August, 1996

Quantum-chemical Descriptors for Estimating the Acute Toxicity of Electrophiles to the Fathed minnow (Pimephales promelas): An Analysis Based on Molecular Mechanisms

Stoyan Karabunarliev, *European Chemicals Bureau, Italy*
Ovanes G. Mekenyan, *Bourgas University of Technology*
Walter Karcher, *European Chemicals Bureau, Italy*
Christine L. Russom, *United States Environmental Protection Agency*
Steven P. Bradbury, *United States Environmental Protection Agency*

Quantum-chemical Descriptors for Estimating the Acute Toxicity of Electrophiles to the Fathed minnow (*Pimephales promelas*): An Analysis Based on Molecular Mechanisms

Stoyan Karabunarliev*[a,b], Ovanes G. Mekenyan[b], Walter Karcher[a], Christine L. Russom[c], and Steven P. Bradbury[c]

[a] Joint Research Centre, Environment Institute, European Chemicals Bureau, TP 280, 21020 Ispra (VA), Italy
[b] Department of Physics and Department of Physical Chemistry, Bourgas University of Technology, 8010 Bourgas, Bulgaria
[c] U.S. EPA, National Health and Environmental Effects Research Laboratory, Mid-Continent Ecology Division, 6201 Congdon Boulevard, Duluth, MN 55804, USA

Abstract

Estimating the toxicity of reactive xenobiotics to aquatic organisms requires physicochemical descriptors of passive transport and chemical reactions with nucleophilic biological ligands. Herein, electrophiles whose toxic action is attributed to nucleophilic substitution (S_N), Michael-type addition and Schiff-base formation were examined. Training sets for each molecular mechanism were generated through substructure search applied to chemicals in a fathead minnow (*Pimephales promelas*) database. Based on a delineation of compounds by a presumed molecular mechanism, relationships between modes of toxic action, potency (96-hour LC_{50} values) and mechanistically-appropriate quantum-chemical descriptors were explored. Monohalo-C(sp^3) function which may give rise to S_N reactivity was encountered in 35 compounds. The inclusion of E_{LUMO}, a nonspecific electrophilicity descriptor, to the generic LC_{50} – hydrophobicity relation increased the explained variance from $r^2 = 36\%$ to 69%. Eighteen potential Michael-type acceptors, mainly acrylates, were identified by the presence of a localized CC double bond at an α, β position to a polar group. Due to different modes of action, the toxic potency of these chemicals varies almost independently of hydrophobicity ($r^2 = 0.12$). Two additional electronic descriptors that are consistent with the likely molecular mechanism provide a multivariate QSAR with $r^2 = 0.78$. Forty-five aldehydes and 3 formamides comprised the training set associated with probable Schiff-base mechanism of toxicity. The results suggest a marginal increase of toxic potency from that expected due to narcosis for more electrophilic carbonyl groups. Overall, it was concluded that regressions based on data sets that combine reactive chemicals with narcotics typically require an electronic descriptor in addition to hydrophobicity, even if the compounds all contain a common electrophilic moiety related to the putative specific reaction mechanism. However, without the generation of additional toxicity data from chemical sets that incorporate a broader range of electronic and steric character, it will likely remain extremely difficult to develop a quantitative ability to predict the potency of electrophilic compounds.

Key words: Acute fish toxicity, electrophiles, mode of toxic action, molecular mechanism, substructural screens, quantum-chemical descriptors

* to receive all correspondence

Abbreviations

LC_{50}	medium lethal concentration
K_{ow}	1-octanol/water partition coefficient
E_{LUMO}/E_{HOMO}	energy of the lowest unoccupied/highest occupied molecular orbital
Q_X	net charge for atom X
A_X/D_X	acceptor/donor superdelocalizability index for atom X
B_{X-Y}	bond order for the covalent bond XY
Q^2	squared correlation coefficient of predictions by "leave one out" procedure

1 Introduction

Initial chemical effect assessments for aquatic organisms typically incorporate the use of acute toxicity estimates. Through the associated development of quantitative structure-activity relationships (QSARs) during the early 1980s [1, 2], it became well-established that the majority of industrial organic chemicals (excluding pesticides and pharmaceutical agents) elicit their acute toxic effects through a narcosis mechanism [3]. The findings of Könnemann [1] and Veith *et al.* [2] established that the potency of narcotics was entirely dependent upon the xenobiotics' hydrophobicity. With subsequent experimental studies and modelling efforts it has been generally accepted that these relationships represent the minimum, or baseline, toxicity that a compound can elicit in the absence of a more specific mode of toxic action (e. g., see review by Lipnick [4]). With the development of additional well-defined toxicity data sets (e. g., [5, 6]), which in some instances include complementary joint toxic action, physiological and behavioral analyses (e. g., see [7, 8]), it has become obvious that some industrial chemicals are significantly more toxic than would be predicted from narcosis because they are capable of acting through different modes of toxic action. As a consequence, the process of QSAR development and selection in ecological effect assessments represents a major area of uncertainty, where errors associated with mode of action considerations can result in 10 to 1000 fold errors in toxic potency estimates.

Traditionally, the selection of QSARs has been based on a "chemical class" perspective; however, research completed over the past several years has been addressing the need to establish chemical

0931-8771/96/0408-0302 $10.00+.25/0

similarity in the context of common mechanisms or modes of toxic action (e. g., see [6–9]). Many of these recent efforts to establish qualitative structure activity relationships for common modes of toxic action have centered on the use of two-dimensional (2-D) substructural features. For example, Veith and Broderius [10] discussed structural requirements for separating nonpolar (baseline or Type I) from polar (Type II) narcotics and numerous authors have proposed structural fragments that can be associated with electrophilic compounds [11, 12], and associated molecular mechanisms [12, 13]. Incorporation of substructural fragment rules into expert systems to classify compounds by modes of toxic action have also been reported [6, 8].

Although substructural fragments can be used to qualitatively identify potential electrophilic compounds [6, 8, 12, 13], the quantitative prediction of toxicity, consistent with a proposed or known molecular mechanism, remains a challenging problem. Several studies have been published that address the incorporation of quantum-chemical descriptors in mode of toxic action and/or potency prediction (e. g., see [14–16]). These investigations have generally been restricted to relatively small congeneric data sets. However, there have been some efforts reported where the acute toxicity of compounds in larger data sets have been reasonably well-resolved in terms of lipophilicity and soft electrophilicity [17]. Although the successful use of quantum-chemical descriptors in predictive aquatic toxicology research are being realized, these findings must be balanced against the need to assess large sets of compounds in a computationally-efficient, but toxicologically-defensible, manner. As a consequence, there remains a need to refine the specificity of 2-D substructural fragments used to identify electrophiles and, in the context of different molecular mechanisms, improve quantitative toxicity models. In the current investigation, relationships between 2-D and quantum-chemical descriptors and the acute toxicity of 101 compounds to the fathead minnow (*Pimephales promelas*) were studied in the context of mechanisms attributed to nucleophilic substitution (S_N), Michael-type addition and Schiff-base formation.

2 Materials and Methods

2.1 Toxicological and Chemical Descriptor Data

A toxicity database that contains 96-hour LC_{50} values based on flow-through exposures and measured toxicant concentrations was used in the study. The toxicity collection of 666 discrete organic non-ionic chemicals was generated under a common set of experimental conditions by the same laboratory [5]. For replicate bioassays on a given chemical, the average LC_{50} value was used. Additional toxicodynamic data has been summarized [7, 8] for a subset of these chemicals and included into the database. The latter was used to refine and evaluate relationships between LC_{50} values and electrophilic substructures and associated properties.

The 2-D structures of the chemicals, in SMILES notation [18], were converted to three-dimensional (3-D) structures through the use of CORINA [19] and subsequently optimized by means of MOPAC 7 [20]. The AM1 all-valence electron semiempirical Hamiltonian [21] was used. No molecular mechanics correction for peptide linkage was employed. The geometry optimization was performed in Cartesian coordinates, without any constraints and using the keyword "precise". All molecules were processed without any failure, both in the initial 3-D generation and the consequent quantum-chemical optimization.

Net atomic charges Q_X; bond orders B_{X-Y} [22]; acceptor (nucleophilic) and donor (electrophilic) superdelocalizability indices [23], A_X and D_X, were used as quantum-chemical descriptors of electronic structure. These quantities were calculated in correspondence with Mulliken population analysis by taking into account the differential overlap between atomic orbitals. Equations 1 and 2 were used to derive the superdelocalizability indices:

$$A_X = \sum_{i}^{vac} \sum_{\alpha}^{X} \frac{C_{i\alpha} C_{i\alpha}}{E_i - E_{ref}} \tag{1}$$

$$D_X = \sum_{i}^{occ} \sum_{\alpha}^{X} \frac{C_{i\alpha} C_{i\alpha}}{E_{ref} - E_i} \tag{2}$$

Here C_i are the eigenvectors, premultiplied by the inverse square root of the overlap matrix, E_i are the MO levels, and α pertains to the atomic orbitals of site X. The superdelocalizability indices reflect the propensities of atomic sites to stabilize a negative/positive charge. In recent QSAR studies [15, 16] employing A_X and D_X were based on chemical-specific individual energy reference levels (i. e., a unique $E_{ref} = 1/2\ (E_{HOMO} + E_{LUMO})$ for each compound). In the present study, calculations were performed with a fixed E_{ref} value [14, 17] to more systematically exploit the potential of the superdelocalizability indices to differentiate site-specific acceptor/donor properties not only within a single molecule, but also between sites within different chemicals. Accordingly, E_{ref} was set equal to the average HOMO-LUMO midgap level for all the chemicals of the database; i. e., $E_{ref.} = aver\ \{1/2\ (E_{HOMO} + E_{LUMO})\}$.

Quantum-chemical descriptors, log K_{ow} estimates [24], and 96-hour LC_{50} values for the compounds were stored in OASIS [25] file format.

2.2 QSAR analyses

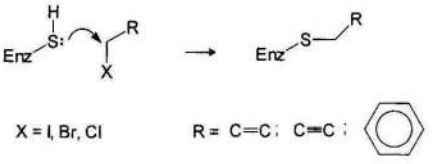
To identify sets of compounds hypothesized to react with target nucleophiles in biomacromolecules [12, 13] through nucleophilic substitution, Michael-type addition or Schiff-base formation, a general substructure recognition technique was employed. The relevant 2-D screens were defined by means of a linear notation [26]. Selection rules based on Boolean logic were used. The "electrophile" search strategies used in this study were less restrictive than those previously described [6, 8] and typically selected narcotics as well as electrophiles according multireference toxicodynamic characterization [7, 8] aimed at the assignment of prevailing mode of toxic action. Such an approach was deliberately employed with the aim to determine whether more simplified substructure search, combined with subsequent use of quantum-chemical descriptor could provide a more quantitative assessment of mode of toxic action and potency.

Multivariate linear regression was undertaken using the statistical module within OASIS [25]. Stepwise variable selection was performed, with a linear relationship between log LC_{50} and log K_{ow} used as the initial regression. Acceptance/rejection of additional electronic descriptors was based on a 70% confidence level. The significance of any additional descriptors was assessed by the statistical probability of the F-value for explained variance with k + 1 versus k independent variables. The 95% t-test confidence limits for the free parameters of the regressions are also provided.

3 Results and Discussion

3.1 Nucleophilic Substitution

In many instances, significant excess toxicity has been attributed to S_N reaction involving a halogen or other suitable leaving group [12, 13]. Chemicals acting by this type of mechanism are considered to undergo covalent binding with sulfhydryl, amino, and other nucleophilic ligands found in biological macromolecules. Pronounced excess toxicity is exhibited by those electrophiles in which the leaving group is at an α-position to a double bond (allylic activation), a triple bond (propargylic activation) or an aromatic ring (benzylic activation) [13].



(I)

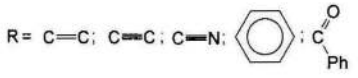
However, other groups in proximity of the halo-carbon, such as phenacyl, carboxyl, and amide [13, 14], could also introduce π-activation. Due to the variety of potential S_N activating moieties, a substructural screen based on an sp^3 carbon site covalently bound to a single halogen and any combination of carbons and hydrogens was used: $X–CY_3$, with X = I, Br, Cl and Y = C, H. Only I, Br and Cl were considered as leaving groups, while F, C≡N and OH were

Table 1. Potential S_N electrophiles

No.	CAS no.	Chemical name	a)	b)	c)	d)	e)
1	–	2-chloroethyl-n-cyclohexyl carbamate	R	3.769	4.010	2.46	0.835
2	78875	1,2-dichloropropane	N	2.949	3.448	1.99	1.087
3	79005	1,1,2-trichloroethane	N	3.214	4.894	2.05	0.174
4	96139	2,3-dibromo-1-propanol	R	3.487	4.363	0.63	0.188
5	96184	1,2,3-trichloropropane	N	3.398	3.955	1.98	0.760
6	106945	1-bromopropane	N	3.262	3.882	2.10*	0.834
7	107062	1,2-dichloroethane	N	2.862	3.907	1.48*	0.676
8	107073	2-chloroethanol	R	3.152	2.427	0.03*	1.293
9	107142	chloroacetonitrile	BA	4.748	4.058	0.45*	0.343
10	109648	1,3-dibromopropane	R	4.980	4.581	1.99	0.360
11	109659	1-bromobutane	N	3.572	4.122	2.75*	0.830
12	110565	1,4-dichlorobutane	N	3.391	3.385	2.24	1.185
13	111251	1-bromohexane	N	4.680	4.508	3.80*	0.823
14	111831	1-bromooctane	N	5.363	4.912	4.89*	0.814
15	126727	2,3-dibromo-1-propanol,phosphate (3:1)	–	6.200	6.144	3.51	−0.293
16	127004	1-chloro-2-propanol	R	2.586	2.312	0.14	1.391
17	142289	1,3-dichloropropane	N	3.001	3.543	2.00*	1.029
18	542756	1,3-dichloropropene	RA	5.667	4.327	1.60	0.434
19	623256	1,4-bis(chloromethyl)benzene	RA	6.652	5.895	3.27	−0.188
20	627305	3-chloro-1-propanol	–	2.072	2.282	0.01	1.381
21	628762	1,5-dichloropentane	N	3.746	3.405	2.76	1.292
22	629049	1-bromoheptane	N	5.086	4.705	4.36*	0.826
23	760236	3,4-dichloro-1-butene	RA	4.180	4.023	1.97	0.714
24	822866	trans-1,2-dichlorocyclohexane	N	3.920	3.765	3.18*	1.157
25	1204213	2-bromo-1-(2,5-dimethoxyphenyl)ethanone	RA	6.581	6.133	2.39	−0.543
26	1871574	3-chloro-2-chloromethyl-1-propene	RA	5.818	3.869	1.56	0.719
27	5407045	3-chloro-n,n-dimethyl-propanamine,hydrochloride	?	2.961	2.446	0.66	1.425
28	7250671	1-(2-chloroethyl)pyrrolidinehydrochloride	N	2.941	2.743	1.43	1.411
29	10293068	[1(R)-endo]-(+)-3-bromocamphor	?	3.528	5.191	2.99	0.199
30	14064109	diethyl chloromalonate	?	5.311	5.441	2.59	−0.054
31	15972608	alachlor	N	4.732	5.721	3.52*	−0.019
32	23184669	n-(butoxymethyl)-2-chloro-n-(2,6-diethylphenyl)acetamide	–	6.031	5.923	5.58	0.324
33	27304138	1,2,4,5,6,7,8,8-octachloro-2,3-epoxy-3a,4,7,7a-tetrahydro-, exo-,endo-4,7-methanoindan	−A	8.247	6.705	3.72	−0.605
34	30030252	chloromethyl styrene	RA	5.692	6.076	3.43	0.268
35	34723825	2-(bromomethyl)tetra-hydro-2h-pyran	N	2.941	3.622	1.61	0.889

a) Mode of toxic action [8]: R – reactive, N – narcosis, B – respiratory blocker, ? – unresolved, – untested, A – presence of π-activation; b) log($1/LC_{50}$. mol/l) observed; c) log $1/LC_{50}$ from Eq. (3); d) log K_{ow} from CLogP [24], StarList measured values marked by *; e) E_{LUMO} [eV].

not. Previous reports [11, 13] have suggested that F and C≡N are not suitable leaving groups in S_N displacements. The hydroxy group was excluded because with the relevant allylic and propargylic alcohols in the data set, the electrophilicity is presumed to be due to metabolic activation to the corresponding aldehyde [13, 14].



(2)

Based on the substructural screen, a set of 35 compounds were identified (Table 1) and include non-polar narcotics as well as reactive toxicants [8]. Although the LC_{50} database contains some iodinated compounds, none of them meets the substructural criteria. Figure 1a depicts a toxicity versus hydrophobicity plot for these chemicals. Those compounds acting via narcosis comply to the baseline toxicity – hydrophobicity relation. In general, halides identified as reactive toxicants are more potent than would be predicted from the baseline model. Closer inspection of the data suggests that excess toxicity can be attributed to S_N reactions for compounds with

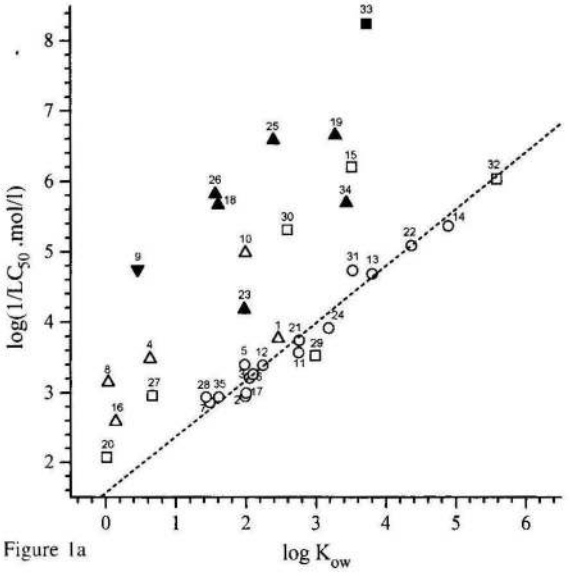


Figure 1a

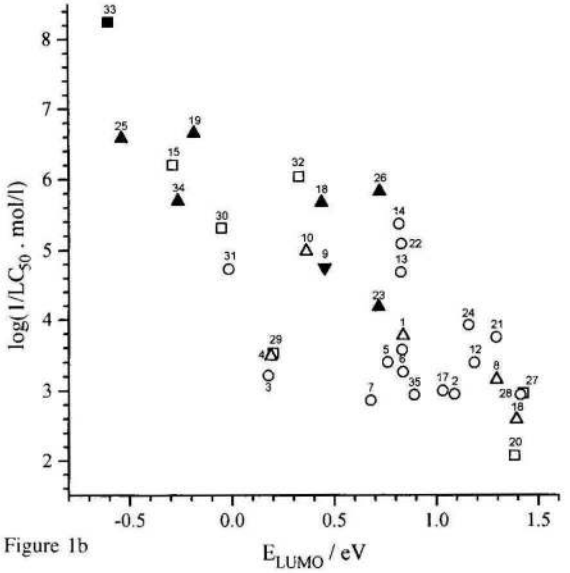


Figure 1b

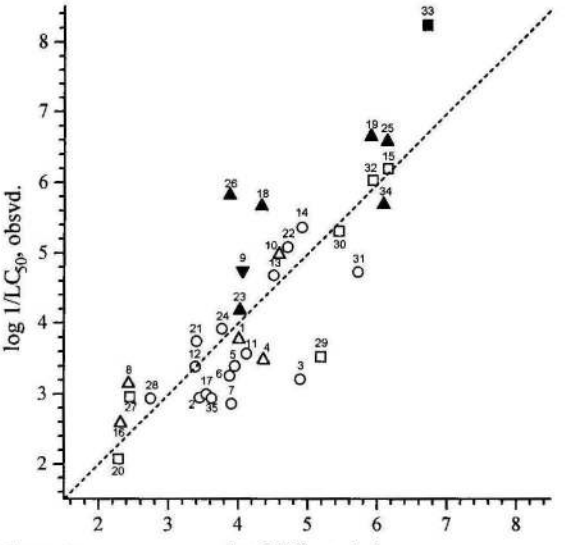


Figure 1c

Figure 1a–c. Plots of observed log $1/LC_{50}$ values for potential S_N electrophiles (see Table 1) against **a)** log K_{ow} (dotted line is the baseline toxicity regression for the chemicals of the entire database, resolved [8] as non-polar narcotics); **b)** E_{LUMO}; and **c)** log $1/LC_{50}$ calculated from Eq. 3. (circles – narcotics; up triangles – reactive chemicals; down triangle – respiratory blokker; squares – unresolved mode of toxic action or untested chemicals; solid symbols indicate the presence of π-activation).

a halogenated carbon site at an α-position to the following activating substrate groups:

Diethyl chloromalonate (30) with the halo-carbon adjacent to two ester bonds exhibits excess toxicity at a level comparable with that of two other non-halogenated malonates in the database. Hence, the elevated toxicity is likely due to the diester moiety [27], rather than to chlorine as a potential leaving group in a S_N reaction. In contrast to other bromoalkanes, both the excess toxicity and mode of action of 1,3-dibromopropane imply an effective S_N mechanism, in spite of the lack of π-activation. The three α-haloalcohols (4, 8 and 16), which are also more toxic than baseline narcotics and exhibit a mode of action consistent with chemical reactivity [8], likely are metabolically activated to the corresponding carbonyl metabolites [28, 29]. Apart from the allylic activated chloro-carbon functions, one of the most potent toxicants contains an epoxy group (33). The mechanism of epoxide toxicity [15, 30] may be distinct from the S_N reactivity associated with the other compounds in this data set. The excess toxicity of the organophosphate (15) may be due to neurotoxicity caused by inhibition of acetylcholinesterase.

Due to the presence of direct-acting electrophiles in the data set of the 35 halogenated compounds, log K_{ow} explained only 36% of the variance of log LC_{50}, while E_{LUMO}, a nonspecific electrophilicity descriptor, explained 60% of the variance (Figure 1b). A single multivariate model combined hydrophobicity and E_{LUMO}, as summarized in Eq. 3:

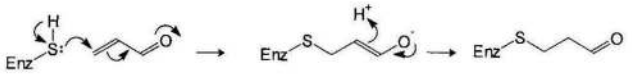
$$\log 1/LC_{50} = -1.559(\pm 0.337)E_{LUMO} + 0.358(\pm 0.106)\log K_{ow} + 4.430(\pm 0.283) \quad (3)$$

$n = 35 \quad r^2 = 0.69 \quad s^2 = 0.67 \quad F = 35.64 \quad Q^2 = 0.41$

A comparison of observed vs. predicted log LC_{50} values is presented in Figure 1c. Although Eq. 3 confirms the general trend that more electrophilic chemicals are more toxic, it does not reveal any particular features consistent with the putative S_N mechanism of toxicity for some of the chemicals. Indeed, E_{LUMO} does not provide a strict distinction between reactive and inert chemicals (Figure 1b). No site-specific electronic descriptors pertaining to the common 2-D fragment associated with nucleophilic substitutions were found that improved the toxicity regressions. Several critical factors may be involved that could not be adequately assessed due to the limited number of compounds. For example, bromine is a better leaving group than chlorine and steric factors, such as branching at the active halo-carbon, can influence reaction rates.

3.2 Michael-type Addition

The pronounced toxicity of compounds with a CC unsaturated bond



(3)

α, β to a polar group, R, can be explained by Michael-type addition to nucleophilic macromolecular sites.

Several prominent examples of such toxicants are arcolein, acrylamide and α-naphthoquinone [13]. To identify the potential Michael-type acceptors in the database, a 2-D search was performed for the fragments C=C–R and C≡C–R, with R = C=O, C≡N, NO_2 and SO_2. The search resulted in 29 substances, none of which contained a nitro or a sulfone as a polar R group. The original search criteria were apparently too unrestrictive, since some of the chemicals selected were molecules in which the α, β-(R) unsaturated bond involved a stable, 6 π-electron heterocycle (e. g. 2-furancarboxaldehyde, uracil) or in which the active β position was shielded (e. g.; isophorone). Based on additional refinements to the 2-D search, chemicals with an unsaturated CC bond not involved in aromatic stabilization and with at least one β-proton were identified and are given in Table 2. Most of the resulting compounds were acrylates with a terminal C=C bond and the ester moiety as a polar group. Along with naphthalenedione (36), the active C_β position is branched for (E,E)-1-[5-(1,3-benzodioxol-5-yl)-1-oxo-2,4-pentadienyl]piperidine (39) and dibutyl fumarate (41). In fact for dibutyl fumarate (41), the unsaturated bond is symmetrically surrounded by two ester functions. The acrylates from the data set have been classified either as reactive toxicants or as chemicals invoking a narcosis-like toxicity syndrome [8], albeit they are more toxic than baseline narcotics. The variable mode of toxic action within this group of potential Michael-type acceptors is consistent with the lack of correlation between log LC_{50} and log K_{ow}, which explained only 12% of the variation in the toxicity data (Figure 2a). However, electrophilicity descriptors such as acceptor superdelocalizabilities for the common chemical pattern and E_{LUMO} were more highly correlated with the toxicity endpoint. The most significant multivariate regression obtained is summarized in Eq. 4 (Figure 2b):

$$\log 1/LC_{50} = 28.638(\pm 0.525)A_R + 81.330(\pm 0.306)B_{\alpha\text{-}R} + 0.359(\pm 0.142)\log K_{ow} - 89.090(\pm 0.290) \quad (4)$$

$$n = 18 \quad r^2 = 0.78 \quad s^2 = 0.33 \quad F = 16.49 \quad Q^2 = 0.43$$

where, A_R is the acceptor superdelocalizability for the carbon site of the polar group R, and $B_{\alpha\text{-}R}$ is the bond-order of the $C_\alpha C_R$ single bond. The participation of both quantum-chemical descriptors is consistent from a mechanistic standpoint. During the intermediate stage of the Michael addition (Diagram 3), a $C_\alpha C_R$ double bond is formed and the bond order $B_{\alpha\text{-}R}$ quantifies the proximity of C_α–C_R to C_α=C_R. The acceptor superdelocalizability of C_R is also relevant since this site stabilizes an unpaired electron in the transition from C_α–C_R to C_α=C_R.

Table 2. Potential Michael-type acceptors

No.	CAS no.	Chemical name	a)	b)	c)	d)	e)	f)
36	58275	2-methyl-1,4-naphthalenedione	R	6.195	6.306	0.646	0.936	2.20*
37	79061	2-propenamide	–	2.814	3.215	0.548	0.945	−0.67*
38	80626	methyl methacrylate	–	2.587	3.599	0.541	0.943	1.38*
39	94622	(E,E)-1-[5-(1,3-benzodioxol-5-yl)-1-oxo-2,4-pentadienyl]piperidine	N	4.561	4.810	0.547	0.950	2.70
40	96059	allyl methacrylate	R	5.105	3.635	0.542	0.942	1.57
41	105759	dibutyl fumarate	E	5.553	5.900	0.570	0.950	3.91
42	106638	isobutyl acrylate	R	4.786	4.812	0.538	0.955	2.22*
43	107028	2-propenal(acrolein)	R	6.518	5.978	0.596	0.959	−0.01*
44	140885	ethyl acrylate	–	4.602	4.632	0.540	0.956	1.32*
45	818611	2-hydroxyethyl acrylate	R	4.384	4.516	0.553	0.957	−0.21*
46	868779	2-hydroxyethyl methacrylate	E	2.758	3.377	0.543	0.944	0.47*
47	999611	2-hydroxypropyl acrylate	R	4.589	4.571	0.546	0.958	0.35*
48	2370630	2-ethoxyethyl methacrylate	E	3.757	3.405	0.536	0.942	1.40
49	2455245	tetrahydrofurfuryl methacrylate	E	3.691	3.394	0.535	0.943	1.30
50	2495376	benzyl methacrylate	E	4.577	4.343	0.547	0.944	2.82
51	2499958	hexyl acrylate	?	5.146	5.367	0.539	0.957	3.39
52	3066715	cyclohexyl acrylate	R	5.018	4.776	0.535	0.954	2.78
53	4655349	isopropyl methacrylate	E	3.528	3.530	0.535	0.941	2.25*

a) Mode of toxic action according [8]: R – reactive, N – nonpolar narcosis, E – ester narcosis, ? – unresolved, – – untested; b) log($1/LC_{50}$. mol/l) observed; c) log($1/LC_{50}$) from Eq. 4; d) A_R [eV^{-1}]; e) $B_{\alpha\text{-}R}$; f) log K_{ow} from CLogP [24], StarList measured values marked by *.

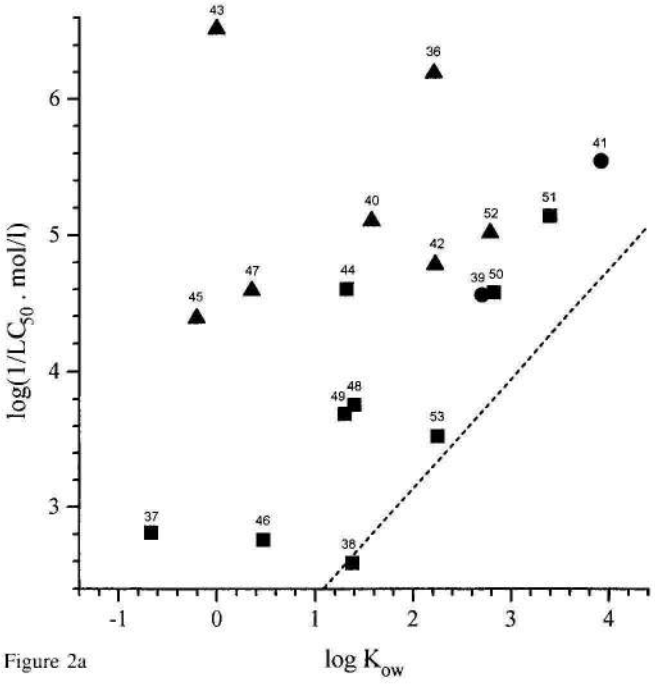


Figure 2a

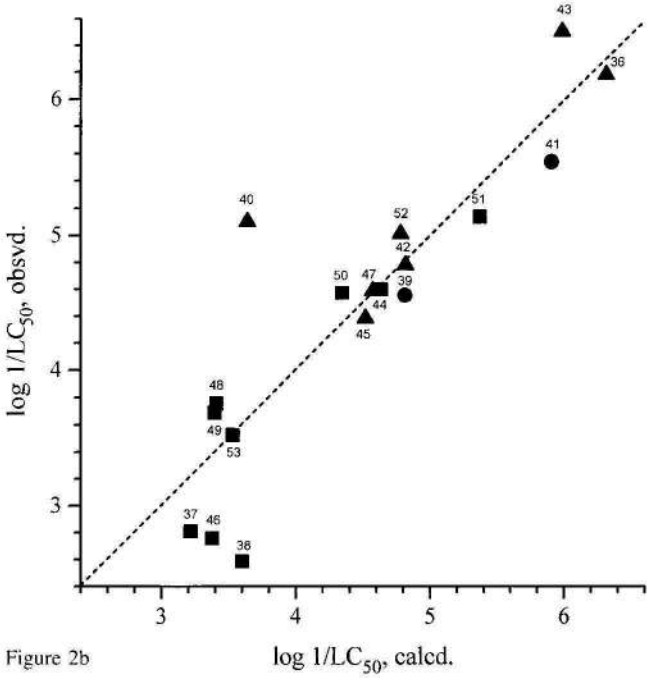


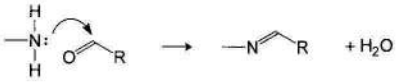
Figure 2b

Figures 2a–b. Plots of observed log $1/LC_{50}$ values for potential Michael-type acceptors (see Table 2) against **a)** log K_{ow} (dotted line is the toxicity regression for the nonpolar narcotics of the database) and **b)** log $1/LC_{50}$ calculated from Eq. 4 (circles – narcotics; triangles – reactive toxicants; squares – unresolved or untested mode of toxic action).

Two prominent outliers, allyl methacrylate (40), and acrolein (43), were associated with the analysis. Allyl methacrylate is the only compound in the data set which has a terminal CC double bond, distinct from a second activated linkage. Acrolein is the single aldehyde within the series and the aldehyde function may be introducing excess toxicity through a different molecular mechanism.

Schiff-base formation

Aldehydes are thought to elicit their toxic effects as electrophiles [12] via the formation of Schiff bases [13] with the amino groups in biological macromolecules:



(4)

Representative examples of aldehydes that elicit excess toxicity via Schiff-base formation are acetaldehyde, butyraldehyde and substituted benzaldehydes [13]. As a screen for electrophiles capable of reacting through Schiff-base formation, chemicals containing the function O=CH–R were selected. Forty-six aldehydes (R = C, H) and 3 formamides (R = N) were identified (see Table 3). Consistent with a previous study [11], 24 out of the 43 aldehydes in the database exhibited excess toxicity of more than 0.8 log units; however, there was no apparent differentiation between aromatic and aliphatic aldehydes. As has been repeatedly documented, acrolein was extremely toxic, with excess toxicity of approximately 4 log units more than predicted assuming baseline narcosis and as a consequence was excluded from the regression analysis. Within the remaining series of compounds, 31 aldehydes have been found to act as reactive toxicants, while 10 have been identified as narcotics [8]. However, the distinction between reactive and narcotic aldehydes with respect to excess toxicity was not pronounced (Figure 3a). Unlike the nucleophilic substitution and Michael-type addition electrophiles, the toxicity of aldehydes, and aldehydes and formamides combined, more closely followed a linear log K_{ow} regression ($r^2 = 0.55$ and 0.45, respectively). Similar hydrophobicity-dependent toxic potency relationships of aldehydes have been reported for guppies [31]. Some statistical improvement in the regressions was obtained with the inclusion of a second descriptor related to the electrophilicity of the carbonyl group. For the aldehyde data set, the two-factor regression with the highest correlation incorporated the charge of the carbonyl oxygen, Q_O, with log K_{ow}, as summarized in Eq. 5:

$$\log 1/LC_{50} = 0.466(\pm 0.059)\log K_{ow} + 12.702(\pm 0.457)Q_O + 7.285(\pm 0.145) \quad (5)$$

$n = 45 \quad r^2 = 0.60 \quad s^2 = 0.23 \quad F = 31.35 \quad Q^2 = 0.31$

Similar levels of statistical significance had other two-variable regressions involing A_C or A_O (the acceptor superdelocalizability for the carbon or oxygen) as electronic descriptors for the carbonyl moiety. The positive correlation of these electronic descriptors (Q_O, A_C or A_O) suggests that the toxic effect depends somewhat on the electrophilic nature of the whole carbonyl group; however, their relation with the likely Schiff-base mechanism of toxic action remains subtle. The inclusion of the 3 formamides in the data set confirms that Q_O is the most suitable complementary descriptor. For the combined series, Q_O is included as a second significant factor at confidence level above 90%. The resulting regression is summarized in Eq. 6:

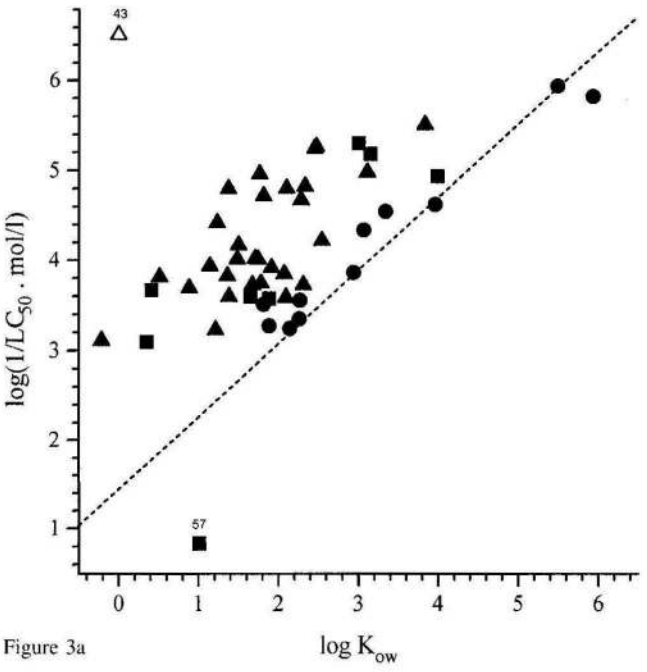


Figure 3a

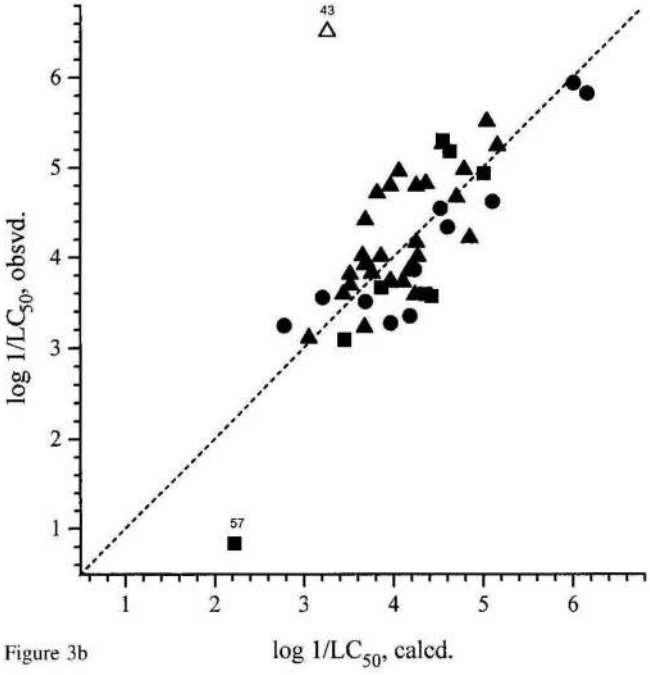


Figure 3b

Figures 3a–b. Plots of observed log $1/LC_{50}$ values for aldehydes and formamides (see Table 3) against **a)** log K_{ow} (dotted line is the toxicity regression for nonpolar narcotics) and **b)** log $1/LC_{50}$ calculated from Eq. 6 (circles – narcotics; solid triangles – reactive toxicants; squares – unresolved or not evaluated mode of toxic action; open triangle – acrolein).

$$\log 1/LC_{50} = 0.496(\pm 0.063)\log K_{ow} + 19.404(\pm 0.474)Q_O + 9.340(\pm 0.153) \quad (6)$$

$n = 48 \quad r^2 = 0.66 \quad s^2 = 0.27 \quad F = 44.42 \quad Q^2 = 0.35$

Observed and calculated log $1/LC_{50}$ values for the combined data set are given in Table 3 and plotted in Figure 3b. The results for this data set do not reveal any pronounced effect of Schiff-base formation as a mechanism of action on the overall toxicity, with log K_{ow} explaining a significant part of the toxicity variance. In the data set studied, the influence of the aldehyde function on toxicity may be attenuated in that phenol and nitro moieties were encountered in 16 of the 35 benzaldehydes. Phenols, anilines and nitrobenzenes are typically more toxic than would be expected from baseline narcosis [6, 10, 11], and, in the case of phenols and anilines, seemingly act through a different mode of narcosis [7, 8, 10].

4 Conclusion

Guided by hypotheses on three molecular mechanisms of electrophilic reactivity, generalized substructure recognition techniques, combined with the use of quantum-chemical descriptors, were used to screen a large heterogeneous chemical structure set and predict 96-hour LC_{50} values for the fathead minnow. The results of this exercise highlight several important issues concerning the differentiation of xenobiotics from both a toxicodynamic and chemical class perspective. Clearly, any strict 2-D classification scheme based on the occurrence of characteristic structural patterns will be imperfect since toxicodynamic classes, as well as chemical classes, can overlap. Thus, 2-D screens for nucleophilic substitution, Michael-type addition and Schiff-base formation all resulted in sets of compounds that comprised baseline narcotics in addition to reactive toxicants. The presence of an electrophilic function does not necessarily imply reactive mode of toxic action because of the general trend toward prevailing narcosis mechanism with increasing lipophilicity of the chemicals [32]. While more specific 2-D screens and other chemical rules could be envisioned in an attempt to better resolve the compounds from a toxicodynamic perspective, the resulting set of criteria quickly becomes unmanageable in the context of developing a rule-based system designed to screen large numbers of noncongeneric structures. Limitations in 2-D structural screens are also noted from a chemical class perspective. An example in this respect is acrolein, which from a 2-D perspective could act through a Michael-type addition or Schiff-base formation, although it seems to act as outlier in both groups. In some cases the differentiation of molecular mechanisms and toxicodynamic responses become interrelated. For example, the evaluation of hydroxy and nitro-substituted benzaldehyde toxicity must be considered in the context of Schiff-base formation and electrophilic reactivity, as well as the potential involvement of polar narcosis mechanisms. As an alternative to strict 2-D divisions of chemicals in molecular mechanism classes, a fragment-additive approach might be more suitable. Differentiation of electrophilic xenobiotics through the identification of global or local quantum-chemical descriptor ranges may be also possible [17].

The results of this effort further documented that regressions based on data sets that combine reactive chemicals with narcotics typically require an electronic descriptor in addition to log K_{ow}, even though the compounds may all contain a common 2-D structural fragment associated with a specific reaction mechanism. An example in this respect can be drawn from the group of chemicals that were identified through the 2-D screening for Michael-type acceptors. In this group of narcotics and reactive toxicants LC_{50}

Table 3. Aldehydes and formamides

No.	CAS no.	Chemical name	a)	b)	c)	d)	e)
54	50000	formaldehyde	–	3.095	3.452	0.35*	−0.312
55	66251	hexanal	R	3.745	3.964	1.78*	−0.323
56	67367	*p*-phenoxybenzaldehyde	N	4.634	5.100	3.96	−0.320
57	68122	n,n-dimethylformamide	–	0.839	2.222	1.01*	−0.393
58	75070	ethanal	R	3.112	3.052	−0.22	−0.318
59	90028	salicylaldehyde	R	4.725	3.810	1.81*	−0.331
60	90595	3,5-dibromosalicylaldehyde	R	5.518	5.035	3.83	−0.320
61	95012	2,4-dihydroxybenzaldehyde	R	4.023	3.650	1.71	−0.337
62	96173	2-methylbutyraldehyde	R	3.936	3.700	1.14	−0.320
63	98011	furfurale	–	3.669	3.864	0.41*	−0.293
64	100107	*p*-dimethylaminobenzaldehyde	N	3.514	3.687	1.81*	−0.338
65	100527	benzaldehyde	R	4.017	3.855	1.48*	−0.320
66	104881	4-chlorobenzaldehyde	R	4.805	4.248	2.10*	−0.316
67	110623	valeraldehyde	R	3.825	3.755	1.36	−0.323
68	120218	4-(diethylamino)benzaldehyde	N	3.870	4.235	2.94	−0.338
69	121324	3-ethoxy-4-hydroxybenzaldehyde	N	3.278	3.967	1.88	−0.325
70	121335	vanillin	R	3.228	3.673	1.21*	−0.323
71	122032	*p*-isopropyl benzaldehyde	N	4.350	4.600	3.07	−0.323
72	123159	2-methylvaleraldehyde	R	3.727	3.959	1.67	−0.320
73	123728	butanal	R	3.691	3.516	0.88*	−0.323
74	148538	*o*-vanillin	R	4.802	3.959	1.37*	−0.312
75	387451	2-chloro-6-fluorobenzaldehyde	R	4.227	4.844	2.54	−0.297
76	446526	*o*-fluorobenzaldehyde	R	4.964	4.056	1.76	−0.317
77	454897	α, α, α-trifluoro-*m*-tolualdehyde	R	5.269	4.541	2.47*	−0.310
78	500221	3-pyridinecarboxaldehyde	R	3.815	3.510	0.51	−0.313
79	529204	*o*-tolualdehyde	N	3.356	4.183	2.26*	−0.324
80	552896	*o*-nitrobenzaldehyde	R	4.016	4.266	1.74*	−0.306
81	555168	4-nitrobenzaldehyde	R	4.175	4.247	1.50	−0.301
82	590863	isovaleraldehyde	R	4.423	3.682	1.23	−0.323
83	613456	2,4-dimethoxybenzaldehyde	R	3.917	3.679	1.91	−0.341
84	635938	5-chlorosalicylaldehyde	?	5.308	4.545	3.00	−0.324
85	653372	pentafluorobenzaldehyde	R	5.251	5.154	2.45	−0.278
86	708769	4,6-dimethoxy-2-hydroxybenzaldehyde	R	4.832	4.358	2.33	−0.316
87	761659	n,n-dibutylformamide	N	3.246	2.777	2.14	−0.393
88	874420	2,4-dichlorobenzaldehyde	R	4.988	4.781	3.11	−0.314
89	1761611	5-bromosalicylaldehyde	?	5.189	4.622	3.15	−0.324
90	2973764	5-bromovanillin	R	3.588	4.238	2.09	−0.316
91	3944761	2,3-dimethylvaleraldehyde	R	3.853	4.166	2.07	−0.320
92	4460860	2,4,5-trimethoxybenzaldehyde	R	3.598	3.436	1.38	−0.340
93	6361213	2-chloro-5-nitrobenzaldehyde	R	4.680	4.699	2.28	−0.297
94	10031820	*p*-ethoxybenzaldehyde	R	3.728	4.104	2.31	−0.329
95	17754904	4-(diethylamino)salicylaldehyde	N	4.557	4.519	3.34	−0.334
96	42454068	5-hydroxy-2-nitrobenzaldehyde	?	3.601	4.351	1.65	−0.299
97	61096842	4-(hexyloxy)-*m*-anisaldehyde	?	4.947	5.007	3.99	−0.325
98	69770236	3-(4-tert-butylphenoxy)benzaldehyde	N	5.837	6.155	5.93	−0.316
99	71862027	3'-chloro-*o*-formotoluidide	N	3.561	3.206	2.27	−0.374
100	79124768	3-(3,4-dichlorophenoxy)benzaldehyde	N	5.950	6.002	5.49	−0.312
101	98434345	5-bromo-2-nitrovanillin	?	3.576	4.423	1.88*	−0.301
43	107028	2-propenal (acrolein)	R	6.518	3.255	−0.01*	−0.313

a) Mode of toxic action [8]: R – reactive, N – narcosis acting, ? – unresolved, – – untested; b) $\log(1/LC_{50} \cdot mol/l)$ observed; c) $\log(1(LC_{50})$ from Eq. 6; d) log K_{ow} from CLogP [24], StarList measured values marked by *; e) Q_o [a.u.]. Acrolein 43 excluded from the correlation sample.

values varied almost independently of log K_{ow}. For chemical data sets comprised of a spectrum of different modes of toxic action, quantum-chemical descriptors can be useful in an attempt to derive adequate correlations. However, their use in this type of regressions must be accompanied by a determination that the electronic properties are appropriate for the presumed molecular mechanism associated with the 2-D selection criteria. In addition, it must be acknowledged that these regressions do not necessarily imply that a statistically acceptable "goodness-of-fit" indicates members of the data set act through the same mode of toxic action. On the contrary, it is more likely that certain electronic descriptors provide a crude distinction between reactive and non-reactive chemicals with a common electrophilic moiety. In the current study, general trends in electrophilic character, mode of toxic action and potency

were identified within specified 2-D classes. However, there remain several important issues concerning the integration of quantum-chemical descriptors in QSAR analyses. For example, in modelling the set of halides within the context of a nucleophilic substitution mechanism, the role of steric hindrance in attenuating the reaction requires additional study.

Consistent with the above mentioned limitations, the use of quantum-chemical approaches in QSAR has been predominately limited to highly-specific congeneric series. Using mechanistically-based 2-D selection techniques, an initial attempt to expand the applicability of quantum-chemical approaches to more heterogeneous data sets was encouraging. Still, it will likely continue to be extremely challenging to develop a quantitative ability to predict the potency of electrophilic compounds without the generation of additional toxicity data from chemical sets that are selected to incorporate an appropriate range of electronic and steric character.

Disclaimer

Mention of trade names or specific products or approaches does not constitute endorsement on the part of the U.S. Environmental Protection Agency.

Acknowledgement

We are grateful to Dr. J. Sadowski and Prof. J. Gasteiger (Universität Erlangen-Nürnberg) for providing the 3-D model builder, CORINA, on a temporary basis. Two of the authors (S.K. and O.M.) acknowledge research grants from the European Commission. This work has been accomplished in the framework of a collaborative agreement between the USEPA and the ECB.

5 References

[1] Könemann, H., *Toxicology 19*, 209–221 (1981).
[2] Veith, G.D., Call, D.J. and Brooke, L.T., *Can. J. Fish. Aquat. Sci. 40*, 743–748 (1983).
[3] Franks, N.R. and Lieb, W.R., *Environ. Health. Persp. 87*, 199–205 (1990).
[4] Lipnick, R.L., in: Karcher, W. and Devillers, J. (Eds.), *Practical Applications of QSARs in Environmental Chemistry and Toxicology*, Kluwer, Dordrecht 1990, pp. 281–294.
[5] Geiger, D.L., Brooke, L.T. and Call, D.J. (Eds.), *Acute Toxicities of Organic Chemicals to Fathead Minnows (Pimephales promelas)*, Center for Lake Superior Environmental Studies, University of Wisconsin-Superior: Superior, WI, 1990; Vol. 5; also Vols. 1–4.
[6] Verhaar, H.J.M., van Leeuwen, C.J. and Hermens, J.L.M., *Chemosphere 25*, 471–491 (1992).
[7] Bradbury, S.P., *SAR QSAR Environ. Res. 2*, 89–104 (1994).
[8] Russom, C.L., Bradbury, S.P., Broderius, S.J., Drummond, R.A. and Hammermeister, D.E., *Environ. Toxicol. Chem.* (submitted).
[9] Cronin, M.T.D. and Dearden, J.C., *Quant. Struct.-Act. Relat. 14*, 1–7 (1995).
[10] Veith, G.D. and Broderius, S.J., *Environ. Health Perspect. 87*, 207–211 (1990).
[11] Jäckel, H. and Nendza, M., *Aquatic Toxicol. 29*, 305–314 (1994).
[12] Hermens, J.L.M., *Environ. Health Perspect. 87*, 219–225 (1990).
[13] Lipnick, R.L., *Sci. Total Environ. 109/110*, 131–153 (1991).
[14] Mekenyan, O.G., Veith, G.D., Bradbury, S.P. and Russom, C.L., *Quant. Struct.-Act. Relat. 12*, 132–136 (1993).
[15] Eriksson, L., Verhaar, H.J.M., Sjöström, M. and Hermens, J.L.M., *Quant. Struc.-Act. Relat. 12*, 357–366 (1993).
[16] Verhaar, H.J.M., Eriksson, L., Sjöström, M., Schüürmann, G., Seinen, W. and Hermens, J.L.M., *Quant. Struct.-Act. Relat. 13*, 133–143 (1994).
[17] Veith, G.D. and Mekenyan, O.G., *Quant. Struct.-Act. Relat. 12*, 349–356 (1993).
[18] Weininger, D., *J. Chem. Inf. Comput. Sci. 28*, 31–36 (1988).
[19] Sadowski, J. and Gasteiger, J., *Chem. Rev. 93*, 2567–2581 (1993); Sadowski, J., Gasteiger, J., Klebe, G., *J. Chem. Inf. Comput. Sci. 34*, 1000–1008 (1994).
[20] Stewart, J.J.P., MOPAC: A general molecular orbital package (Version 7), Q.C.P.E. no. 455 (1995).
[21] Dewar, M.J.S., Zoebisch, E.G., Healy, E.F. and Stewart, J.J.P., *J. Am. Chem. Soc. 107*, 3902–3909 (1985).
[22] Gopinathan, M.S. and Jug, K., *Theor. Chim. Acta 63*, 497–503 (1983).
[23] Fukui, K., Yonezawa, T. and Nagata, C., *Bull. Chem. Soc. Japan 27*, 423–427 (1957).
[24] CLogP version 3.4, Pomona College Medicinal Chemistry Project, Claremont, CA: Version 3.53 for VAX-11 by UPDRIVE software (1988).
[25] Mekenyan, O.G., Karabunarliev, S. and Bonchev, D., *Comput. & Chem. 14*, 193–200 (1990); Mekenyan, O.G., Karabunarliev, S., Ivanov, J. and Dimitrov, D., *Comput. & Chem. 18*, 173–187 (1994).
[26] Karabunarliev, S., Ivanov, J. and Mekenyan, O.G., *Comput. & Chem. 18*, 189–193 (1994).
[27] Veith, G.D., Defoe, D. and Knuth, M., *Drug Metab. Rev. 15*, 1295–1303 (1984).
[28] Henschler, D., in: Anders, M.W. (Ed.), *Bioactivation of foreign compounds*, Academic Press, Orlando, FL 1985, pp. 317–347.
[29] Broderius, S.J., in: Mayes, M.A. and Barron, M.G. (Eds.), *Aquatic Toxicology and Risk Assessment: 14th Volume*, ASTM STP 1124, American Society for Testing and Materials, Philadelphia, PA. 1991, pp. 107–127.
[30] Deneer, J.W., Sinnige, T.L., Seinen, W. and Hermens, J.L.M., *Aquatic Toxicol. 13*, 195–204 (1988).
[31] Deneer, J.W., Seinen, W., Hermens, J.L.M., *Aquatic Toxicol. 12*, 185–192 (1988).
[32] Veith, G.D., Lipnick, R.L. and Russom, C.L., *Xenobiotica 19*, 555–565 (1989).

Received on December 13th, 1995;
accepted on March 27th, 1996.