Theoretical Descriptors for the Correlation of Aquatic Toxicity of Environmental Pollutants by Quantitative Structure-Toxicity Relationships

Alan R. Katritzky* and Douglas B. Tatham

Florida Institute of Heterocyclic Compounds, Department of Chemistry, University of Florida, Gainesville, Florida 32611-7200

Uko Maran

Department of Chemistry, University of Tartu, 2 Jakobi Street, Tartu 51014, Estonia

Received February 8, 2001

Quantitative structure-toxicity relationships were developed for the prediction of aqueous toxicities for *Poecilia reticulata* (guppy) using the CODESSA treatment. A two-parameter correlation was found for class 1 toxins with $R^2 = 0.96$, and a five-parameter correlation was found for class 2 toxins with $R^2 = 0.92$. A five-parameter correlation for class 3 toxins had $R^2 = 0.85$. The correlations for class 4 toxins were less satisfactory. All the descriptors utilized are calculated solely from the structures of the molecules, which makes it possible to predict unavailable or unknown toxins.

INTRODUCTION

Numerous organic chemicals can be environmental pollutants. Because of this, during the past decade a great deal of effort has been put into the study of the relationships between a compound's structure and its toxicity. Attempts have been made to classify chemicals according to the mechanism of their toxicity and to screen them for their environmental risk assessment. Several studies relate toxic activity to the partition coefficient. Early papers by Overton[1] and Meyer [2] reported a correlation between the olive oil−water partition coefficient and narcosis of simple compounds. Later, Hansch and co-workers[3] developed relationships between the biological activities and the hydrophobic, electronic, and steric properties of compounds. The hydrophobic interaction is usually expressed by the octanol−water partition coefficient (log *P*). Most of the QSAR applications to toxicities have been developed for congeneric sets, but non-congeneric sets of compounds have also been documented in several reviews.[4−7]

THE CLASSIFICATION OF ORGANIC ENVIRONMENTAL POLLUTANTS

The research of Könemann and co-workers[8] on inert narcotic pollutants led to the so-called baseline toxicity concept,[9] which expresses the minimal toxic effect exerted by a chemical. In the framework of this concept, the toxicity is related to the octanol−water partition coefficient[4,10,11] expressed as the median lethal concentration (LC_{50}). Chemicals with toxicities in line with the baseline concept are classified as inert and are not as interacting with specific receptors in an organism. In aquatic toxicity, this mode of action is called narcosis.[12] It is considered to be completely nonspecific, depending solely on the compounds hydrophobicity, i.e., whether the compound is as toxic as its hydrophobicity indicates. However, many classes of compounds are more toxic than predicted by the baseline toxicity concept,[4] and these are now discussed.

(i) One group includes phenols, anilines, and nitrogen heterocycles. These compounds produce toxicity syndromes similar to those from inert narcotic pollutants but have greater toxicities.[13,14] They act by a so-called "polar narcosis" mechanism, which is associated with the presence of a strong hydrogen bond donor (polar) group in the molecule. Other examples include nonspecifically acting uncouplers of oxidative phosphorylation.[15] Nonpolar and polar narcotic pollutants are also classified, according to their mode of action, into narcosis I and narcosis II compounds.[16] It is estimated that about 60% of the industrial chemicals entering the aquatic environment exert toxicity by means of narcosis.[10] For narcosis, the lethal concentrations of a compound are often close for similar species (e.g. for the guppy and the fathead minnow), which is not surprising considering the nonspecific nature of narcotic toxins.

(ii) Another group of toxins consists of those that either react unselectively with common biomolecular chemical substructures, such as nucleophilic sites, or are bioactivated (metabolized into more toxic species).[17−19]

(iii) A final group of chemicals exhibit toxicity due to specific interactions with particular receptor molecules.[20]

Hermens et al. recently devised a rule-based system to classify individual compounds into four classes along lines similar to those described above: class 1 (inert chemicals: nonpolar narcosis), class 2 (less inert chemicals: polar narcosis), class 3 (reactive chemicals), and class 4 (specifically acting compounds, such as pesticides).[21] These rules rely on the presence or absence of certain structural or substructural features to assign the compound to one of the four classes. It is possible to calculate either an expected effect concentration (LC_{50}) or an expected range of the

* Corresponding author phone: (352)392-0554; fax: (352)392-9199; e-mail: katritzky@chem.ufl.edu.

10.1021/ci010011r CCC: $20.00
Published on Web 08/03/2001

Table 1. Examples of QSAR Models on Toxicities

no.	species[a]	compounds	value[b]	N	class[c]	R^2	s	#D	F	descriptors	ref
1	TA	alcohols, ketones, aromatic hydrocarbons	C^{-1}	18	1	0.968	0.227	1	*d*	LogP	24
2	GU	alcohols, ketones, alkyl halides, sub. benzenes	LC_{50}^{-1}	50	1	0.988	0.237	1	*d*	LogP	8
3	FM	alcohols, ketones, alkyl halides, sub. benzenes	LC_{50}	65	1	*d*	*d*	1	*d*	Log P, log(P + 1)	10
4	FM	phenols, anilines	LC_{50}	39	2	0.900	0.22	1	*d*	Log P	16
5	CI	phenols, anilines	LC_{50}^{-1}	95	2	0.818	0.320	1	419	Log P	25
6	GU	aldehydes	LC_{50}^{-1}	14	3	0.852	0.19	1	*d*	Log P	26
7	CI	aldehydes	IGC_{50}^{-1}	14	3	0.961	0.168	1	294	Log P	27
8	GU	organic halides	LC_{50}^{-1}	15	3	0.914	0.39	2	*d*	Log P, log $(2484 + k^{-1})$	28
9	FM	ketones, esters, alcohols, nitriles	LC_{50}	57	1/2/3	0.85	*d*	2	*d*	Log P, σ^d	29
10	FM	ketones, esters, alcohols, nitriles	LC_{50}	57	1/2/3	0.85	*d*	2	*d*	Log P, σ_I	29
11	FM	anilines, phenols, benzenes	LC_{50}^{-1}	114	1/2/3	0.81	0.436	2	245	Log P, E_{LUMO}	30
12	FM	anilines, phenols, benzenes	LC_{50}^{-1}	114	1/2/3	0.81	0.436	2	239	Log P, S^N_{av}	31, 30
13	FM	acrylates	LC_{50}^{-1}	18	3	0.78	0.57	3	17	A_R, $B_{\alpha-R}$. log *P*	33, 34
14	FM	benzenes	LC_{50}^{-1}	122	1/2/3/4	0.831	0.40	2	291	Log P, A_{max}	35
15	GU	benzenes	LC_{50}^{-1}	64	1/2/3/4	0.87	0.27	3	138	Log *P*, $A_{max,\ Dmax}$	34, 35
16	MP	ketones, aldehydes	EC_{50}	19	1/2/3	0.892	0.46	2	75	Log P, E_{LUMO}	36
17	CI	nitrobenzenes	IGC_{50}^{-1}	42	1/2/3	0.897	0.229	2	180	Log P, A_{max}	37
18	CI	nitrobenzenes	IGC_{50}^{-1}	42	1/2/3	0.881	0.246	2	154	Log P, E_{LUMO}	37
19	FM, GU	alcohols, ketones, alkyl halides, phenols, anilines	LC_{50}	172	1/2	0.901	*d*	2	*d*	Log P, Q^-	22
20	GU	alcohols, ketones, alkyl halides, phenols, anilines	LC_{50}	19	1/2	0.97	*d*	5	*d*	Log P. Q^-, Q^+, E_{LUMO}, E_{HOMO}	38
21	SN	alcohols, ketones, alkyl halides, phenols, anilines	LC_{50}	19	1/2	0.97	*d*	5	*d*	Log P. Q^-, Q^+, E_{LUMO}, E_{HOMO}	38
22	WF	alcohols, ketones, alkyl halides, phenols, anilines	LC_{50}	15	1/2	0.95	*d*	5	*d*	Log P. Q^-, Q^+, E_{LUMO}, E_{HOMO}	38
23	MB	chloroanilines, -phenols, -benzenes	EC_{50}	48	1/2	0.89	0.302	3	55	$^1\kappa$, α_Z, C_{Z-0}	42
24	MB	chlorinated organic compounds	EC_{50}	80	1/2/4	0.77	0.5	3	91	TLSER	43
25	WF	amines, chlorobenz. organotins organophos.	EC_{50}^{-1}	49	1/2/4	0.817	*d*	7	*d*	6 WHIM desc. & MW	44
26	TA	alcohols, ketones, alkyl halides, sub. benzenes	C^{-1}	84	1	0.947	0.246	4	351	LFER	45

[a] Abbreviation of species: TA, tadpole (*Rana temporaria*); GU, guppy (*Poecilia reticulata*); FM, fathead minnows (*Pimephales promelas*); CI, cilate (*Tetrahymena pyriformis*); MP, marine prokaryote (*Vibrio fischeri*); SN, snail (*Lymnaea stagnalis*); WF, water flea (*Daphnia magna*); MB, marine bacterium (*Photobacterium phosphoreum*). [b] In logarithmic scale. [c] Indicates class of the compound. [d] Value missing in original article.

possible effect concentrations for chemicals belonging to one of these four classes based on the octanol−water partition coefficient of the compound. An important drawback, which is inherent to all rule-based classification systems, is that it is impossible to classify compounds that do not fit the existing rules, even if they have structural features that would indicate toxicity.[22]

Moreover, any unambiguous division of compounds into discrete classes is hardly feasible (or rational) since overlap is inevitable and the qualitative variation of toxic action does not necessarily parallel the chemical notion of different classes.[23] In addition, some substances, such as hexane, are less toxic than would be expected because they are rapidly metabolized. Other toxins with log $P \geq 6$ (e.g. hexachlorobenzene) can have LC_{50} values that are higher than expected because their toxicity exceeds their water solubility.[8] Other substances have a very low membrane affinity (low log *P*) yet are quite toxic because they produce osmotic effects[24] rather than narcosis.

SURVEY OF PUBLISHED STRUCTURE-TOXICITY CORRELATIONS

Table 1 lists published quantitative structure-toxicity relationships, selected to exemplify the uses of experimentally derived log *P* values to estimate toxicity and ways in which log *P* has been combined with other descriptors. The first group of correlations in Table 1 (#1−7) involves structure-toxicity models, which correlate the toxicities for small fish solely with log *P*. Models #1−3 follow the baseline concept and accurately predict toxicities for class 1 compounds: alcohols, ketones, alkanes, aromatic hydrocarbons, and alkyl halides.[8,10,24]

Log *P* can be used as the sole descriptor for the correlation of toxicities higher than baseline. Veiht and Broderius showed that polar narcotics could be successfully correlated in the set (#4) of 39 phenols and anilines[16] for fathead minnows. A similar but larger set (#5) of 95 phenols and anilines was successfully used by Schultz et al. to correlate and predict the 60 h growth inhibition of class 2 compounds for *Tetrahymnea pyriformis*.[25] This study demonstrated the validity of growth inhibition assays as well as demonstrating the strong correlation between the toxicity of polar narcotics and log *P*. The same study[25] also showed that phenol and aniline derivatives containing phenylazo-, dinitro-, four or five halo-substituents, or one nitro with two or more halo-substituents were outliers and behaved as respiratory uncouplers rather than polar narcotics. The quality of the equation for set #5 was enhanced ($R^2 = 0.904$) by adding a second descriptor, $\sum\sigma$, the summation of the electronic parameter.

Hermens and co-workers found (#6) that the guppy 14-day toxicity of aldehydes (class 3) correlates with log *P*.[26] Adding the corresponding reaction rate constants to the model did not improve the model. Considerably better results for simple aliphatic aldehydes were obtained by Schultz et al.[27] for *Tetrahymnea pyriform* (#7).

Models #8−10 in Table 1 used additional experimentally derived descriptors together with log *P*. Hermens et al. correlated 14-day guppy toxicities for 15 reactive organic halides (#8) with log *P* values and their reaction rates with 4-(4-nitrobenzyl)pyridine as a standard nucleophile.[28] Purdy found that the toxicity for fathead minnows (#9) was predicted with greater accuracy if either the Taft σ^* or the σ_I (polar and inductive) substituent constant was included in the correlation.[29] This provides (in both cases) the following improvements in R^2 values: for ketones, from 0.82 to 0.96; for esters, from 0.77 to 0.91; for alcohols, from 0.84 to 0.93; and most significantly for nitriles, from 0.15 to 0.85. The R^2 value of the combined data set improved from 0.59 (using only log *P*) to 0.85 for the correlation using both log *P* and the substituent constants (#9 and 10). Highly carbanionic species, which can be transported readily across the gills and then metabolized into more reactive substances, were outliers.

Structure-toxicity models (#11−22) in the third part of Table 1 combine log *P* with one or more additional theoretical molecular descriptor(s), generally derived from quantum-chemical calculations. Veith and Mekenyan predicted the toxicities for a set of 114 benzenes, anilines, and phenols using log *P* together with either the LUMO energy (#11) or the average acceptor superdelocalizability (S^N_{av}) calculated for the conjugated π-bonds in the molecule (#12).[30,31] The two models (#11 and 12) are of similar quality, which is understandable because the superdelocalizability index[32] is derived from the LUMO energy. The advantage of using the reactivity index is that it reveals the type of the reaction and possible reaction site. The combination of log *P* and the superdelocalizability index allows compounds with different modes of action to be modeled in one equation. The authors also showed how one could distinguish between the compounds with different mode of action on a "toxicity response plane" using the average acceptor superdelocalizability.

Karanbunarliev et al.[33,34] used quantum chemically derived acceptor superdelocalizabilities for the carbon of the polar group to which the electron transfer occurs (A_R), the bond order for the bridging single bond that acquires double bond character at the intermediate step of the reaction ($B_{\alpha-R}$), and log *P* to predict the acute aquatic toxicity for fathead minnows for a set of 18 Michael-type acceptors, mostly acrylates, with C═O as the polar group (#13). Acceptor superdelocalizability described most of the variance, and the authors showed that quantum chemical descriptors are compatible with the accepted mechanism. Karanbunarliev et al. also analyzed the toxicities of the fathead minnow (#14) in a diverse set of 122 substituted benzenes.[35] In this study, they used maximal acceptor superdelocalizability (A_{max}) for the π-sites of the benzene ring in addition to log *P*. The same group also analyzed (#15) the toxicity of 64 substituted benzenes toward guppies.[34,35] The maximal acceptor superdelocalizability (A_{max}) and the maximal donor superdelocalizability (D_{max}) for the π-sites of the benzene ring were used in addition to log *P*.

Cronin et al. performed a series of studies combining quantum chemical descriptors with log *P*. They analyzed the toxicity of a diverse set of 19 compounds including alkanones (nonreactive compounds with baseline narcosis mechanism), aldehydes (Schiff base-forming mechanism of electrophilicity), and alkenals (Michael-type acceptor mechanism of electrophilicity)[36] to *Vibrio fischeri* using the so-called response-surface approach.[31] The relative reactivity of the compounds was described by including a quantum chemical parameter, the energy of the lowest unoccupied molecular orbital (E_{LUMO}) (#16). They also described each subset with separate correlation equations, involving log *P* as a single descriptor. In subsequent work (#17) they analyzed the acute toxicity of a diverse group of 42 alkyl- and halogen-substituted nitro- and dinitrobenzenes, expressed as the 50% growth inhibition concentration against *Tetrahymena pyriformis*.[37] The best correlation equation was obtained by relating toxicity to log *P* and the maximum acceptor superdelocalizability on the benzene ring (A_{max}). Superdelocalizability, an atom based descriptor, is the most significant parameter of equation #17, describing 80% of the variance. Another model (#18) of similar quality ($R^2 = 0.881$) was obtained with E_{LUMO}, a descriptor based on the entire molecule.

Recently, significant contributions to the QSAR studies of acute aquatic toxicity have been made by Hermens and co-workers, who addressed the relationship between class 1 and class 2 narcotic pollutants. They applied a partial least squares (PLS) and multilinear regression (MLR) analysis to a set of 172 compounds (#19) to model the combined toxicity of *Pimephales promelas* and/or *Poecilia reticulata*.[22] A pool of 11 descriptors (log *P* and quantum-chemical descriptors) provided a model of four significant latent variables with $R^2 = 0.928$. Multilinear regression analysis of this set revealed that log *P* represents 87% of the variance in the data set and the best two-parameter correlation equation involved the most negative charge (Q^-) on any non-hydrogen atom (#19). Based on the PLS and MLR analysis they concluded that the hydrogen bond acceptor and (less importantly) donor capacities are the prime determinants for differentiating between class 1 and class 2 compounds.

Hermens and co-workers extended the range of species and constructed unified correlation equations (#20−22) for the water flea (*Daphnia magna*), guppy (*Poecilia reticulata*), and pond snail (*Lymnaea stagnalis*).[38] In addition to the conventional hydrophobicity term (log *P*), equations (#20−22) each include four descriptors for the ability of a molecule to form hydrogen bonds: a hydrogen bonding acceptor term (Q^-, the most negative partial charge on any non-hydrogen atom of the molecule), a hydrogen bonding donor term (Q^+, the most positive partial charge on a hydrogen atom), and E_{HOMO} and E_{LUMO} to account for the covalent contribution to the hydrogen bond. The training sets gave high quality correlation equations for three species (#20−22). The validation sets resulted in good correlations for 143 compounds to *Poecilia reticulata* ($R^2 = 0.93$); for four compounds to *Lymnaea stagnalis* with $R^2 = 0.88$; and for 52 compounds to *Daphnia magna* ($R^2 = 0.96$).

For all three species, the most important descriptor was log *P*, followed by E_{LUMO}. Hermens group concluded[38] that pollutants with high log *P*, low E_{LUMO}, and high absolute charges, i.e., electronegative hydrophobic chemicals with charged atoms, are highly toxic. Q^+ seemed to be more important than Q^- (which was used in the model #19, Table 1). However, from models #20−22, it is not clear that a hydrogen bond is actually involved. This study,[38] together

with other studies carried out by the same group,[39,40] indicates that more chemicals than those recognized as class 1 act by the same mode of action attributed to class 1 chemicals. Therefore, Hermens and co-workers suggested that the definition of baseline toxicity based on log *P* should be redefined based on membrane-water partition coefficients (log K_{MW})[38] and used a data set of 11 class 2 and eight class 1 compounds to test this suggestion.[39] Plots of toxicity versus log *P* distinguish classes 1 and 2, but this distinction disappears when toxicity is plotted against log K_{MW} ($R^2 = 0.98$ after the removal of quinoline, a significant outlier). Recently, Hermens group reported that for class 2 chemicals the partitioning behavior to phospholipids is higher than to octanol, whereas no difference is observed for class 1.[41]

Models (#23−26) in the final section of Table 1 do not include log *P* as a descriptor but use mainly theoretical molecular descriptors to predict the toxicity. Gombar and Enslein (#23), who found it difficult to calculate some log *P* values, produced a three-descriptor correlation from a regression analysis using 23 topological, polarizability, and atomic charge descriptors ($R^2 = 0.80$) to determine toxicity without using log *P*.[42] Their data set (#23) comprised 48 chlorobenzenes, chloroanilines, and chlorophenols and was described by the first-order shape index, effective polarizability, and the sigma atomic charges. Brüggemann and co-workers applied TLSER parameters to correlate the acute toxicities for *Photobacterium phosphoreum* to a set (#24) of 80 chlorinated organic compounds included narcosis I, narcosis II, and specifically acting toxins.[43] They found that molecular volume was the most important parameter, followed by the dipolarity/polarizability term and the basic electrostatic term. They also found better correlations for describing subsets of the compounds.[43]

Todeschini et al. applied weighted holistic invariant molecular (WHIM) descriptors to analyze (#25) the EC_{50} toxicity to *Daphnia magna* for 49 compounds:[44] seven descriptors covered 81% of variance in the current diverse set of compounds.

Interesting work (#26) by Abraham et al. utilized linear free energy relationship (LFER) analysis to model Overton's original data on tadpole narcosis.[45] The solute excess molar refraction, solute dipolarity/polarizability, the solute hydrogen-bond basicity, and solute volume gave remarkably good results (#26) for the set of 84 compounds. Evidently, the solute hydrogen-bond basicity markedly reduces the solute narcotic activity and solute excess molar refraction (somewhat) and solute volume (greatly) increases the solute narcotic activity, as previously found with smaller sets.

AIMS AND OBJECTIVE OF THE PRESENT WORK

As can be seen from Table 1, log *P* has almost always been used as an anchor descriptor for QSAR estimations of toxicity, except for correlations #23−26. In conjunction with log *P*, other descriptors correct for electronic interactions with the surrounding media. The majority of these descriptors are obtained from semiempirical (all use AM1 parametrization) quantum chemical calculations, particularly positive/negative partial charges, and bond orders. More extensive treatments have utilized HOMO/LUMO energies, and various modifications of these energies, in the form of reactivity indexes. Reactivity indexes were particularly useful for the analysis of data sets that include reactive and specifically acting compounds. The correlations of Table 1 generally use experimentally determined values of log *P*, with some exceptions where theoretical values were used to fill the gaps in available experimental values. The success of the treatments utilizing theoretical molecular descriptors (Table 1: #23−25) encouraged us to look for the correlations for the toxicities of classes 1−4 compounds using only theoretically derived descriptors.

In recent years, we have worked extensively on the development of methodology for a general QSAR/QSPR approach and its applications. Our current approach is coded as the CODESSA software package. CODESSA combines diverse methods for quantifying the structural information about the molecule with advanced statistical analysis to establish molecular structure−property/activity relationships. CODESSA enables the calculation of a large number of quantitative descriptors based solely on the molecular structural information[46−48] and codes this chemical information into mathematical form. CODESSA has been applied successfully to predict a variety of physical, chemical, and biological properties of compounds,[49] ranging from boiling points to complex properties of surfactants and polymers. The current work is also encouraged by our QSAR analysis of genotoxicity of aromatic and heteroaromatic amines,[50] which provided a successful QSAR model using only theoretically derived descriptors.

In the current study we exclude experimentally determined log *P* values from estimations of toxicity and utilize solely structure-based descriptors. We have gathered structure based theoretical descriptors into a large database and have derived QSAR models for the toxicities of each of the classes (1−4) of toxic compounds. The primary purpose of the present work is to establish QSAR models of aquatic toxicities for a diverse set of classes (1−4) compounds.

DATA SET AND METHODOLOGY

The data set for classes 1, 2, 3, and 4 toxins contains 293 compounds (Table 2) with toxicity data (LC_{50}) and experimental *P* values for *Poecilia reticulata* that were obtained from several sources. Items 1−161 were taken from a compilation by Ramos, Vaes, Verhaar, and Hermens,[38] compounds 162−242 from an earlier paper by Verhar et al.,[21] the substituted phenols, 242−254, from Saarikoskis paper,[51] and the remaining data from another paper from the same group.[22] Our total set included 90 class 1 compounds, 121 class 2 compounds (amines, anilines, phenols, and nitroaromatics, several of which approached class 3 toxins in Verhar scheme,[21] 51 class 3 compounds (epoxides, aldehydes, miscellaneous halogen containing compounds with the halogens positioned alpha to vinyl or carbonyl groups, and aromatic nitriles or phenols with four or more halogen substituents), and 31 class 4 specifically reacting chemicals (mostly phosphates or sulfates).

The molecular structures from the data set (Table 2) were drawn and preoptimized using the MM+ molecular mechanics method included in HyperChem (version. 5.1).[52] Final structural optimizations were performed using AM1[53] parametrizations and eigenvector following geometry optimization procedure[54] within the semiempirical quantum-chemical program MOPAC 6.0.[55] Gradient norms of 0.01 kcal/Å were

Table 2. Data Set

no.	class	structure name	Log[c] LC_{50} exp.	Log[a] LC_{50}	Δa	Log[b] LC_{50}	Δb	$logP_{exp}$	$logP_C$
1	1	1,1,1-trichloroethane	3.00	2.71	−0.29	2.30	−0.70	2.49	2.25
2	1	1,1,2,2-tetrachloroethane	2.23	2.16	−0.07	1.77	−0.46	2.39	2.60
3	1	1,1,2-trichloroethane	2.82	2.72	−0.10	1.98	−0.84	1.89	2.24
4	1	1,1-dichloroethane	3.31	3.27	−0.04	2.63	−0.68	1.79	1.88
5	1	1,2,3,4-tetrachlorobenzene	0.65	0.43	−0.22	0.86	0.21	4.64	4.28
6	1	1,2,3,5-tetrachlorobenzene	0.57	0.43	−0.15	0.86	0.29	4.65	4.29
7	1	1,2,3-trichlorobenzene	1.11	1.13	0.02	1.27	0.16	4.13	3.76
8	1	1,2,3-trichloropropane	2.45	2.24	−0.21	1.70	−0.75	1.98	2.69
9	1	1,2,4,5-tetrachlorobenzene	0.15	0.42	0.27	0.84	0.69	4.6	4.29
10	1	1,2,4-trichlorobenzene	1.17	1.12	−0.05	1.25	0.08	4.05	3.77
11	1	1,2-dichlorobenzene	1.60	1.83	0.23	1.74	0.14	3.43	3.23
12	1	1,2-dichloroethane	3.06	3.27	0.21	2.55	−0.51	1.48	1.88
13	1	1,2-dichloropropane	3.01	2.80	−0.21	2.27	−0.74	1.98	2.33
14	1	1,2-ethanediol	5.90	6.07	0.17	5.11	−0.79	−1.36	−1.30
15	1	1,3,5-trichlorobenzene	1.26	1.11	−0.15	1.24	−0.02	4.18	3.78
16	1	1,3-dichlorobenzene	1.72	1.82	0.10	1.73	0.01	3.52	3.24
17	1	1,3-dichloropropane	2.87	2.80	−0.07	2.21	−0.66	2.00	2.33
18	1	1,4-dichlorobenzene	1.44	1.82	0.38	1.71	0.27	3.44	3.24
19	1	1,4-dimethoxybenzene	2.93	2.72	−0.21	2.36	−0.57	2.03	2.07
20	1	1-butanol	4.37	4.30	−0.07	3.80	−0.57	0.88	0.82
21	1	1-chlorobutane	3.02	2.90	−0.12	2.52	−0.50	2.64	2.41
22	1	1-decanol	1.19	1.49	0.30	1.92	0.73	4.57	3.50
23	1	1-dodecanol	0.74	0.56	−0.18	1.22	0.48	5.13	4.39
24	1	1-hexanol	2.98	3.36	0.38	3.25	0.27	2.03	1.72
25	1	1-nonanol	1.60	1.96	0.36	2.27	0.67	4.26	3.06
26	1	1-octanol	2.02	2.43	0.41	2.61	0.59	3.00	2.61
27	1	1-undecanol	0.79	1.02	0.23	1.57	0.78	4.53	3.95
28	1	2-(2-ethoxyethoxy)ethanol	5.30	5.01	−0.29	3.93	−1.37	−0.54	−0.71
29	1	2,2,2-trichloroethanol	3.31	3.55	0.24	2.67	−0.64	1.42	1.02
30	1	2,3,4-trimethoxyacetophenone	2.92	2.53	−0.39	2.24	−0.68	1.63	1.51
31	1	2,4,5-trichlorotoluene	0.94	0.66	−0.28	0.71	−0.23	4.78	4.21
32	1	2,4-dichloroacetophenone	1.8	1.87	0.07	1.03	−0.77	2.73	2.74
33	1	2,4-dichlorotoluene	1.46	1.36	−0.10	1.14	−0.32	4.24	3.68
34	1	2,6-dimethoxytoluene	2.13	2.26	0.13	2.02	−0.11	2.87	2.51
35	1	2-butanone	4.65	4.45	−0.20	3.26	−1.39	0.29	0.66
36	1	2-butoxyethanol	3.92	4.19	0.27	3.54	−0.38	0.83	0.50
37	1	2-decanone	1.57	1.64	0.07	1.54	−0.03	3.73	3.34
38	1	2-ethoxyethanol	5.26	5.12	−0.14	4.16	−1.10	−0.1	−0.39
39	1	2-hydroxy-4-methoxyacetophenone	2.52	2.65	0.13	2.08	−0.44	1.98	1.82
40	1	2-isopropoxyethanol	4.72	4.65	−0.07	3.96	−0.76	0.05	0.06
41	1	2-methoxyethanol	5.36	5.59	0.23	4.43	−0.93	−0.77	−0.84
42	1	2-methyl-2,4-pentanediol	4.96	4.19	−0.77	4.37	−0.59	−0.68	0.49
43	1	2-octanone	2.45	2.57	0.12	2.17	−0.28	2.37	2.45
44	1	2-phenoxyethanol	3.40	3.34	−0.06	2.82	−0.58	1.16	1.34
45	1	2-propanol	5.16	4.76	−0.40	4.27	−0.89	0.05	0.38
46	1	2-xylene	2.52	2.30	−0.22	1.97	−0.55	3.12	3.06
47	1	3,3-dimethyl-2-butanone	2.94	3.52	0.58	3.18	0.24	1.20	1.55
48	1	3,4-dichlorotoluene	1.40	1.36	−0.04	1.14	−0.26	4.06	3.68
49	1	3-chlorotoluene	2.16	2.06	−0.10	1.64	−0.52	3.28	3.15
50	1	3-furanmethanol	3.72	3.95	0.23	2.96	−0.76	0.3	0.91
51	1	3-methyl-2-butanone	4.01	3.98	−0.03	3.21	−0.81	0.84	1.11
52	1	3-pentanol	4.05	3.83	−0.22	3.62	−0.43	1.21	1.27
53	1	3-pentanone	4.26	3.98	−0.28	3.04	−1.22	0.85	1.11
54	1	3-xylene	2.55	2.30	−0.25	1.97	−0.58	3.20	3.06
55	1	4-chlorotoluene	1.67	2.06	0.39	1.63	−0.04	3.33	3.15
56	1	4-methyl-2-pentanone	3.71	3.51	−0.20	2.92	−0.79	1.31	1.55
57	1	4-xylene	2.52	2.30	−0.22	1.97	−0.55	3.15	3.06
58	1	5-nonanone	2.34	2.11	−0.23	1.81	−0.53	2.97	2.89
59	1	6-methyl-5-hepten-2-one	2.84	2.73	−0.11	2.18	−0.66	1.82	2.28
60	1	acetone	5.10	4.92	−0.18	3.69	−1.41	−0.24	0.21
61	1	acetophenone	3.13	3.29	0.16	2.15	−0.98	1.58	1.67
62	1	benzene	2.91	3.23	0.32	3.09	0.18	2.13	2.17
63	1	benzophenone	1.93	1.05	−0.88	1.19	−0.74	3.18	3.84
64	1	butyldigol	3.85	4.08	0.23	3.29	−0.56	0.56	0.18
65	1	chlorobenzene	2.23	2.52	0.29	2.32	0.09	2.89	2.71
66	1	chloroform	2.93	3.19	0.26	2.41	−0.52	1.97	1.79
67	1	cyclohexanol	3.85	3.60	−0.25	3.18	−0.67	1.23	1.45
68	1	cyclohexanone	3.73	3.75	0.02	2.59	−1.14	0.81	1.29
69	1	dibutyl ether	2.40	2.42	0.02	2.24	−0.16	3.21	2.62
70	1	dichloromethane	3.54	3.75	0.21	2.59	−0.95	1.25	1.43
71	1	diethyl ether	4.46	4.28	−0.18	3.58	−0.88	0.87	0.84
72	1	diethyleneglycol	5.76	5.96	0.20	4.94	−0.82	−1.3	−1.62
73	1	diisopropyl ether	2.96	3.35	0.39	3.22	0.26	1.52	1.73
74	1	dipentyl ether	1.31	1.48	0.17	1.57	0.26	4.04	3.52
75	1	diphenyl ether	1.38	1.35	−0.03	1.47	0.09	4.21	3.57

Table 2 (Continued)

no.	class	structure name	Log[c] LC_{50} exp.	Log[a] LC_{50}	Δa	Log[b] LC_{50}	Δb	$logP_{exp}$	$logP_C$
76	1	ethanol	5.44	5.23	−0.21	4.29	−1.15	−0.31	−0.07
77	1	furan	2.96	3.58	0.62	3.03	0.07	1.34	1.70
78	1	hexachloroethane	0.81	1.04	0.23	1.53	0.72	4.14	3.33
79	1	iso-butanol	4.29	4.30	0.01	3.96	−0.33	0.76	0.82
80	1	methanol	5.94	5.70	−0.24	4.56	−1.38	−0.77	−0.52
81	1	pentachlorobenzene	−0.15	−0.26	−0.11	−0.38	−0.23	5.18	4.80
82	1	pentachloroethane	1.74	1.61	−0.13	1.42	−0.32	3.22	2.96
83	1	*t*-butanol	4.68	4.29	−0.39	4.21	−0.47	0.35	0.83
84	1	*t*-butylmethyl ether	3.91	3.81	−0.10	3.64	−0.27	0.94	1.29
85	1	tetrachloroethene	1.98	2.34	0.36	1.56	−0.42	3.40	2.41
86	1	tetrachloromethane	2.64	2.63	−0.01	2.65	0.01	2.83	2.16
87	1	tetrahydrofuran	4.48	4.52	0.04	3.26	−1.22	0.47	0.58
88	1	toluene	2.87	2.77	−0.10	2.28	−0.59	2.78	2.61
89	1	trichloroethene	2.58	2.89	0.31	1.32	−1.26	2.61	2.06
90	1	triethyleneglycol	5.65	5.85	0.20	4.73	−0.92	−1.24	−1.94
91	2	1,2-dimethylpropylamine	3.51	3.30	−0.21	3.54	0.03	1.10	1.37
92	2	1,3-dihydroxybenzene	2.96	2.53	−0.43	2.94	−0.02	0.80	1.15
93	2	1-adamantaneamine	2.22	2.06	−0.16	1.41	−0.81	1.44	2.80
94	2	1-amino-2-propanol	4.52	4.58	0.06	4.47	−0.05	0.96	−0.76
95	2	1-methylheptylamine	1.60	1.89	0.29	2.42	0.82	2.82	2.71
96	2	1-naphthanol	1.50	1.47	−0.03	1.61	0.11	2.84	2.83
97	2	2-(2-butoxyethoxy)ethanol	3.85	3.81	−0.04	3.29	−0.56	0.56	0.18
98	2	2,2-dimethylpropylamine	3.74	3.41	−0.33	3.67	−0.07	1.19	1.36
99	2	2,3,4,5-tetrachloroaniline	0.19	0.61	0.42	0.24	0.05	4.56	3.13
100	2	2,3,4,5-tetrachlorophenol	0.48	0.40	−0.08	0.29	−0.19	4.21	3.76
101	2	2,3,4-trichloroaniline	0.85	1.22	0.37	0.59	−0.26	3.68	2.61
102	2	2,3,5,6-tetrachloroaniline	0.07	0.63	0.56	0.20	0.13	4.46	3.13
103	2	2,3,5,6-tetrachlorophenol	0.74	0.41	−0.33	0.25	−0.49	3.88	3.77
104	2	2,3,5-trichlorophenol	1.08	0.89	−0.19	1.01	−0.07	3.58	3.25
105	2	2,3,6-trichloroaniline	1.27	1.18	−0.09	0.56	−0.71	3.32	2.61
106	2	2,3,6-trichlorophenol	1.44	0.94	−0.50	0.97	−0.47	3.77	3.25
107	2	2,3,6-trimethylphenol	1.79	1.72	−0.07	2.26	0.47	2.92	2.99
108	2	2,3-dichloronitrobenzene	1.34	1.66	0.32	1.15	−0.19	3.05	3.09
109	2	2,3-dimethylnitrobenzene	1.61	1.93	0.32	1.53	−0.08	2.83	2.92
110	2	2,4,5-trichloroaniline	1.08	1.10	0.02	0.58	−0.50	3.69	2.62
111	2	2,4,5-trichlorophenol	0.80	0.89	0.09	0.65	−0.15	3.80	3.25
112	2	2,4,6-tribromophenol	1.30	0.47	−0.83	−0.75	−2.05	3.92	4.07
113	2	2,4,6-trichlorophenol	1.06	0.96	−0.10	0.96	−0.10	4.03	3.26
114	2	2,4-dichloroaniline	1.59	1.70	0.11	1.02	−0.57	2.91	2.09
115	2	2,4-dichloronitrobenzene	1.54	1.57	0.03	1.13	−0.41	3.09	3.10
116	2	2,4-dichlorophenol	1.41	1.45	0.04	1.78	0.37	3.17	2.72
117	2	2,4-dimethylphenol	2.14	1.93	−0.21	2.49	0.35	2.30	2.54
118	2	2,5-dichloroaniline	1.01	1.67	0.66	1.36	0.35	2.92	2.09
119	2	2,5-dichloronitrobenzene	1.41	1.60	0.19	1.14	−0.27	2.90	3.10
120	2	2,5-dichlorophenol	1.42	1.39	−0.03	1.76	0.34	3.06	2.72
121	2	2,6-dichlorophenol	1.68	1.45	−0.23	1.74	0.06	2.84	2.72
122	2	2,6-diisopropylaniline	1.94	1.29	−0.65	1.23	−0.71	3.18	3.70
123	2	2,6-dimethylphenol	2.25	2.12	−0.13	2.54	0.29	2.36	2.54
124	2	2-allylphenol	2.04	2.12	0.08	2.16	0.12	2.55	2.83
125	2	2-aminoethanol	4.54	4.71	0.17	4.58	0.04	−1.31	−1.21
126	2	2-chloro-4-methylphenol	2.40	1.86	−0.54	2.09	−0.31	2.65	2.63
127	2	2-chloro-4-nitroaniline	2.07	1.60	−0.47	1.07	−1.00	2.06	2.14
128	2	2-chloro-6-nitrotoluene	1.48	1.74	0.26	0.96	−0.52	3.09	3.00
129	2	2-chloroaniline	1.69	2.24	0.55	2.23	0.54	1.9	1.56
130	2	2-chloronitrobenzene	2.28	2.00	−0.28	1.67	−0.61	2.24	2.56
131	2	2-chlorophenol	1.94	2.00	0.06	2.61	0.67	2.24	2.19
132	2	2-ethylaniline	2.79	2.66	−0.13	2.19	−0.60	1.93	1.92
133	2	2-methoxyethylamine	3.84	4.54	0.70	4.15	0.31	−0.67	−0.74
134	2	2-methylaniline	2.88	2.68	−0.20	2.58	−0.31	1.32	1.47
135	2	2-methylphenol	2.23	2.34	0.11	2.71	0.48	1.95	2.10
136	2	2-nitroaniline	1.85	2.15	0.30	2.09	0.24	1.85	1.60
137	2	2-nitrotoluene	2.41	2.28	−0.13	1.76	−0.65	2.30	2.47
138	2	2-phenylphenol	1.24	1.00	−0.24	1.71	0.47	3.09	3.56
139	2	2-*tert*-butyl-4-methylphenol	1.10	1.40	0.30	1.87	0.77	3.80	3.88
140	2	3,3-dimethylbutylamine	3.78	3.01	−0.77	3.34	−0.44	1.72	1.81
141	2	3,4,5-trichlorophenol	0.92	0.89	−0.03	0.66	−0.26	4.28	3.24
142	2	3,4-dichloroaniline	1.61	1.65	0.04	1.04	−0.57	2.69	2.09
143	2	3,4-dimethylnitrobenzene	1.79	1.83	0.04	1.33	−0.46	2.91	2.92
144	2	3,4-dimethylphenol	2.08	1.96	−0.12	2.52	0.44	2.23	2.55
145	2	3,5-dichloroaniline	1.38	1.62	0.24	1.02	−0.36	2.90	2.10
146	2	3,5-dichloronitrobenzene	1.42	1.48	0.06	1.12	−0.30	3.13	3.10
147	2	3,5-dichlorophenol	1.22	1.37	0.15	1.76	0.54	3.63	2.73
148	2	3-benzyloxyaniline	1.66	1.70	0.04	1.03	−0.63	2.77	2.88
149	2	3-chloroaniline	2.02	2.19	0.17	2.20	0.18	1.88	1.56
150	2	3-chloronitrobenzene	1.99	1.95	−0.04	1.65	−0.34	2.46	2.57

Table 2 (Continued)

no.	class	structure name	Log[c] LC_{50} exp.	Log[a] LC_{50}	Δa	Log[b] LC_{50}	Δb	$logP_{exp}$	$logP_C$
151	2	3-chlorophenol	1.70	1.92	0.22	2.59	0.89	2.59	2.19
152	2	3-ethylaniline	2.35	2.47	0.12	2.17	−0.18	1.88	1.92
153	2	3-methoxyphenol	2.78	2.76	−0.02	2.60	−0.18	1.58	1.61
154	2	3-methylaniline	2.53	2.65	0.12	2.58	0.05	1.44	1.47
155	2	3-methylphenol	2.52	2.41	−0.11	2.73	0.21	1.96	2.10
156	2	3-nitroaniline	2.76	2.20	−0.56	2.05	−0.71	1.37	1.61
157	2	3-nitrophenol	1.93	2.54	0.61	2.44	0.51	2.00	1.52
158	2	3-nitrotoluene	2.35	2.28	−0.07	1.75	−0.60	2.42	2.47
159	2	4-(n-methoxymethyl)aminophenol	3.73	3.71	−0.02	2.86	−0.87	0.48	0.21
160	2	4-amino-2-nitrophenol	2.36	2.36	0.00	2.72	0.36	0.96	1.09
161	2	4-bromoaniline	2.44	1.93	−0.51	1.41	−1.03	2.26	1.83
162	2	4-butylaniline	1.84	1.91	0.07	1.52	−0.32	3.05	2.81
163	2	4-chloro-2-nitrotoluene	1.56	1.74	0.18	1.16	−0.40	3.05	3.01
164	2	4-chloro-3,5-dimethylphenol	1.34	1.59	0.25	1.58	0.24	3.45	3.08
165	2	4-chloro-3-methylphenol	1.67	1.91	0.24	2.09	0.42	3.10	2.63
166	2	4-chloroaniline	2.33	2.13	−0.20	2.46	0.13	1.88	1.56
167	2	4-chloronitrobenzene	1.58	1.93	0.35	1.62	0.04	2.39	2.57
168	2	4-chlorophenol	1.82	1.92	0.10	2.58	0.76	2.39	2.19
169	2	4-decylaniline	−0.58	−0.35	0.23	−0.43	0.15	6.09	5.49
170	2	4-ethoxy-2-nitroaniline	2.15	2.06	−0.09	1.71	−0.44	2.38	2.00
171	2	4-ethylaniline	2.48	2.51	0.03	2.16	−0.32	1.92	1.92
172	2	4-ethylphenol	1.93	2.26	0.33	2.33	0.40	2.58	2.55
173	2	4-hexyloxyaniline	1.22	1.11	−0.11	1.13	−0.09	3.64	3.21
174	2	4-methoxyphenol	2.95	2.75	−0.20	2.81	−0.14	1.34	1.61
175	2	4-methylaniline	2.28	2.54	0.26	2.57	0.29	1.39	1.47
176	2	4-methylphenol	2.26	2.32	0.06	2.72	0.46	1.94	2.10
177	2	4-*n*-butylphenol	1.53	1.67	0.14	1.70	0.17	3.56	3.44
178	2	4-nitroaniline	2.77	2.07	−0.70	1.70	−1.07	1.39	1.60
179	2	4-nitrophenol	2.01	2.53	0.52	2.11	0.10	1.96	1.52
180	2	4-nitrotoluene	2.33	2.16	−0.17	1.74	−0.59	2.37	2.47
181	2	4-nonylphenol	−0.2	−0.27	−0.07	0.10	0.30	6.21	5.67
182	2	4-*n*-pentylphenol	0.88	1.22	0.34	1.40	0.52	4.09	3.89
183	2	4-octylaniline	−0.23	−0.48	−0.25	0.24	0.47	5.03	5.59
184	2	4-phenoxyphenol	1.42	1.53	0.11	1.99	0.57	3.35	3.06
185	2	4-phenylazophenol	0.76	0.97	0.21	1.08	0.32	3.96	3.32
186	2	4-propylphenol	1.91	1.84	−0.07	2.00	0.09	3.20	3.00
187	2	4-*tert*-butylphenol	1.54	1.78	0.24	2.12	0.58	3.31	3.44
188	2	4-*t*-pentylphenol	1.19	1.32	0.13	1.73	0.54	3.83	3.89
189	2	amylamine	3.31	2.84	−0.47	3.28	−0.03	1.49	1.36
190	2	aniline	3.09	2.74	−0.35	3.17	0.08	0.94	1.03
191	2	benzylamine	2.98	2.94	−0.04	2.63	−0.35	1.09	1.48
192	2	butylamine	3.56	3.27	−0.29	3.55	−0.01	0.97	0.92
193	2	decylamine	0.82	0.86	0.04	1.64	0.82	4.10	3.59
194	2	dodecylamine	−0.27	0.07	0.34	0.94	1.21	5.16	4.49
195	2	ethylamine	3.70	4.17	0.47	4.08	0.38	−0.13	0.02
196	2	heptylamine	2.28	2.02	−0.26	2.65	0.37	2.57	2.26
197	2	hexylamine	2.75	2.44	−0.31	2.96	0.21	2.06	1.81
198	2	*N*,*N*-dimethylaniline	2.67	2.86	0.19	2.19	−0.48	2.31	1.94
199	2	nitrobenzene	3.03	2.50	−0.53	2.35	−0.68	1.85	2.03
200	2	nonylamine	1.18	1.23	0.05	1.99	0.81	3.57	3.15
201	2	octylamine	1.60	1.64	0.04	2.32	0.72	3.04	2.70
202	2	pentafluoroaniline	2.31	1.95	−0.36	2.87	0.56	1.86	1.48
203	2	phenol	2.55	2.51	−0.04	3.34	0.79	1.46	1.66
204	2	propylamine	3.72	3.69	−0.03	3.81	0.09	0.48	0.47
205	2	quinoline	2.37	2.62	0.25	1.04	−1.33	2.03	1.78
206	2	*s*-butylamine	3.58	3.60	0.02	3.65	0.07	0.74	0.92
207	2	*t*-octylamine	2.28	2.19	−0.09	3.07	0.79	2.69	2.71
208	2	tridecylamine	−0.46	−0.33	0.13	0.59	1.05	5.10	4.93
209	2	undecylamine	0.09	0.45	0.36	1.30	1.21	4.63	4.04
210	2	α,α,α,4-tetrafluoro-2-methylaniline	2.22	2.23	0.01	2.53	0.31	2.51	1.88
211	2	α,α,α,4-tetrafluoro-3-methylaniline	2.23	2.23	0.00	2.49	0.26	2.51	1.89
212	3	1,2,7,8-diepoxyoctane	1.67	1.38	−0.29	1.97	0.30	1.18	0.99
213	3	1,2-epoxybutane	2.66	2.36	−0.30	3.07	0.41	0.76	0.63
214	3	1,2-epoxydecane	1.32	1.54	0.22	1.23	−0.09	3.94	3.31
215	3	1,2-epoxydodecane	0.78	1.20	0.42	0.53	−0.25	5.00	4.20
216	3	1,2-epoxyhexane	2.27	2.19	−0.08	2.53	0.26	1.82	1.52
217	3	1,2-epoxyoctane	1.91	1.89	−0.02	1.90	−0.01	2.88	2.41
218	3	1,3-butadienediepoxide	1.49	1.70	0.21	3.09	1.60	−0.48	−0.80
219	3	1,3-dichloropropene	0.66	0.50	−0.16	1.25	0.59	1.41	2.16
220	3	1,4-dichloro-2-butene	−0.16	0.65	0.81	1.42	1.58	1.94	2.61
221	3	1-chloro-2,4-dinitrobenzene	−0.19	−0.05	0.14	1.16	1.35	2.20	2.42
222	3	1-chloro-2-butene	1.82	1.24	−0.58	2.02	0.20	2.05	2.25
223	3	2,2′-dichlorodiethyl ether	2.54	2.09	−0.45	2.16	−0.38	1.81	1.56
224	3	2,3,4,6-tetrachlorophenol	0.67	0.82	0.15	0.25	−0.42	4.45	3.77
225	3	2,3-dichloropropene	1.01	0.25	−0.77	1.38	0.37	1.99	2.16

Table 2 (Continued)

no.	class	structure name	Log[c] LC_{50} exp.	Log[a] LC_{50}	Δa	Log[b] LC_{50}	Δb	$logP_{exp}$	$logP_C$
226	3	2,4,α-trichlorotoluene	0.08	0.01	−0.07	0.41	0.33	3.87	4.04
227	3	2,5-dinitrophenol	1.00	0.87	−0.13	2.07	1.07	1.75	1.37
228	3	2-butenal	0.90	1.94	1.04	2.55	1.65	0.20	0.50
229	3	2-ethylbutanal	1.89	1.72	−0.17	2.72	0.83	1.49	1.56
230	3	2-furaldehyde	2.04	2.04	0.00	2.63	0.59	0.81	1.47
231	3	2-methylpropanal	2.57	2.51	−0.06	3.26	0.69	0.43	0.66
232	3	2-*s*-butyl-4,6-dinitrophenol	0.17	0.08	−0.09	0.93	0.76	3.33	3.16
233	3	3,4,5,6-tetrachloro-2-hydroxyphenol	1.00	0.78	−0.22	1.00	0.00	4.29	3.24
234	3	3,4,5-trichloro-2,6-dimethoxyphenol	1.12	0.93	−0.19	1.34	0.22	3.74	3.13
235	3	3,4,5-trichloro-2-methoxyphenol	1.03	1.01	−0.02	0.89	−0.14	3.77	3.19
236	3	3-chloro-1-butene	1.85	1.25	−0.60	2.04	0.19	1.93	2.25
237	3	3-cyclohexene-1-carboxaldehyde	1.01	1.21	0.20	1.90	0.89	1.34	1.57
238	3	3-methylbutanal	2.19	2.18	−0.01	3.07	0.88	0.96	1.11
239	3	4,5-dichloro-2-methoxyphenol	1.40	1.52	0.12	1.24	−0.16	3.26	2.67
240	3	4-dinitrobenzylbromide	−0.30	−0.54	−0.24	0.36	0.66	2.41	3.08
241	3	allyl chloride	1.20	1.04	−0.16	2.02	0.82	1.53	1.80
242	3	benzaldehyde	1.57	1.40	−0.17	2.69	1.12	1.49	1.95
243	3	benzyl chloride	0.49	0.75	0.26	1.60	1.11	2.48	2.97
244	3	butanal	2.28	2.41	0.13	3.05	0.77	0.53	0.66
245	3	chloroacetone	0.88	1.58	0.70	2.63	1.75	0.45	0.57
246	3	cyclohexanecarboxaldehyde	1.91	1.32	−0.59	2.29	0.38	1.68	1.73
247	3	decanal	1.31	1.28	−0.03	1.47	0.16	3.71	3.34
248	3	epibromohydrin	0.77	1.06	0.29	1.86	1.09	0.38	0.79
249	3	epichlorohydrin	0.85	1.58	0.73	2.53	1.68	0.58	0.54
250	3	ethanal	2.90	2.73	−0.17	3.20	0.30	−0.53	−0.23
251	3	glycidol	2.83	2.74	−0.09	3.85	1.02	−0.92	−1.05
252	3	heptanal	1.89	1.83	−0.06	2.38	0.49	2.12	2.00
253	3	hexachlorobutadiene	−0.20	0.17	0.37	−0.04	0.16	4.63	3.84
254	3	hexanal	1.99	2.03	0.04	2.65	0.66	1.59	1.56
255	3	methanal	2.96	2.82	−0.14	2.97	0.01	−0.75	−0.67
256	3	octanal	1.79	1.66	−0.13	2.09	0.30	2.65	2.45
257	3	pentachlorophenol	0.22	0.20	−0.02	0.04	−0.18	5.15	4.28
258	3	pentanal	2.18	2.22	0.04	2.88	0.70	1.06	1.11
259	3	propanal	2.41	2.61	0.20	3.22	0.81	0.00	0.22
260	3	propylene oxide	2.74	2.53	−0.21	3.38	0.64	0.23	0.18
261	3	styrene oxide	1.77	1.72	−0.05	1.82	0.05	1.43	1.63
262	3	α,α-dichloro-*m*-xlyene	−0.16	0.05	0.21	0.84	1.00	2.87	3.78
263	4	1,3-dinitrobenzene	1.75	1.57	−0.18	2.04	0.29	1.53	1.89
264	4	4-hexylresorcinol	0.72	1.33	0.61	1.55	0.83	4.10	3.82
265	4	8-hydroxyquinoline	−0.99	0.37	1.36	1.49	2.48	2.12	1.27
266	4	acrylamide	2.69	1.92	−0.77	3.17	0.48	−1.04	−1.10
267	4	allyl alcohol	1.16	1.70	0.54	3.32	2.16	−0.03	0.21
268	4	azinphos-methyl[c]	−0.74	−0.44	0.30	−0.94	−0.20	2.76	1.80
269	4	bromophos[c]	0.09	−0.29	−0.38	−0.45	−0.54	5.21	4.49
270	4	chlorothion[c]	−0.19	0.78	0.97	0.68	0.87	3.63	3.01
271	4	cyanophos[c]	1.75	1.08	−0.67	0.96	−0.79	2.71	2.34
272	4	decamethrin[c]	−2.34	−2.16	0.18	−3.28	−0.94	4.40	6.22
273	4	dicapthon[c]	0.43	0.46	0.03	0.44	0.01	3.72	3.02
274	4	dieldrin[c]	−1.78	−2.16	−0.38	−2.52	−0.74	5.30	5.12
275	4	disulfiram[c]	−1.65	−1.85	−0.20	−1.04	0.61	3.88	4.11
276	4	ethyl acrylate	0.87	0.91	0.04	2.36	1.49	0.88	1.17
277	4	etrimfos[c]	1.09	1.10	0.01	1.34	0.25	3.67	2.23
278	4	fenitrothion[c]	1.00	0.70	−0.30	0.75	−0.25	3.47	3.47
279	4	fenthion[c]	0.89	0.38	−0.51	0.02	−0.87	4.17	4.01
280	4	fluoroacetamide	2.88	2.17	−0.71	3.70	0.82	−1.04	−1.28
281	4	iodofenphos[c]	0.32	−0.51	−0.83	−0.65	−0.97	5.51	4.73
282	4	lethane[c]	0.76	0.78	0.02	1.18	0.42	1.68	0.90
283	4	lindane	−0.69	−0.50	0.19	−0.39	0.30	4.14	4.85
284	4	methidathion[c]	−0.96	−0.29	0.67	−0.15	0.81	2.50	1.24
285	4	methylisocyanothion	0.23	0.00	−0.23	1.95	1.72	3.58	1.39
286	4	methylparathion[c]	0.61	1.28	0.67	1.35	0.74	3.04	2.48
287	4	phenthoate[c]	−0.99	−1.07	−0.08	−0.86	0.13	3.96	4.29
288	4	phosmet[c]	−0.12	−0.24	−0.12	−0.79	−0.67	2.81	1.84
289	4	proclonol[c]	0.01	−0.64	−0.65	−0.77	−0.78	5.50	5.48
290	4	ronnel[c]	0.00	−0.09	−0.09	0.13	0.13	5.07	4.22
291	4	rotenone[c]	−0.84	−0.68	0.16	−1.16	−0.32	3.95	3.07
292	4	thiomedon[c]	1.53	1.52	−0.01	1.60	0.07	−1.87	−0.77
293	4	α-endosulfan[c]	−2.74	−2.39	0.35	−2.53	0.21	3.83	2.47

[a] Predicted with equations from Table 3A−D. [b] Predicted with general equation (Table 3E); all concentrations in log μmol/L. [c] Structures of compounds are provided in the Supporting Information.

used to calculate electronic, geometric, and energetic parameters for the isolated molecules. After optimization, the CODESSA program was used to calculate five types of molecular descriptors: constitutional, topological, geometri-

cal, electrostatic, and quantum-chemical.[47,56] Up to 941 descriptors (the precise number depended on the atomic constitution of the molecule) were calculated for each structure in the seven sets studied. The best multilinear regression (BMLR) procedure[47,57,58] was used to select the best two-parameter regression model; the best multiparameter regression models were selected based on the highest R^2 value in a forward stepwise regression procedure.[59] The correlation equations were constituted from the selected noncollinear descriptors according to the maximum value of the *Fisher* criteria and the highest cross-validated correlation coefficient (R^2_{cv}).[46,47,58]

Considering log *P* has been an anchor descriptor in QSAR toxicity analysis (particularly of narcotic toxins), we incorporated log P_C into our pool of descriptors. Calculated values of log *P* were determined using the molecular size based approach of Bodor et al.[60] implemented in the QLogP software package. These calculated values of log *P* are listed along with the experimental values in Table 2. The QLogP program combines the 3D molecular size with a parameter (*N*) that indicates the number and types of polar functional groups.[61] The *N* value roughly correlates with the hydrogen bonding capability of the molecule. The QLogP program provides reliable calculated values for log *P*, thus eliminating any dependence on experimental data. Indeed we found a good correlation ($R^2 = 0.91$) between the experimental and calculated log *P* values for the 293 structures used in this study (see Supporting Information for plot).

The latest version of QlogP (2.01 beta) was used for the calculations. The program was modified by Peter Buchwald to allow Mopac output files to be directly calculated rather than first converting the structures to an alternative format. The QlogP program was originally optimized for structures produced by Alchemy, which uses slightly different bond length constants than Mopac. As a result, the phosphorus/sulfur bonds of the Mopac optimized structures were not properly recognized. Structures with P−S bonds were converted to HyperChem input files, and the geometry of (only) the P−S bonds was minimized using MM+. The resulting structures were properly identified by QlogP and provided high quality log *P* values.

RESULTS AND DISCUSSION

For the 90 class 1 compounds the best correlation equation derived had two parameters ($R^2 = 0.9551$, Figure 1 and Table 3A). The model is dominated by log P_C, which in a single-parameter correlation has a coefficient of $R^2 = 0.9307$. The additional descriptor used for the two-parameter correlation, *difference in third-order charged partial surface areas (DPSA-3)*, is related to the positive and negative charge distribution and the respective surface areas and can correct for slight polarities in class 1 compounds. The results for class 1 compounds clearly follow previous findings with log *P* as the dominant contributor. The prominence of log *P* in the correlation for class 1 compounds is not surprising considering that these toxins are defined as substances which follow the baseline toxicity model.

Table 4 gives the correlation of each toxin class with log P_C. The class 2 compounds only follow the baseline moderately so that toxicity trends can be predicted. Class 3 and class 4 toxins have poor correlations to log *P*, which illustrates the need for improved correlations for predicting toxicity. As stated earlier, there is often an overlap between classes, and since some class 1 compounds possess some degree of polarity, the remaining descriptor serves as a correction.

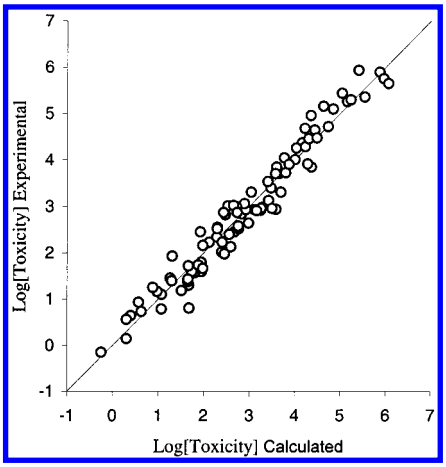

Figure 1. Class 1 toxins: two-parameter equation for 90 data points.

The best five-parameter correlation equation for the 121 class 2 compounds had $R^2 = 0.9184$ (Figure 2 and Table 3B). According to the *t*-test values, the most important descriptor is again log P_C, which alone describes the 79% of the variance. The strong appearance of log *P* in the correlation is supported by the *relative negative charge*. The descriptor reflects the way in which the most highly charged negative atom is related to overall negative charge of the molecule (charge of most negative atom/sum of total negative charge). The descriptor indicates how evenly the negative charge is distributed in the molecule and can be related to the polar behavior of class 2 compounds. The third descriptor stresses the importance of hydrogen bonding capability in the prediction of toxicity for class 2 compounds. This fact deviates somewhat from the current opinion that the toxicity of class 2 compounds should depend only on polar interactions and hydrophobicity. The appearance of a hydrogen bonding parameter reveals that the classes can overlap and that class 2 compounds can have other important interactions besides polar. The *fractional hydrogen donor charged surface area* is the surface area multiplied by the corresponding partial charge of the compound that is capable of donating a hydrogen in an interaction with the environment. The descriptor takes into account all of the hydrogen atoms in the compound that have a net atomic charge higher than 0.185. The resulting sum is then divided by the total of the partial surface areas of those hydrogen atoms. The remaining descriptors, *XY shadow* and *YZ shadow*, are calculated from their respective projection of the shadows on the 2D plane according to the orientation in the space. They are the natural shadow indexes and reflect the relationship of the size of the class 2 molecules with the acute toxicity and are therefore reflecting the importance of nonspecific interaction on class 2 compounds.

Again, there is an overlap of classes and several of the class 2 structures could arguably be switched into class 3. The results of class 2 are largely as expected but also show some new trends. Alongside log P_C, the additional descriptors are related to polar attractions, hydrogen bonding, and the

Table 3. Correlation Equations

	X	DX	R^2	S^2	*t*-test	descriptors
			A: Class 1 Toxins, Two-Paramater Equation (Figure 1)			
0	5.4262e + 00	1.0152e − 01			53.4488	intercept
1	−9.5096e − 01	2.2531e − 02	0.9307	0.1469	−42.2067	$\log P_C$
2	−3.9252e − 02	5.7175e − 03	0.9551	0.0964	−6.8652	DPSA-3 difference in CPSAs (PPSA3-PNSA3) [Zefirov's PC]
			0.9520			R^2_{cv}
			0.9552			validation
			B: Class 2 Toxins, Five-Parameter Equation (Figure 2)			
0	5.5586e + 00	3.3452e − 01			16.6169	intercept
1	−7.7786e − 01	4.2801e − 02	0.7918	0.2109	−18.1736	$\log P_C$
2	−3.1694e + 00	4.3840e − 01	0.8237	0.1350	−7.2294	RNCG relative negative charge (QMNEG/QTMINUS) [semi-MO PC]
3	−1.6815e + 01	2.7092e + 00	0.8690	0.1801	−6.2067	FHDCA fractional HDCA (HDCA/TMSA) [semi-MO PC]
4	−3.0729e − 02	5.2077e − 03	0.9022	0.1017	−5.9005	XY shadow
5	3.3784e − 02	7.0597e − 03	0.9184	0.0855	4.7855	YZ shadow
			0.9083			R^2_{cv}
			0.9083			validation
			C: Class 3 Toxins, 5-Paramater Equation (Figure 3)			
0	5.1611e − 01	4.8257e − 01			1.0695	intercept
1	3.4085e + 00	3.5674e − 01	0.3870	0.5128	9.5545	FPSA-1 fractional PPSA (PPSA-1/TMSA) [Zefirov's PC]
2	−8.3705e − 02	1.0429e − 02	0.5568	0.2248	−8.0265	number of single bonds
3	−1.6196e − 01	2.6434e − 02	0.7423	0.3785	−6.1271	final heat of formation/# of atoms
4	−9.7312e − 01	2.0651e − 01	0.8048	0.1740	−4.7123	average information content (order 0)
5	−6.8854e + 00	1.8425e + 00	0.8510	0.1357	−3.7370	min partial charge (Qmin) [Zefirov's PC]
			0.8201			R^2_{cv}
			0.8286			validation
			D: Class 4 Toxins, Four-Paramater Equation (Figure 4)			
0	1.1803e + 00	3.1761e − 01			3.7163	intercept
1	−2.9870e − 02	2.7621e − 03	0.4316	1.0526	−10.8142	ALFA polarizability (DIP)
2	−9.9238e + 00	1.5529e + 00	0.5651	0.8342	−6.3904	FNSA-3 fractional PNSA (PNSA-3/TMSA) [semi-MO PC]
3	1.2302e − 01	1.9287e − 02	0.7351	0.5269	6.3787	count of H-donors sites [Zefirov's PC]
4	9.2159e − 01	2.0913e − 01	0.8484	0.3132	4.4067	number of benzene rings
			0.7745			R^2_{cv}
			0.6740			validation
			E: Combined Classes 1, 2, 3, and 4 Toxins, Seven-Paramater Equation (Figure 5)			
0	1.0588e + 01	6.6201e − 01			15.9942	intercept
1	−9.1219e − 01	1.1202e − 01	0.4840	1.0649	−19.0652	Kier & Hall index (order 1)
2	1.0981e + 00	4.7801e − 01	0.6509	0.7230	9.8026	topographic electronic index (all pairs) [Zefirov's PC]
3	−4.2188e + 00	4.7764e − 01	0.6871	0.6504	−8.8326	average structural information content (order 1)
4	−4.5212e + 00	6.4418e − 01	0.7219	0.5799	−7.0185	av bond order of a C atom
5	−1.6695e − 02	2.9377e − 03	0.7577	0.5071	−5.6831	complementary information content (order 1)
6	−1.6170e + 00	2.8528e − 01	0.7871	0.4472	−5.6681	min net atomic charge
7	3.2016e + 00	9.4810e − 01	0.7953	0.4315	3.3769	HACA-2/SQRT(TMSA) [semi-MO PC]
			0.7834			R^2_{cv}
			0.7834			validation
			F: Combined Classes 1 and 2 Toxins, Six-Paramater Equation (Figure 6)			
0	−1.6859e + 01	2.2906e + 00			−7.3602	intercept
1	−9.0189e − 01	2.0836e − 02	0.8279	0.2935	−43.2848	$\log P_C$
2	2.4579e + 01	2.3402e + 00	0.8538	0.2506	10.5028	max sigma-sigma bond order
3	−4.4841e − 01	4.6057e − 02	0.8803	0.2060	−9.7360	average information content (order 1)
4	−1.2806e − 01	1.3414e − 02	0.8970	0.1783	−9.5466	HA dependent HDSA-2 [semi-MO PC]
5	−2.9038e + 00	3.3371e − 01	0.9218	0.1359	−8.7015	molecular volume/XYZ box
6	2.6633e + 01	3.5049e + 00	0.9391	0.1064	7.5988	HACA-1/TMSA [Zefirov's PC]
			0.9340			R^2_{cv}
			0.9362			validation
			G: Combined Classes 1, 2, and 3 Toxins, Eight-Parameter Equation (Figure7)			
0	6.7685e + 00	2.3961e − 01			28.2478	intercept
1	−8.4498e − 01	3.1644e − 02	0.6281	0.6223	−26.7028	$\mathrm{Log}P_C$
2	−2.3431e + 01	1.0118e + 00	0.6720	0.4777	−12.9011	max sigma-pi bond order
3	−1.2475e + 00	1.1780e − 01	0.7167	0.4341	−10.5904	max bond order of a C atom
4	−7.2412e − 01	7.0311e − 02	0.7435	0.4276	−10.2989	number of N atoms
5	−1.5384e − 01	2.2262e − 02	0.7483	0.3034	−6.9103	Kier flexibility index
6	−8.3394e − 02	1.5722e − 02	0.7519	0.4232	−5.3044	RPCS relative positive charged SA (SAMPOS*RPCG) [Zefirov's PC]
7	7.4607e − 01	1.4966e − 01	0.8228	0.2847	4.9850	FPSA-2 fractional PPSA (PPSA-2/TMSA) [semi-MO PC]
8	3.2480e − 02	7.7193e − 03	0.8344	0.5442	4.2077	PNSA-3 atomic charge weighted PNSA [Zefirov's PC]
			0.8183			R^2_{cv}
			0.8255			validation

size of the molecule. The last three terms are important because class 2 chemicals are attracted to the phospholipid head of the cell membrane which results in a higher accumulation of the chemical in the cell membrane, thus producing a higher narcotic effect than would be predicted by log *P* alone.

For the 51 class 3 compounds, the five-parameter correlation equation with $R^2 = 0.8510$ is considered to be the

Table 4. Correlation of Each Toxin Class with log P_C

	R^2	coefficient	F
class 1	0.9307	−0.956	1182.7
class 2	0.7918	−0.778	452.6
class 3	0.4727	−0.465	43.9
class 4	0.3494	−0.411	15.57

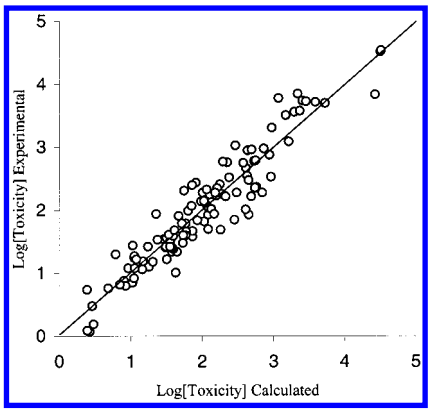

Figure 2. Class 2 toxins: five-parameter equation for 121 data points.

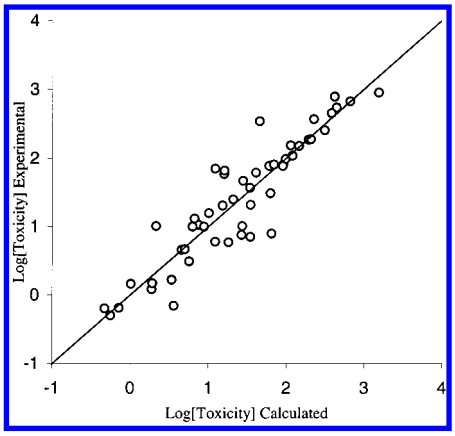

Figure 3. Class 3 toxins: five-parameter equation for 51 data points.

most satisfactory (Figure 3 and Table 3C). Removal of the outlier, 2-butenal, gives an improved R^2 of 0.8713. Significantly, log P_C was not among the five descriptors, which is a further indication of the difference between class 3 toxins and the narcotic toxins. There is a high probability that the effect of log P_C is averaged to zero because of the variety of modes of action possible within a data set. This does not necessarily mean that hydrophobic interaction is no longer important. While log P is certainly still important, it is no longer the best descriptor, and other descriptors should be applied for the description of hydrophobicity in such sets of compounds.

The *fractional partial positive surface area* (*FPSA* − partial positive surface area weighted by total molecular surface area) reflects positive charge distribution in the molecule through the respective surface area. It shows how charge is distributed in the molecule relative to the total surface area of the molecules. This descriptor can be loosely related to hydrogen bonding capability and reactivity as well. The next descriptor, *number of single bonds*, has a negative coefficient. This is reasonable since less aliphatic molecules are typically more reactive. The *final heat of formation/# of atoms* is a quantum chemical descriptor that quantifies

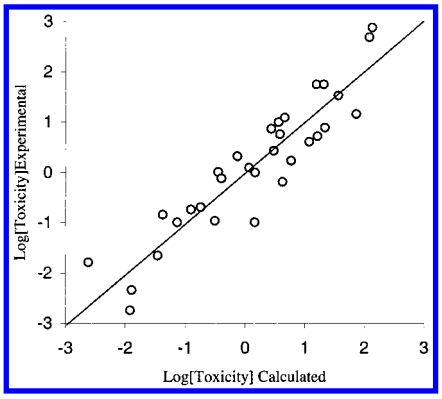

Figure 4. Class 4 toxins: four-parameter equation for 31 data points.

reactive bonds in the molecule. The *average information content (zeroth order)* accounts for the diversity of the atomic constitution of the compounds. Because the descriptor is zeroth order, it does not describe branching. Instead, the descriptor simply groups the atoms into distinct classes that reflect the complexity of the molecule. The *minimum partial charge (Qmin)* divides the data set into the three main reactive groups: aldehydes, Michael acceptors, and epoxides.

Due to the variety of different modes of action, the class 3 compounds are less easy to interpret than classes 1 or 2. Our data set of 51 compounds contains many examples with the same functional group but varying in chain length. This may distort the results and attribute a more than realistic significance to the size and the content of specific atoms. This is evidenced by the minimum partial charge descriptor. If we treat the class 3 compounds in a combined group with class 1 and 2 (see the following text), the larger number of molecules blurs the presence of the specific functional groups. None of the reactivity indices were shown to be significant in the correlation, but this is not surprising considering that the indices calculated by CODESSA indicate the reactivity at specific atomic sites rather than overall reactivity, and the particular reactive atomic site varies throughout the data set. We find that reactivity indices specific to a particular atomic species, i.e., the reactivity indices in the current version of CODESSA, are inadequate for predicting overall reactivity. If new descriptors, similar to the parameter (*N*) used in the calculation of QlogP, were developed, they could possibly indicate the presence of highly reactive functional groups and thus predict the expected high toxicity. Unfortunately, such descriptors would have the same shortcomings as classification rules, and only recognized functional groups would be predicted accurately. An alternate approach would utilize descriptors based on the magnitudes of rate constants for particular electrophiles and nucleophiles; such an approach was used successfully by Hermens.[18]

The four-parameter correlation equation for the 31 class 4 compounds (Figure 4 and Table 3D) consists of one constitutional, one quantum chemical, and two electrostatic descriptors. The *alfa polarizability* represents the effect (polarization) on a molecule produced by an external electric field in a medium and therefore reflects the interaction between the medium and the embedded molecule.[62] It has been shown that polarizability values are related to the hydrophobicity and that the higher order polarizability terms characterize the electrophilic properties of the molecule.[56]

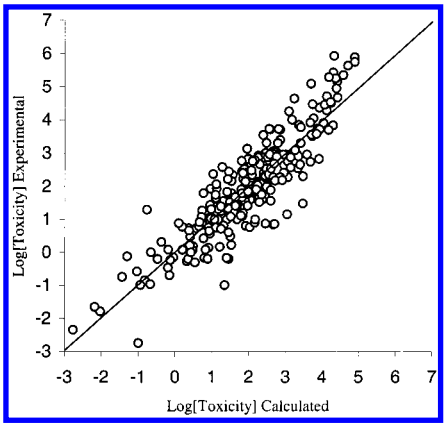

Figure 5. Combination of all toxin classes: seven-parameter equation for 293 data points.

The *third-order fractional partial negative charged surface area (FNSA-3)* shows the negative charge distribution on the respective surface area and is normalized by the total molecular surface area. The charge distribution is calculated from AM1 parametrization and can be related to the hydrogen acceptor ability of the molecule. The *count of H-donors sites [Zefirov's PC]* distinguishes the molecules according to the number of hydrogen donor sites that are capable of donating a hydrogen to the surrounding media. The *number of benzene rings* distinguishes the molecules with 1 and 2 benzene rings from other molecules and shows their different behavior. The occurrence of this descriptor is consistent with our previous work[50] where we showed that the number of rings is a major determinant of mutagenicity (genotoxicity) for aromatic and heteroaromatic amines. Also Hatch and Colvin[63] recently found that the mutagenicity of aromatic and heterocyclic amines depends mainly on the size of the aromatic ring system. Both studies conclude that the size of the ring system can affect the mutagenicity at various steps of the mechanism, but it most likely affects the penetration through the biomembranes.

The interpretation of the class 4 toxins is also difficult. The descriptors found are either indicators of reactivity or hydrogen bonding and bulk effects. However, the relevancy of some of the descriptors is probably affected by the small size of the data set. Because of this, the equation should only be used to predict the toxicity of structures similar to those in the training set. Benzene rings frequently occur in biologically active materials because they can align with the active sites of many enzymes, thus their presence can easily be expected to contribute to site-specific toxicity. Hydrogen bonding is also frequently a factor in enzyme substrate binding. The unsatisfactory quality of our model for predicting toxicities of class 4 compounds is primarily due to the variety of different modes of action that occur with the compounds in the data set. Despite this, the descriptors in the model represent the reactive and site-specific (interactions related to hydrogen bonding and bulk) properties of the compounds.

The best overall correlation equation for the full set of 293 compounds (Figure 5 and Table 3E) involved seven descriptors ($R^2 = 0.7953$). The *Keir and Hall Index* (first order) describes the valence connectivity of the molecule of the first order (coordination sphere). The *average structural information content (first order)* and *complimentary information content (first order)* reflect the branching of the molecule; also at the first coordination sphere. The last two descriptors are defined on the basis of the Shannon information theory. The calculated value for each molecule reflects how information rich the molecule is. "Information rich" describes how many different atoms there are in the molecule and how diverse the branching of these atoms is at the first valence level (coordination sphere). In essence, both descriptors give us information concerning how many atoms with similar connectivity patterns are in the molecule. The *average structural information content* also depends on the number of atoms in the molecule, and it arranges the molecules in the order of compound size.[64] The descriptor may be thought of as a normalized information content, with the maximum information content as the normalization factor. The *complimentary information content* takes into account the deviation of information content from the maximum information content. In other words, it represents the difference between the maximum possible complexity of a graph and the realized topological information of the chemical species as defined by the information content.[65]

A quantum chemical descriptor, *average bond order of a C atom*, reflects the hybridization and functional groups connected to the carbon skeleton and thus are loosely related to reactivity. Charge distribution is directly accounted for by the *topographic electronic index (all pairs)* using the charge distribution scheme described by Zefirov and coworkers,[66] and *minimum net atomic charge* is obtained from the Mulliken charge distribution scheme of quantum chemical calculations. The remaining descriptor *HACA-2/SQRT(T-MSA)* shows the H-bond acceptor capabilities as well as the negative charge distribution. The descriptor considers all amino, cyano, hydroxyl, carbonyl, carboxyl, thiol group, and aromatic nitrogen atoms. It sums the products of every partial surface area and its respective partial charge. The sum is normalized using the square root of total molecular surface area. The descriptor excludes nitrogen atoms connected to oxygen, double-bonded sulfur atoms, or COO groups.

When the toxicities of individual classes are calculated using this global correlation equation, we get R^2 values of 0.8705, 0.7894, 0.6326, and 0.7042 for classes 1, 2, 3, and 4, respectively. Considering the complexity of the data set, it is hard to expect any general conclusion regarding the mechanism of mutagenicity. However the model does indicate that hydrogen bonding and charge distribution are important factors for toxicity. The model also shows that topological descriptors are useful in the description of large and diverse data sets. The model can still be used to estimate trends in toxicity even if they do not explicitly show chemical structural trends that can be related to the toxicity.

Correlations were also performed to predict the toxicity of combined sets: classes 1 and 2 and classes 1, 2, and 3. A six-parameter correlation (Table 3F, Figure 6) was chosen for its ability to predict either a combined group of classes 1 and 2 toxins ($R^2 = 0.9391$) or only class 1 toxins ($R^2 = 0.9578$) or only class 2 toxins ($R^2 = 0.8998$). The log P_C was the most important descriptor. The quantum chemical descriptor, *max sigma-sigma bond order* describes the valency of the molecule and particularly the presence of the functional groups in the molecule. The descriptor has its lowest value for tetrachloromethane and its highest value for 2,6-diisopropylaniline; arranging the molecules according to the sigma-sigma bonding contribution in the molecule.

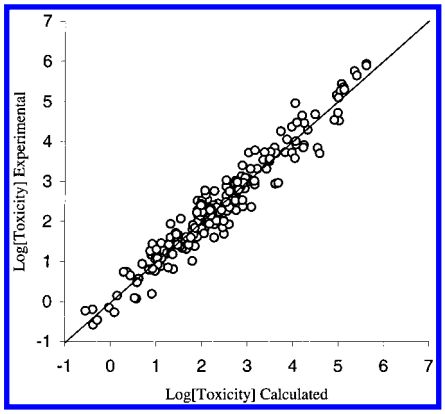

Figure 6. Class 1 and 2 toxins combined: six-parameter equation for 211 data points.

The *average information content (order 1)* is similar to the constitutional descriptors in the overall model. It considers the branching of the molecule through the similarity of connectivity patterns on the molecular graph. The two electrostatic descriptors, *hydrogen bond acceptor dependent hydrogen bond donor surface area (HDCA-2)* and *hydrogen bond acceptor charged surface area/total molecular surface area (HACA-1/TMSA)*, describe the hydrogen bonding acceptor and donor properties of the compounds, respectively. Finally, the equation was supported by the geometrical descriptor *molecular volume/XYZ box*, which describes the bulk related properties and, by normalizing the descriptor with a unit box, shows how compact the molecule is.

The correlation of the combined set of 261 structures composed of classes 1, 2, and 3 toxins resulted in an eight-parameter equation (Table 3G) with $R^2 = 0.8344$. The log P_C values were a significant descriptor for the combined set. This is not unexpected considering the majority of the data set consisted of classes 1 and 2 toxins. The quantum chemical descriptors, *max sigma-pi bond order* and *maximum bond order of a carbon*, both indicate reactivity. *Max sigma-pi bond order* covers the different reactive properties of epoxides in comparison with other compounds dividing data set into two different groups. *Maximum bond order of a carbon* accounts for the different behavior of purely aliphatic compounds from aromatic systems and compounds that have at least one full double bond. The constitutional descriptor, *number of N atoms*, results from the large number of amines in the data set. The descriptor divides the data set into three groups, with zero, one, and two nitrogen(s) in compound. The final topological descriptor, *Kier flexibility index*, describes the shape of the molecule. The three electrostatic descriptors, *RPCS relative positive charged SA (SAMPOS*RPCG) [Zefirov's PC]*, *FPSA-2 molecule weighted fractional positive charged surface area*, and *PNSA-3 atomic weighted partial negative surface area* describe the charge distribution in the molecule. They can be loosely related to the hydrogen bonding acceptor and donor ability and also to the reactivity.

The individual and combined narcotic sets produced results that are consistent with the work of Ramos et al.[38] The class 1 toxins can be predicted by log *P*. The occurrence of the *DSPA-3* descriptor was interesting and may correct for minor differences between experimental and calculated log *P* values as well as demonstrate overlap between the classes 1 and 2. The class 2 toxins can be predicted by log *P*, polarity, and hydrogen bonding descriptors. The descriptors used in the equation for predicting the combined classes 1 and 2 sets are similar to those used in the equation for class 2 toxins. The dominant descriptor is log *P*, and the additional descriptors are related to hydrogen bonding, bulk, and the compounds' polarity.

All correlations described in this study were cross-validated using internal validation sets. In each case, the full set of structures was divided into three groups: structures 1, 4, 7, etc. formed group I, structures 2, 5, 7, etc. formed group II, and structures 3, 6, 8, etc. formed group III. Groups I and II were then combined to form set A, groups II and III to form set B, and groups I and III to form set C. The descriptors used in the original correlation equation were saved to a descriptor set labeled "verify".

Correlations were performed on set A, set B, and set C each using the same verify descriptor set. The partition coefficients of the descriptors for sets A, B, and C were recorded. The toxicities for group III were then predicted using the descriptor partition coefficients from set A (generated from groups I and II—see above), group I was predicted using the partition coefficients from set B, and group II toxicities were predicted using the partition coefficients from set C. The toxicities calculated from groups A, B, and C were combined, and a correlation was then performed between them and the experimental toxicity values. The resulting R^2 values were then compared to the R^2 values from the original correlations. The cross-validation R^2 values were close to the original R^2 values (Table 3) for correlations for classes 1 and 2 and the combined sets. The validations were acceptable for classes 3 and 4 toxins. The poorer validation results for classes 3 and 4 can be attributed to the small data sets.

CONCLUSIONS

The present work represents an initial attempt to correlate toxicities using only calculated descriptors. This would make it possible to avoid the problems mentioned earlier concerning classification rules. It has been shown, in this study and others, that accurately describing the toxicity of class 1 and 2 compounds with a single correlation is quite feasible. The reason is that the mode of action is the same, a reversible accumulation of the toxin within the cell membrane that results in distortion and disruption of function. The correlations and conclusions, resulting from this study, are in strong agreement with work performed by Ramos et al.[38] A significant advantage to the correlations produced in this study is that they are completely independent from experimental data, which allows toxicities to be calculated for any structure that can be drawn.

The correlations for class 3 toxins were moderately successful, but the predictions are limited to the types of functional groups found in the data set. The ability to predict class 3 toxins using a single correlation is difficult since there are many possible modes of action within an organism. Additionally, the differences in toxicity from one species to another can be large since the toxicity mode may or may not vary between the organisms. Reasonable correlations can be found for a given species if the class 3 compounds are subcategorized by reaction type. Descriptors for electrophilicity may prove to be an indicator of toxicity class as well

as the degree of toxicity. Unfortunately, the relationship between toxicity and electrophilicity is limited by the environment; once electrophilicity increases beyond a point the substance will likely react with other nucleophiles before reaching any organisms.

The correlation that was developed for a combined set of classes 1, 2, and 3 is not accurate enough to be able to assign toxicity values, especially for class 3 toxins, but it can be used for predicting trends in toxicity. The correlation used to predict the combined set of four classes was not very satisfactory for predicting the individual classes, especially for class 3 toxins. It is unlikely that any single equation can be developed to accurately predict toxicities with such a great variation in the modes of action.

The correlations for class 4 toxins will only be valid for structures that are very similar to the insecticides used in this study. Class 4 compounds with the same basic structure, i.e., benzene with a variety of inert substituents, can be predicted with some accuracy. Unfortunately, different substituents can cause molecules with the same basic structure to affect an organism by different modes of action, which can lead to large errors in predicted toxicity. Additionally, the substituent can sometimes become the most reactive portion of the molecule, which will also lead to invalid predictions.

The agreement between this work and work by others (see introduction) is significant. The main approach up to now has been to develop descriptors, based on the theoretical mode of action of the toxin, to predict toxicities. Our approach was to produce the best correlation possible from an extremely large pool of descriptors and to determine later whether the correlation can be justified by the theoretical mode of action. The level of agreement between this study and those performed by others is a strong indicator that the narcosis mode of action can now be modeled satisfactorily.

ACKNOWLEDGMENT

We thank Dr. J. Hermens for providing the majority of the toxicity data used in this study. We also thank Dr. N. Bodor and Dr. P. Buchwald for providing the beta copy of QLogP used for log *P* calculations. One of the authors is thankful to the Estonian Science Foundation (Grant no. 4571) for the support.

Supporting Information Available: Figures of class 1, 2, and 3 toxins combined (Figure 7) and correlation between log P and Log P_c (Figure 8) and structures of compounds with trivial names. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Overton, E. Uber die Osmotischen Eigenschaften der Zelle in Ihrer Bedeutung für die Toxikologie und Pharmakologie (Mit Besonderer Berücksichtigung der Ammoniake und Alkaloide). *Vierteljahrsschrift d. Naturf. Ges Zürich* **1896**, 385−406.

(2) Meyer, H. Zur Theorie der Alkoholnarkose. *Arch. Exp. Pathol. Pharmakol.* **1899**, *42*, 109−118.

(3) Hansch, C.; Fujita, T. $\rho-\sigma-\pi$ Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616−1626.

(4) Hermens, J. L. M. Quantitative Structure−Activity Relationships of Environmental Pollutants. In *The Handbook of Environmental Chemistry*; Hutzinger, O., Ed.; Springer-Verlag: Berlin, 1989; Vol. 2, Part E, pp 111−162.

(5) Cronin, M. T. D.; Dearden, J. C. Review QSAR in Toxicology. 1. Prediction of Aquatic Toxicity. *Quant. Struct.-Act. Relat.* **1995**, *14*, 1−7.

(6) Hansch, C.; Leo, A. *Exploring QSAR, Fundamentals and Applications in Chemistry and Biology*; American Chemical Society: Washington, DC, 1995; Chapter 6, p 349.

(7) Schultz, T. W.; Sinks, G. D.; Bearden, A. P. QSAR in Aquatic Toxicology: A Mechanism of Action Approach Comparing Toxic Potency to *Pimephales promelas*, *Tetrahymena pyriformis*, and *Vibrio fisheri*. In *Comparative QSARs*; Devillers, J., Ed.; Taylor and Francis: London; 1997; pp 52−109.

(8) Könemann, H. Quantitative Structure−Activity Relationships in Fish Toxicity Studies. I. Relationship for 50 Industrial Pollutants. *Toxicology* **1981**, *19*, 209−221.

(9) Lipnick, R. L. Practical Applications of Quantitative Structure−Activity Relationships (QSAR). In *Environmental Chemistry and Toxicology;* Karcher, W., Devillers, J., Eds.; Kluwer Academic: Dordrecht, 1990; pp 281−293.

(10) Veith, G. D.; Call, D. J.; Brooke, L. T. Structure-Toxicity Relationships for the Fathead Minnow, *Pimephales Promelas*: Narcotic Industrial Chemicals. *Can. J. Fish. Aquat. Sci.* **1983**, *40*, 743−748.

(11) Lipnick, R. L. Base-Line Toxicity Predicted by Quantitative Structure−Activity Relationships as a Probe for Molecular Mechanism of Toxicity. In *Probing Bioactive Mechanism*; Magee, P. S., Henry, D. R., Block, J. H., Eds.; American Chemical Society: Washington, DC; 1989; Vol. 413, pp 366−389.

(12) Franks, N. R.; Lieb, W. R. Mechanism of General Anesthesia. *Environ. Health. Perspect.* **1990**, *87*, 199−205.

(13) Schultz, T. W.; Holcombe, G. W.; Phipps, G. L. Relationships of Quantitative Structure−Activity to Comparative Toxicity of Selected Phenols in the *Pimephales Promelas* and *Tetrahymena Pyriformis* Test Systems. *Ecotoxicol. Environ. Saf.* **1986**, *12*, 146−153.

(14) Veith, G. D.; Broderius, S. J. Rules for Distinguishing Toxicants that Cause Type I and Type II Narcosis Syndromes. *Environ. Health Perspect.* **1990**, *87*, 207−211.

(15) Terada, H. Uncouplers of Oxidative Phosphorylation. *Health Perspect.* **1990**, *87*, 213−218.

(16) Veith, G. D.; Broderius, S. J. Structure-Toxicity Relationships for Industrial Chemicals Causing Type (II) Narcosis Syndrome. In *QSPR in Environmental Toxicology-II*; Kaiser, K. L. E., Ed.; D. Reidel Publishing Company: Dordrecht; 1987; p 385.

(17) Lipnick, R. L.; Johnson, D. E.; Gilford, J. H.; Bickings, C. K.; Newsome, L. D. Comparison of Fish Toxicity Screening Data for 55 Alcohols with the Quantitative Structure−Activity Relationship Predictions of Minimum Toxicity for Nonrteactive Nonelectrolyte Organic Compounds. *Environ. Toxicol. Chem.* **1985**, *4*, 281−296.

(18) Hermens, J. L. M. Electrophiles and Acute Toxicity to Fish. *Environ. Health Perspect.* **1990**, *87*, 219−225.

(19) Lipnick, R. L. Outliers: Their Origin and Use in the Classification of Molecular Mechanism of Toxicity. *Sci Total Environ.* **1991**, *109/110*, 131−153.

(20) Ariëns, E. J. Receptors: a Tool in Drug Development. In *Innovative Approaches in Drug Research;* Harms, A. F., Ed.; Elsevier Science Publishers: Amsterdam, 1986; pp 9−22.

(21) Verhaar, H. J. M.; van Leeuwen, C. J.; Hermens, J. L. M. Classifying Environmental Pollutants. 1: Structure−Activity Relationships for Prediction of Aquatic Toxicity. *Chemosphere* **1992**, *25*, 471−491.

(22) Verhaar, H. J. M.; Ramos, E. U.; Hermens, J. L. M. Classifying Environmental Pollutants. 2: Separation of Class 1 (Baseline Toxicity) and Class 2 ('Polar Narcosis') Type Compounds Based on Chemical Descriptors. *J. Chemometr.* **1996**, *10*, 149−162.

(23) Brüggemann, R.; Altschuh, J.; Matthies, M. QSAR for Estimating Physicochemical Data. In *Practical Applications of Quantitative Structure−Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Kracher, E., Devillers, J., Eds.; Kluwer Academic Publishers: London; 1990; pp 197−212.

(24) Lipnick, R. L. A Quantitative Structure−Activity Relationship Study of Overton's Data on the Narcosis and Toxicity of Organic Compounds to the Tadpole, *Rana Temporaria*. In *Aquatic Toxicology and Environmental Fate*; Suter, G. W., Lewis, M. A., Eds.; American Society for Testing and Materials: Philadelphia, 1989; Vol. 11, p 468.

(25) Schultz, T. W.; Lin, D. T.; Wilke, T. S.; Arnold. L. M. Quantitative Structure−Activity Relationships for the Tetrahymena Pyriformis Population Growth Endpoint: a Mechanisms of Acton Approach. In *Practical Applications of Quantitative Structure−Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W., Devillers, J., Eds.; Kluwer Academic: Dordrecht, 1990; pp 241−262.

(26) Deneer, J. W.; Seinen, W.; Hermens, J. M. L. The Acute Toxicity of Aldehydes to the Guppy. *Aquat. Toxicol.* **1988**, *12*, 185−192.

(27) Schultz, T. W.; Bryant, S. E.; Lin, D. T. Structure-Toxicity Relationships for *Tetrahymena*: Aliphatic Aldehydes. *Bull. Environ. Contam. Toxicol.* **1994**, *52*, 279−285.

(28) Hermens, J.; Busser, F.; Leeuwanch, P.; Musch, A. Quantitative Correlation Studies Between the Acute Lethal Toxicity of 15 Organic Halides to the Guppy (*Poecilla Reticulata*) and Chemical Reactivity Towards 4-Nitrobenzylpyridine. *Toxicol. Environ. Chem.* **1985**, *9*, 219−236.

(29) Purdy, R. Taft Sigma and Sigma$_1$ Constants Improve Log Octanol/Water Partition Coefficient Based QSAR for Fathead Minnow Toxicity. In *QSAR in Environmental Toxicology − II*; Kaiser, K. L. E., Ed.; D. Reidel: Dordrecht, 1987; pp 271−280.

(30) Veith, G. D.; Mekenyan, O. G. A QSAR Approach for Estimating the Aquatic Toxicity of Soft Electrophiles [QSAR for Soft Electrophiles]. *Quant. Struct.-Act. Relat.* **1993**, *12*, 349−356.

(31) Mekenyan, O. G.; Veith, G. D. Relationships Between Descriptors for Hydrophobicity and Soft Electrophilicity in Predicting Toxicity. *SAR QSAR Environ. Res.* **1993**, *1*, 335−344.

(32) Fukui, K. Reactivity Indices. In *Theory of Orientation and Stereoselection*; Fukui, K., Ed.; Springer-Verlag: New York, 1975; pp 34−39.

(33) 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: An Analysis Based on Molecular Mechanisms. *Quant. Struct.-Act. Relat.* **1996**, *15*, 302−310.

(34) Karcher, W.; Karabunarliev, S. The Use of Computer Based Structure−Activity Relationships in the Risk Assessment of Industrial Chemicals. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 672−677.

(35) Karabunarliev, S.; Mekenyan, O. G.; Karcher, W.; Russom, C. L.; Bradbury, S. P. Quantum-Chemical Descriptors for Estimating the Acute Toxicity of Substituted Benzenes to the Guppy (*Poecilia reticulata*) and Fathead Minnow (*Pimephales promelas*). *Quant. Struct.-Act. Relat.* **1996**, *15*, 311−320.

(36) Cronin, M. T. D.; Schultz, T. W. Structure-Toxicity Relationships for Three Mechanisms of Action of Toxicity to *Vibrio fischeri*. *Ecotoxicol. Environ. Saf.* **1998**, *39*, 65−69.

(37) Cronin, M. T. D.; Gregory, B. W.; Schultz, T. W. Quantitative Structure−Activity Analyses of Nitrobenzene Toxicity to *Tetrahymena Pyriformis*. *Chem. Res. Toxicol.* **1998**, *11*, 902−908.

(38) Ramos, E. U.; Vaes, W. H. J.; Verhaar, H. J. M.; Hermens, J. L. M. Quantitative Structure−Activity Relationships for the Aquatic Toxicity of Polar and Nonpolar Narcotic Pollutants. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 845−852.

(39) Vaes, W. H. J.; Ramos, E. U.; Verhaar, H. J. M.; Hermens, J. L. M. Acute Toxicity of Nonpolar Versus Polar Narcosis: Is There a Difference? *Environ. Toxicol. Chem.* **1998**, *17*, 1380−1384.

(40) Vaes, W. H. J.; Ramos, E. U.; Verhaar, H. J. M.; Clamer, C. J.; Hermens, J. L. M. Understanding and Estimating Membrane/Water Partition Coefficients: Approaches to Derive Quantitative Structure Property Relationships (QSPR). *Chem. Res. Toxicol.* **1998**, *11*, 847−854.

(41) Vaes, W. H. J.; Ramos, E. U.; Hamwijk, C.; van Holsteijn, I.; Blaauboer, B. J.; Seinen, W.; Verhaar, H. J. M.; Hermens; J. L. M. Solid-Phase Microextraction as a Tool to Determine Membrane/Water Partition Coefficients and Bioavailable Concentrations in in Vitro Systems. *Chem. Res. Toxicol.* **1997**, *10*, 1067−1072.

(42) Gombar, V. K.; Enslein, K. Topological Shape and Electronic Descriptors and Their Correlation with Toxicity to *Photobacterium Phosphoreum*. *In Vitro Toxicol.* **1988/89**, *2*, 117−127.

(43) Sixt, S.; Altschuh, J.; Brüggemann, R. Quantitative Structure-Toxicity Relationships for 80 Chlorinated Compounds Using Quantum Chemical Descriptors. *Chemosphere* **1995**, *30*, 2397−2414.

(44) Todeschini, R.; Vighi, M.; Provenzani, R.; Finizio, A.; Gramatica, P. Modeling and Prediction by Using WHIM Descriptors in QSAR Studies: Toxicity of Heterogeneous Chemicals on *Daphnia Magna*. *Chemosphere* **1996**, *32*, 1527−1545.

(45) Abraham, M. H.; Rafols, C. Factors That Influence Tadpole Narcosis. An LFER Analysis. *J. Chem. Soc., Perkin Trans. 2* **1995**, 1843−1851.

(46) Katritzky, A. R.; Lobanov, V. S.; Karelson, M. *Comprehensive Descriptors for Structural and Statistical Analysis, Reference Manual, Version 2.0*; 1994.

(47) Katritzky, A. R.; Lobanov, V.; Karelson, M. QSPR: The Correlation and Quantitative Prediction of Chemical and Physical Properties from Structure. *Chem. Soc. Rev.* **1995**, 279−287.

(48) Katritzky, A. R.; Karelson M.; Lobanov, V. QSPR as a Means of Predicting and Understanding Chemical and Physical Properties in Terms of Structure. *Pure Appl. Chem.* **1997**, *69*, 245−248.

(49) Karelson, M.; Maran, U.; Wang, Y.; Katritzky, A. R. QSPR and QSAR Models Derived Using Large Molecular Descriptor Spaces. A Review of CODESSA Applications. *Collect. Czech. Chem. Commun.* **1999**, *64*, 1551−1571.

(50) Maran, U.; Karelson, M.; Katritzky, A. R. A Comprehensive QSAR Treatment of the Genotoxicity of Heteroaromatic and Aromatic Amines. *Quant. Struct.-Act. Relat.* **1999**, *18*, 3−10.

(51) Saarikoski, J.; Viluksela, M. Relation Between Physicochemical Properties of Phenols and Their Toxicity and Accumulation in Fish. *Ecotoxicol. Environ. Saf.* **1982**, *6*, 501−512.

(52) *HyperChem 5.1, Molecular modeling package*; Hypercube Software: Gainesville, FL, 1997.

(53) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. AM1: A New General Purpose Quantum Mechanical Molecular Model. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.

(54) Baker, J. An Algorithm for the Location of Transition States. *J. Comput. Chem.* **1986**, *7*, 385−395.

(55) Stewart, J. J. P. *MOPAC Program Package*; QCPE: 1989; No 455.

(56) Karelson, M.; Lobanov, V. S.; Katritzky, A. R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* **1996**, *96*, 1027−1043.

(57) Katritzky, A. R.; Lobanov, V. S.; Karelson, M.; Murugan, R.; Grendze, M. P.; Toomey, J. E., Jr. Comprehensive Descriptors For Structural and Statistical Analysis.1 Correlations Between Structure and Physical Properties and Substituted Pyridines. *Rev. Roum. Chim.* **1996**, *41*, 851−867.

(58) Katritzky, A. R.; Mu, L.; Lobanov, V. S.; Karelson, M. Correlation of Boiling Points with Molecular Structure. 1. A Training Set of 298 Diverse Organics and a Test Set of 9 Simple Inorganics. *J. Phys. Chem.* **1996**, *100*, 10400−10407.

(59) Draper, N. R.; Smith, H. *Applied Regression Analysis*; Wiley: New York, 1966.

(60) Bodor, N.; Buchwald, P. Molecular Size Based Approach to Estimate Partition Properties for Organic Solutes. *J. Phys. Chem.* **1997**, *101*, 3404−3412.

(61) Edward, J. T. Calculation of Octanol−Water Partition Coefficients of Organic Solutes from Their Molecular Volumes. *Can. J. Chem.* **1998**, *76*, 1294−1303.

(62) Kurtz, H. A.; Stewart, J. J. P.; Dieter, K. M. Calculation of the Nonlinear Optical Properties of Molecules. *J. Comput. Chem.* **1990**, *11*, 82−87.

(63) Hatch, F. T.; Colvin, M. E. Quantitative Structure−Activity (QSAR) Relationships of Mutagenic Aromatic and Heterocyclic Amines. *Mut. Res.* **1997**, *376*, 87−96.

(64) Roy, A. B.; Basak, S. C.; Harriss, D. K.; Magnuson, V. R. Neighborhood Complexities and Symmetry of Chemical Graphs and Their Biological Applications. In *Mathematical Modelling in Science and Technology*; Avula, X. J. R., Kalman, R. E., Liapis, A. I., Rodin, E. Y., Eds.; Pergamon Press: New York, 1983; pp 745−750.

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

(66) Zefirov, N. S.; Kirpichenok, M. A.; Izmailov, F. F.; Trofimov, M. I. Scheme for the Calculation of the Electronegativities of Atoms in a Molecule in the Framework of Sanderson's Principle. *Dokl. Akad. Nauk SSSR* **1987**, *296*, 883−887.

CI010011R