4 was used to predict the log *P* values of the compounds of the test set (Table 1), the r_{pred}^2 value was found to be 0.992 (Table 7).

The results of relations of log *P* values of different subsets of the combined set with the molecular descriptor (DCW) are given in Table 8. It is observed that the descriptor could explain the variance of log *P* values to the extent of 97.9% for alcohols ($n = 51$), 98.7% for amines ($n = 25$), 99.9% for acids ($n = 4$), 99.8% for esters ($n = 20$), 84.6% for ethers ($n = 12$), 96.8% for hydrocarbons ($n = 34$), and 98.5% for ketones ($n = 15$). However, in the case of halocarbons ($n = 6$), only 66.4% of the variance was explained. The average of the absolute values of the residuals was less than or equal to 0.10 in all cases. When all compounds of the combined sets ($n = 168$) were considered (Table 8), the molecular descriptor could explain 98.5% of the variance. Out of 168 compounds, ethylene glycol, decanoic acid, ethyl cyclopropyl ether, chloroform, 1-chlorobutane, cycloheptane, cyclooctane, acetone, and 1-ethynylcyclohexanol acted as outliers in the case of the modeling of all compounds (combined set) with the molecular descriptor.

The same data set was modeled previously[22] using molecular connectivity ($^1\chi^v$), molecular negentropy, and TAU indices. The statistical quality of the QSPR relation obtained in the present paper considering all the compounds ($n = 168$) is better than the relations obtained previously.[22]

QSPR for Set 2 Compounds. The results of the relations of log *P* values of set 2 compounds with the molecular descriptor (DCW) are given in Table 9. When compounds of the training set ($n = 69$) were considered, the following relation was obtained:

$$\log P = 0.173\text{DCW} \quad (5)$$

From Table 9, it can be observed that the above equation could predict and explain 99.6% of the variance of the log *P* values of the training set. As the intercept of the equation was insignificant, it was set to the value zero. Out of 69 compounds, 1,1,1-trichloroethane, 1,2-diphenylethane, benz[*b*]anthracene, pentachloroethane, and *n*-butylbenzene acted as outliers. Equation 5 was applied on the compounds of the training set and test set to calculate the log *P* values of set 2 compounds as shown in Table 2. When all compounds of the test set ($n = 70$) were considered (Table

Table 2. Optimized Molecular Descriptor and Observed and Calculated log *P* Values of Diverse Functional Compounds (Set 2)

Sl. no.	compound name	molecular descriptor (DCW)	log *P* obs.[a]	log *P* calc.[b]	log *P* res.[b]	Sl. no.	compound name	molecular descriptor (DCW)	log *P* obs.[a]	log *P* calc.[b]	log *P* res.[b]
				Training Set							
1	1,1,1-trichloroethane	15.874	2.481	2.748	−0.267	36	9,10-dimethylanthracene	33.556	5.788	5.808	−0.020
2	1,1,2,2-tetrachloroethane	16.030	2.644	2.775	−0.131	37	acenaphthene	24.196	4.070	4.188	−0.118
3	1,2,3,5-tetramethylbenzene	27.532	4.738	4.765	−0.027	38	adamantane	23.356	3.982	4.043	−0.061
4	1,2,3-trichlorobenzene	24.720	4.281	4.279	0.002	39	anthracene	25.988	4.490	4.498	−0.008
5	1,2,4,5-tetramethylbenzene	27.532	4.738	4.765	−0.027	40	benz[*b*]anthracene	30.718	5.664	5.317	0.347
6	1,2-dibromobenzene	21.528	3.588	3.726	−0.138	41	benzene	12.396	2.142	2.146	−0.004
7	1,2-dichlorobenzene	20.612	3.568	3.568	0.000	42	benz[*a*]pyrene	35.448	6.124	6.136	−0.012
8	1,2-diphenylethane	29.796	4.888	5.157	−0.269	43	benzo[*b*]fluorene	30.718	5.399	5.317	0.082
9	1,3-dichlorobenzene	20.612	3.568	3.568	0.000	44	benzo[g,h,i]perylene	38.112	6.584	6.597	−0.013
10	1,3-dimethylnaphthalene	26.760	4.614	4.632	−0.018	45	carbontetrachloride	16.581	2.875	2.870	0.005
11	1,4-dibromobenzene	21.528	3.868	3.726	0.142	46	chrysene	32.784	5.664	5.675	−0.011
12	1,4-dichlorobenzene	20.612	3.568	3.568	0.000	47	cyclohexane	19.416	3.354	3.361	−0.007
13	1,4-dimethylnaphthalene	26.760	4.614	4.632	−0.018	48	cyclooctane	25.888	4.472	4.481	−0.009
14	1-butene	12.893	2.266	2.232	0.034	49	dibenz[a,h]anthracene	39.580	6.838	6.851	−0.013
15	1-chloroheptane	24.261	4.110	4.199	−0.089	50	diethyl sulfide	10.945	1.900	1.894	0.006
16	1-chloronaphthalene	23.300	4.029	4.033	−0.004	51	dimethyl sulfide	4.819	0.842	0.834	0.008
17	1-chloropropane	11.317	1.994	1.959	0.035	52	ethyl chloride	8.081	1.465	1.399	0.066
18	1-hexene	19.365	3.324	3.352	−0.028	53	fluorobenzene	13.172	2.285	2.280	0.005
19	1-methylbenz[*a*]anthracene	36.568	6.313	6.329	−0.016	54	hexamethylbenzene	35.100	6.036	6.075	−0.039
20	1-methylnaphthalene	22.976	3.965	3.977	−0.012	55	iodobenzene	18.845	3.265	3.262	0.003
21	2,2′,4-trichlorobiphenyl	35.648	6.169	6.170	−0.001	56	pentachlorobenzene	32.936	5.707	5.701	0.006
22	2,2,4-trimethylpentane	25.797	4.536	4.465	0.071	57	pentachloroethane	19.371	3.627	3.353	0.274
23	2,3-dimethylnaphthalene	26.760	4.614	4.632	−0.018	58	pentamethylbenzene	31.316	5.387	5.420	−0.033
24	2,4′-dichlorobiphenyl	31.540	5.456	5.459	−0.003	59	phenanthrene	25.988	4.490	4.498	−0.008
25	2,4,6-tetrachlorobiphenyl	35.648	6.169	6.170	−0.001	60	tetrachloroethylene	17.424	3.020	3.016	0.004
26	2,5-dichlorobiphenyl	31.540	5.456	5.459	−0.003	61	trichloroethylene	13.066	2.267	2.262	0.005
27	2,6-dichlorobiphenyl	31.540	5.456	5.459	−0.003	62	*m*-xylene	19.964	3.440	3.456	−0.016
28	2-chloronaphthalene	23.300	4.029	4.033	−0.004	63	*n*-octane	28.452	4.926	4.925	0.001
29	2-chlorotoluene	20.288	3.504	3.512	−0.008	64	*n*-heptane	25.216	4.397	4.365	0.032
30	2-methylanthracene	29.772	5.139	5.153	−0.014	65	*n*-butylbenzene	25.888	4.738	4.481	0.257
31	2-methylhexane	24.247	4.267	4.197	0.070	66	*n*-pentane	18.744	3.339	3.244	0.095
32	2-methylpentane	21.011	3.738	3.637	0.101	67	*p*-xylene	19.964	3.440	3.456	−0.016
33	4-chlorotoluene	20.288	3.504	3.512	−0.008	68	*tert*-butylbenzene	24.202	4.118	4.189	−0.071
34	5-methylchrysene	36.568	6.313	6.329	−0.016	69	*trans*-1,2-dichloroethylene	8.708	1.514	1.507	0.007
35	7-methylbenz[*a*]anthracene	36.568	6.313	6.329	−0.016						
				Test Set							
70	1,2,3,4-tetrachlorobenzene	28.828	4.994	4.990	0.004	105	9-methylanthracene	29.772	5.139	5.153	−0.014
71	1,2,3,4-tetramethylbenzene	27.532	4.738	4.765	−0.027	106	benz[*a*]anthracene	32.784	5.664	5.675	−0.011
72	1,2,3,5-tetrachlorobenzene	28.828	4.994	4.990	0.004	107	benzo[*a*]fluorene	30.718	5.399	5.317	0.082
73	1,2,3-trimethylbenzene	23.748	4.089	4.110	−0.021	108	benzo[*b*]fluoranthene	35.448	6.124	6.136	−0.012
74	1,2,4,5-tetrachlorobenzene	28.828	4.738	4.990	−0.252	109	benzo[*e*]pyrene	35.448	6.124	6.136	−0.012
75	1,2,4-trichlorobenzene	24.720	4.281	4.279	0.002	110	benzo[*j*]fluoranthene	35.448	6.124	6.136	−0.012
76	1,2,4-trimethylbenzene	23.748	4.089	4.110	−0.021	111	benzo[*k*]fluoranthene	35.448	6.124	6.136	−0.012
77	1,2-dibromoethane	8.042	1.738	1.392	0.346	112	biphenyl	23.324	4.030	4.037	−0.007
78	1,2-dichloroethane	7.126	1.458	1.233	0.225	113	bromobenzene	16.962	3.005	2.936	0.069
79	1,3,5-trichlorobenzene	24.720	4.281	4.279	0.002	114	cholanthrene	37.788	6.418	6.541	−0.123
80	1,3,5-trimethylbenzene	23.748	4.089	4.110	−0.021	115	chlorobenzene	16.504	2.855	2.857	−0.002
81	1,4,5-trimethylnaphthalene	30.544	5.263	5.287	−0.024	116	cycloheptane	22.652	3.913	3.921	−0.008
82	1,5-dimethylnaphthalene	26.760	4.614	4.632	−0.018	117	cyclohexene	17.076	2.810	2.956	−0.146
83	1-chlorobutane	14.553	2.523	2.519	0.004	118	cyclopentane	16.180	2.795	2.801	−0.006
84	1-chlorohexane	21.025	3.581	3.639	−0.058	119	cyclopentene	13.840	2.251	2.396	−0.145
85	1-chloropentane	17.789	3.052	3.079	−0.027	120	dibenz[a,j]anthracene	39.580	6.838	6.851	−0.013
86	1-ethylnaphthalene	26.212	4.494	4.537	−0.043	121	ethylbenzene	19.416	3.320	3.361	−0.041
87	1-isopropyl-4-methylbenzene	25.467	4.368	4.408	−0.040	122	fluoranthene	28.652	4.950	4.959	−0.009
88	1-methylfluorene	28.876	4.874	4.998	−0.124	123	fluorene	23.922	4.225	4.141	0.084
89	1-pentene	16.129	2.795	2.792	0.003	124	fluorotrichloromethane	13.249	2.435	2.293	0.142
90	12-methylbenz[*a*]anthracene	36.568	6.313	6.329	−0.016	125	hexachlorobenzene	37.044	6.420	6.412	0.008
91	2,2′,4,5-tetrachlorobiphenyl	39.756	6.882	6.881	0.001	126	isopropylbenzene	21.683	3.719	3.753	−0.034
92	2,3,4,5-tetrachlorobiphenyl	39.756	6.882	6.881	0.001	127	naphthalene	19.192	3.316	3.322	−0.006
93	2,4,5-tetrachlorobiphenyl	35.648	6.169	6.17	−0.001	128	perylene	35.448	6.124	6.136	−0.012
94	2,6-dimethylnaphthalene	26.760	4.614	4.632	−0.018	129	pyrene	28.652	4.950	4.959	−0.009
95	2-chlorobiphenyl	27.432	4.743	4.748	−0.005	130	toluene	16.180	2.791	2.801	−0.010
96	2-chlorophenanthrene	30.096	5.203	5.209	−0.006	131	triphenylene	32.784	5.664	5.675	−0.011
97	2-methylbutane	17.775	3.209	3.077	0.132	132	*n*-decane	34.924	5.984	6.045	−0.061
98	2-methylnaphthalene	22.976	3.965	3.977	−0.012	133	*n*-nonane	31.688	5.455	5.485	−0.030
99	2-methylphenanthrene	29.772	5.139	5.153	−0.014	134	n-undecane	38.160	6.513	6.605	−0.092
100	3-chlorotoluene	20.288	3.504	3.512	−0.008	135	*n*-butane	15.508	2.810	2.684	0.126
101	5,6-dimethylchrysene	40.352	6.962	6.984	−0.022	136	*n*-hexane	21.980	3.868	3.804	0.064
102	6-methylbenzo[*e*]pyrene	39.232	6.773	6.791	−0.018	137	*n*-propylbenzene	22.652	3.849	3.921	−0.072
103	6-methylchrysene	36.568	6.313	6.329	−0.016	138	*o*-xylene	19.964	3.440	3.456	−0.016
104	7-ethylbenz[*a*]anthracene	39.804	6.842	6.89	−0.048	139	*tert*-amylbenzene	27.438	4.647	4.749	−0.102

[a] From ref 44. [b] From eq 5.

Table 3. Optimized Correlation Weights for Different Local Invariants (for Set 1 Compounds)

invariant type	local invariant	optimized weight
a_k	C	−0.295
	H	−0.503
	O	−4.028
	F	0.478
	N	−4.764
	Cl	4.244
	I	4.880
	Br	3.265
NNC_k	100	0.367
	110	0.513
	211	1.312
	220	1.635
	301	4.893
	310	4.484
	312	2.494
	320	−0.574
	321	1.420
	403	2.482
	410	2.138
	412	1.209
	413	2.878
	421	−0.227
	422	2.120
	430	−2.014
	431	0.554
	440	−0.026

Table 4. Optimized Correlation Weights for Different Local Invariants (for Set 2 Compounds)

invariant type	local invariant	optimized weight
a_k	H	0.666
	C	0.269
	F	−1.073
	S	−0.272
	Cl	2.259
	Br	2.717
	I	4.600
NNC_k	110	0.216
	220	0.005
	310	3.493
	311	0.728
	312	−0.130
	320	3.430
	321	0.915
	330	1.063
	400	6.412
	403	−0.372
	410	3.662
	411	1.914
	412	−0.945
	413	1.603
	422	1.203
	431	−0.166
	440	−1.283

Table 5. Calculation of the Molecular Descriptor for Methanol[a]

atom no.	atom type	NNC	CW(a_k)	CW(NNC_k)
1	C	403	−0.295	2.482
2	H	110	−0.503	0.513
3	H	110	−0.503	0.513
4	H	110	−0.503	0.513
5	O	211	−4.028	1.312
6	H	100	−0.503	0.367

[a] $DCW(a_k,NNC_k) = \sum CW(a_k) + \sum CW(NNC_k) = -0.635$.

9), the molecular descriptor could explain 99.7% of the variance. Out of 70 compounds, 1,2,4,5-tetrachlorobenzene, 1,2-dibromoethane, 1,2-dichloroethane, cyclohexene, and cyclopentene acted as outliers. When eq 5 was used to predict the log P values of the compounds of the test set (Table 2), the r_{pred}^2 value was found to be 0.997 (Table 9). When all compounds of the combined sets ($n = 139$) were considered (Table 9), the molecular descriptor could explain 99.7% of the variance. Out of 139 compounds, 1,1,1-trichloroethane, 1,2-diphenylethane, benz[*b*]anthracene, pentachloroethane, *n*-butylbenzene, 1,2,4,5-tetrachlorobenzene, 1,2-dibromoethane, 1,2-dichloroethane, and cyclopentene acted as outliers in the case of the modeling of all compounds (combined set) with the molecular descriptor. The same data set was modeled previously by Basak et al.[44] The statistical quality of the relation generated in the present study considering all the compounds ($n = 139$) is better than that reported by Basak et al.[44]

Table 6. Relations of log P Values of Different Subsets of the Training Set (Set 1 Compounds) with the Optimized Molecular Descriptor (DCW)[a]

type of compound	regression coefficient		statistics		
	β (se)	α (se)	r_a^2 (r)	r^2 (s)	F (AVRES)
alcohols	0.287	−0.460	0.980	0.981	1343.5
($n = 28$)	(0.008)	(0.046)	(0.990)	(0.135)	(0.084)
amines	0.279	−0.437	0.987	0.988	843.8
($n = 12$)	(0.010)	(0.067)	(0.994)	(0.110)	(0.096)
acids	0.291	−0.361	0.999	0.999	2535.6
($n = 4$)	(0.006)	(0.050)	(1.000)	(0.066)	(0.044)
esters	0.266	−0.452	0.989	0.991	437.5
($n = 6$)	(0.013)	(0.087)	(0.995)	(0.082)	(0.047)
acids and esters	0.286	−0.474	0.984	0.986	555.0
($n = 10$)	(0.012)	(0.092)	(0.993)	(0.160)	(0.132)
ethers	0.145	0.492	0.452	0.589	4.3
($n = 5$)	(0.070)	(0.413)	(0.768)	(0.148)	(0.089)
halocarbons	0.122	0.963	0.664	0.732	10.9
($n = 6$)	(0.037)	(0.344)	(0.855)	(0.149)	(0.101)
hydrocarbons	0.268	−0.384	0.919	0.924	158.9
($n = 15$)	(0.021)	(0.268)	(0.961)	(0.161)	(0.101)
ketones	0.249	−0.241	0.983	0.986	415.9
($n = 8$)	(0.012)	(0.090)	(0.993)	(0.128)	(0.092)
all[b]	0.270	−0.377	0.980	0.980	3995.0
($n = 84$)	(0.004)	(0.035)	(0.990)	(0.160)	(0.104)

[a] Model eq: $\log P = \alpha + \beta$*DCW(a,NNC). [b] Leave-one-out cross-validation statistics: $q^2 = 0.978$, SDEP = 0.164.

Table 7. Relations of log P Values of Different Subsets (Set 1 Compounds) of the Test Set with the Optimized Molecular Descriptor (DCW)[a]

type of compound	regression coefficient		statistics		
	β (se)	α (se)	r_a^2 (r)	r^2 (s)	F (AVRES)
alcohols	0.268	−0.342	0.974	0.975	832.9
($n = 23$)	(0.009)	(0.065)	(0.988)	(0.112)	(0.061)
amines	0.284	−0.404	0.988	0.989	991.7
($n = 13$)	(0.009)	(0.056)	(0.994)	(0.096)	(0.071)
esters	0.274	−0.512	0.998	0.998	7225.5
($n = 14$)	(0.003)	(0.036)	(0.999)	(0.060)	(0.034)
acids and esters	0.274	−0.512	0.998	0.998	7225.5
($n = 15$)	(0.003)	(0.036)	(0.999)	(0.060)	(0.034)
ethers	0.295	−0.479	0.974	0.978	222.8
($n = 7$)	(0.020)	(0.125)	(0.989)	(0.072)	(0.051)
hydrocarbons	0.270	−0.392	0.996	0.996	4774.0
($n = 19$)	(0.003)	(0.044)	(0.998)	(0.036)	(0.027)
ketones	0.272	−0.432	0.999	0.999	4701.1
($n = 7$)	(0.004)	(0.024)	(0.999)	(0.016)	(0.012)
all[b]	0.267	−0.362	0.992	0.992	9771.3
($n = 84$)	(0.003)	(0.024)	(0.996)	(0.093)	(0.063)

[a] Model eq: $\log P = \alpha + \beta$*DCW(a,NNC). [b] Prediction statistics: $r_{pred}^2 = 0.992$.

Table 8. Relations of log *P* Values of Different Subsets of the Combined Set (Set 1 Compounds) with the Optimized Molecular Descriptor (DCW)[a]

type of compound	regression coefficient		statistics		
	β (se)	α (se)	r_a^2 (r)	r^2 (s)	F (AVRES)
alcohols	0.280	−0.428	0.979	0.980	2371.5
($n = 51$)	(0.006)	(0.037)	(0.990)	(0.126)	(0.074)
amines	0.280	−0.413	0.987	0.987	1813.6
($n = 25$)	(0.007)	(0.043)	(0.994)	(0.103)	(0.086)
acids	0.291	−0.361	0.999	0.999	2535.6
($n = 4$)	(0.006)	(0.050)	(1.000)	(0.066)	(0.044)
esters	0.272	−0.497	0.998	0.998	8125.1
($n = 20$)	(0.003)	(0.030)	(0.999)	(0.064)	(0.037)
acids and esters	0.273	−0.461	0.992	0.992	2785.8
($n = 24$)	(0.005)	(0.051)	(0.996)	(0.127)	(0.085)
ethers	0.253	−0.181	0.846	0.860	61.5
($n = 12$)	(0.032)	(0.198)	(0.927)	(0.136)	(0.099)
halocarbons	0.122	0.963	0.664	0.732	10.9
($n = 6$)	(0.037)	(0.344)	(0.855)	(0.149)	(0.101)
hydrocarbons	0.269	−0.378	0.968	0.969	991.4
($n = 34$)	(0.009)	(0.102)	(0.984)	(0.106)	(0.060)
ketones	0.253	−0.291	0.985	0.986	894.5
($n = 15$)	(0.008)	(0.058)	(0.993)	(0.096)	(0.058)
all	0.269	−0.370	0.985	0.985	11089.1
($n = 168$)	(0.003)	(0.022)	(0.993)	(0.130)	(0.084)

[a] Model eq: $\log P = \alpha + \beta$*DCW(a,NNC).

Table 9. Relations of log *P* Values of Set 2 Compounds with the Optimized Molecular Descriptor (DCW)[a]

type of compound	regression coefficient		statistics		
	β (se)	α (se)	r_a^2 (r)	r^2 (s)	F (AVRES)
training[b]	0.173		0.996	0.996	167461.8
($n = 69$)	(0.000)		(0.998)	(0.090)	(0.047)
test[c]	0.169	0.088	0.997	0.997	25863.2
($n = 70$)	(0.001)	(0.029)	(0.999)	(0.073)	(0.048)
all	0.171	0.050	0.997	0.997	40015.1
($n = 139$)	(0.001)	(0.023)	(0.998)	(0.082)	(0.047)

[a] Model eq: $\log P = \alpha + \beta$*DCW(a,NNC). [b] Cross-validation statistics: $q^2 = 0.996$, SDEP = 0.090. [c] Prediction statistics: $r_{pred}^2 = 0.997$.

Overview. The present analysis shows that the optimization of a correlation weights scheme can generate a statistically acceptable model for lipophilicity of diverse functional aliphatic and aromatic compounds. Moreover, the scheme does not require complex calculation of diverse descriptors and statistical analysis for proper selection of descriptors and intercorrelation among them. Thus, the scheme merits further assessment on exploring QSPR/QSAR of different physicochemical properties/biological activity data using different local invariants to justify its suitability in modeling studies. Since the statistical characteristics of the lipid-water partition coefficient model based on the weighting of the kinds of atoms together with NNC values are better than the ones based on the weighting of the kinds of atoms together with the Morgan extended connectivity values, NNCs should be examined as a useful alternative to the Morgan extended connectivity in QSPR analysis.

REFERENCES AND NOTES

(1) Selassie, C. D. In *Burger's Medicinal Chemistry and Drug Discovery*; Abraham, D. J., Ed.; Wiley: New Jersey, 2003; Vol. 1, pp 1−48.
(2) Franke, R. *Theoretical Drug Design Methods*; Elsevier: Amsterdam; 1984; pp 30−79.
(3) Taylor, P. J. In *Comprehensive Medicinal Chemistry*; Hansch, C.; Sammes, P. G.; Taylor, J. B., Eds.; Pergamon Press: Oxford, 1990; Vol. 4, pp 241−294.
(4) Martin, A.; Bustamante, P.; Chun, A. H. C. *Physical Pharmacy*; Lippincott, Williams and Wilkins: Baltimore, 2001; pp 251−283.
(5) Debnath, A. K. In *Combinatorial Library Design and Evaluation*; Ghose, A. K., Viswanadhan, V. N., Eds.; Marcel Dekker: New York, 2001; pp 73−129.
(6) Jurs, P. C.; Dixon, S. L.; Egolf, L. M. In *Chemometric Methods in Molecular Design*; Waterbeemd, H. van de, Ed.; VCH: Weinheim, 1995; Vol. 2, pp 15−37.
(7) Rekker, R. F. *The Hydrophobic Fragment Constants. Its Derivation and Applications. A Means of Characterizing Membrane Systems*; Elsevier: Amsterdam, 1977.
(8) el Tayar, N.; Tsai, R. S.; Testa, B.; Carrupt, P. A.; Leo, A. Partitioning of Solutes in Different Solvent Systems: The Contribution of Hydrogen Bonding Capacity and Polarity. *J. Pharm. Sci.* **1991**, *80*, 590−598.
(9) Hansch, C.; Leo, A. *Exploring QSAR. Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995.
(10) Ghose, A. K.; Crippen, G. M. Atomic Physicochemical Parameters for Three-Dimensional-Structure-Directed Quantitative Structure−Activity Relationships. 2. Modeling Dispersive and Hydrophobic Interactions. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 21−35.
(11) Ghose, A. K.; Viswanandhan, V. N.; Wendoloski, J. J. Prediction of Hydrophobic Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods. *J. Phys. Chem.* **1998**, *102*, 3762−3772.
(12) Bodor, N.; Gabanyi, Z.; Wong, C.-K. A New Method for The Estimation of Partition Coefficient. *J. Am. Chem. Soc.* **1989**, *111*, 3783−3786.
(13) Klopman, G.; Wang, S. A Computer Automated Structure Evaluation (CASE) Approach to the Calculation of Partition Coefficients. *J Comput. Chem.* **1991**, *12*, 1025−1032.
(14) Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, I.; Matsushita, Y. Simple Method of Calculating Octanol/Water Partition Coefficient. *Chem. Pharm. Bull. (Tokyo)* **1992**, *40*, 127−130.
(15) Suzuki, T.; Kudo, Y. Automatic Log P Estimation Based on Combined Additive Modeling Studies. *J. Comput. Aided Mol. Des.* **1990**, *4*, 155−198.
(16) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research;* Academic Press: New York, 1976.
(17) Saxena, A. K. Physicochemical Significance of Topological Parameters: Molecular Connectivity Index and Information Content. Part 2. Correlation Studies with Molar Refractivity and Lipophilicity. *Quant. Struct.-Act. Relat.* **1995**, *14*, 142−150.
(18) Mandloi, M.; Sikarwar, A.; Sapre, N. S.; Karmakar, S.; Khadikar, P. V. A Comparative QSAR Study Using Wiener, Szeged and Molecular Connectivity Indices. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 57−62.
(19) Vaes, W. H.; Ramos, E. U.; Verhaar, H. J.; Cramer, C, J.; Hermens, J. L. Understanding and Estimating Membrane-Water Partition Coefficients: Approaches to Derive Quantitative Structure−Property Relationships. *Chem. Res. Toxicol.* **1998**, *11*, 847−854.
(20) Estrada, E.; Molina, E. 3D Connectivity Indices in QSPR/QSAR Studies. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 791−797.
(21) Du, Q.; Arteca, G. A. Modeling Lipophilicity from the Distribution of Electrostatic Potential on A Molecular Surface. *J. Comput. Aided Mol. Des.* **1996**, *10*, 133−144.
(22) Roy, K.; Saha, A. Comparative QSPR Studies with Molecular Connectivity, Molecular Negentropy and TAU Indices. Part II: Lipid-Water Partition Coefficient of Diverse Functional Acyclic Compounds. *Internet Electron. J. Mol. Des.* **2003**, *2*, 288−305, http://www.biochempress.com.
(23) Randic, M. On Computation of Optimal Parameters for Multivariate Analysis of Structure−Property Relationship. *J. Comput. Chem.* **1991**, *12*, 970−980.
(24) Randic, M. Novel Graph Theoretic Approach to Heteroatoms in Quantitative Structure−Activity Relationships. *Chemom. Intell. Lab. Syst.* **1991**, *10*, 213−227.
(25) Randic, M. Resolution of Ambiguities in Structure−Property Studies by Use of Orthogonal Descriptors. *J. Chem. Inf. Comput. Sci.* **1991**, *31*, 311−320.
(26) Randic, M. Similarity Based on Extended Basic Descriptors. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 686−692.
(27) Estrada, E. Graph Theoretical Invariants of Randic Revisited. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1022−1025.
(28) Amic, D.; Beslo, D.; Lucic, D.; Nikolic, S.; Trinajstic, N. The Vertex-Connectivity Index Revisited. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 819−822.
(29) Randic, M.; Basak, S. C. Optimal Molecular Descriptors Based on Weighted Path Numbers. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 261−266.
(30) Randic, M.; Basak, S. C. In *Some Aspects in Mathematical Chemistry*; Sinha, D. K., Basak, S. C., Mohanty, R. K., Basumallick, I. N., Eds.; Visva-Bharati University Press: Santiniketan; 1999; p 24.

(31) Toropov, A. A.; Toropova, A. P. Optimization of Correlation Weights of the Local Graph Invariants: Use of the Enthalpies of Formation of Complex Compounds for the QSPR Modeling. *Russ. J. Coord. Chem.* **1998**, *24*, 81−85.

(32) Toropov, A. A.; Toropova, A. P.; Voropaeva, N. L.; Ruban, I. N.; Rashidova, S. Sh. Generalized Zero-Order Molecular Connectivity Index: Enthalpies of Crystalline Aquo and Ammino Complexes in QSPR Modeling. *Russ. J. Coord. Chem.* **1998**, *24*, 525−529.

(33) Toropov, A. A.; Voropaeva, N. L.; Ruban, I. N.; Rashidova, S. Sh. Quantitative Structure−Property Relationships for Binary Polymer Solvent Systems: Correlation Weighing of the Local Invariants of Molecular Graphs. *Polym. Sci. Ser. A* **1999**, *41*, 975−985.

(34) Krenkel, G.; Castro, E. A.; Toropov, A. A. Improved molecular descriptors to calculate boiling points based on the optimization of correlation weights of local graph invariants. *J. Mol. Struct. (THEOCHEM)* **2001**, *542*, 107−113.

(35) Mercader, A.; Castro, E. A.; Toropov, A. A. Calculation of Total Molecular Electronic Energies from Correlation Weighting of Local Graph Invariants. *J. Mol. Model.* **2001**, *7*, 1−5.

(36) Mercader, A.; Castro, E. A.; Toropov, A. A. QSPR Modeling of the Enthalpy of Formation from Elements by Means of Correlation Weighting of Local Invariants of Atomic Orbital Molecular Graphs. *Chem. Phys. Lett.* **2000**, *330*, 612−623.

(37) Krenkel, G.; Castro, E. A.; Toropov, A. A. Improved Molecular Descriptors Based on the Optimization of Correlation Weights of Local Graph Invariants. *Int. J. Mol. Sci.* **2001**, *2*, 57−65, http://www.mdpi.org/ijms.

(38) Marino, D. J. G.; Perruzo, P. J.; Castro, E. A.; Toropov, A. A. QSAR Carcinogenic Study of Methylated Polycyclic Aromatic Hydrocarbons Based on Topological Descriptors Derived from Distance Matrices and Correlation Weights of Local Graph Invariants. *Internet Electron. J. Mol. Des.* **2002**, *1*, 115−133, http://www.biochempress.com.

(39) Duchowicz, P.; Castro, E. A.; Toropov, A. A. Improved QSPR Analysis of Standard Entropy of Acyclic and Aromatic Compounds Using Optimized Correlation Weights of Linear Graph Invariants. *Comput. Chem.* **2002**, *26*, 327−332.

(40) Toropov, A. A.; Duchowicz, P.; Castro, E. A. Structure-Toxicity Relationships for Aliphatic Compounds Based on Correlation Weighting of Local Graph Invariants. *Int. J. Mol. Sci.* **2003**, *4*, 272−283, http://www.mdpi.org/ijms.

(41) Perruzo, P. J.; Marino, D. J. G.; Castro, E. A.; Toropov, A. A. QSPR Modeling of Lipophilicity By Means of Correlation Weights of Local Graph Invariants. *Internet Electron. J. Mol. Des.* **2003**, *2*, 334−347, http://www.biochempress.com.

(42) Leo, A.; Hansch, C.; Elkins, D. Partition Coefficients and Their Uses. *Chem. Rev.* **1971**, *71*, 525−616.

(43) Hansch, C.; Quinlan, J. E.; Lawrence, G. L. The Linear Free Energy Relationship Between Partition-Coefficients and The Aqueous Solubility of Organic Liquids. *J. Org. Chem.* **1968**, *33*, 347−350.

(44) Basak, S. C.; Niemi, G. J.; Veith, G. D. Optimal Characterization of Structures for Prediction of Propertes. *J. Math. Chem.* **1990**, *4*, 185−205.

(45) The program for optimization of correlation weights was developed in PASCAL by A. A. Toropov.

(46) The GW-BASIC programs *RRR98*, *KRPRES1* and *KRPRES2* were developed by Kunal Roy (1998) and standardized using known data sets.

(47) Snedecor, G. W.; Cochran, W. G. *Statistical Methods*; Oxford & IBH Publishing Co. Pvt. Ltd.: New Delhi, 1967; pp 381−418.

(48) Kier, L. B.; Hall, L. H. In *Advances in Drug Research*; Testa, B., Ed.; Academic Press: New York, 1992; Vol. 22, pp 1−38.

(49) Wold, S.; Eriksson, L. In *Chemometric Methods in Molecular Design*; Waterbeemd, H. van de, Ed.; VCH: Weinheim, 1995; p 312.

CI034200G