QSAR Analysis of the Toxicity of Aromatic Compounds to *Chlorella vulgaris* in a Novel Short-Term Assay

Tatiana I. Netzeva, John C. Dearden, Robert Edwards,† Andrew D. P. Worgan, and Mark T. D. Cronin*

School of Pharmacy and Chemistry, Liverpool John Moores University, Byrom Street, Liverpool, L3 3AF, England

Received September 2, 2003

The use of alternative toxicity tests and computational prediction models is widely accepted to fill experimental data gaps and to prioritize chemicals for more expensive and time-consuming assessment. A novel short-term toxicity test using the alga *Chlorella vulgaris* was utilized in this study to produce acute aquatic toxicity data for 65 aromatic compounds. The compounds tested included phenols, anilines, nitrobenzenes, benzaldehydes and other poly-substituted benzenes. The toxicity data were employed in the development of quantitative structure−activity relationships (QSARs). Using multiple regression (MLR) and partial least squares (PLS) analyses, statistically significant, transparent and interpretable QSARs were developed using a small number of physicochemical descriptors. A two-descriptor model was developed using MLR ($\log(1/EC_{50}) = 0.73 \log K_{ow} - 0.59 E_{lumo} - 1.91$; $n = 65$, $r^2 = 0.84$, $r^2_{CV} = 0.82$, $s = 0.43$) and a four-descriptor model using PLS ($\log(1/EC_{50}) = 0.40 \log K_{ow} - 0.23 E_{lumo} + 9.84 A_{max} + 0.20\ ^0\chi^v - 5.40$; $n = 65$, $r^2 = 0.86$, $q^2 = 0.84$, RMSEE = 0.40). The latter model was obtained by stepwise elimination of variables from a set of 102 calculated descriptors. Both models were validated successfully by simulating external prediction through the use of complementary subsets. The two factors, which were identified as being critical for the acute algal toxicity of this set of compounds were hydrophobicity and electrophilicity.

INTRODUCTION

The adoption in the European Union of the White Paper on the Strategy for a Future Chemical Policy[1] and the subsequent changes in the legislation may require toxicity assessment of 30 100 chemical substances in a tight time frame. Basic information for all chemicals marketed in the European Union in volumes greater than 1 tonne per year is required before the end of 2012 with shorter deadlines for the chemicals with higher production/importation volumes.[2] The high expected cost of the Registration, Evaluation and Authorization of CHemicals (REACH) legislation has stimulated discussion regarding the development and validation of a large number of rapid, sensitive and cost-effective in vitro toxicity tests[3] as well as computational prediction models to fill the existing data gaps.[4−8]

The use of short-term toxicity assays and computational models over their traditional counterparts is preferred for many reasons including ease of use, speed and relatively low cost. In addition, they pose fewer problems with regard to the stability of chemicals and potential losses during the testing due to volatility. The acute toxicity, assessed in short and low-cost unicellular tests, is also considered to be a surrogate for the prediction of toxicity to higher aquatic organisms.[9]

The 15-minute algal assay utilized in this study is a rapid method to obtain acute toxicity data. It is based upon the premise that all living organisms, including algae, contain nonspecific esterases, the activity of which can be assessed by the measurement of the disappearance of an ester, or the appearance of the product. As a substrate of esterases in algae, fluorescein diacetate is used in this assay, and the fluorescence of the product (fluorescein) is measured spectrofluorimetrically.[10] This test has been evaluated successfully through the use of quantitative structure activity relationships (QSARs). It has also been found to be an excellent predictor of the toxicity of chemicals to other species.[10]

QSARs have been applied widely to prioritize untested chemicals for more intensive and costly experimental evaluations.[8] There are a number of advantages to the use of QSARs to predict toxicity. Key among these are the facts that the prediction of toxicity can be made from chemical structure alone and the methods are easily automated.[4] However, to gain regulatory acceptance the QSAR models should meet a number of criteria, which are currently subject of intensive discussion in the regulatory and regulated communities.[11] Further, for regulatory use, QSARs will require to undergo some form of validation procedure.[12]

Benzene derivatives comprise a significant component of the pollutant burden on the environment. The toxicity of these compounds can arise from a multitude of mechanisms of toxic action including a range of narcoses as well as reactive mechanisms in which the compounds are able to form covalent bonds with biological macromolecules.[13,14] Although highly desirable,[15] the development of QSARs on the basis of mechanism of action is often a difficult task. Thus, development of models that accurately predict toxicity

* Corresponding author phone: + 44 (0) 151 231 2066; fax: + 44 (0) 151 231 2170; e-mail: m.t.cronin@livjm.ac.uk.

† Current address: Hawkins & Associates, 18-20 Manchester Road, Wilmslow, Cheshire, SK9 1BG, England.

10.1021/ci034195g CCC: $27.50
Published on Web 11/08/2003

without requiring the initial identification of an exact mechanism can be beneficial.[13]

The aims of this study, therefore, were to evaluate a series of aromatic compounds in a novel rapid toxicity test utilizing the alga *Chlorella vulgaris*. Further, the information from the assay was used to develop transparent and interpretable QSAR models for the prediction of the toxicity of chemicals from structure alone, without a priori assignment of mechanism of action.

MATERIALS AND METHODS

Chemicals Tested. A total of 65 aromatic chemicals representing several mechanisms of toxic action was considered in this study. These are listed in Table 1. The data set is chemically heterogeneous and includes phenols, anilines, nitrobenzenes, and benzaldehydes as well as compounds with more than one functional group on the benzene ring. The data set does not include chemicals which are anticipated to elicit their toxicity via target-specific mechanisms such as enzyme inhibition. All chemicals were purchased from Aldrich Chemical Co., Poole, Dorset, England, with chemical purity greater than 95%. Chemicals were not repurified prior to use.

Toxicological Assessment. Toxicity data ($\log(1/EC_{50})$) were determined in a biochemical assay utilizing the unicellular alga *C. vulgaris*. Algae in the logarithmic phase of their growth cycle were used. All toxicological analyses were performed in a buffer solution at a pH of 6.9 and temperature of between 25 °C and 30 °C. Assays were conducted following the protocol described by Worgan et al.[10] with a 15-minute static design. The disappearance of fluorescein diacetate (FDA) was accounted by spectrofluorimetrical measurement of fluorescein (the product of hydrolysis)[16] at an excitation wavelength of 465 nm and an emission wavelength to 515 nm. Range-finding experiments were performed in order to determine the highest and lowest concentrations required to produce a dose−response relationship ranging from 100% inhibition of enzyme activity to no observed toxicological effect. Blank buffer solution was utilized as a control, and the relative responses to it were used to generate the dose−response curve. The 50% effective concentration was estimated by Probit analysis using the SPSS software (SPSS Inc, Chicago, IL). The average EC_{50} (mM) was taken from at least three separate analyses.

Molecular Descriptors. A total of 102 molecular descriptors were calculated. A full list of descriptors is given in Table 2. Hydrophobicity was quantified by the logarithm of the 1-octanol/ water partition coefficient (log K_{ow}). The hydrophobicity values were measured or estimated (the measured value was preferred when available) by the ClogP for Windows ver. 1.0.0 software (BIOBYTE Corp., Claremont, CA). Quantum chemical descriptors were obtained in MOPAC93 (J. J. P. Stewart, (1993) Fujitsu Ltd., Windows 95/98/NT/2k adaptation and MO indices − J. Kaneti (1988−1994) MO-QC) using the AM1 Hamiltonian. Initially SMILES strings were converted into 3-D structures by the CORINA conformation analysis software, as implemented in the TSAR ver. 3.3 molecular spreadsheet (Accelrys Ltd., Oxford, England). The 3D structures underwent energy minimization utilizing full geometry optimization with the AM1 Hamiltonian in the VAMP module of TSAR. The optimized structures were exported to MOPAC93 for the calculation of quantum-mechanical indices. An additional 16 structural and topological descriptors were calculated using TSAR and 76 descriptors using QSARis ver. 1.1 (SciVision-Academic Press, San Diego, CA) software. Descriptors containing more than 60 zero values (∼92%) were excluded from the analysis.

QSAR Development. QSARs were developed using multiple linear regression (MLR) as implemented in MINITAB ver. 13.1 (Minitab Inc., State College, PA) and partial least squares (PLS) analysis, as implemented in SIMCA-P ver. 9.0 (Umetrics AB, Umeå, Sweden). Initially, an empirical approach for selection of variables was applied. This was to select descriptors known to be successful in the modeling of acute aquatic toxicity endpoints. Further development, using PLS analysis, employed stepwise elimination of variables keeping at each step only those descriptors with the highest standardized coefficients and low error associated with them. The quality of the MLR models was assessed by the regression coefficient (r^2), cross-validated regression coefficient (r^2_{CV}), standard error of the estimate (s) and the Fisher's criterion (F). The error of the coefficients is shown in parentheses in the QSARs. The statistical fit of the PLS model was assessed by the cumulative sum of squares (SS) of all descriptors participating in the model explained by all components (r^2(X)), cumulative SS of the toxicity explained by the model (r^2(Y)), cumulative q^2(Y) for all components, and the root-mean-square error of the fit (RMSEE), which corresponds to the standard error in the multiple regression analysis. The number of significant principal components was determined by the cross-validated q^2(Y). A negative contribution of a component to the cumulative q^2(Y) was used as an indication that the component is not significant. The PLS model is presented as the raw coefficients and those obtained after scaling and centring of the descriptors and scaling of toxicity. The former are given for transparency and to facilitate the prediction of toxicity of new compounds. The latter are useful for assessment of the variable importance and interpretation of the model.

Model Validation. Both MLR and PLS models were validated initially by the default cross-validation procedures implemented in the software used for their development. For the MLR models the MINITAB software allows automatic leave-one-out cross-validation. The default options for cross-validation in SIMCA software, including 7 cross-validation rounds and maximum iterations of 200, were accepted. In both cross-validation techniques each compound was omitted once and only once.

A second validation procedure simulating external prediction from complementary subsets[17] was introduced to enable the comparison between the models. The compounds were ordered by increasing toxicity and then subdivided into two equalized complementary subsets by taking alternate chemicals. Separate models were evaluated for *Group 1* and *Group 2*. Consequently, the toxicity of the compounds in *Group 1* was predicted by the model for *Group 2* and vice versa. Thus each chemical participated in the evaluation of a model and also had predicted a toxicity value. Finally, the predicted toxicity values for both groups were merged in a single column and correlated with observed toxicity.

Table 1. Chemical Abstracts Service (CAS) Number, Chemical Name, Acute Algal Toxicity and Calculated Descriptors for the Studied Aromatic Compounds

CAS	name	log(1/EC_{50}) (mM)	log K_{ow}	E_{lumo} (eV)	E_{homo} (eV)	Q_{Hmax} (au)	Q_{min} (au)	A_{max} (1/eV)	$^0\chi^v$
108-95-2	phenol	−1.46	1.47[d]	0.398	−9.114	0.2172	−0.2526	0.3013	3.834
62-53-3	aniline	−1.34	0.90[d]	0.639	−8.522	0.1824	−0.3273[b]	0.2944	3.964
100-66-3	anisole	−1.09	2.11[d]	0.483	−9.004	0.1481[a]	−0.2115	0.2999	4.795
367-12-4	2-fluorophenol	−1.08	1.71[d]	0.013	−9.271	0.2295	−0.2434	0.3092	4.135
348-54-9	2-fluoroaniline	−1.05	1.26[d]	0.266	−8.649	0.1944	−0.3222[b]	0.3021	4.265
108-39-4	3-cresol	−1.01	1.96[d]	0.396	−9.015	0.2167	−0.2534	0.3004	4.757
150-76-5	4-methoxyphenol	−0.97	1.34[d]	0.313	−8.648	0.2154	−0.2526	0.2981	5.165
95-55-6	2-hydroxyaniline	−0.91	0.62[d]	0.474	−8.357	0.2203	−0.3116	0.2961	4.334
90-05-1	2-methoxyphenol	−0.88	1.32[d]	0.392	−8.783	0.2346	−0.2502	0.3009	5.165
87-62-7	2,6-dimethylaniline	−0.87	1.84[d]	0.595	−8.364	0.1820	−0.3235[b]	0.2944	5.809
100-52-7	benzaldehyde	−0.81	1.47[d]	−0.435	−10.003	0.1542[a]	−0.2891	0.3169	4.372
95-48-7	2-cresol	−0.81	1.95[d]	0.396	−8.965	0.2163	−0.2521	0.3006	4.757
90-02-8	2-hydroxybenzaldehyde	−0.80	1.81[d]	−0.434	−9.500	0.2325	−0.2995	0.3176	4.742
98-95-3	nitrobenzene	−0.78	1.85[d]	−1.068	−10.562	0.1709[a]	−0.3585	0.3179	4.650
106-44-5	4-cresol	−0.66	1.94[d]	0.429	−8.881	0.2168	−0.2527	0.3001	4.757
95-65-8	3,4-dimethylphenol	−0.65	2.23[d]	0.436	−8.804	0.2164	−0.2529	0.2992	5.679
104-87-0	4-tolualdehyde	−0.65	1.99[c]	−0.430	−9.701	0.1538[a]	−0.2917	0.3158	5.295
94-71-3	2-ethoxyphenol	−0.62	1.85[d]	0.422	−8.742	0.2349	−0.2511	0.3001	5.872
24964-64-5	3-cyanobenzaldehyde	−0.57	1.18[d]	−0.917	−10.325	0.1612[a]	−0.2800	0.3254	5.242
99-08-1	3-nitrotoluene	−0.50	2.42[d]	−1.017	−10.200	0.1706[a]	−0.3595	0.3168	5.573
106-48-9	4-chlorophenol	−0.42	2.39[d]	0.095	−9.125	0.2200	−0.2483	0.3082	4.891
97-02-9	2,4-dinitroaniline	−0.36	1.72[c]	−1.475	−9.918	0.2728	−0.3919	0.3606	6.337
106-41-2	4-bromophenol	−0.35	2.59[d]	0.020	−9.189	0.2213	−0.2470	0.3104	5.721
106-40-1	4-bromoaniline	−0.33	2.26[d]	0.218	−8.632	0.1922	−0.3343[b]	0.3038	5.851
108-42-9	3-chloroaniline	−0.31	1.88[d]	0.263	−8.732	0.1888	−0.3300[b]	0.3119	5.021
2495-37-6	benzyl methacrylate	−0.21	2.53[d]	0.079	−9.697	0.1387[a]	−0.3608	0.3200	7.617
618-87-1	3,5-dinitroaniline	0.03	1.89[d]	−1.780	−9.885	0.2126	−0.3504	0.3384	6.337
89-98-5	2-chlorobenzaldehyde	0.06	2.33[d]	−0.683	−9.912	0.1608[a]	−0.2860	0.3280	5.429
540-38-5	4-iodophenol	0.16	2.91[d]	0.024	−9.243	0.2213	−0.3024	0.3109	6.292
4748-78-1	4-ethylbenzaldehyde	0.16	2.52[c]	−0.423	−9.747	0.1537[a]	−0.2910	0.3158	6.002
58-27-5	2-methyl-1,4-naphthoquinone	0.16	2.20[d]	−1.493	−10.217	0.1605[a]	−0.2746	0.3309	7.203
88-69-7	2-isopropylphenol	0.17	2.88[d]	0.408	−8.979	0.2200	−0.2575	0.3005	6.334
626-43-7	3,5-dichloroaniline	0.24	2.90[d]	−0.042	−8.933	0.1948	−0.3328[b]	0.3187	6.077
603-71-4	1,3,5-trimethyl-2-nitrobenzene	0.25	3.22[c]	−0.857	−9.890	0.1456[a]	−0.3640	0.3130	7.418
608-31-1	2,6-dichloroaniline	0.26	2.82[d]	−0.006	−8.720	0.2113	−0.3375[b]	0.3143	6.077
88-18-6	2-*tert*-butylphenol	0.29	3.27[d]	0.407	−8.989	0.2199	−0.2576	0.3003	7.041
95-50-1	1,2-dichlorobenzene	0.37	3.43[d]	−0.142	−9.602	0.1510[a]	−0.1214[b]	0.3190	5.577
99-65-0	1,3-dinitrobenzene	0.38	1.49[d]	−1.911	−11.428	0.2072[a]	−0.3477	0.3446	5.837
51-28-5	2,4-dinitrophenol	0.40	1.67[d]	−1.807	−10.807	0.2377	−0.3568	0.3647	6.207
100-25-4	1,4-dinitrobenzene	0.41	1.47[d]	−2.208	−11.344	0.1861[a]	−0.3420	0.3469	5.837
99-61-6	3-nitrobenzaldehyde	0.45	1.47[d]	−1.404	−10.883	0.1750[a]	−0.3558	0.3318	5.559
99-30-9	2,6-dichloro-4-nitroaniline	0.64	2.80[d]	−1.096	−9.362	0.2465	−0.3640	0.3449	7.264
121-14-2	2,4-dinitrotoluene	0.70	1.98[d]	−1.841	−11.030	0.2080[a]	−0.3525	0.3448	6.759
3531-19-9	6-chloro-2,4-dinitroaniline	0.80	2.46[c]	−1.667	−9.960	0.2760	−0.3886	0.3704	7.393
99-28-5	2,6-dibromo-4-nitrophenol	0.81	3.57[d]	−1.452	−10.216	0.2454	−0.3537	0.3540	8.794
89-61-2	2,5-dichloronitrobenzene	0.97	3.03[d]	−1.296	−10.218	0.1813[a]	−0.3455	0.3485	6.764
94-62-2	piperine	0.97	2.70[c]	−0.767	−8.695	0.1572[a]	−0.3629	0.3135	12.716
939-97-9	4-*tert*-butylbenzaldehyde	1.00	3.32[c]	−0.391	−9.717	0.2245[a]	−0.2918	0.3153	7.795
634-93-5	2,4,6-trichloroaniline	1.11	3.69[d]	−0.240	−8.761	0.2176	−0.3418[b]	0.3208	7.134
83-42-1	2-chloro-6-nitrotoluene	1.17	3.09[d]	−1.219	−10.126	0.1758[a]	−0.3614	0.3294	6.630
5388-62-5	4-chloro-2,6-dinitroaniline	1.19	2.46[c]	−1.895	−9.778	0.2727	−0.3838	0.3722	7.393
528-29-0	1,2-dinitrobenzene	1.23	1.69[d]	−1.840	−11.338	0.1795[a]	−0.3374	0.3514	5.837
100-00-5	1-chloro-4-nitrobenzene	1.25	2.39[d]	−1.344	−10.475	0.1763[a]	−0.3553	0.3401	5.707
128-37-0	2,6-di-*tert*-butyl-4-methylphenol	1.45	5.89[c]	0.464	−8.732	0.2186	−0.2600	0.2984	11.171
3481-20-7	2,3,5,6-tetrachloroaniline	1.48	4.47[d]	−0.560	−8.994	0.2221	−0.3452[b]	0.3305	8.190
609-89-2	2,4-dichloro-6-nitrophenol	1.50	3.07[c]	−1.431	−9.957	0.2400	−0.3668	0.3508	7.133
83-38-5	2,6-dichlorobenzaldehyde	1.50	3.08[c]	−0.473	−9.905	0.2304[a]	−0.2472	0.3296	6.485
96-76-4	2,4-di-*tert*-butylphenol	1.60	4.36[d]	0.431	−8.926	0.2192	−0.2598	0.2995	10.248
87-86-5	pentachlorophenol	1.69	5.12[d]	−0.978	−9.574	0.2386	−0.2220	0.3432	9.117
89-69-0	1,2,4-trichloro-5-nitrobenzene	1.88	3.47[d]	−1.536	−10.305	0.1853[a]	−0.3437	0.3543	7.820
6284-83-9	1,3,5-trichloro-2,4-dinitrobenzene	1.89	2.97[c]	−2.037	−11.158	0.1880[a]	−0.3134	0.3852	9.006
1689-82-3	phenylazophenol	2.16	3.96[c]	−0.768	−8.849	0.2212	−0.2467	0.3099	8.038
not found	4-(dibutylamino)benzaldehyde	2.18	5.06[c]	−0.097	−8.391	0.2314[a]	−0.3068	0.3105	10.985
117-18-0	2,3,5,6-tetrachloronitrobenzene	2.34	4.38[d]	−1.419	−10.224	0.1786[a]	−0.3204	0.3595	8.877
608-71-9	pentabromophenol	3.10	4.85[c]	−1.193	−9.684	0.2415	−0.2413	0.3457	13.267
minimum		−1.46	0.62	−2.21	−11.43	0.14	−0.39	0.29	3.83
maximum		3.10	5.89	0.64	−8.36	0.28	−0.12	0.39	13.27

[a] Molecules which cannot participate in hydrogen bonding as donors according to TSAR software. [b] Molecules which cannot participate in hydrogen bonding as acceptors according to TSAR software. [c] Calculated log K_{ow}. [d] Measured log K_{ow}.

Table 2. Descriptors Calculated for the Compounds Listed in Table 1

software	descriptors calculated
Clog P	logarithm of the octanol−water partition coefficient (log K_{ow})
MOPAC	energy of the highest occupied molecular orbital (E_{homo}), energy of the lowest unoccupied molecular orbital (E_{lumo}), electronegativity ($EN = (E_{homo} + E_{lumo})/2$), absolute hardness ($AH = (E_{homo} - E_{lumo})/2$), maximum acceptor ($A_{max}$) and donor ($D_{max}$) superdelocalizability, maximum positive partial charge (Q_{max}), maximum positive partial charge on a hydrogen atom (Q_{Hmax}), maximum negative partial charge (Q_{min})
TSAR	dipole moment (Dip), polarizability (Pol), molecular volume (MV), molecular surface area (MSA), inertia moments (IM_1, IM_2 and IM_3 size, and IM_1, IM_2 and IM_3 length), Wiener and Balaban topological indices, ellipsoidal volume (MV_{elipp}), molecular refraction (MR), number of hydrogen-bond donors (N_{Hdon}) and acceptors (N_{Hacc})
QSARis	Kier simple, valence and delta molecular connectivity indices ($^0\chi$, $^1\chi$, $^2\chi$, $^3\chi_p$, $^4\chi_p$, $^5\chi_p$, $^6\chi_p$, $^7\chi_p$, $^8\chi_p$, $^9\chi_p$, $^{10}\chi_p$, $^3\chi_c$, $^4\chi_{pc}$, $^6\chi_{pc}$, $^0\chi^v$, $^1\chi^v$, $^2\chi^v$, $^3\chi^v_p$, $^4\chi^v_p$, $^5\chi^v_p$, $^6\chi^v_p$, $^7\chi^v_p$, $^8\chi^v_p$, $^9\chi^v_p$, $^{10}\chi^v_p$, $^3\chi^v_c$, $^4\chi^v_{pc}$, $^6\chi^v_{pc}$, $\Delta^0\chi$, $\Delta^1\chi$, $\Delta^2\chi$, $\Delta^3\chi_p$, $\Delta^4\chi_p$, $\Delta^5\chi_p$, $\Delta^6\chi_p$, $\Delta^7\chi_p$, $\Delta^8\chi_p$, $\Delta^9\chi_p$, $\Delta^{10}\chi_p$, $\Delta^0\chi^v$, $\Delta^1\chi^v$, $\Delta^2\chi^v$, $\Delta^3\chi^v_p$, $\Delta^4\chi^v_p$, $\Delta^5\chi^v_p$, $\Delta^6\chi^v_p$, $\Delta^7\chi^v_p$, $\Delta^8\chi^v_p$, $\Delta^9\chi^v_p$, $\Delta^{10}\chi^v_p$), E-state indices ($SsCH_3$, $SssCH_2$, SdsCH, SaaCH, SaasC, $SsNH_2$, SddsN, SsOH, SdO, SssO, SsCl, SHHBd, SHBa), kappa and kappa alpha shape indices (^{0}k, ^{1}ka, ^{2}k, ^{3}k, ^{1}ka, ^{2}ka and ^{3}ka), the sum of absolute values of the charges on each atom (ABSQ), the sum of absolute values of the charges on the nitrogen and oxygen atoms in a molecule ($ABSQ_{on}$), ovality, dipolar descriptors (Q_s, Q_v, Q_{sv})

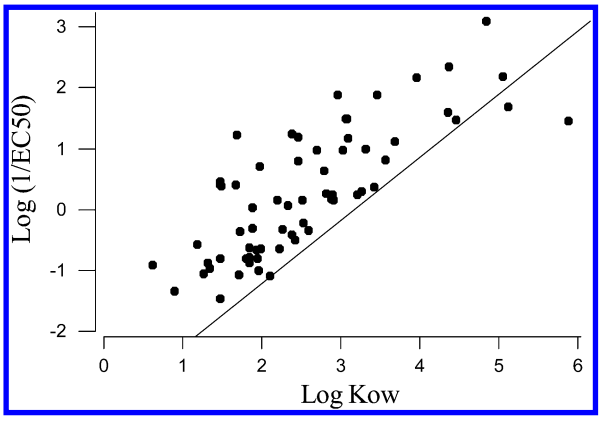

Figure 1. Plot of toxicity ($\log(1/EC_{50})$) against hydrophobicity ($\log K_{ow}$) for all compounds considered in the study. Also shown is the QSAR ($\log(1/EC_{50}) = 1.04\ (0.07)\ \log K_{ow} - 3.28\ (0.11)$; $n = 10$, $r^2 = 0.96$) for nonpolar narcosis (baseline toxicity) developed by Worgan et al.[10]

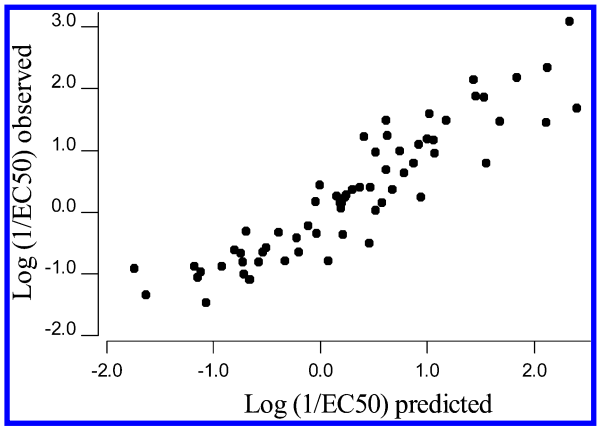

Figure 2. Plot of observed algal toxicity against that predicted by the 2-descriptor MLR model (eq 2).

RESULTS

Toxicity data to *C. vulgaris* for a total of 65 aromatic compounds are presented in Table 1. Compounds exhibited a wide range of algal toxicity (from −1.46 to 3.10 on a (mM) logarithmic scale). Hydrophobicity, quantified by log K_{ow}, ranged from 0.62 to 5.89, and the reactivity, quantified by E_{lumo}, ranged from −2.21 to 0.64. The minimum and maximum values of selected descriptors are listed in Table 1.

In an initial attempt to develop QSARs, an empirical approach was applied to the selection of descriptors. This was to select descriptors that are known to model the two of the main features of toxicity, namely penetration through the cell membrane (hydrophobicity: log K_{ow}) and reactivity (electrophilicity: E_{lumo}) inside the cell.[18]

Analysis of relationship between algal toxicity and hydrophobicity resulted in a statistically significant equation, however, with a relatively low coefficient of determination:

$$\log(1/EC_{50}) = 0.757\ (0.076)\ \log K_{ow} - 1.653\ (0.211)$$

$$n = 65,\ r^2 = 0.613,\ r^2_{CV} = 0.581,\ s = 0.660,\ F = 100 \qquad (1)$$

A plot of $\log(1/EC_{50})$ versus log K_{ow} for all compounds is presented in Figure 1. Figure 1 illustrates also the QSAR for baseline toxicity in the same algal assay as described by Worgan et al.[10] As it can be seen from Figure 1, almost all aromatic compounds (with exception of 2,6-di-*tert*-butyl-4-methylphenol and pentachlorophenol) have toxicity greater than those predicted by the baseline. This result is in agreement with the observations made by Veith and Broderius,[19] Cronin and Schultz,[20] and Bearden and Schultz[21] that the aromatic compounds acting by a narcosis mechanism of action (polar narcosis) have toxicity in excess to that of the aliphatic compounds acting by the same mechanism of toxic action (nonpolar narcosis).

The statistical criteria of the hydrophobicity-dependent QSAR (eq 1) were considerably improved by including the reactivity term E_{lumo}:

$$\log(1/EC_{50}) = 0.731\ (0.049)\ \log K_{ow} - 0.590\ (0.063)\ E_{lumo} - 1.914\ (0.140)$$

$$n = 65,\ r^2 = 0.839,\ r^2_{CV} = 0.819,\ s = 0.429,\ F = 161 \qquad (2)$$

There were no statistical outliers to this relationship. The plot of observed algal toxicity against that predicted by eq 2 is shown in Figure 2.

It is should also be noted that E_{lumo} was not the only quantum chemical descriptor that, in combination with hydrophobicity, improved the QSAR for algal toxicity. Statistically significant models were also obtained by including either maximum acceptor superdelocalizability (A_{max}) or

Table 3. Coefficients in the 5-Descriptor PLS Model ($n = 65$, $r^2(X) = 0.662$, $r^2((Y) = 0.829$, $q^2((Y) = 0.788$, RMSEE = 0.442)

descriptor	raw coefficients	CS[a] coefficient	error of the CS coefficient
log K_{ow}	0.729	0.754	0.121
E_{lumo}	−0.408	−0.329	0.071
E_{homo}	−0.162	−0.125	0.093
Q_{Hmax}	4.346	0.137	0.192
Q_{min}	−0.633	−0.032	0.152
const	−4.433	0.276	

[a] Coefficient, obtained after scaling and centering of the variables.

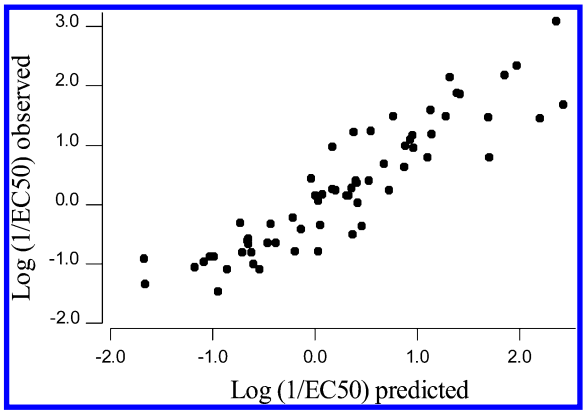

Figure 3. Plot of observed algal toxicity against that predicted by the 5-descriptor PLS model (Table 3).

electronegativity (EN) in addition to log K_{ow} ($r^2 = 0.822$ and $r^2 = 0.815$, respectively). These similar results can be explained by the high correlation between these quantum chemical descriptors (E_{lumo} and A_{max} r = −0.909, E_{lumo} and EN $r = 0.967$, A_{max} and EN r = −0.862). However, relatively low regression coefficients were obtained using E_{homo}, AH, D_{max} and charge-related indices calculated by MOPAC93 in combination with log K_{ow}.

An attempt to further improve eq 2 was performed by the application of an approach used by Ramos et al.[22,23] to model acute aquatic toxicity. It relies on the assumption that the hydrophobicity expressed as log K_{ow} and hydrogen bond capabilities are important factors which determine the toxicity. To account for possible hydrogen bonding, four descriptors (E_{lumo}, E_{homo}, Q_{Hmax} and Q_{min} as defined in Table 2) were proposed. Ramos et al.[22,23] believed that E_{lumo} and E_{homo} describe the covalent contribution to hydrogen bonding donation and acceptance respectively, while Q_{Hmax} and Q_{min} account for the ionic contribution. This approach requires PLS analysis as a statistical technique due to the expected intercorrelation between the descriptors. The result of the PLS modeling of the algal toxicity with these five descriptors is shown in Table 3. The plot of observed algal toxicity against that predicted by the 5-descriptor PLS model is shown in Figure 3. The model was derived after the extraction of two significant principal components (the cumulative $q^2(Y)$ was 0.714 and 0.788 after the extraction of the first and second principal component, respectively).

As it can be seen from Table 3, there was no improvement of the statistical fit and predictivity of the 5-descriptor PLS model as compared to eq 2. In addition, two of the descriptors (Q_{Hmax} and Q_{min}) had error higher than the coefficient of the descriptor itself, and the error of a third descriptor (E_{homo}) was comparable to the value of its coefficient. The apparent lack of significance of some descriptors in the 5-descriptor PLS model motivated a search of other descriptors which can predict the algal toxicity more successfully.

Table 4. Coefficients in the 4-Descriptor PLS Model ($n = 65$, $r^2(X) = 0.930$, $r^2((Y) = 0.858$, $q^2((Y) = 0.843$, RMSEE = 0.403)

descriptor	raw coefficients	CS[a] coefficient	error of the CS coefficient
log K_{ow}	0.404	0.418	0.071
E_{lumo}	−0.233	−0.188	0.051
A_{max}	9.842	0.212	0.051
$^0\chi^v$	0.204	0.384	0.074
const	−5.399	0.276	

[a] Coefficient, obtained after scaling and centering of the variables.

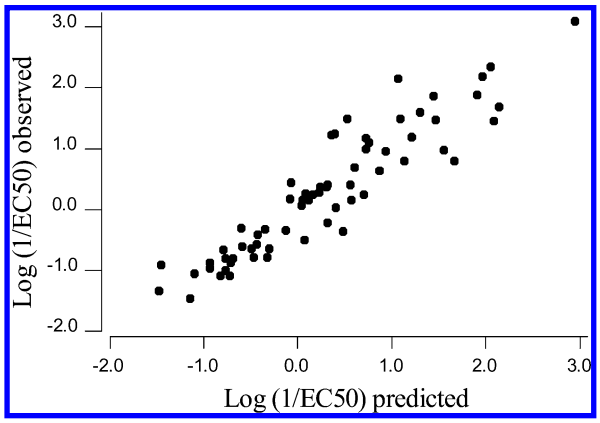

Figure 4. Plot of observed algal toxicity against that predicted by the 4-descriptor PLS model (Table 4).

Table 5. Intercorrelation Coefficients (r) between the Variables Used in Eq 2, Table 3 and Table 4

	log K_{ow}	E_{lumo}	E_{homo}	Q_{Hmax}	Q_{min}	A_{max}
E_{lumo}	−0.055					
E_{homo}	0.116	0.863				
Q_{Hmax}	0.174	−0.025	0.203			
Q_{min}	0.143	0.556	0.372	−0.029		
A_{max}	0.152	−0.909	−0.753	0.197	−0.509	
$^0\chi^v$	0.797	−0.272	−0.017	0.170	−0.103	0.316

An iterative approach for the selection of variables, from the calculated set of 102 descriptors was applied. A similar approach was used to model acute toxicity of aliphatic compounds to the ciliate *Tetrahymena pyriformis*, and a significant reduction in the number of variables was obtained.[24] A stepwise elimination of descriptors with low importance and significance was performed until only the most important (with highest absolute values of the coefficients) and most significant (highest coefficient/error ratio) descriptors remained. Every effort was made to keep the number of descriptors as low as possible. The results of this procedure are listed in Table 4. Four descriptors were included in a final PLS model, which was created from two principal components (the cumulative $q^2(Y)$ was 0.837 and 0.843 after the extraction of the first and second principal component respectively). The plot of observed algal toxicity against that predicted by the 4-descriptor PLS model is shown in Figure 4. Table 5 shows the intercorrelation coefficients between the variables included in the PLS modeling.

The predictivity of the three models (eq 2, Table 3 and Table 4) was assessed in a uniform validation exercise. It was performed on complementary subsets (*Group 1* and

Table 6. Statistics of the Relationship between Observed and Predicted Algal Toxicity ($\log(1/EC_{50})_{Obs} = a \log(1/EC_{50})_{Pred} + b$) in the Leave-50%-Out Validation Procedure

	MLR 2-descriptor model	PLS 5-descriptor model	PLS 4-descriptor model
r^2_{pred}	0.837	0.820	0.848
s_{pred}	0.428	0.450	0.413
a	1.000	1.012	0.966
b	0.002	−0.014	0.005

Group 2), formed as 50% of the compounds in the study. *Group 1* contained 33 compounds (starting with the first compound listed in Table 1), and *Group 2* had 32 due to the odd number of the chemicals in the series. After each group was used for model development, the toxicity of the chemicals from the complementary group was predicted. Finally, the predicted toxicity of all 65 compounds was correlated with the observed toxicity in an equation: $\log(1/EC_{50})_{obs} = a \log(1/EC_{50})_{pred} + b$. The coefficient of determination (r^2_{pred}), the standard error of prediction (s_{pred}), the slope (a) and the intercept (b) are shown in Table 6. All the three models had a slope near to one and intercept near to zero, which is considered an important prerequisite for good predictive ability.[25] However, the r^2_{pred} and s_{pred} confirm the superior predictivity of the 4-descriptor PLS models (Table 4) compared to other two models.

DISCUSSION

The aromatic compounds form an important group of environmental pollutants for which toxicity data are required. This study has provided a novel set of algal toxicity values and a number of computational models to predict toxicity from structure alone. They were developed on a relatively large heterogeneous set of aromatic compounds, for which several different mechanisms of toxic action are anticipated, without a priori identification of the exact molecular mechanism.

Although there are not yet formal, precise guidelines for the development of ecotoxicological QSARs, a number of essential and desirable characteristics have been identified by Schultz and Cronin.[15] Essential and desirable characteristics include reliable ecotoxicological data; high quality, interpretable and reproducible descriptors, of a number and type consistent with the endpoint being modeled; and the use of a statistical procedure allowing the development of a rigorous and transparent mathematical model. In addition, the authors emphasize the need of validation and identification of the chemical domain of the model, since the QSARs are only valid within their respective domains.[8]

The algal toxicity data utilized in this study meet most of the accepted criteria as being "high quality" toxicity data.[15,26,27] These data were measured in the same laboratory, by a single worker utilizing a consistent protocol. It should be noted, however, that toxicity data were not determined to GLP standards, nor is the test protocol "standardized" to, for instance, OECD criteria. Despite these factors, there was relatively low experimental error associated with the data (in the region of 0.2−0.5 log units, depending on chemical), and the toxicity data were validated by interspecies correlation with toxicity data to other aquatic organisms.[10,26]

The applicability domain of the data set is well characterized in terms of physicochemical properties and structural classes considered. The latter include phenols, anilines, benzaldehydes, and nitrobenzenes as well as alkyl-substituted phenols, halogenated phenols and anilines, nitro-substituted phenols and anilines and halogenated nitrobenzenes. The diversity in chemical structure resulted in a large range of properties, which determine the physicochemical domain of the models. The physicochemical domain is defined by the minimum and maximum values of the descriptors participating in each individual QSAR, and no prediction should be made for a compound with a value outside of the training set of the model.

The reduction in the number of variables used in a QSAR is also an important issue. Inclusion of a large number of descriptors might improve the statistical fit of the model but leads to spurious, unstable or invalid models.[27] The ratio of observations to variables should be as high as possible, and at least 1:5[28] although those with a more conservative nature may prefer a ratio of at least 1:10.[29] From a mechanistic point of view, inclusion of a variable into a QSAR suggests that the variable is related to how the chemistry influences biological activity, and since there is not an infinite number of controlling factors in a QSAR, it would be expected to have no more variables than factors controlling the biological activity.[27]

Following the recommendations of Cronin and Schultz,[27] an attempt to use mechanistically sound descriptors was made in the modeling process. Two mechanistic hypotheses were tested. The first one is based upon the premise that the acute aquatic toxicity of nonspecifically acting compounds is controlled by hydrophobicity and electrophilicity[14,30−34] and the second one that hydophobicity and hydrogen bonding capacity determine the magnitude of toxicity.[22,23] Subsequently, descriptors reflecting these factors were selected for model development. While undoubtedly log K_{ow} is the most popular descriptor used for accounting of hydrophobicity, there is large variability in the type of descriptors used to account for electrophilicity and hydrogen-bonding ability. Despite that Cronin et al.[35] demonstrated the superior performance of A_{max} for modeling the acute toxicity of aromatic compounds compared to E_{lumo}, the latter was more successful in this study for development of the 2-descriptor MLR model. E_{lumo} was also found to be a superior descriptor of reactivity as compared to EN. Four descriptors (E_{lumo}, E_{homo}, Q_{Hmax}, Q_{min}) accounting for the ability of the molecules to donate or accept hydrogen in hydrogen-bond formation were selected directly as proposed by Ramos et al.[22,23] for development of a model using PLS analysis.

The comparison between the 2-descriptor MLR model (eq 2) and the 5-descriptor PLS model (Table 3) revealed that the former has better statistical fit and lower standard error of estimation than the latter. In addition, both descriptors in eq 2 were highly significant, while the significance of three of the descriptors in Table 3 was doubtful. There could be a number of reasons why E_{homo}, Q_{Hmax}, and Q_{min} were not successful in the modeling of the acute algal toxicity. One possible reason could be the fact that the descriptors used to represent the hydrogen-bonding capacity are not the best choice for that purpose. Another reason for the low performance of E_{homo}, Q_{Hmax}, and Q_{min} in Table 3 could be the chemical structure of the compounds used in this study. As

it can be seen from Table 1, 27 compounds are not able to donate, and 10 compounds are not able to accept, hydrogen in a hydrogen bond according to the structural rules implemented in the TSAR ver. 3.3 software. These include the following atom types for hydrogen-bond donors: O−H, N−H and S−H, and for hydrogen-bond acceptors: −O−, C═O, −C≡N, −N═, >N− excluding N−C (sp^2), −S−H and C═S. Subsequently, there is no firm theoretical ground for the application of an hypothesis for the importance of hydrogen bonding in case of a heterogeneous aromatic data set unless it is not restricted to certain chemical classes of molecules, able to donate or accept a hydrogen bond.

The stepwise elimination procedure for the selection of variables to model acute algal toxicity (Table 4) resulted in a 4-descriptor model with better statistical fit and lower error than the model presented in Table 3. Additionally, all independent variables in Table 4 are significant with coefficients much greater than the corresponding error. The descriptors, selected in additional to log K_{ow} and E_{lumo}, confirm the meaning of the first two descriptors used for development of eq 2. Log K_{ow} and ${}^0\chi^v$ reflect the readiness of molecules to penetrate through biological membranes. ${}^0\chi^v$ is a topological index and is believed to account for molecular size[36,37] with a correction for the presence of π- and lone pair electrons. According to Hall and Kier[38] the π- and lone pair electrons are more exposed and more interactive than the σ-electrons, and hence more important to noncovalent intermolecular interactions such as those occurring with molecular transport through membranes. Both E_{lumo} and A_{max} account for electrophilicity. E_{lumo} reflects the reactivity of the whole molecule and depends mainly on the number and type of heteroatoms as well as on the type of the polarized group, while A_{max} depends more on the arrangement of the heteroatoms and characterizes the electron-acceptor ability of a molecule at particular atom.[32]

The simulated external validation using complementary subsets allowed comparative assessment of the stability and predictivity of the models developed in this study. As shown in Table 6, the 2-descriptor MLR model (eq 2) and the 4-descriptor PLS model (Table 4) demonstrated stability and good predictivity, comparable with the statistical fit of the original models. The 5-descriptor PLS model (Table 3) showed somewhat lower stability and predictivity than the other two models which was expected bearing in mind the insignificance of three of the descriptors. It is interesting to note that the leave-50%-out validation procedure appeared less demanding than the default PLS validation procedure implemented in SIMCA ver. 9.0 software and the poorer the model the higher the difference between r^2_{pred} and q^2(Y).

CONCLUSION

This study utilized a novel set of algal toxicity values for the development of computational models to predict toxicity. Two hypotheses relating to the factors controlling acute algal toxicity in a relatively large, heterogeneous data set were tested. The hypothesis based on hydrophobicity and hydrogen bonding was not confirmed, probably due to the relatively large structural domain of the studied data set. The results of the QSAR analysis indicated that for this diverse data set the two main factors controlling toxicity are hydrophobicity and electrophilicity. Statistically significant, transparent and interpretable models were developed using MLR and PLS analyses. The PLS hydrophobicity/electrophilicity model demonstrated slightly better statistical fit and predictivity. However, the choice of a model for predicting algal toxicity depends on the availability of the descriptors and the accuracy of the prediction required by the user.

ACKNOWLEDGMENT

This work was supported in part by the European Union IMAGETOX Research Training Network (HPRN-CT- 1999-00015).

REFERENCES AND NOTES

(1) Anon, 2001. White Paper on a Strategy for a Future Chemical Policy. Brussels: Commission of the European Communities. Web site: http://europa.eu.int/comm/ environment/chemicals/whitepaper.htm (Last update: 17/06/2003)

(2) Worth, A. P.; Balls, M. Background. *ATLA* **2002**, *30*, Supplement 1, 1−10.

(3) Combes, R.; Balls, M.; Bansil, L.; Barrat, M.; Bell, D.; Botham, P.; Broadhead, C.; Clothier, R.; George, E.; Fentem, J.; Jackson, M.; Indans, I.; Loizou, G.; Navaratnam, V.; Pentreath, V.; Phillips, B.; Stemplewski, H.; Stewart, J. An assessment of progress in the use of alternatives in toxicity testing since the publication of the report of the Second FRAME Toxicity Committee 1991. *ATLA* **2002**, *30*, 365−406.

(4) Cronin, M. T. D. The current status and future applicability of quantitative structure−activity relationships (QSARs) in predicting toxicity. *ATLA* **2002**, *30*, Supplement 2, 81−84.

(5) Cronin, M. T. D.; Walker, J. D.; Jaworska, J. S.; Comber, M. H. I.; Watts, C. D.; Worth, A. P. Use of QSARs in international decision-making frameworks to predict ecologic effects and environmental fate of chemical substances. *Environ. Health Persp.* **2003**, *111*, 1376−1390.

(6) Cronin, M. T. D.; Jaworska, J. S.; Walker, J. D.; Comber, M. H. I.; Watts, C. D.; Worth, A. P. Use of QSARs in international decision-making frameworks to predict health effects of chemical substances. *Environ. Health Persp.* **2003**, *111*, 1391−1401.

(7) Walker, J. D. Applications of QSARs in toxicology: a US Government perspective. *J. Mol. Struct. (THEOCHEM)* **2003**, *622*, 167−184.

(8) Walker, J. D.; Carlsen, L.; Jaworska, J. Improving opportunities for regulatory acceptance of QSARs: the importance of model domain, uncertainty, validity and predictability. *QSAR Comb. Sci.* **2003**, *22*, 346−350.

(9) Schultz, T. W. TETRATOX: The *Tetrahymena pyriformis* population growth impairment endpoint. A surrogate for fish lethality. *Toxicol. Methods* **1997**, *7*, 289−309.

(10) Worgan, A. D. P.; Dearden, J. C.; Edwards, R.; Netzeva, T. I.; Cronin, M. T. D. Evaluation of a novel short-term algal toxicity assay by the development of QSARs and inter-species relationships for narcotic chemicals. *QSAR Comb. Sci.* **2003**, *22*, 204−209.

(11) Worth, A. P. ECVAM's activities on computer modelling and integrated testing. *ATLA* **2002**, *30*, Supplement 2, 133−137.

(12) Worth, A. P.; Cronin, M. T. D.; Van Leeuwen, C. J. A framework for promoting acceptance and regulatory use of (quantitative) structure−activity relationships. In: *Predicting Chemical Toxicity and Fate*; Cronin, M. T. D., Livingstone, D. J., Eds.; Taylor & Francis: London; 2004; in press.

(13) Schultz, T. W. Structure-toxicity relationships for benzenes evaluated with *Tetrahymena pyriformis*. *Chem. Res. Toxicol.* **1999**, *12*, 1262−1267.

(14) Cronin, M. T. D.; Schultz, T. W. Development of quantitative structure−activity relationships for the toxicity of aromatic compounds to *Tetrahymena pyriformis*: comparative assessment of the methodologies. *Chem. Res. Toxicol.* **2001**, *14*, 1284−1295.

(15) Schultz, T. W.; Cronin, M. T. D. Essential and desirable characteristics of ecotoxicity quantitative structure-toxicity relationships. *Environ. Toxicol. Chem.* **2003**, *22*, 599−607.

(16) Leszczynska, M.; Oleszkiewic, J. A. Application of fluoresceine diacetate hydrolysis as an acute toxicity test. *Environ. Technol.* **1996**, *17*, 79−85.

(17) Aptula, A. O.; Netzeva, T. I.; Valkova, I. V.; Cronin, M. T. D.; Schultz, T. W.; Kühne R.; Schüürmann, G. Multivariate discrimination between modes of toxic action of phenols. *Quant. Struct.-Act. Relat.* **2002**, *21*, 12−22.

(18) McFarland, J. W. On the parabolic relationship between drug potency and hydrophobicity. *J. Med. Chem.* **1970**, *13*, 1092−1196.
(19) Veith, G. D.; Broderius, S. J. Rules for distinguishing toxicants that cause type I and type II narcosis syndrome. *Environ. Health Persp.* **1990**, *87*, 207−211.
(20) Cronin, M. T. D.; Schultz, T. W. Validation of *Vibrio fisheri* acute toxicity data: mechanism of action-based QSARs for nonpolar narcotics and polar narcotic phenols. *Sci. Total Environ.* **1997**, *204*, 75−88.
(21) Bearden, A. P.; Schultz, T. W. Comparison of *Tetrahymena* and *Pimephales* toxicity based on mechanism of action. *SAR QSAR Environ. Res.* **1998**, *9*, 127−153.
(22) 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.
(23) Ramos, E. U.; Vaes, W. H. J.; Mayer, P.; Hermens, J. L. M. Algal growth inhibition of *Chlorella pyrenoidosa* by polar narcotic pollutants: toxic cell concentrations and QSAR modelling. *Aquat. Toxicol.* **1999**, *46*, 1−10.
(24) Netzeva, T. I.; Schultz, T. W.; Aptula, A. O.; Cronin, M. T. D. Partial least squares modelling of the acute toxicity of aliphatic compounds to *Tetrahymena pyriformis*. *SAR QSAR Environ. Res.* **2003**, *14*, 265−283.
(25) Golbraikh, A.; Tropsha, A. Beware of q^2! *J. Mol. Graphics Mod.* **2002**, *20*, 269−276.
(26) Netzeva, T. I.; Worgan, A. D. P.; Dearden, J. C.; Edwards, R.; Cronin M. T. D. Toxicological evaluation and QSAR modelling of aromatic amines to *Chlorella vulgaris*. *Aquat. Toxicol.* **2003**, submitted for publication.
(27) Cronin, M. T. D.; Schultz, T. W. Pitfalls in QSAR. *J. Mol. Struct. (THEOCHEM)* **2003**, *622*, 39−51.
(28) Topliss, J. G.; Costello, R. J. Chance correlations in structure−activity studies using multiple regression analysis. *J. Med. Chem.* **1972**, *15*, 1066−1068.
(29) Schultz, T. W.; Netzeva, T. I.; Cronin, M. T. D. Selection of data sets for QSARs: analyses of *Tetrahymena* toxicity from aromatic compounds. *SAR QSAR Environ. Res.* **2003**, *14*, 59−81.
(30) Mekenyan, O. G.; Veith G. D. A relationship between descriptors for hydrophobicity and soft electrophilicity in predictive toxicology. *SAR QSAR Environ. Res.* **1993**, *1*, 335−344.
(31) 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.
(32) Cronin, M. T. D.; Aptula, A. O.; Duffy, J. C.; Netzeva, T. I.; Rowe, P. H.; Valkova I. V.; Schultz, T. W. Comparative assessment of methods to develop QSARs for the prediction of the toxicity of phenols to *Tetrahymena pyriformis*. *Chemosphere* **2002**, *49*, 1201−1221.
(33) Cronin, M. T. D.; Dearden, J. C.; Duffy, J. C.; Edwards, R.; Manga, N.; Worth, A. P.; Worgan, A. D. P. The importance of hydrophobicity and electrophilicity descriptors in mechanistically based QSARs for toxicological endpoints. *SAR QSAR Environ. Res.* **2002**, *13*, 167−176.
(34) Schultz, T. W.; Cronin, M. T. D.; Netzeva, T. I.; Aptula, A. O. Structure-toxicity relationships for aliphatic chemicals evaluated with *Tetrahymena pyriformis*. *Chem Res. Toxicol.* **2002**, *15*, 1602−1609.
(35) Cronin, M. T. D.; Manga, N.; Seward, J. R.; Sinks, G. D.; Schultz, T. W. Parametrisation of electrophilicity for the prediction of the toxicity of aromatic compounds. *Chem. Res. Toxicol.* **2001**, *14*, 1498−1505.
(36) Dearden, J. C.; Bradburne, S. J. A.; Cronin, M. T. D.; Solanki, P. The physical significance of molecular connectivity. In: *Proceedings of the Third International Workshop on Quantitative Structure−Activity Relationships in Environmental Chemistry*; Turner, J. E., England, M. W., Schultz, T. W., Kwaak, N. J., Eds.; Oak Ridge: Knoxville, TN, 1988; pp 43−50.
(37) Netzeva, T. I. Whole molecule, and atom based, topological descriptors. In: *Predicting Chemical Toxicity and Fate*; Cronin, M. T. D., Livingstone, D. J., Eds.; Taylor & Francis: London; 2004; in press.
(38) Hall, L. H.; Kier, L. B. Issues in representation of molecular structure. The development of molecular connectivity. *J. Mol. Graphics Mod.* **2001**, *20*, 4−18.

CI034195G