Prediction of Aquatic Toxicity: Use of Optimization of Correlation Weights of Local Graph Invariants

Andrey Andreevich Toropov[†] and Terry Wayne Schultz*[,‡]

Vostok Holding Innovation Co. 700000, Sadik Azim 4th Street, 15 Tashkent, Uzbekistan, and College of Veterinary Medicine, The University of Tennessee, A205, 2407 River Drive, Knoxville, Tennessee 37996-4543

Received June 19, 2002

Quantitative structure−activity relationships (QSARs) were developed for three sets of toxicity data. Chemicals in each set represented a number of narcoses and electrophilic mechanisms of toxic action. A series of quantitative structure-toxicity models correlating toxic potency with a number of optimization of correlation weights of local graph invariants were developed. In the case of the toxicity of a heterogeneous set of benzene derivatives to *Tetrahymena pyriformis*, the QSARs were based on the Descriptor of Correlation Weights (DCW) using atoms and extended connectivity (EC) graph invariants. The model [log (IGC_{50}^{-1}) = 0.0813 DCW(a_k,^3EC_k) + 2.636; $n = 157$, $r^2 = 0.883$, $s = 0.27$, $F = 1170$, Pr > $F = 0.0001$] based on third-order EC of 89 descriptors was observed to be best for the benzene data. However, fits for these data of > 0.800 were achieved ECs with as few as 23 variables. The relationship between the toxicity predicted by this model and experimental toxicity values for the test set [obs. log(IGC_{50}^{-1})) = 0.991 (pred. (log(IGC_{50}^{-1})) − 0.012; $n = 60$, $r^2 = 0.863$, $s = 0.28$, $F = 372$, Pr > $F = 0.0001$] is excellent. The utility of the approach was demonstrated by the model [log (IGC_{50}^{-1}) = 0.1744(DCW (a_k, ^2EC) − 3.505; $n = 39$, $r^2 = 0.900$, $s = 0.35$, $F = 333$, Pr > $F = 0.0001$] for the toxicity data for *T. pyriformis* exposed to halo-substituted aliphatic compounds and the model [log (IC_{50}^{-1}) = 0.1699(DCW (a_k, ^2EC)) − 2.610; $n = 66$, $r^2 = 0.901$, $s = 0.31$, $F = 583$, Pr > $F = 0.0001$] for the *Vibrio fischeri* toxicity data.

INTRODUCTION

Molecular descriptor[1] is a generic term that applies to any parameter used as an independent variable in the formulation of a quantitative structure−activity relationship (QSARs). A review of the literature shows that hundreds of such descriptors of structure have been used in modeling toxicity. As an example, the QSARIS software (SciVision, 200 Wheeler Road, Burlington, MA) is able to generate over 400 2-D and 3-D descriptors per chemical. Molecular descriptors, and therefore QSAR models, may be classified according to the method used to encode the structural information—empirical (measured or estimated), quantum chemical, nonempirical, or computer graphic. Nonempirical descriptors are typically structural property based on chemical topology and include topological indices.[2] With such indices, the structure of a substance is represented by a number or a set of numbers that are derived from the application of chemical graph theory.[3]

The advantage of topological properties is that they are a direct, simple description of molecular structure; the disadvantage is those chemical topologies have no direct mechanistic meaning to toxicology.[4] There is an increasing discussion as to whether, or not, parameters for fundamental properties such as hydrophobicity need be included in QSAR.[4] Some workers[5−8] have developed approaches for modeling ecotoxic potency without the use of such properties. These workers use a number of atom- and fragment-based, structural, and topological indices to encode molecular structure. The implied assumption is that these descriptors *in toto* account for hydrophobicity and other fundamental molecular properties.

A graph invariant is a topological property, which is conserved by molecular isomorphism. Therefore, an inclusive graph invariant is one, which is the same for two or more graphs (i.e., molecules) and represents a degree of molecular similarity. Since structurally alike compounds have "alike" physicochemical properties, and therefore biological activity, graph invariant properties have the potential to be useful descriptors of chemical toxicity. Among the graph invariant methodologies, the optimization of correlation weights of local graph invariants (OCWLGI) has shown promise in modeling toxic endpoints.[9]

Among the databases that have significance in the development and validation of QSARs in aquatic toxicity is that for population growth impairment to the ciliated protozoan *Tetrahymena pyriformis*.[10] This data set, TETRATOX, lists relative toxicity on a wide variety of industrial organic chemicals including substituted benzenes.[11] These data originated from a single laboratory using a standard protocol[12] which has been validated.[13]

The aims of this investigation were to examine the predictive ability of the OCWLGI for modeling aquatic

* Corresponding author phone: (865)974-5826; fax: (865)974-5640; e-mail: tschultz@utk.edu.
[†] Vostok Holding Innovation Co.
[‡] The University of Tennessee.

10.1021/ci025555n CCC: $25.00
Published on Web 01/31/2003

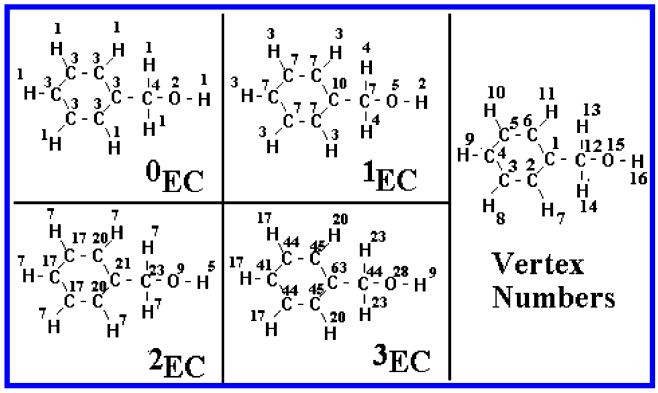

Figure 1. $^{x}EC_k$ and vertex numbers of LHFG of benzyl alcohol.

toxicity of benzene derivatives, demonstrate the effectiveness of the approach via analyses of two other toxicity data sets, and compare the models presented herein with results obtained using empirical and quantum chemical descriptors.

METHODS AND MATERIALS

Biological Data. Toxicity data [50% growth inhibitory concentration (log IGC_{50}^{-1})] for 217 substituted benzenes tested in the 2-d *Tetrahymena pyriformis* population growth impairment assay and representing the neutral narcosis, polar narcosis, respiratory uncoupling, and various electro(nucleo)-philic mechanisms of toxic action were evaluated. Data were taken from Schultz[11] where the benzenes were selected a priori to ensure uniform distribution over the maximum range of values for both hydrophobicity and electrophilicity. Population growth impairment testing with *T. pyriformis* (strain GL-C) was conducted following the protocol described by Schultz.[12] Briefly, this assay is static in design and uses 40-h population density quantitated spectrophotometrically at 540 nm as its endpoint. The IGC_{50} was determined for each compound tested by using the Probit Analysis of Statistical Analysis System (SAS) software.[14] The *y*-values were absorbencies normalized as percentage of control. The X-values were the toxicant concentrations in mg/L. In an effort to demonstrate the utility of the approach, two other toxicity data sets, those of Acker et al.[15] and Cronin et al.,[16] were also modeled.

Molecular Descriptors. The descriptors used in this investigation were confined to selected topological indexes.[17] Specifically, the descriptors are functions of correlation weights (CWs) of local graph invariants [i.e., Descriptor of Correlation Weights (DCW)]. They are calculated from the general formula

$$\mathrm{DCW}\,(a_k,{}^{x}\mathrm{EC}_k) = \sum_{k=1}^{N} [\mathrm{CW}\,(a_k) + \mathrm{CW}\,({}^{x}\mathrm{EC}_k)] \quad (1)$$

where n is the total number of vertices in the graph; a_k is the chemical element that occupies the kth vertex of the graph; CW (a_k) is the correlation weight present in the molecule (or in the graph) given the kind of a_k (a_k = H, C, O, etc.); $^{x}EC_k$ is the xth order (x = 0, 1, 2, and 3); the Morgan extended connectivity value[18] of the kth vertex; and CW ($^{x}EC_k$) is the correlation weight of the $^{x}EC_k$ presence. The Morgan extended connectivity of zero order (^{0}EC) is the vertex degree of the labeled hydrogen filled graphs (LHFGs). The $^{x}EC_k$ (i.e., value of the local invariant on kth vertex) is calculated with the recurrent formula:

$$^{x}\mathrm{EC}_k = \sum {}^{x-1}\mathrm{EC}_j\,(k, j)\ \mathrm{edge} \quad (2)$$

In other words, eq 2 is the sum of the Morgan extended connectivity of the ($x-1$)th order over all neighbors of the kth vertex in the LHFG.

More specifically, the algorithm of the OCWLGI is described as follows. By reading all structures (i.e., molecular graphs) of a given training set, a list of all local graph invariants present in the training set is defined. Chemical elements, which are images of graph vertices, are part of the list. Other local invariants are values of Morgan extended connectivity; the greater the order of Morgan extended connectivity, the greater number of such local graph invariants.

For each local graph invariant (i.e., chemical element or numeric values of Morgan extended connectivity), CW is determined initially by setting the starting values of all CWs to 1.0. The regular order of number of local invariants (i.e., 1,2,3,4,5,6,7,...,n) is replaced by a random sequence (e.g., {7, 2, 1, 5, 3, 4,,6, n, ... n'}). Subsequently, a starting correlation coefficient (r_1) between toxicity values and descriptor of eq 1 values on the training set is calculated. As explained in the next paragraph, values of the r_1 should be saved in variable r_{start}.

Then, in a generated random sequence, each local invariant correlation weight (CWi) was modified with the follow algorithm:

1. ΔCWi =0.1 * CWi;
2. CWi = CWi + ΔCWi;
3. Calculating of r_2 [i.e., correlation coefficient between toxicity and descriptor of eq 1] after change of the CWi value;
4. if ($r_2 - r_1$) > 0.001, then $r_1 = r_2$; go to 2;
5. CWi = CWi − ΔCWi;
6. ΔCWi = − 0.5 * ΔCWi;
7. If absolute values (ΔCWi) > 0.01 then go to 2.

Table 1. Statistical Characteristics of the Training Set (n = 157) Models

LI[a]	NLI[b]	slope C1	intercept C0	r^2	s	F
Optimization of the CWs[c] on DCW[d] (a_k, $^{0}EC_k$)						
^{0}EC-1	13	0.268	−1.341	0.734	0.414	427
^{0}EC-2	13	0.256	−1.281	0.735	0.413	430
^{0}EC-3	13	0.257	−1.352	0.732	0.415	424
Optimization of the CWs[c] on DCW[d] (a_k, $^{1}EC_k$)						
^{1}EC-1	23	0.181	−1.053	0.817	0.343	692
^{1}EC-2	23	0.133	−1.468	0.809	0.350	657
^{1}EC-3	23	0.173	−1.468	0.806	0.354	642
Optimization of the CWs[c] on DCW[d] (a_k, $^{2}EC_k$)						
^{2}EC-1	41	0.135	−1.944	0.820	0.341	703
^{2}EC-2	41	0.109	−1.908	0.816	0.343	689
^{2}EC-3	41	0.112	−1.910	0.811	0.348	666
Optimization of the CWs[c] on DCW[d] (a_k, $^{3}EC_k$)						
^{3}EC-1	89	0.130	−3.066	0.883	0.274	1170
^{3}EC-2 (best result)	89	0.081	−2.636	0.883	0.274	1170
^{3}EC-3	89	0.096	−2.482	0.883	0.274	1170

[a] LI = the type of graph invariant. [b] NLI = the of number parameters of optimization. [c] CWs = correlation weights. [d] DCW = descriptor of correlation weights.

Table 2. Optimal Correlation Weights of Three Probes Run of the Optimization of Correlation Weights of Local Graph Invariants with Descriptor of Correlation Weights (a_k ^3EC_k)

local graph invariants	probe 1	probe 2	probe 3	local graph invariants	probe 1	probe 2	probe 3
Atom Kinds, i.e., a_k							
H	0.059	−0.267	0.000	S	2.720	1.923	0.837
C	0.355	0.732	0.527	Cl	4.120	6.226	5.472
N	1.344	1.622	2.521	Br	5.000	7.210	6.866
O	2.490	3.881	3.320	J	6.973	11.536	10.364
F	2.490	3.541	3.380				
Morgan Extended Connectivity of Third Order, i.e., ^3EC_k							
7	9.521	18.311	16.479	48	0.508	0.599	0.663
8	2.191	2.039	1.115	49	0.190	0.113	0.418
9	−1.096	−2.050	−2.294	50	0.697	0.911	0.711
10	0.198	−0.244	−0.194	51	−0.330	−0.636	0.259
12	−2.969	−7.422	−5.277	52	−0.489	−0.921	−0.693
13	0.052	0.199	−0.304	53	0.991	1.096	1.569
14	2.884	4.651	2.523	54	1.571	2.167	2.386
15	1.810	2.414	1.287	55	0.420	0.079	0.883
16	1.854	3.214	3.369	56	0.565	0.858	0.952
17	1.313	2.031	1.250	57	−0.422	−0.515	−0.533
18	1.681	1.421	1.394	58	1.826	3.228	2.578
19	0.541	1.196	0.305	59	−0.175	−1.118	0.237
20	0.843	1.556	0.840	60	1.611	0.860	0.114
21	1.644	2.802	1.984	61	4.688	6.592	6.437
22	0.664	1.854	0.898	62	1.831	2.539	2.203
23	−1.044	−1.380	−1.407	63	1.758	1.277	0.713
24	1.011	1.538	0.952	64	−0.734	−1.996	−1.162
25	0.778	1.483	1.001	65	1.758	1.758	0.352
26	−0.394	−0.308	−0.766	66	1.071	0.345	0.806
27	1.401	1.978	1.461	67	0.942	0.606	0.705
28	0.453	0.596	1.099	68	2.441	3.735	2.568
29	0.646	0.900	−0.011	69	2.377	1.195	2.615
30	−0.609	−0.021	−0.325	70	0.109	−0.025	0.132
31	−0.515	−0.466	−0.802	71	−0.605	−0.427	−0.646
32	1.888	1.295	1.069	72	1.625	0.407	1.285
33	−0.101	−0.243	−0.198	73	2.298	4.104	3.098
34	−0.734	−0.202	0.302	74	0.696	1.675	−0.093
35	−1.490	−0.194	−0.268	75	−0.307	0.766	0.131
36	1.859	1.709	2.407	76	4.219	6.592	4.104
37	−1.999	−2.424	−3.444	77	0.478	0.342	0.041
38	4.828	4.150	4.316	79	−2.119	−1.938	−2.388
39	1.588	2.076	1.190	82	−0.024	0.125	0.996
40	1.926	2.549	2.007	86	4.395	7.141	6.848
41	1.690	2.153	1.868	88	0.338	−0.060	−0.274
42	2.051	3.154	2.188	90	1.909	2.393	0.452
43	0.948	1.102	1.112	95	−0.861	−1.994	−1.735
44	1.395	1.751	1.758	98	−2.969	−1.836	−3.388
45	0.913	1.373	1.322	99	0.943	1.289	0.095
46	0.308	0.373	0.435	110	0.314	2.139	0.277
47	0.050	0.098	0.253	113	0.283	0.009	0.156

Then steps of 1−7 are carried out for all CWs. One can repeat this algorithm from point of generation of random sequence. It is clear that a number of such cycles of optimization must be limited. As criterion for stopping the process, the following was used:

$$(r_1 - r_{start}) < 0.001$$

Having derived the values of the optimal CWs, one can calculate values of eq 1 for both the training set and validation set; then by the least-squares method of regression analysis, one can calculate the toxicity model:

$$\text{toxicity} = C0 + C1 * \text{DCW}\ (a_k, {}^xEC_k) \qquad (3)$$

Predictive potential of eq 3 may be estimated with data on the validation set.

Table 3. Calculation of the DCW (a_k, ^3EC_k) with Correlation Weights of Second Probe of Optimization Based on the Chemical Elements (a_k) and Morgan Extended Connectivity of Third Order (^3EC_k)

no.	a_k	^3EC_k	CW[a]	CW (^3EC_k)	CW[a] + CW (^3EC_k)
1	C	63	0.7320	1.2770	2.0090
2	C	45	0.7320	1.3730	2.1050
3	C	44	0.7320	1.7510	2.4830
4	C	41	0.7320	2.1530	2.8850
5	C	44	0.7320	1.7510	2.4830
6	C	45	0.7320	1.3730	2.1050
7	H	20	−0.2670	1.5560	1.2890
8	H	17	−0.2670	2.0310	1.7640
9	H	17	−0.2670	2.0310	1.7640
10	H	17	−0.2670	2.0310	1.7640
11	H	20	−0.2670	1.5560	1.2890
12	C	44	0.7320	1.7510	2.4830
13	H	23	−0.2670	−1.3800	−1.6470
14	H	23	−0.2670	−1.3800	−1.6470
15	O	28	3.8810	0.5960	4.4770
16	H	9	−0.2670	−2.0500	−2.3170

[a] CW = correlation weights. [b] Figure 1 presents numbers of the atoms/vertices, DCW (a_k,^3EC_k) = 23.289.

As an example, Figure 1 presents the Morgan extended connectivity of zero (^0EC), first (^1EC), second (^2EC), and third (^3EC) orders as well as the vertex numbers in LHFG of the model compound benzyl alcohol.

Model Development and Validation. With all three data sets, toxicants were divided into a training set and a test set. The separation of compounds into a training set and test set was done randomly, but by such a manner that all local invariants, which happened in all the compounds under consideration, were present in the structures (graphs) of the training set.

With the training set data, simple regression equations in the form of eq 3 were calculated by the Least Squares method. As the base of the OCWLGI model building, the descriptors of eq 1 calculated with the Morgan extended connectivity from zero to third orders were used for optimization.

RESULTS AND DISCUSSION

Table 1 lists the results of the optimization experiments. On each version of the DCW (a_k, ^xEC_k), three experiments on OCWLGI were carried out. The results taken *in toto* show that such OCWLGI gives models stable statistical characteristics. An examination of the results displayed in Table 1 reveals that the best statistical characteristics have been obtained with the DCW (a_k, ^3EC_k).

Table 2 lists the results of three probes of the optimization of the CWs on the DCW (a_k, ^3EC_k). Table 1 demonstrates that at best, the DCW (a_k, ^3EC_k) probes is the second one.

Calculation of the DCW (a_k, ^3EC_k) on the model compound benzyl alcohol is demonstrated in Table 3 (Note Figure 1 presents the vertex numeration.).

Table 4 lists the DCW (a_k, ^3EC_k) value for each compound in the training set calculated with CWs of probe 2. In addition, Table 4 lists the experimental determined toxicity values as well as those calculated from the EC3-2 equation listed in Table 1.

Table 4. Values of the Descriptor of Correlation Weights (a_k, ^3EC_k) Calculated with Correlation Weights of Probe 2 (Table 2)—Experimental and Calculated Aquatic Toxicity of the Training Set of Substituted Benzenes

compounds	DCW (a_k, ^3EC_k)[a]	log(IGC_{50}^{-1})[b] exptl	log(IGC_{50}^{-1})[b] calcd	compounds	DCW (a_k, ^3EC_k)[a]	log(IGC_{50}^{-1})[b] exptl	log(IGC_{50}^{-1})[b] calcd
benzyl alcohol	23.289	−0.830	−0.747	3-chlorophenol	38.851	0.870	0.519
benzylamine	31.119	−0.240	−0.110	3-chloro-5-methoxyphenol	38.666	0.760	0.504
(sec)phenethyl alcohol	24.669	−0.660	−0.634	2,5-dichloroaniline	39.001	0.580	0.531
3-phenyl-1-propanol	35.817	−0.210	0.272	3,5-dichloroaniline	39.001	0.710	0.531
1-phenyl-2-butanol	30.342	−0.160	−0.173	bromobenzene	35.371	0.750	0.236
methoxybenzene	33.857	−0.100	0.113	2,4-dichlorophenol	45.344	1.040	1.046
3-phenyl-1-butanol	31.964	0.010	−0.041	4-chloro-3,5-dimethylphenol	41.431	1.200	0.728
4-ethylbenzyl alcohol	31.365	0.070	−0.090	4-bromotoluene	41.078	1.000	0.700
(+-)−2-phenyl-2-butanol	33.541	0.060	0.087	3,4-dichlorotoluene	46.587	1.070	1.148
4-phenylbutanol	32.842	0.120	0.030	1-bromo-4-ethylbenzene	41.761	1.100	0.755
α,α-diethylbenzenepropanol	31.736	−0.070	−0.060	3-nitroaniline	37.585	0.030	0.416
ethoxybenzene	35.947	0.100	0.282	3-nitroanisole	41.040	0.720	0.697
methylbenzene	33.601	0.250	0.092	2-nitrotoluene	36.148	0.260	0.299
5-phenyl-1-pentanol	42.851	0.420	0.844	4-nitrotoluene	38.710	0.650	0.507
4-methylanisole	36.062	0.250	0.292	4-ethoxy-2-nitroaniline	44.317	0.760	0.963
1,1-diphenyl-2-propanol	42.095	0.750	0.782	α,α,α−4-tetrafluoro-m-toluidine	44.113	0.770	0.946
1,4-dimethylbenzene	34.304	0.250	0.149	4-nitrophenetole	41.212	0.830	0.711
benzophenone	42.582	0.870	0.822	1,2-dimethyl-3-nitrobenzene	39.060	0.560	0.536
(+-)−1,2-diphenyl-2-propanol	43.669	0.800	0.910	1,2-dichlorobenzene	40.880	1.000	0.684
6-phenyl-1-hexanol	42.578	0.870	0.822	3,5-dichlorophenol	45.344	1.560	1.046
isopropylbenzene	40.992	0.690	0.693	2,4,5-trichloroaniline	45.494	1.300	1.059
biphenyl	45.374	1.050	1.049	1,4-dibromobenzene	42.848	0.680	0.844
n-butylbenzene	45.456	1.250	1.056	5-hydroxy-2-nitrobenzaldehyde	39.475	0.330	0.569
n-amylbenzene	48.521	1.790	1.305	2-nitrophenol	39.626	0.670	0.582
aniline	26.015	−0.230	−0.525	2-nitroaniline	34.082	0.080	0.131
4-methylaniline	30.414	−0.050	−0.167	4-chlorobenzaldehyde	35.766	0.400	0.268
3-methylaniline	28.881	−0.280	−0.292	pentafluorobenzaldehyde	48.313	0.820	1.288
2-methylaniline	27.852	−0.160	−0.376	2,4,6-trichlorophenol	51.837	1.410	1.574
2-ethylaniline	30.054	−0.220	−0.197	1,2,4-trichlorobenzene	47.373	1.100	1.211
2-penthyl-3-butyn-2-ol	29.826	−0.180	−0.215	2,4,6-tribromophenol	54.789	1.910	1.814
3-ethylaniline	34.340	−0.030	0.152	1,3,5-trichlorobenzene	47.373	0.870	1.211
4-ethylaniline	33.883	0.030	0.115	pentafluoroaniline	45.055	0.260	1.023
4-butoxyaniline	42.215	0.610	0.792	1-fluoro-4-nitrobenzene	38.119	0.100	0.459
2,6-diethylaniline	36.244	0.310	0.307	4-chloro-2-nitrotoluene	42.641	0.820	0.827
4-pentyloxyaniline	43.547	0.970	0.900	2-chloro-6-nitrotoluene	42.641	0.680	0.827
2,6-diisopropylaniline	41.127	0.760	0.704	2,3,4,5-tetrachloroaniline	51.987	1.960	1.587
4-butylaniline	45.055	1.070	1.023	1,2,4,5-tetrachlorobenzene	53.866	2.000	1.739
4-methoxyphenol	28.213	−0.140	−0.346	phenyl isothiocyanate	49.686	1.410	1.399
benzaldehyde	29.273	−0.200	−0.260	3,5-dibromosalycilaldehyde	49.542	1.650	1.388
phenol	32.358	−0.350	−0.009	2,3,4,6-tetrachlorophenol	58.330	2.180	2.102
3-methoxyphenol	32.173	−0.330	−0.024	2,3,4,5-tetrachlorophenol	58.330	2.720	2.102
acetophenone	30.756	−0.050	−0.140	2-chloro-5-nitrobenzaldehyde	47.336	0.530	1.208
4-methylphenol	32.549	−0.160	0.006	1-chloro-4-nitrobenzene	40.804	0.430	0.677
2-methylphenol	31.826	−0.290	−0.053	4-chloro-3-nitrophenol	46.881	1.270	1.171
propiophenon	32.424	0.050	−0.004	1-bromo-2-nitrobenzene	41.788	0.750	0.757
2-ethylphenol	34.668	0.160	0.179	1-bromo-3-nitrobenzene	41.788	1.030	0.757
4-ethylphenol	34.990	0.210	0.205	3-trifluoromethyl-4-nitrophenol	53.883	1.650	1.741
3-ethylphenol	36.453	0.290	0.324	2-bromo-1-methyl-5-nitrobenzene	44.654	1.160	0.990
allylphenol	34.693	0.330	0.181	3,4,5,6-tetrabromo-2-methylphenol	61.734	2.570	2.379
2,3,6-trimethylphenol	38.272	0.280	0.472	pentachlorophenol	64.823	2.070	2.630
2,4,6-trimethylphenol	36.247	0.420	0.307	1,3-dinitrobenzene	45.881	0.760	1.090
3,4,5-trimethylphenol	43.624	0.930	0.907	2,5-dichloronitrobenzene	47.297	1.130	1.205
2,3,5-trimethylphenol	39.543	0.360	0.575	2,3-dichloronitrobenzene	47.297	1.070	1.205
4-isopropylbenzaldehyde	40.155	0.670	0.625	3,5-dichloronitrobenzene	47.297	1.130	1.205
valerophenone	37.679	0.560	0.423	3,4-dichloronitrobenzene	47.297	1.160	1.205
4-propylphenol	35.817	0.640	0.272	pentabromophenol	69.743	2.660	3.030
iodobenzene	39.697	0.500	0.587	2,6-dinitroaniline	43.753	0.840	0.917
4-(tert)butylphenol	39.231	0.910	0.549	3,4-dinitrobenzyl alcohol	42.913	1.090	0.849
4-hexyloxyaniline	51.143	1.380	1.518	1,4-dinitrobenzene	46.016	1.300	1.101
4-(tert)pentylphenol	47.864	1.230	1.251	2-bromo-4,6-dinitroaniline	51.782	1.240	1.570
4-penthyloxybenzaldehyde	46.805	1.180	1.165	1,2,3-trichloro-4-nitrobenzene	53.790	1.510	1.733
octanophenone	54.193	1.890	1.766	1,3,5-trichloro-2-nitrobenzene	53.790	1.430	1.733
4-nonylphenol	62.175	2.470	2.415	2,4-dinitrophenol	47.254	1.060	1.202
1,3-dihydroxybenzene	28.549	−0.650	−0.319	2,5-dinitrophenol	49.114	1.040	1.353
thiobenzamide	34.444	0.090	0.160	4-(tert)butyl-2,6-dinitrophenol	57.870	1.800	2.065
4-chloroaniline	32.508	0.050	0.003	1,2,3-trifluoro-4-nitrobenzene	45.735	1.890	1.078
2-chloroaniline	32.508	−0.170	0.003	2,3,4,5-tetrachloronitrobenzene	60.283	1.780	2.261
benzyl chloride	40.094	0.100	0.620	1,5-difluoro-2,4-dinitrobenzene	53.497	2.080	1.709
2-chloro-4-methylaniline	36.907	0.180	0.361	1-iodo-2,4-dinitrobenzene	57.684	2.120	2.050
4-chloroanisole	40.350	0.600	0.640	1-fluoro-2,4-dinitrobenzene	49.689	1.710	1.400
olivetol	51.609	1.310	1.556	pentafluoronitrobenzene	53.351	2.430	1.697
salicylaldehyde	34.588	0.420	0.172	1-bromo-2,4-dinitrobenzene	53.358	2.310	1.698
2-hydroxy-4-methoxyacetophenone	35.334	0.550	0.233	1,5-dichloro-2,3-dinitrobenzene	55.364	2.420	1.861
4-chlorophenol	38.851	0.540	0.519	1-chloro-2,4-dinitrobenzene	52.374	2.160	1.618
chlorobenzene	34.387	−0.130	0.156	1,3,5-trichloro-2,4-dinitrobenzene	65.360	2.190	2.674
phenethyl bromide	41.761	0.500	0.755	4-chloro-3,5-dinitrobenzonitrile	57.688	2.660	2.050
4-chloro-3-methylphenol	41.535	0.800	0.737	1,4-dinitrotetrachlorobenzene	71.988	2.820	3.213
4-chlorobenzophenone	49.075	1.500	1.350	3-amino benzyl alcohol	23.688	−1.130	−0.714
5-bromovanillin	43.334	0.620	0.883				

[a] Correlation weights (CWs) of local graph invariants (i.e., Descriptor of Correlation Weights (DCW). [b] IGC_{50}^{-1} = 50% growth inhibitory concentration.

Table 5. Values of the Descriptor of Correlation Weights (a_k, ^3EC_k) Calculated with Correlation Weights of Probe 2 (Table 2)—Experimental and Calculated Aquatic Toxicity of the Test Set of Substituted Benzenes

compounds	DCW (a_k, ^3EC_k)[a]	log(IGC_{50}^{-1})[b] exptl	log(IGC_{50}^{-1})[b] calcd
benzene	27.894	−0.120	−0.372
1-phenyl-1-butanol	31.410	−0.010	−0.086
thioanisole	31.899	0.180	−0.047
4-biphenylmethanol	42.751	0.920	0.836
4-ethylbiphenyl	56.326	1.970	1.939
3-ethoxy-4-hydroxybenzaldehyde	41.026	0.020	0.695
3-methylphenol	35.042	−0.080	0.209
2,4-dimethylphenol	32.040	0.140	−0.035
butyrophenone	38.534	0.300	0.493
heptanophenone	46.211	1.560	1.117
4-chlorobenzamide	37.478	0.160	0.407
3-chloroaniline	32.508	0.220	0.003
2-tolunitrile	30.566	−0.240	−0.155
4-hexylresorcinol	52.075	1.800	1.594
2-amino-5-chlorobenzonitrile	36.563	0.440	0.333
3,4-dichloroaniline	39.001	0.560	0.531
nitrobenzene	34.311	0.140	0.149
3-nitrotoluene	37.177	0.420	0.382
1,2-dimethyl-4-nitrobenzene	42.422	0.590	0.809
2-chloro-4-nitroaniline	44.213	0.750	0.955
2,4,5-trichlorophenol	51.837	2.100	1.574
1-chloro-3-nitrobenzene	40.804	0.730	0.677
2,3,5,6-tetrachloroaniline	51.987	1.760	1.587
4,5-dichloro-2-nitroaniline	47.068	1.660	1.187
2,4-dichloro-6-nitroaniline	47.068	1.260	1.187
1-chloro-2-nitrobenzene	40.804	0.680	0.677
1-fluoro-3-iodo-5-nitrobenzene	49.922	1.090	1.419
3-nitrobenzonitrile	39.128	0.450	0.541
1,2-dinitrobenzene	42.378	1.250	0.805
2,4-dinitrotoluene	50.688	0.870	1.481
2,4-dichloronitrobenzene	47.297	0.990	1.205
2,4-dinitroaniline	44.305	0.720	0.962
3-chloro-4-fluoro-1-nitrobenzene	44.612	0.800	0.987
6-chloro-2,4-dinitroaniline	50.798	1.120	1.490
1,2,4-trichloro-5-nitrobenzene	53.790	1.530	1.733
2,6-dinitrophenol	46.960	0.830	1.178
4,6-dinitro-2-methylphenol	53.285	1.730	1.692
2,3,5,6-tetrachloronitrobenzene	60.283	1.820	2.261
2,3,4,6-tetrafluoronitrobenzene	49.543	1.870	1.388
1,2-dichloro-4,5-dinitrobenzene	55.364	2.210	1.861
1,3-dinitro-2,4,5-trichlorobenzene	65.360	2.600	2.674
4-ethyl benzyl alcohol	31.365	0.070	−0.090
3,4-dimethylaniline	34.126	−0.160	0.134
4-isopropylaniline	36.897	0.220	0.360
4-hydroxyphenethyl alcohol	28.356	−0.830	−0.335
4-hydroxypropiophenone	33.632	0.050	0.094
4-(tert.)-penthylphenol	47.864	1.230	1.251
2-(4-chlorophenyl)ethylamine	31.390	0.140	−0.088
vanillin	35.857	−0.030	0.275
4-iodophenol	44.161	0.850	0.950
4-bromobenzophenone	50.059	1.260	1.430
4-methyl-2-nitroaniline	39.918	0.370	0.605
2,4-dinitrotoluene	50.688	0.870	1.481
2,4-dibromophenol	47.312	1.400	1.206
4-bromo-2,6-dichlorophenol	52.821	1.780	1.654
pentafluorophenol	51.398	1.630	1.539
4-nitrobenzyl chloride	45.203	1.180	1.035
4-chloro-6-nitro-3-methylphenol	47.757	1.630	1.243
2,6-diiodo-4-nitrophenol	60.973	1.810	2.317
2,4-dichloro-6-nitrophenol	52.612	1.750	1.637

[a] Correlation weights (CWs) of local graph invariants (i.e., Descriptor of Correlation Weights (DCW). [b] IGC_{50}^{-1} = 50% growth inhibitory concentration.

Table 5 lists the DCW (a_k, ^3EC_k) value for each compound in the test set calculated with CWs of probe 2. In addition, Table 5 lists the experimental determined toxicity values as well as those calculated from the EC3-2 equation listed in Table 1.

Table 6. Statistical Characteristics of the OCWLGI Models of the Data from Akers et al.[15]

kind of LIs	no. of LIs	training set $n = 23$ r^2	s	F	validation set $n = 16$ r^2	s	F	all structures $n = 39$ r^2	s	F
^0EC	9	0.5376	0.704	24	0.4680	0.888	12	0.5074	0.773	38
^1EC	14	0.8942	0.337	178	0.8544	0.462	82	0.8765	0.387	263
^2EC	21	0.9119	0.307	217	0.8851	0.406	108	0.9009	0.346	337
^3EC	35	0.9140	0.304	223	0.8804	0.414	103	0.8999	0.348	333

Table 7. Correlation Weights of the DCW (a_k, ^2EC_k) in Modeling of Log ($1/IGC_{50}$) Data from Akers et al.[15]

local invariant	correlation weight of local invariants
Chemical Elements, a_k	
H	0.065
C	0.067
N	3.462
O	−0.010
Cl	0.051
Br	3.400
Morgan Extended Connectivity of Second Order (^2EC_k)	
5	3.888
7	0.990
8	−0.385
10	0.369
11	0.016
13	0.002
17	6.400
20	2.803
0022	0.939
23	1.173
24	−0.932
25	3.950
26	−1.276
28	1.246
29	−1.205

The relationship between the aquatic toxicity predicted by the DCW (a_k, ^3EC_k) model obtained with the data on the training set (without any information on the test set) and the observed or experimental toxicity values for the test set is presented in eq 4:

$$\text{obs. } (\log(IGC_{50}^{-1})) = 0.991\ (\text{pred. } (\log(IGC_{50}^{-1})) - 0.012 \quad (4)$$

$$n = 60,\ r^2 = 0.863,\ s = 0.28,\ F = 372,\ \Pr > F = 0.0001$$

The slope near unity and the intercept near zero, demonstrates a one to one fit between observed and predicted values. Test compounds with large residual values include 3-ethyoxy-4-hydroxybenzaldehyde, 2,4,5-trichlorophenol, and 2,4-dinitrotoluene.

As noted in Table 1, when using correlation weights of local graph invariants as descriptors and simple linear regression analysis, the best fitted QSAR was for the LI EC3-2. While this QSAR based on 157 derivatives used 89 variables, it was not in violation of the minimum 5:1 ratio of chemicals to descriptors suggested by Topliss and Costello[19] as the 89 variables were consolidated into a single descriptor.

Table 8. Values of the DCW (a_k, ^2EC_k) in Modeling of Log ($1/IGC_{50}$) Data from Akers et al.[15] as Well as Experimental and Calculated Data with Eq 7 Values of Toxicity Endpoint under Consideration

compounds	DCW (a_k, ^3EC_k)	log(IGC_{50}^{-1}) exptl	log(IGC_{50}^{-1}) calcd
Training Set			
2-chloroethanol	12.563	−1.420	−1.314
2,2-dichloroethanol	12.549	−0.990	−1.316
1-chloro-2-propanol	14.062	−1.490	−1.053
3-chloro-2,2-dimethyl-1-propanol	14.816	−0.780	−0.921
3-chloro-1,2-propanediol	10.359	−1.630	−1.698
4-chloro-1-butanol	13.613	−0.760	−1.131
6-chloro-1-hexanol	17.975	−0.270	−0.370
2-bromoethanol	15.912	−0.850	−0.730
2,2,2-tribromoethanol	22.582	0.110	0.433
1,3-dibromo-2-propanol	20.746	0.570	0.113
3-bromo-2,2-dimethyl-1-propanol	18.165	−0.460	−0.337
3-bromo-1,2-propanediol	13.708	−1.210	−1.114
6-bromo-1-hexanol	21.324	0.010	0.214
chloroacetonitrile	26.719	0.850	1.155
dichloroacetonitrile	26.705	0.970	1.152
2-chloropropionitrile	16.037	−0.860	−0.708
4-chlorobutyronitrile	14.651	−0.930	−0.950
bromoacetonitrile	30.068	2.230	1.739
2-bromopropionitrile	19.386	0.630	−0.124
3-bromopropionitrile	19.386	−0.500	−0.124
5-bromovaleronitrile	20.436	−0.210	0.059
1-bromopentane	23.628	0.480	0.616
1-bromooctane	30.171	1.870	1.757
Validation Set			
2,2,2-trichloroethanol	12.535	−0.460	−1.319
1,3-dichloro-2-propanol	14.048	−0.790	−1.055
3-chloro-1-propanol	11.177	−1.400	−1.556
8-chloro-1-octanol	22.337	0.490	0.391
1-bromo-2-propanol	17.411	−1.190	−0.469
3-bromo-1-propanol	14.526	−0.930	−0.972
8-bromo-1-octanol	25.686	1.040	0.975
trichloroacetonitrile	26.691	1.880	1.150
3-chloropropionitrile	16.037	−1.120	−0.708
5-chlorovaleronitrile	17.087	−0.630	−0.525
7-chloroheptanitrile	21.449	0.290	0.236
dibromoacetonitrile	33.403	2.400	2.320
4-bromobutyronitrile	18.000	−0.880	−0.366
7-bromoheptanitrile	24.798	0.510	0.820
1-bromohexane	25.809	0.940	0.996
1-bromoheptane	27.990	1.490	1.376

This high quality model

$$\log(IGC_{50}^{-1}) = 0.0813\ (DCW) + 2.636 \qquad (5)$$

$n = 157$, $r^2 = 0.883$, $s = 0.27$, $F = 1170$, $\Pr > F = 0.0001$

estimates the relative aquatic toxicity of substituted benzenes without the a priori identification of the mechanism of action. Moreover, eq 5 compares favorably with the QSAR

$$\log(IGC_{50}^{-1}) = 0.50\ (\log K_{ow}) + 9.85\ (S_{max}) - 3.47 \qquad (6)$$

$n = 197$, $r^2 = 0.816$, $s = 0.34$, $F = 429$, $\Pr > F = 0.0001$

which correlates toxic potency [i.e., log(IGC_{50}^{-1})] with hydrophobicity quantified by the 1-octanol/water partition coefficient (log K_{ow}) and electrophilic reactivity quantified by the molecular orbital parameter, maximum superdelocalizability (S_{max}).[11]

Table 9. Statistical Characteristics of the OCWLGI Models of Toxicity Data from Cronin et al.[20]

kind of LIs	no. of LIs	training set $n = 44$ r^2	s	F	validation set $n = 22$ r^2	s	F	all structures $n = 66$ r^2	s	F
^0EC	10	0.6077	0.609	65	0.7504	0.548	60	0.6570	0.585	123
^1EC	18	0.7401	0.495	120	0.8577	0.441	121	0.7780	0.475	224
^2EC	31	0.8972	0.311	367	0.9104	0.325	203	0.9011	0.314	583
^3EC	69	0.9330	0.252	584	0.8788	0.429	145	0.9043	0.318	605

Table 10. Correlation Weights of the DCW (a_k, ^2EC_k) in Modeling of Log ($1/IGC_{50}$) Data from Cronin et al.[16]

local invariant	correlation weight of local invariants
Chemical Elements, a_k	
H	−0.091
C	0.713
N	1.688
O	−0.702
Cl	1.336
Br	3.275
Morgan Extended Connectivity of Second Order (^2EC_k)	
4	2.136
5	2.856
6	−0.679
7	0.376
8	−1.169
9	1.014
10	0.337
11	−3.805
12	0.422
13	1.764
15	5.900
16	5.475
17	7.400
18	1.949
19	0.099
20	2.549
21	0.086
22	1.337
23	0.939
24	0.152
25	−1.607
26	−0.072
27	0.792
28	1.304
29	−0.329

The OCWLGI-based model presented herein is simple and provides a statistical fit slightly better than the physicochemical-based QSAR derived by Schultz.[11] However, the tradeoff for the improved fit is that the OCWLGI-based QSAR is not as transparent or mechanistically comprehensible.

It should be noted that catechols and benzoquinones were not included in the data sets evaluated herein. Catechols are capable of undergoing tautomerization to more electrophilic semiquinones.[10] Benzoquinones is based on their specific structure and one-electron reduction potential reactive via a series of mixed mechanisms.[20] Since both of these subclasses are small and can be easily identified from molecular structure, these deletions do not significantly detract from the application of the model.

In an effort to further demonstrate the utility of the OCWLGI approach, models for two other toxicity data sets[15,16] were calculated. While neither the original toxicity

Table 11. Values of the DCW (a_k, ^2EC_k) in Modeling of Log ($1/IC_{50}$) Data from Cronin et al.[16] as Well as Experimental and Calculated Data with Eq 8 Values of Endpoint under Consideration

		$\log(IC_{50}^{-1})$	
compounds	DCW (a_k, ^2EC_k)	exptl	calcd
Training Set			
2,2,2-trichloroethanol	11.540	−0.460	−0.640
2,2,2-tribromoethanol	17.357	0.110	0.348
1,3-dichloro-2-propanol	9.451	−0.790	−0.995
chloroacetonitrile	22.041	0.850	1.144
2-chloroethanol	8.686	−1.420	−1.125
3-chloropropionitrile	11.880	−1.000	−0.583
methyl-5-bromovaleriate	18.131	−0.080	0.479
3,4-hexanrdione	13.086	−0.010	−0.378
2-undecanone	21.859	1.500	1.113
1-chloro-2-propanol	8.024	−1.490	−1.238
6-chloro-1-hexanol	11.613	−0.270	−0.628
3-bromo-2,2-dimethyl-1-propanol	12.630	−0.460	−0.455
3-chloro-1-propanol	8.008	−1.400	−1.240
3-chloro-1,2-propandiol	5.475	−1.630	−1.671
2-bromopropionitrile	13.819	0.630	−0.253
dibromoacetonitrile	28.773	2.400	2.288
5-bromopentyl acetate	17.037	0.290	0.294
ethyl-2-bromopropionate	18.247	1.060	0.499
ethyl-2-bromoisobutyrate	16.812	0.150	0.255
ethyl-2-bromovaleriate	18.149	0.700	0.483
ethyl-2-bromohexanoate	20.658	0.860	0.909
2,4-pentadione	11.949	−0.270	−0.571
2,4-octanedione	16.460	0.130	0.196
2,3-butanedione	12.756	−0.230	−0.434
2,3-pentadione	15.093	−0.160	−0.037
2,4-nonanedione	18.969	0.510	0.622
3,5-heptanediene	13.999	−0.380	−0.223
acetone	3.903	−2.200	−1.938
2-butanone	5.828	−1.750	−1.611
2-decanone	19.350	0.580	0.687
butyraldehyde	11.435	−0.380	−0.658
valeraldehyde	11.673	−0.020	−0.618
heptaldehyde	16.691	0.000	0.235
undecylaldehyde	26.727	1.690	1.940
2-butenal	19.749	0.700	0.754
2-hexanal	20.170	0.760	0.826
2-heptanal	20.408	1.050	0.866
4-bromobutyronitrile	13.141	−0.470	−0.368
4-chlorobutyronitrile	11.202	−0.930	−0.698
ethyl-4-bromobutyrate	17.911	−0.030	0.442
ethyl-3-bromopropionate	18.247	0.130	0.499
DL-methyl-2-bromobutyrate	17.893	1.020	0.439
2,5-hexanedione	7.900	−1.400	−1.259
2,3-hexanedione	15.334	−0.210	0.004
Validation Set			
1-bromo-2-propanol	9.963	−1.190	−0.908
8-chloro-1-octanol	16.631	0.490	0.225
6-bromo-1-hexanol	13.552	0.010	−0.299
8-bromo-1-octanol	18.570	1.040	0.554
2-bromoethanol	10.625	−0.850	−0.796
4-chloro-1-butanol	6.595	−0.760	−1.481
2-chloropropionitrile	11.880	−0.860	−0.583
7-bromoheptonitrile	16.746	0.510	0.244
7-chloroheptonitrile	14.807	0.290	−0.085
3-bromopropionitrile	13.819	−0.500	−0.253
5-bromovaleronitrile	11.728	−0.210	−0.608
ethyl-5-bromovalerate	15.654	0.220	0.059
ethyl-6-bromohaxnoate	20.658	0.590	0.909
2,3-heptanedione	15.572	0.040	0.045
2-pentanone	9.076	−1.220	−1.059
2-tridecanone	26.877	2.120	1.965
octylaldehyde	19.200	0.450	0.661
dodecylaldehyde	29.236	1.760	2.366
2-pentanal	18.573	0.660	0.555
2-octenal	22.917	1.200	1.293
2-nonenal	25.426	1.600	1.719
2-decenal	27.935	1.850	2.145

data nor the DCW descriptor data are shown, the statistics of the models are presented.

Table 6 presents the statistical characteristics of the OCWLGI models for *Tetrahymena* toxicity [$\log(1/IGC_{50})$] for 39 halogen substituted aliphatic chemicals[15] representing both covalent and noncovalent mechanisms of toxic action.[21]

In the case of modeling these halo-aliphatic compounds, the DCWs for atoms in combination with ^2EC_k [eq 7]

$$\log(1/IGC_{50}) = 0.1744 * DCW\ (a_k, {}^2EC_k) - 3.505 \quad (7)$$

give a model with quality fit (see Table 6).

Statistical characteristics of the OCWLGI model presented in eq 7 is better than those reported by Akers et al.[19] ($r^2 = 0.696$, $s = 0.623$, $F = 40$) for the regression model of the toxicity of all 39 compounds based on the 1-octanol/water partition coefficient ($\log K_{ow}$) and quantum chemical parameter energy of the Lowest Unoccupied Molecular Orbitals (E_{LUMO}). However, the statistics for eq 7 are similar to those reported when the statistical outliers are removed[15] ($n = 34$, $r^2 = 0.915$, $s = 0.297$, $F = 162$). Table 7 lists values of optimal correlation weights for calculating of the DCW (a_k, ^2EC_k). Table 8 presents values of the DCW (a_k, ^2EC_k), experimental, and calculated with eq 7 values of $\log(1/IGC_{50})$.

Similarly, Table 9 lists the statistical characteristics of the OCWLGI models for *Vibrio fischeri* toxicity data ($\log 1/IC_{50}$) for 66 aliphatic chemicals.[16] Again, the tested substances represent a variety of chemical classes and mechanism of toxic action.[21]

In the case of modeling these *V. fischeri* data, the DCWs for atoms in combination with ^2EC [eq 8] are as follows:

$$\log(1/IC_{50}) = 0.1699 * DCW\ (a_k, {}^2EC_k) - 2.601 \quad (8)$$

Statistical characteristics of the OCWLGI models presented in eq 8 is better than those reported by Cronin et al.[20] ($r^2 = 0.814$, $s = 0.527$, $F = 143$) for the regression model of the toxicity of all 66 compounds based on hydrophobicity ($\log K_{ow}$) and electrophilicity (E_{LUMO}). Moreover, the statistics for eq 8 are better than those reported when the statistical outliers were removed[20] ($n = 63$, $r^2 = 0.846$, $s = 0.462$, $F = 171$).

CONCLUSIONS

The present study indicates that optimization correlation weights of local graph invariants produce reasonably well models of the aquatic toxicity endpoints under consideration. In the case of the benzene data, the most appropriate OCWLGI was based on third order of Morgan extended connectivity, and in the case of QSAR modeling of toxicity endpoints for aliphatic chemicals, the more appropriate correlation weighting of Morgan extended connectivity was that of the second order.

REFERENCES AND NOTES

(1) Karcher, W. Basic concepts and aims of QSAR studies. In *Quantitative Structure/Activity Relationships (QSAR) in Toxicology*; Coccini, T., Giannoni, L., Karcher, W., Manzo, L., Roi, R., Eds.; Commission of the European Communities: Luxembourg, 1992; pp 5−25.

(2) Sabljic, A. Topological indices and environmental chemistry. In *Practical Application of Quantitative Structure−Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W.,

Devillers, J., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1990; pp 61−82.
(3) Basak, S. C. A nonempirical approach to predicting molecular properties using graph-theoretic invarients. In *Practical Application of Quantitative Structure−Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W., Devillers, J., Eds.; Kluwer Academic Publishers: Dordrecht, The Netherlands, 1990; pp 83−104.
(4) Cronin, M. T. D.; Schultz, T. W. Development of quantitative structure−activity relationships for the toxicity of aromatic compounds to *Tetrahymena pyriformis*: Comparative assessment of methodologies. *Chem. Res. Toxicol.* **2001**, *14*, 1284−1295.
(5) Burden, F. R.; Winkler, D. A. A quantitative structure−activity relationships model for the acute toxicity of substituted benzenes to *Tetrahymena pyriformis* using Bayesian-regularized neural networks. *Chem. Res. Toxicol.* **2000**, *13*, 436−440.
(6) Agrawal, V. K.; Khadikar, P. V. QSAR prediction of toxicity of nitrobenzenes. *Bioorg. Med. Chem.* **2001**, *9*, 3035−3040.
(7) Basak, S. C.; Mills, D. R.; Balaban A. T.; Gute, B. D. Prediction of mutagenicity of aromatic and heteroaromatic amines from structure: A hierarchical QSAR approach. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 671−678.
(8) Estrada, E.; Uriarte, E. Quantitative structure-toxicity relationships using TOPS-MODE. 1. Nitrobenzene toxicity to *Tetrahymena pyriformis*. *SAR QSAR Environ. Res.* **2001**, *12*, 309−324.
(9) Toropov, A. A.; Toropova, A. P. Modeling of lipophilicity by means of correlation weighting of local graph invariants. *Theochem.* **2001**, *538*, 197−199.
(10) Schultz, T. W.; Sinks, G. D.; Cronin, M. T. D. Identification of mechanism of toxic action of phenols to *Tetrahymena pyriformis* from molecular structure. In *Quantitative Structure−Activity Relationships in Environmental Sciences- VII*; Fredenslund, F. C., Schüürmann, G., Eds.; SETAC Press: FL, 1997a; pp 329−342.
(11) Schultz, T. W. Structure-toxicity relationships for benzenes evaluated with *Tetrahymena pyriformis*. *Chem. Res. Toxicol.* **1999**, *12*, 1262−1267.
(12) Schultz, T. W. TETRATOX: The *Tetrahymena pyriformis* population growth impairment endpoint. A surrogate for fish lethality. *Toxicol. Methods* **1997**, *7*, 289−309.
(13) Larsen, J.; Schultz, T. W.; Rasmussen, L.; Hoofman, R.; Pauli, W. Progress in an ecotoxicological standard protocol with protozoa: Results from a pilot ring test with *Tetrahymena pyriformis*. *Chemosphere* **1997**, *35*, 1023−1041.
(14) SAS Institute Inc. (Statistical Analysis System) *SAS/STAT User's Guide*, 4th ed.; North Carolina, 1989; version 6, Vol. 2, p 846.
(15) Akers, K. S.; Sinks, G. D.; Schultz, T. W. Structure-toxicity relationships for selected halogenated aliphatic chemicals. *Environ. Toxicol. Pharmacol.* **1999**, *7*, 33−39.
(16) Cronin, M. T. D.; Bower, G. S.; Sinks, G. D.; Schultz, T. W. Structure-toxicity relationships for aliphatic compounds encompassing a variety of mechanisms of toxic action to *Vibrio fischeri*. *SAR QSAR Environ. Res.* **2000**, *11*, 301−312.
(17) Toropov, A. A.; Voropaeva, N. L.; Runab, I. N.; Rashidiva, S. Sh. Quantitative structure−property relationships for binary polymer−solvent systems: Correlation weighting of the local invariants of molecular graphs. *Polym. Sci. Ser.* **1999**, *A41*, 975−985.
(18) Bonchev, D.; Kier, L. B. Topological atomic indices and the electronic charges in alkanes. *J. Math. Chem.* **1992**, *9*, 75−85.
(19) Topliss J. G.; Costello, R. J. Chance correlations in structure−activity studies using multiple regression analysis. *J. Med. Chem.* **1972**, *15*, 5, 1066−1068.
(20) Schultz, T. W.; Sinks, G. D.; Cronin, M. T. D. Quinone-induced aquatic toxicity to *Tetrahymena*: Structure−activity relationships. *Aquat. Toxicol.* **1997b**, *39*, 267−278.
(21) Karabunarliev, S.; Mekenyan, O. G.; Karcher, W.; Russom, C. L.; Bradbury, S. P. Quantum-chemical descriptors for estimating the acute toxicity of electrophiles to the fathead minnow (*Pimephales promelas*): An analysis based on molecular mechanisms. *Quant. Struct.-Act. Relation.* **1996**, *15*, 302−310.

CI025555N