EUROPEAN COMMISSION
DIRECTORATE-GENERAL
Joint Research Centre

Institute for Health and Consumer Protection
Toxicology and Chemical Substances Unit
I-21020 Ispra (VA) Italy

Comparative Review of QSARS for Acute Toxicity

Iglika Lessigiarska, Andrew P. Worth, Tatiana I. Netzeva

2005

EUR 21559 EN

A great deal of additional information on the
European Union is available on the Internet.
It can be accessed through the Europa server
(http://europa.eu.int)

EUR 21559 EN

Printed in Italy

CONTENTS

LIST OF ABBREVIATIONS

CoMFA	comparative molecular field analysis
CoMSIA	comparative molecular shape indices analysis
E-state	electrotopological state
EC_{50}	concentration of a compound that causes 50% effect relative to a control
F	Fisher statistic
GETAWAY	geometry, topology and atomic weights assembly descriptors
H-bonding	hydrogen bonding
IC_{50}	concentration of a compound that causes 50% inhibition of cell growth
HQSAR	holographic quantitative structure-activity relationship
LC_{50}	concentration of a compound that causes 50% lethality of the animals in a test batch
LNO	leave-n-out statistical procedure
LOO	leave-one-out statistical procedure
n	number of data points used to develop a QSAR model
OCWLGI	optimisation of correlation weights of local graph invariants
p	level of statistical significance
pEC_{50}	negative logarithm of EC_{50}
PCA	principal component analysis
PLS	partial least squares
PRESS	predicted residual sum of squares
R^2	square of the regression coefficient of determination
R_A^2	square of the coefficient of determination adjusted for the degrees of freedom
R^2_{cv}	square of the cross-validated coefficient of determination from a leave-one-out statistical procedure
R_{LNO}	square of the cross-validated coefficient of determination from a leave-one-out statistical procedure
R^2_{pred}	analogue of the square of the regression coefficient of determination, obtained on the basis of predictions for a training set
QSAR	quantitative structure-activity relationship
s	standard error of estimate
s_{cv}	cross-validated standard error of prediction from a leave-one-out statistical

	procedure
SAR	structure-activity relationship
TI	topological indices
TLSER	Theoretical linear solvation energy relationship
WHIM	Weighted holistic invariant mole

PART I. QSARS FOR TOXICITY TO AQUATIC ORGANISMS

Abstract

A large number of QSARs for acute toxicity have been developed and published. Toxicity to a wide variety of biological systems was investigated, including bacteria, algae, *Daphnia*, fish, fungi, plants, and mammals. The aim of this paper is to review and compare some of these QSARs. The review focuses on recently-published QSAR models (since 2000), which were developed using more traditional statistical methods (regression analysis, partial least squares analysis), rather than neural network approaches, for example. In the first part of the review QSARs for toxicity to aquatic organisms (aquatic bacteria, protozoa, algae, *Hydraozoa*, *Daphnia*, fish, and amphibians) are presented.

I.1. Introduction

Development of QSARs for acute toxicity is a widely explored research area. QSARs can contribute to understanding potential mechanisms of toxic action of chemical compounds. They provide for an easy tool to estimate acute toxicity of chemicals, benefiting from time-effectivness and low financial costs. Furthermore, under the proposal for the future EU legislation for chemicals and chemical products, called REACH (Registration, Evaluation and Authorisation of Chemicals; EC, 2001; EC, 2003), it is foreseen that there will be an increased use of QSARs for toxicity evaluation and risk assessment of chemicals. Together with *in vitro* methods QSARs for toxicity would contribute to a replacement, reduction and refinement of animal use, as required by the Directive on the Protection of Laboratory Animals (EC, 1986).

QSARs for acute toxicity to a wide variety of biological systems, including bacteria, protozoa, algae, *Daphnia*, fish, fungi, plants, and mammals, have been developed and published. The aim of this paper is to review and compare some of these QSARs. The review focuses on recently-published QSAR models (since 2000), with the exception of neural network models. Recent reviews on the application of neural networks for developing QSARs for acute aquatic and health toxicological endpoints are given in Kaiser (2003a) and Kaiser (2003b). The first part of the review includes QSARs for toxicity to aquatic organisms – aquatic bacteria, protozoa, algae, *Hydraozoa*, *Daphnia*, fish, and amphibians.

I.2. QSARs for toxicity to aquatic bacteria

A number of QSARs for toxicity to bacteria were found in the literature. The most widely investigated species was *Vibrio fischeri* (formerly called *Photobacterium phosphoreum*). Cronin et al. (2000) developed QSARs for the toxicity of aliphatic compounds to the marine bacterium *V. fischeri* (Microtox® assay was performed, toxicity was expressed as the negative logarithm the concentration of the test substance (mmol/l) required for 50% reduction of the light emission of bioluminescent bacteria, *V. fischeri*, after 15 min exposure; pT_{15} values). The QSARs used the logarithm of the octanol-water partition coefficient, logP, and energy of the lowest unoccupied molecular orbital, LUMO as molecular descriptors (so called response-surface approach, Cronin and Schultz, 2001).

Separate QSARs were developed for haloalcoholes, halonitriles, bromoesters and diones. Hydrophobicity-dependent QSARs were found for haloalcohols and halonitriles:

Haloalcohols:

$$pT_{15} = 1.15 \log P - 1.34 \quad \text{(I.2.1)}$$
$$n = 15,\ R^2 = 0.908,\ s = 0.388,\ F = 139$$

Halonitriles:

$$pT_{15} = 1.54 \log P - 1.34 \quad \text{(I.2.2)}$$
$$n = 11,\ R^2 = 0.875,\ s = 0.388,\ F = 71$$

The relationship improved marginally by inclusion of LUMO:

$$pT_{15} = 1.61 \log P - 0.501\ LUMO - 1.19 \quad \text{(I.2.3)}$$
$$n = 11,\ R^2 = 0.954,\ s = 0.236,\ F = 104$$

The correlations between the toxicity of bromoesters and diones and logP alone had worse statistical fits. The following eqiuations were obtained combining logP and LUMO:

Bromoesters:

$pT_{15} = 0.473\ logP - 0.521\ LUMO + 0.146$ (I.2.4)
$n = 11, R^2 = 0.873, s = 0.127, F = 35.3$

Diones:

$pT_{15} = 1.18\ logP - 0.859\ LUMO - 0.281$ (I.2.5)
$n = 10, R^2 = 0.924, s = 0.234, F = 55.8$

Aditionally, data for toxicity of 19 alkanones, alkanals, and alkenals (taken from Cronin and Schultz, 1998) were included to obtaine a general QSAR model for toxicity to *V. fischeri*.

$pT_{15} = 0.790\ logP - 0.541\ LUMO - 0.527$ (I.2.6)
$n = 66, R^2 = 0.814, s = 0.527, F = 143$

This model includes aliphatic compounds with different mechanism of action. It is based on descriptors that allow for more clear interpretation compared to the models derived by Khadikar et al. (2002), namely logP as a descriptor of hydrophobicity, and LUMO as a descriptor of electrophilicity. However, Cronin et al. (2000) noted that the model is relevant for the prediction of aliphatic compounds only, because the values of LUMO for aromatic and aliphatic compounds are not strictily comparabel, as the π-bondiong in the aromatic ring decreases LUMO substantially even for relatively unreactive compounds.

In an interspecies correlation, the toxicity of these compounds to *V. fischeri* compared well to the toxicity (50% population growth inhibition) to the ciliate *Tetrahymena pyriformis* (n = 664, $R^2 = 0.795$).

Agrawal and Khadikar (2002) developed QSARs for toxicity to *V. fischeri* (expressed as $pEC_{50} = -logEC_{50}$), based on 39 chemicals by using molecular connectivity indices and indicator variables as predictors. Indices used were the first-order valence-connectivity index ($^1\chi^v$) and the equalized electronegativity (χ_{eq}). $^1\chi^v$ is a descriptor of size, shape, and symmetry of a molecule, while χ_{eq} accounts for the electronegativity effect of the substituents

(Agrawal and Khadikar, 2002). The toxicity data were taken from Zhao et al. (1998a). The original data set consisted of 75 compounds with classifications of the mechanism of toxicity, however the criteria for selecting the 39 investigated compounds were not presented in the paper of Agrawal and Khadikar (2002). Most of the investigated chemicals possessed non-polar and polar narcotic mechanisms of toxic action, but some of the compounds could also act by electrophilic mechanisms (for example benzaldehydes, Russom et al., 1997).

The two-parametric model based on the combination of $^1\chi^v$ and χ_{eq} is the following:

$$pEC_{50} = 0.745\ ^1\chi^v + 4.60\ \chi_{eq} - 9.23 \qquad \text{(I.2.7)}$$

$n = 39$, $R = 0.793$, $R_A^2 = 0.628$, $s = 0.608$, $F = 30.4$, $Qf = 1.245$

In this paper, n represents the number of compounds used to derive the equation, s is the standard error of estimation, R is the coefficient of determination, R_A is the coefficient of determination adjusted for the degrees of freedom, and F is the Fisher statistic. Qf is the quality factor equal to R/s. R_A takes into account the number of independent variables in the regression model. Whereas R^2 will increase with addition of independent variables to the model, R_A^2 will decrease if an added variable does not reduce the unexplained variation enough to compensate the loss of degrees of freedom (Khadikar et al., 2002).

Adding an indicator variable for polychloro substitution on aromatic rings (three or more chlorine atoms) (Ip_5) resulted in a model with improved statistical parameters:

$$pEC_{50} = 0.788\ ^1\chi^v + 3.93\ \chi_{eq} + 0.735\ Ip_5 - 7.78 \qquad \text{(I.2.8)}$$

$n = 39$, $R = 0.814$, $R_A^2 = 0.663$, $s = 0.634$, $F = 20.9$, $Qf = 1.324$

It should be noted that χ_{eq} takes a narrow range of values between 2.21 and 2.87 (a range of 0.66). This range is comparable to the standard error of the coefficients for this descriptor, on the basis of which one could question the significance of the descriptor.

Both the R-value and the standard error of prediction (s) increased from the first to the second equation. According to the authors, the use of R or s alone is not enough for assessing the quality of the model, so the simultaneous use of R and s should be used, preferably in the

form of the quality factor Qf. Qf increased from 1.245 to 1.324, suggesting the higher statistical quality of the second model. However, a difference of 0.1 is not enough to make a definite conclusion.

The authors concluded that the positive coefficients of $^{1}\chi^{v}$ and χ_{eq} in the equations suggest that the degree of saturation, the presence of hetero-atoms, and the electronegativity of the substituents are determining factors of the toxicity of the investigated compounds.

The same group (Khadikar et al., 2002) investigated toxicity to *V. fischeri* of 20 benzene derivatives using the Padmakar–Ivan (PI) index, which had been previously introduced by the same group, and some indicator variables. The index PI was defined as follows: if e is a given atom (edge) in a molecule, connected by the bonds (vertices) u and v, n_{eu} is the number of atoms lying closer to u than to v and n_{ev} is the number of atoms lying closer to v than to u. PI is a summation of n_{eu} and n_{ev} for all atoms e in the molecule.

The measured toxicity was represented as the concentration values causing a 50% inhibition of bacterial bioluminescence after 15 min exposure (EC_{50}, mol/l). The EC_{50} values were converted into pEC_{50} (= -logEC_{50}) units.

Khadikar et al. (2002) derived regression equations for logP and pEC_{50} on the basis of the PI index and the following indicator parameters: IP_1, indicating for the presence of two halogens in the aromatic nucleus; IP_2, accounting for the presence of poly-halogen substitution (three or more halogens); and IP_3 indicating mono-halogen substitution.

The correlation equation between pEC_{50} and PI was the following:

$$pEC_{50} = 0.030\ PI + 2.61 \qquad \text{(I.2.9)}$$

$n = 20$, $R = 0.877$, $F = 59.9$, $s = 0.366$, $Qf = 2.395$, $R^2_{cv} = 0.594$, $s_{cv} = 0.431$

In this paper, R^2_{cv} represents the square of the cross-validated coefficient of determination and s_{cv} is the cross-validated standard error of prediction from a leave-one-out statistical procedure.

The best equation obtained by Khadikar et al. (2002) for calculating toxicity was the following:

$pEC_{50} = 0.0237\ PI + 0.799\ IP_1 + 0.695\ IP_2 + 0.426\ IP_3 + 2.45$ (I.2.10)

$n = 20$, $R = 0.926$, $R_A^2 = 0.820$, $s = 0.331$, $F = 22.7$, $Qf = 2.798$, $R^2_{cv} = 0.835$, $s_{cv} = 0.331$

However, Khadikar et al. (2002) observed a strong correlation between pEC_{50} and logP ($R = 0.919$). This suggests that an one-parameter QSAR using simple structural descriptor like logP can be obtained, with a high value of the correlation coefficient R, similar to that of Equation I.2.10.

A weak side of the presented model is that PI and IP_2 intercorrelate with high intercorrelation coefficient ($R = 0.762$). The authors argue that PI and IP_2 can remain together in the proposed model because they give different types of information, as IP_2 is used as an indicator parameter to account for the structural features not covered in the PI index. However, the high intercorrelation coefficient does not support this suggestion.

The authors note that it is well established that substitution by halogen atoms in benzene increases toxicity due to partitioning, and they evaluate their results as not significantly adding to the existing toxicity knowledge but as showing that the introduced PI index could be used for modelling the pEC_{50} values of the benzene derivatives.

Lin et al. (2002) investigated the toxicity of 74 mixtures of 8 halogenated benzenes to the marine bacterium, *V. fischeri*, and derived QSARs using mixture C_{18}-Empore™ disk-water partition coefficients (K_{MD}). Toxicity was measured by quantifying the decrease in light emission from the bacteria as a result of exposure the test chemicals for 15 min.

Concentrations required to reduce light emission by 50% were reported as an EC_{50} value in units mol/l. The toxicity of mixtures was calculated using the following equation (Preston et al., 2000):

$$EC_{50M} = \frac{C_M}{\frac{C_A}{EC_{50A}} + \frac{C_B}{EC_{50B}} + \cdots}, \qquad \text{(I.2.11)}$$

where EC_{50} is the effective concentration required to bring about a 50% decrease in light output, C is the concentration of a substance, and subscripts A, B, etc., and M are the individual chemicals and mixture, respectively.

Lin et al. (2002) used the equation introduced by Verhaar et al (1995a) to predict the C_{18}-containing Empore™ disks-water partition coefficient for a mixture:

$$K_{MD} = \frac{W}{V} \times \left\{ \sum_{i=1}^{n} \frac{Q^0_{water,i}}{1 + (W/VK_{SDi})} \Big/ \left(\sum_{i=1}^{n} Q^0_{water,i} - \sum_{i=1}^{n} \frac{Q^0_{water,i}}{1 + (W/VK_{SDi})} \right) \right\}. \qquad \text{(I.2.12)}$$

where K_{MD} is the C_{18}-Empore™ disk-water partition coefficient for a mixture, W is the volume of solution, V is the volume of hydrophobic phase, Q^0_{water} is the initial amount of chemical i in water, n is the total number of individual chemicals in the mixture, and K_{SDi} is the partition coefficient of individual chemical i.

$LogK_{SD}$ for the individual chemicals were calculated using the equation of Verhaar et al. (1995a) between logP (octanol/warter partition coefficient) and $logK_{SD}$ (C_{18}-Empore™ disk-water partition coefficient of single chemicals):

$$logK_{SD} = 0.995\ logP + 0.70 \qquad \text{(I.2.13)}$$
$$n = 18,\ R^2 = 0.93,\ s = 0.24$$

Lin et al. (2002) derived the following relationship between toxicity and $logK_{SD}$ for the 8 halogenated benzenes, suggesting that the toxicity of the investigated chemicals is mainly dependent on their distribution between water and organic phases:

$$log1/EC_{50} = 0.881\ logK_{SD} + 0.40 \qquad \text{(I.2.14)}$$
$$n = 8,\ R^2 = 0.934,\ s = 0.194,\ F = 84.7,\ p < 0.0001$$

where p is the level of significance of the regression equation.

LogP values were used to calculate $logK_{SD}$ for individual chemicals, and after that these values were used used to calculate C_{18}-Empore™ disk-water partition coefficients ($logK_{MD}$

values) of 74 mixtures by using Equation I.2.12. $LogK_{MD}$ values were correlated with toxicity, suggesting that the toxicity of the mixtures is also based on their total hydrophobicity:

$$\log 1/EC_{50M} = 0.928 \log K_{MD} + 0.224 \quad \text{(I.2.15)}$$

$n = 74$, $R^2 = 0.953$, $s = 0.130$, $F = 1461$, $p < 0.0001$

The predictive capability of the model was tested by predicting the toxicity of 10 other mixtures. The prediction showed good statistics: $R^2 = 0.973$, $s = 0.113$ and $F = 288$ at a level of significance $p < 0.0001$, which are even better than the statistics of the model itself, implying that the mixtures in the test set are very similar to the training-set mixtures.

The same group of authors (Lin et al., 2003b) further extended the study, developing QSARs for toxicity of mixtures based on mixture partition coefficients and mixture Lewis acidity (A) and basicity (B). They investigated toxicity again to the bacterium *V. fischeri* (again expressed as concentrations required to reduce light emission by 50%, EC_{50} in units mol/l). The toxicity of 8 single chemicals (non-polar and polar narcotics) was measured. The toxicity of mixtures was calculated again by the equation of Preston et al. (2000).

Lin et al. (2003b) used again the equation introduced by Verhaar et al. (1995a) to predict partition coefficient for a mixture. By applying the same approach, Lin et al. (2003b) calculated mixture partition coefficients in the following phases: octanol-water (P_M), di-*n*-butyl ether-water (K_{MBW}), cyclohexane-water (K_{MCW}), chloroform-water (K_{MCHW}) and carbon tetrachloride-water (K_{MTW}).

Following the earlier work of Feng et al. (1996) Lewis acidity (A) and basicity (B) of benzene and its derivatives were expressed as:

$$A = \log K_{BW} - \log K_{CW} \quad \text{(I.2.16)}$$

$$B = \log K_{CHW} - \log K_{TW} \quad \text{(I.2.17)}$$

where K_{BW} is the di-n-butyl ether–water partition coefficients, K_{CW} is the cyclohexane–water partition coefficients, K_{CHW} is the chloroform–water partition coefficients, K_{TW} is the carbon tetrachloride–water partition coefficients. Feng et al. (1996) had correlated *A* and *B* with

hydrogen donor and acceptor potencies, and had concluded that Lewis acidity and basicity can be used as descriptors to quantify the effect of hydrogen bonding (H-bonding) for the studied chemicals.

In addition to calculating the Lewis acidity and basicity for individual chemicals, Lin et al. (2003b) also calculated the Lewis acidity and basicity for mixtures A^{MH}, and B^{MH} using mixture partition coefficients K_{MBW}, K_{MCW}, K_{MCHW} and K_{MTW}.

Lin et al. (2003b) derived the following hydrophobicity-based QSAR models for 18 nonpolar-narcotic-chemical mixtures and 18 polar-narcotic-chemical mixtures, respectively:

$$\log 1/EC_{50M} = 0.731 \log P_M + 1.51 \quad \text{(I.2.18)}$$
$$n = 18,\ R^2 = 0.938,\ s = 0.09,\ F = 243,\ p = 0.000$$

$$\log 1/EC_{50M} = 1.01 \log P_M + 1.37 \quad \text{(I.2.19)}$$
$$n = 18,\ R^2 = 0.952,\ s = 1.55,\ F = 318,\ p = 0.000$$

The toxicity of the third type of mixture, composed of nonpolar and polar narcotics together, was best correlated with $\log P_M$ and A^{MH}:

$$\log 1/EC_{50M} = 0.830 \log P_M + 0.527\ A^{MH} + 1.09 \quad \text{(I.2.20)}$$
$$n = 84,\ R^2 = 0.948,\ s = 0.166,\ F = 745,\ p = 0.000$$

Lewis basicity B^{MH} was not included in the QSAR, because its coefficient was found insignificant. According to Feng et al. (1996), Lewis acidity A correlates with the hydrogen donor activity.

The predictive capability of this regression equation was tested on 10 other similar mixtures, randomly composed of the eight single chemicals. The authors found visually, by plotting predicted versus observed values, an agreement between the measured $\log 1/EC_{50M}$ and that predicted by the model. The correlation coefficent between the predicted and measured values was not given.

The same group (Lin et al., 2003a) investigated toxicity to the marine bacterium, *V. fischeri* of binary mixtures between cyanogenic compounds and aldehydes. Toxicity was evaluated as concentration of compounds that causes 50% decrease in light emission from the bacteria as a result of exposure to the test chemicals for 15 min (EC_{50}, in units of mol/l). The mixture toxicity test was conducted in a similar manner as the single chemicals test. The joint effects of the chemicals in mixtures were presented by a sum of toxic units (M) (Chen and Lu, 2002):

$$M = z_1/Z_1 + z_2/Z_2 \qquad \text{(I.2.21)}$$

where z_i is the concentration of the i^{th} toxicant in the mixture and Z_i is its EC_{50} value. z_1 and z_2 were chosen in such a way that applying a mixture of their combination resulted in an exact 50% response, and also that the mixture toxicity tests are conducted at equitoxic ratio, which means that $(z_1/EC50_1) = (z_2/EC50_2)$ (Chen and Lu, 2002).

According to Lin et al. (2003a) the simple addition of the combined effect of compounds in a mixture is characterised by M = 1, antagonism is characterised by M > 1, and synergism by M < 1.

The following QSARs were obtained for binary mixtures of malononitrile in combination with each of 14 aldehydes:

aromatic aldehydes:

$$M = -0.882\ \sigma_p + 0.788 \qquad \text{(I.2.22)}$$

$n = 8,\ R^2 = 0.890,\ s = 0.129,\ F = 48.4$

aliphatic aldehydes:

$$M = -4.88\ \sigma + 0.0015 \qquad \text{(I.2.23)}$$

$n = 6,\ R^2 = 0.922,\ s = 0.099,\ F = 47.1$

where the σ is the Hammett constant.

Hammett constant σ was chosen to derive QSARs on the basis of the suggested mechanism of interaction between malononitrile and the aldehydes. The mechanism included formation of

cyanide ione (CN^-) from malonitrile, which then acts as a nucleophile to attack the positively charged carbon atom of the –CHO group of aldehyde. This mechanism revealed that the charge density of the carbon atom of the –CHO group is important for the joint toxicity between malonitrile and aldehyde. This charge density could be quantified by σ, which is a measure of the electron-donating/electron-withdrawing ability of substituents groups in chemical compounds.

According to the findings of Lipnick (1991), the cyanogenic toxicants investigated by Lin et al., (2003a) form cyanide ion (CN^-) via monooxygenase activation. The mechanism of formation of the ion is the following: under the attack of the monooxygenase enzyme, some carbon atoms (C*) of cyanogenic toxicant take on significant radical character, which readily undergoes perhydroxylation and binds -OH group. This leads to formation of a cyanohydrin type intermediate, which easily generates CN- upon hydrolysis. On the basis of this mechanism Lin et al. (2003a) concluded that the generation of CN^-, and thus the joint toxic effect between cyanogenic toxicants and aldehydes depends on the partial positive charge on the atom C* because of the anionic nature of the -OH group. That is why QSARs were attempted with the positive charge (C*) for the joint effect of acetaldehyde and each of 6 cyanogenic toxicants:

$$M = -0.237\ C^* + 0.0824 \qquad \text{(I.2.24)}$$

$n = 6,\ R^2 = 0.913,\ s = 0.020,\ F = 41.9$

Malonitrile, which was the only compound with two –CN groups (the remaining compounds tested had only one –CN group), was excluded from this equation, because it significantly worsened the correlation (inclusion of malonitrile resulted in equation with $R^2 = 0.711$).

Another study of mixture toxicity was that of Wei et al. (2004). They obtained toxicity to *V. fischeri* (compound concentrations, which cause 50% decrease in the bacterial luminescence after 15 min exposure, EC_{50} values in mmol/l) of 36 substituted aromatic compounds and 33 mixtures prepared from 5 selected compounds (aniline, benzene, 4-chlorophenol, nitrobenzene, phenol). The compounds were divided into non-polar (12 compounds) and polar (22 compounds) narcotics. QSARs were derived using logP values of the compounds ($logP_I$) and mixtures ($logP_M$). Mixture logP values ($logP_M$) were derived using the same equation of Verhaar et al. (1995a) (Equation I.2.12).

The following QSARs were obtained for the non-polar (Equation I.2.25) and polar (Equation I.2.26) narcotics:

$-\log EC_{50} = 1.31 \log P_I - 3.58$ (I.2.25)
$n = 12, R^2 = 0.933, s = 0.195, F = 140$

$-\log EC_{50} = 0.935 \log P_I - 1.36$ (I.2.25)
$n = 22, R^2 = 0.879, s = 0.178, F = 146$

It should be noted that some of the compounds classified as polar narcotics by Wei et al. (2004), might act as well as soft electrophiles (for example 2- and 3-nitrophenol, Schultz et al., 1997). However, a very good correlation between the toxicty and logP was obtained including these compounds. The residuals between predicted and observed toxicities for 2- and 3-nitrophenol were of only approximatelly 0.2 log-units.

The following QSAR was obtained for the toxicity of the mixtures, excluding mixtures containing benzene:

$-\log EC_{50} = 1.30 \log P_M - 2.01$ (I.2.26)
$n = 18, R^2 = 0.941, s = 0.109, F = 257$

A relationship between toxicity and $\log P_M$ for benzene mixtures could not be obtained.

Wei et al combined the 12 non-polar narcotics with the 15 mixtures containing benzene, and obtained the following correlation between the toxicity and logP values of the single compounds and mixtures:

$-\log EC_{50} = 1.11 \log P - 2.90$ (I.2.27)
$n = 27, R^2 = 0.952, s = 0.182, F = 490$

The authors combined also the 22 polar narcotics with the remaining 18 mixtures, obtained from the polar narcotic compounds (aniline, 4-chlorophenol, nitrobenzene, phenol) and obtanied the following relationship:

$-\log EC_{50} = 1.05 \log P - 1.64$ (I.2.28)

$n = 40$, $R^2 = 0.932$, $s = 0.163$, $F = 520$

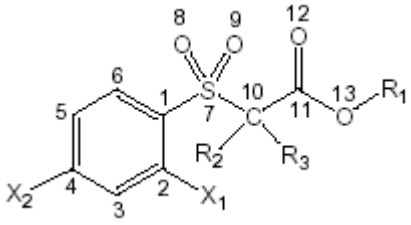
These correlations suggest that important factor for toxicity is the overal logP of the compounds in solution, and not their chemical nature.

The toxicity of mixtures of compounds were also investigated by Yu et al. (2001) who developed QSARs for acute toxicity of mixtures of halogenated benzenes (narcotics) to *V. fischeri*, using mixture octanol-water partition coefficients.

Another investigation of toxicity to *V. fischeri* is that of Liu et al. (2003b) who used Comparative Molecular Field Analysis (CoMFA) and Comparative Molecular Similarity Indices Analysis (CoMSIA) approaches, to develop three-dimensional (3D) QSAR models for the acute toxicity of 56 phenylsulphonyl carboxylates (Figure I.2.1). The toxicity data were collected from Hong et al. (1995), Chen and Wang (1996), and Liu et al. (2001). Liu et al. (2003b) noted that all toxicity data were measured under the same experimental conditions. Toxicity was again expressed as EC_{50} values after a 15 min assay in mol/l units. The CoMFA model derived had the following statistical parameters: steric and electrostatic fields included, number of significant components 6, $R^2 = 0.958$, $s = 0.098$, $F = 118.4$, $R^2_{cv} = 0.823$. The CoMSIA model had the following parameters – steric, electrostatic, hydrophobic, H-bond (including H-bond acceptor and H-bond donor) indices included, number of components 9; $R^2 = 0.933$, $s = 0.129$; $F = 70.8$, $R^2_{cv} = 0.713$.

The contour plot of the CoMFA steric field showed that more bulky substituents at the benzene ring and the 10- and 13-positions increase toxicity (Figure I.2.1). The CoMFA electrostatic field contour map showed that more negatively charged oxygen atoms would increase the biological activity. Also, an electron-withdrawing group with considerable negative charge located at the substituent group of the 10-position would increase toxicity. Contour maps for the CoMSIA analysis were not given in Liu et al. (2003b).

Figure I.2.1. Structures of the 56 phenylsulphonyl carboxilates



It should be noted that the same group of authors published CoMFA analysis of the same endpoint (toxicity to *V. fischeri*) for the same compounds in Liu et al. (2003c). In this paper they restricted the derived CoMFA model (using PLS) to four significant components (instead of six components in Liu et al., 2003b). The model included again steric and electrostatic fields, and the statistical parameters were $R^2 = 0.920$, s = 0.109, F = 1165, $R^2_{cv} = 0.790$. The same contur maps of the CoMFA analysis as in Liu et al. (2003b) were derived.

Ownby and Newman (2003) performed the Microtox® bacterial assay (based on the bacterium *V. fischeri*) to provide 15-min EC_{10}, EC_{20}, EC_{30}, EC_{40}, and EC_{50} values (concentrations resulting in a 10, 20, 30, 40 and 50% decrease in bioluminescence after 15 min of exposure), and concentration-effect relationships for five divalent metal ions: Co^{+2}, Cu^{+2}, Mn^{+2}, Ni^{+2}, Zn^{+2}. On the basis of these data, metal interaction coefficients for each binary pair of metals were calculated. The metal interaction is the random effect due to the i^{th} level of effect of metal 1 and the j^{th} level of effect of metal 2, and is normally distributed.

The following QSAR was obtained for metal interaction coefficients:

$$\text{Metal interaction} = 68.5\ \Delta\sigma_p - 2.63 \qquad \text{(I.2.29)}$$

n = 10, $R^2 = 0.69$, Dev = 14.2, PRESS (predicted residual sum of squares) = 2.1

where $\Delta\sigma_p$ is the difference between the softness indices (σ_p) of the paired metal ions
Dev is the mean relative deviation from perfect fit expressed as a percentage, [(observed-fitted)/observed]*100

The softness index (σ_p) of a metal ion reflects its softness, or the tendency for the outer electron shell to deform (i.e., polarisability), and the ion's tendency to share electrons with ligand donor atoms.

The metal interaction coefficient approaches –1 as the effects of the two metals become independent, and decreases from –1 as the system deviates from independent action. As the difference in the abilities of the metal ions to accept electrons increases (i.e. $\Delta\sigma_p$ increases), the interaction of the metals approached –1.

Liu et al. (2001) developed QSARs for the toxicity of 20 phenylsulphonyl acetates to *V. fischeri*. The toxicity was measured as EC_{50} values (compound concentration that inhibits 50% of the bacterial luminescence after 15 min of exposure). The QSARs were developed by using Theoretical Linear Solvation Energy Relationship (TLSER) descriptors (Wilson and Famini, 1991), representing cavity, dipolarity/polarisability, and H-bonding terms. Cavity was represented by V_{mc}, equal to $V_m/100$, where V_m is the molecular solvent-excluded volume (in A^3). Dipolarity/polarisability was represented by polarisability index (π^*), equal to α/V_m, where α is molecular polarisability (in a.u.; 1 a.u. = 0.1482 A^3). Hydrogen-bond acceptor ability was represented by two terms - the covalent contribution to Lewis basicity, ε_b, equal to the difference in energy between LUMO of water and HOMO of solute, and the electrostatic basicity contribution (q^-) representing the largest negative atomic charge in the solute molecule (the most negatively charged atom will have the greatest interaction with a proton from a neighbouring molecule, Boyd et al., 2001). Analogously, hydrogen bond-donating ability was presented by ε_a – the energy difference between HOMO of water and LUMO of solute, and qH^+ - the largest positive charge of a hydrogen atom in the solute molecule (the most positively charged proton of the molecule will be most attracted to a negatively charged atom of a neighbour, Boyd et al., 2001).

The following QSAR was obtained:

$$\log EC_{50} = -0.467\, V_{mc} - 4.19\, \pi^* - 11.7\, \varepsilon_b + 6.71 \qquad \text{(I.2.30)}$$

$n = 20$, $R_A^2 = 0.868$, $s = 0.094$, $F = 42.78$, $p < 0.001$

The equation contains terms accounting for molecular size and polarisability, suggesting a binding action via dispersion (van der Waals) forces, and ε_b representing the H-bonding potency of chemicals. The bioluminescence reaction in *V. fischeri* is catalysed by the enzyme luciferase. Therefore, the influence of π^* and ε_b may be related to toxicity via inhibition of this enzyme, or of the reduced flavin mononucleotide ($FMNH_2$) donating system, an important coenzyme transferring hydrogen.

Another bacterial species was investigated by Ren and Frymer (2002). They investigated the toxicity of 98 chemicals by using the bioluminescent bacterium strain Shk1. Shk1 is a genetically modified luminescent bacterium whose original strain was a *Pseudomonad* isolated from the activated sludge in an industrial wastewater treatment plant. The determined toxicity was expressed as the concentration of a toxicant required to reduce the emission of bioluminescence by 50% (EC_{50} values, mmol/l). The compounds were divided into the following groups according to their chemical features: aromatics (excluding benzaldehydes and benzoic acids), benzaldehydes and benzoic acids, halogen-substituted aliphatics, alcohols, amines, and phenols. QSARs for the individual chemical groups were derived, using the logarithm of the octanol-water partition coefficient (logP) as a predictor variable. Some chemical groups, such as unsubstituted alkanes, aliphatic acids, ethers, esters, were represented with a small number of compounds, insufficient for performing statistically meaningful analysis. In addition to chemical class-specific QSARs, a QSAR model for all chemicals was also obtained.

The outliers were estimated using the semi-studentised residuals, calculated according to the following equation:

$$e^* = e/s \qquad \text{(I.2.31)}$$

where e* is the semi-studentised residual, e is the residual, and s is the root square of the mean error sum of squares obtained in the regression (standard error of estimate). Compounds that had a semi-studentised residual of four or higher were identified as outliers and removed from the QSARs.

The derived QSARs had the following general form:

$$\log 1/EC_{50} = a \log P + b \quad \text{(I.2.32)}$$

where a is the slope and b is the intercept of the regression line.

The derived QSARs are presented in Table I.2.1, including outlier(s) removed.

Table I.2.1. LogP-dependent QSAR models for organic compounds (reproduced from Ren and Frymer, 2002)

Group of compounds	Slope	Intercept	Outlier(s) removed	n	R	F
Aromatics[a]	0.407	-1.095	-	11	0.94	68.0
Benzaldehyde and benzoic acids	0.403	-0.804	orcinol	11	0.69	8.1
Halogen-substituted aliphatics	0.888	-2.400	1,1,1,2- and 1,1,2,2-tetrachloroethane, tetrachloroethylene	19	0.88	55.8
Alcohols	0.932	-2.401	-	10	0.99	386.2
Amines	0.547	-1.236	Ethanolamine	6	0.98	81.6
Phenols	0.578	-1.119	2,6-dimethoxyphenol	22	0.80	36.3
Overall	0.552	-1.359	2,6-dimethoxyphenol	77	0.77	108.0

[a] Excluding benzaldehydes and benzoic acids.

The authors tried to derive a QSAR for phenols based on the use of the pK_a value (the negative logarithm of the first ionisation constant) as a descriptor, to account for phenol ionisation, but the coefficient of pK_a in the regression equation was found to be insignificant at the 5% significance level.

An overall QSAR model was developed, based on a training set of 78 of the 98 compounds used in the investigation. The remaining 20 compounds were used as a test set to investigate the predictive capacity of the model. The model had a moderate correlation coefficient (R = 0.77), probably resulting from the fact that a diverse range of chemical structures were involved in building the model. Prediction of the toxicity of the remaining 20 compounds showed that the QSAR model had an accuracy of prediction of about one order of magnitude. According to the authors, this accuracy may be acceptable for some applications, considering the inherent variability within and between toxicity test systems.

The authors did not try to derive a narcosis model based on logP including only the compounds acting by narcotic mechanism of toxic action.

I.3. QSARs for toxicity to protozoa

Many authors have investigated toxicity to the ciliated protozoan *Tetrahymena pyriformis* (Arnold et al., 1990; Cajina-Quezada and Schultz, 1990; Cronin and Schultz, 1996; Akers et al., 1999; Balaz and Lukacova, 2002; Cronin et al., 2002a, Schultz et al., 2003b). The toxicity data in these studies were developed in the laboratory of Prof. Terry W. Schultz (College of Veterinary Medicine, The University of Tennessee, 2407 River Drive, Knoxville, TN 37996, http://www.vet.utk.edu/faculty/schultz.shtml). The data were obtained in the same laboratory, which is a necessary precondition for developing of high quality QSARs.

For example Cronin and Schultz (2001) developed QSARs for the toxicity of aromatic compounds to *T. pyriformis*. The toxicity data (the concentration causing 50% inhibition of *T. pyriformis* growth, IC_{50} values in mmol/l) were collected from four previous publications of Cronin et al. (1995), Dearden et al. (1995), Cronin and Schultz (1996), and Cronin et al. (1998). The collected data set consisted of 290 compounds with some duplicate entries and one compound not toxic at saturation. Removing these compounds resulted in a data set of 268 compounds investigated by Cronin and Scultz (2001). The investigated compounds represented a variety of toxic mechanisms of action. Firstly, simple linear QSARs were developed, based on a small number of descriptors selected according to their physico-chemical relevance. This approach is also known as response-surface analysis (Cronin and Schultz, 2001). The descriptors included the logarithm of the octanol–water partition coefficient (logP) and the energy of the lowest molecular orbital (LUMO), which parameterise biouptake and electrophilic reactivity, respectively. The following QSAR was found:

$$\log 1/IC_{50} = 0.568\ \log P - 0.320\ LUMO - 0.871 \qquad \text{(I.3.1)}$$
$$n = 268,\ R^2 = 0.624,\ s = 0.511,\ F = 223,\ R^2_{cv} = 0.617$$

Some of the compounds were found to have toxicity less than predicted. These were likely to be ionised at physiological pH or for which steric hindrance (i.e., alkyl groups adjacent to reactive centres) is likely to be important. Ionised compounds are less toxic than the neutral

form (Zhao et al., 1998b). 2-And 4-substituted aminophenols and similarly substituted nitrobenzenes had toxicity higher than predicted. Such compounds are likely to be transformed *in vivo* to more reactive species (Cronin and Schultz, 2001). Other compounds such as dinitrotetrachlorobenzene have the capability to react by electrophilic mechanisms. Excluding 29 compounds from this equation in QSAR with improved statistical parameters:

$$\log 1/IC_{50} = 0.603\ \log P - 0.330\ LUMO - 1.00 \qquad \text{(I.3.2)}$$
$$n = 239,\ R^2 = 0.800,\ s = 0.335,\ F = 476,\ R^2_{cv} = 0.796$$

To improve the predictivity other structural descriptors and indicator variables devised specifically to account for the outliers were used:

$$\log 1/IC_{50} = 0.633\ \log P - 0.526\ LUMO + 0.721\ I_{2,4AP} - 1.61\ I_{strong\ acid} + 0.314\ N_{H\text{-}don} - 1.39$$
$$n = 268,\ R^2 = 0.780,\ s = 0.393,\ F = 185,\ R^2_{cv} = 0.764 \qquad \text{(I.3.3)}$$

where $I_{2,4AP}$ is an indicator variable for a phenol with a primary amine substituted in either the 2- or 4- position; $I_{strong\ acid}$ is an indicator variable for the presence of a strong acid; $N_{H\text{-}don}$ is the total number of H-bond donors. The positive coefficient for $I_{2,4AP}$ and the negative coefficient for $I_{strong\ acid}$ confirm that 2- and 4-aminosubstituted phenols increase toxicity and strong acids decrease toxicity. According to Cronin and Schultz (2001) the inclusion of the sum of the number of H-bond donor groups is a case of correlation and not causality. The authors note that functional groups such as aromatic amines and hydroxy groups are capable of H-bond donation, so the inclusion of this parameter may simply reflect the increased toxicity of aminophenols.

PLS analysis was also performed on the data set with 28 descriptors. It resulted in a model with three significant components, an R^2 value of 0.790 and an R^2_{cv} value of 0.757. The variables with the greatest standardised loadings in the components were the number of sulphur atoms, which was negatively associated with toxicity in all three components, and the indicator variable for the presence of an acidic group.

The PLS analysis was repeated with the 5 variables that appeared to be important for the toxicity in the derived QSAR. Again, a PLS model with three significant components was identified, having an R^2 value of 0.771 and an R^2_{cv} value of 0.757.

According to Cronin and Schultz (2001) the response-surface approach is simple, transparent and mechanistically interpretable. Its statistical parameters were improved by including of indicator variables and other parameters with mechanistic interpretability. PLS analysis resulted in models with similar statistics, but they were less transparent.

Burden (2001) investigated the toxicity to *T. pyriformis* by using the newly developed Gaussian Processes Method (GPM) for solving regression problems. He used the same toxicity data set of 290 chemicals as Cronin and Schultz (2001) from which he derived a data set of 277 chemicals after removing the duplicate entries and the non toxic compound at saturation. Thus, although derived from the same sources (Cronin et al., 1995; Dearden et al., 1995; Cronin and Schultz, 1996; Cronin et al., 1998), the data set investigated by Burden (2001) had 9 additional chemicals than the toxicity data set investigated by Cronin and Schultz (2001). However, Burden (2001) did not report the investigated compounds and thus no information was given about the extra compounds.

In the GPM the models are developed by using some hyperparameters, derived from the descriptor variables by implementation of the mathematical procedures of gaussian processes (Burden, 2001). The hyperparameters are not easily analysed for the contributions of each of the descriptor variables. The models obtained by GPM are used to predict the bioactivity of new compounds, which are not included in the training data set, and they do not permit for mechanistic explanation. The question for the chemical applicability domain of the developed GPM models was not discussed by Burden (2001).

The molecular indices used in the study of Burden (2001) were: Randic index (R) (Randic, 1975); the valence modification to the Randic index by Kier and Hall (K) (Kier and Hall, 1995); and an atomistic (A) index developed by Burden (Burden, 1996) which was extended in this study by the recognition of aromatic atoms and hydrogen atom donors and acceptors (B). The atomistic indices (A, B) count the numbers of given types of atoms present in the molecule. Also two other indices were used, the first counts the number of rings of various sizes (G), and the second counts some common functional groups (F).

The data set was divided into a training set and a test set by a K-means clustering algorithm clustering on dependent and independent variables together. The test set consisted of 15% of

the total data set. The obtained GPM model included 32 variables and had the following statistics: coefficient of determination $R^2 = 0.81$, standard error of estimate = 0.07, coefficient of determination on the test set $Q^2 = 0.82$, standard error of prediction on the test set = 0.08.

The data set investigated by Cronin and Schultz (2001) was also investigated by Gonzales et al. (2004), who selected 202 aromatic compounds with nitro and cyano groups. They applied the so-called TOPological Substructural MOlecular DEsign (TOPS-MODE) approach. This approach is based on the hydrogen-depleted molecular graphs. In order to account for heteroatoms, bond weights such as bond distances, bond dipoles, or bond polarisabilities, are introduced for the elements of the bond adjacency matrix. The bond adjacency matrix with introduced weights is used to calculate so called spectral moments, defined as the sum of diagonal entries of the powers of the bond matrix (Estrada and Gutierrez, 1999). Linear or nonlinear multivariate statistical technique, such as multiple linear regression analysis, is used to develop QSARs, relating toxicity with the spectral moments as predictor variables.

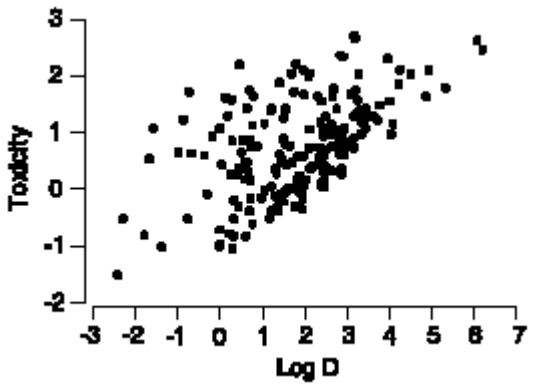
In the study of Gonzales et al. (2004) the dipole moments and hydrophobicity were used as weights in the main diagonal of the bond adjacency matrix. After excluding 12 compounds as statistical outliers (mainly nitrophenols and comounds with aromatic rings contained nitrogen atoms such as pyridines and pyrimidines) a QSAR model was obtained, which included spectral moments of zero (equal to the numer of atoms in a molecule), second and third order that encode information for molecular topology, hydrophobicity, molecular size, and the affinity of the molecules to water. The QSAR model had the following statistics: training set: n = 151, $R^2 = 0.750$, s = 0.376; test set: n = 39, $R^2_{pred} = 0.66$, s = 0.362; the whole data set: n = 190, $R^2 = 0.73$, s = 0.38, $R^2_{cv} = 0.72$.

Similar approaches to those of Cronin and Schultz (2001) were used by Cronin et al. (2002a) to developed QSARs for the toxicity of 200 phenols to *T. pyriformis* (presented as IC_{50} – the concentration in mmol/l causing 50% inhibition of growth), validated by using a test set of 50 phenols. Three approaches were used to develop QSARs. Firstly, response-surface analysis was applied. The descriptors included the logarithm of the octanol–water distribution coefficient (logD) and the energy of the lowest molecular orbital (LUMO). In all the QSAR analyses, logD gave better results than logP, demonstrating the importance of ionisation on the biouptake process for these compounds. Many of the phenols included in this study were

weak acids, which may be partially ionised at the pH of the test system (7.35). This resulted in a low correlation between the logD and logP.

When the toxicity of the phenols was plotted against logD, a baseline effect was observed, meaning that a large number of compounds have a minimal toxicity based on their hydrophobicity (Figure I.3.1). The baseline effect is commonly related to the narcosis mechanism of toxicity (Schultz et al., 2003a). Compounds acting by reactive mechanisms are shown to have toxicity in excess of narcotics.

Figure I.3.1. Plot of toxicity to *T. pyriformis* (presented as log1/IC_{50} values) against logD. 2,4,6-Trinitrophenol is omitted from this plot due to an extremely low logD value (–4.98) (taken from Cronin et al., 2002a).



The following relationship was found by Cronin et al. (2002a) between the toxicity of the phenols to *T. pyriformis* and logD and LUMO:

$$\log 1/IC_{50} = 0.42 \log D - 0.70 \text{ LUMO} - 0.24 \qquad \text{(I.3.4)}$$

$n = 200$, $R^2 = 0.54$, $s = 0.56$, $F = 117$, $R^2_{cv} = 0.53$

Particular sets of compounds were found to be outliers for this equation, resulting in a poor statistical fit of the equation: phenols substituted in the 2- or 4- position by an amino group; phenols substituted in the 2- or 4- position by a nitro group; and hydroquinones. Hydroquinones are believed to be susceptible to oxidation to the quinone. Quinones have a toxicity greater than that expected for polar narcotics. Another group of poorly predicted compounds was the phenols substituted with three or more halogens, which are associated

with the weak respiratory uncoupling mechanism of toxic action (Terada, 1990), which is known to exert toxicity in excess of narcosis. All 40 compounds with these structural features were taken out of the data set, and the model was derived again:

$$\log 1/IC_{50} = 0.53\ \log D - 0.96\ LUMO - 0.58 \qquad \text{(I.3.5)}$$
$$n = 160,\ R^2 = 0.81,\ s = 0.34,\ F = 340,\ R^2_{cv} = 0.80$$

The authors modified the response-surface approach in order to account for the outliers. Indicator variables were added, accounting for the presence (or absence) of the following structural features: substitution by a nitro group in the 2- or 4- position (I_{NO2}); substitution by an amino group in the 2- or 4- position (I_{NH2}); hydroquinones, defined as those compounds substituted by a hydroxy group in the 2- or 4- position (I_{OHOH}); and substitution by three or more halogens ($I_{>3Hal}$). The following QSAR was developed:

$$\log 1/IC_{50} = 0.45\ \log D - 0.53\ LUMO + 0.56\ I_{NO2} + 0.95\ I_{NH2} + 0.97\ I_{OHOH} + 0.68\ I_{>3Hal} - 0.49$$
$$n = 200,\ R^2 = 0.69,\ s = 0.46,\ F = 75,\ R^2_{cv} = 0.67 \qquad \text{(I.3.6)}$$

Twenty compounds were found to be outliers, less toxic than predicted by this equation. These outliers were dominated by phenols with iodine substituents and alkyl substituted aminophenols and nitrophenols. Removal of these 20 compounds resulted in the following QSAR:

$$\log 1/IC_{50} = 0.54\ \log D - 0.57\ LUMO + 0.74\ I_{NO2} + 0.57\ I_{NH2} + 1.02\ I_{OHOH} + 0.85\ I_{>3Hal} - 0.69$$
$$n = 180,\ R^2 = 0.83,\ s = 0.33,\ F = 148,\ R^2_{cv} = 0.82 \qquad \text{(I.3.7)}$$

Another approach applied by Cronin et al. (2002a) to the same data set was stepwise regression, using a data set of 108 physicochemical and topological descriptors. Seven descriptors were identified as important for describing toxicity, resulting in the following QSAR:

$$\log 1/IC_{50} = 0.33\ \log D - 0.45\ LUMO + 0.0028\ MW - 0.020\ P_{NEG} + 0.036\ SsOH - 0.52$$
$$ABSQon + 2.57\ qH^+ - 0.33 \qquad \text{(I.3.8)}$$
$$n = 200,\ R^2 = 0.65,\ s = 0.49,\ F = 54,\ R^2_{cv} = 0.63$$

where MW is molecular weight, P_{NEG} is the negatively charged molecular surface area, SsOH is the electrotopological state index for the hydroxy group, ABSQon is the sum of absolute charges on nitrogen and oxygen atoms, and qH^+ is the largest positive charge on a hydrogen atom.

Some significant outliers from this equation were observed including compounds capable of transformation to the quinone. After removing the outliers, the following significant equation was obtained:

$$\log 1/IC_{50} = 0.38\ \log D - 0.58\ LUMO + 0.0047\ MW - 0.018\ P_{NEG} + 0.050\ SsOH - 0.61\ ABSQon + 2.69\ qH^+ - 0.99 \quad \text{(I.3.9)}$$

$$n = 185,\ R^2 = 0.83,\ s = 0.34,\ F = 128,\ R^2_{cv} = 0.82$$

LogD and LUMO account for biouptake and reactivity, respectively (Cronin et al., 2002a). The interpretation of the remaining variables is difficult. They seem to encode information concerning molecular size (MW) or H-bonding or electron distribution (P_{NEG}, ABSQon, qH^+, SsOH). This equation includes some of the outliers of the equation obtained by the response-surface approach, which makes its domain of applicability broader.

The partial least squares (PLS) statistical procedure was also attempted in order to derive predictive QSAR models. The set of 11 variables found to be useful in both the response-surface and stepwise regression analyses were used in the PLS analysis. After removal of 3 outliers, a PLS model with three significant components, and $R^2 = 0.82$, was obtained.

The models were validated on a test set of 50 compounds. The toxicity was poorly predicted by the models for four compounds: 2,4-diaminophenol, 5-amino-2-methoxyphenol, 6-amino-2,4-dimethylphenol, and methylhydroquinone. These compounds were typical of the outliers removed during model development, and they were excluded from the derived correlations between the observed and predicted toxicity values of the test set compounds. The correlation between the observed and predicted toxicity using the model developed by the response-surface analysis (Equation I.3.5) had correlation coefficient $R^2 = 0.66$. Prediction of the toxicity from the model developed by the modified response-surface analysis (Equation I.3.7) resulted in correlation between the predicted and the observed values with $R^2 = 0.75$.

Stepwise regression analysis (Equation I.3.9) gave $R^2 = 0.78$ for the correlation between the predicted and the observed toxicities. PLS modelling resulted in the best prediction of toxicity of the test set compounds, with correlation between the predicted and observed values having $R^2 = 0.82$.

The authors compared different statistical approaches to the same data set. Their conclusion is that none of the techniques investigated is better than the others. Stepwise regression and PLS give better statistical fit, but the response-surface approach allows for clearer mechanistic interpretability, and for identification and treatment of outliers.

The data set of Cronin et al. (2002a), containing 250 chemicals, was also investigated by Ren (2003). Ren (2003) included data for toxicity to *T. pyriformis* of 206 phenols, without specifying the criterion for selecting the investigated compounds. Toxicity was expressed as $\log 1/IC_{50}$ (logarithm of the inverse 50% growth inhibition concentration, in units of mmol/l). The data set was divided into training set (163 compounds) and test set (43 compounds), and developed mechanism-based QSARs. The phenols were considered to exhibited toxicity via four mechanisms – polar narcosis, respiratory uncoupling, pro-electrophilicity, and soft electrophilicity.

The following structural descriptors were used to develop QSARs: logarithm of the ionisation-corrected octanol-water partition coefficient (logD), the lowest unoccupied molecular orbital (LUMO) energy, the highest occupied molecular orbital (HOMO) energy, molecular weight (MW), the negatively charged molecular surface area (P_{NEG}), the sum of absolute charges on nitrogen and oxygen atoms in a molecule (ABSQon), the largest positive charge on a hydrogen atom (qH^+), the electrotopological state index for the hydroxy group (SsOH), the negative logarithm of the first dissociation constant (pKa), and the number of H-bond donors ($N_{H\text{-}don}$). QSARs were developed using multiple linear regression with backward variable selection. The significance level for a variable to stay in the final model was set at 0.05. Phenols with model residuals higher than ± 2 were considered as statistical outliers.

The following QSARs were obtained:

Polar narcosis:

$\log 1/IC_{50} = -0.10 pK_a + 0.50\ logD - 0.34\ LUMO + 0.0024\ MW - 0.020\ P_{NEG} + 3.69\ qH^+$
$n = 111, R^2 = 0.92, R_A^2 = 0.92, s = 0.27, F = 201, p < 0.0001$ (I.3.10)

The intercept in this equation was found to be insignificant and therefore regression through the origin was performed. The relevance of pKa, LUMO, P_{NEG} and qH^+ for the toxicity of polar narcotics is difficult to interpret, their appearance in the equation resulted from the backward statistical procedure, used for selection of variables.

Respiratory uncoupling:

$\log 1/IC_{50} = 0.40\ logD + 7.34\ qH^+$ (I.3.11)
$n = 14, R^2 = 0.95, R_A^2 = 0.94, s = 0.39, F = 117, p < 0.0001$

Again, the intercept in this equation was insignificant and therefore regression through the origin was performed.

Pro-electrophilicity:

$\log 1/IC_{50} = -0.50\ pK_a + 1.23\ HOMO - 1.01\ N_{H\text{-}don} + 0.79\ logD + 0.083\ P_{NEG} + 0.11\ SsOH + 12.1$ (I.3.12)
$n = 19, R^2 = 0.95, R_A^2 = 0.93, s = 0.21, F = 41.6, p < 0.0001$

Methoxyhydroquinone was a statistical outlier and was removed. This compound had also been found to be an outlier by Cronin et al. (2002a).

Soft electrophilicity:

$\log 1/IC_{50} = 1.44\ LUMO - 0.079\ P_{NEG} + 4.91$ (I.3.13)
$n = 18, R^2 = 0.43, R_A^2 = 0.35, s = 0.42, F = 5.64, p = 0.0149$

No statistical outliers were found. Although statistically significant, the equation is not of good statistical quality. The positive sign of LUMO is unusual since this term normally has a negative sign in QSARs for toxicity (Ren, 2003), as decreasing LUMO would increase electrophilicity and reactivity of the compounds.

Ren (2003) developed a single QSAR without regard to mechanism of toxicity:

$\log 1/IC_{50} = -0.095\ pK_a + 0.38\ HOMO + 0.35\ logD - 0.56\ LUMO + 0.0026\ MW - 0.017\ P_{NEG} - 0.61\ ABSQon + 0.050\ SsOH + 4.22$ (I.3.14)

$n = 162$, $R^2 = 0.73$, $R_A^2 = 0.72$, $s = 0.43$, $F=52.1$, $p < 0.0001$

statistical outlier – methoxyhydroquinone.

The toxicity of the phenols in the test set was predicted by using the mechanism-based QSARs and the single QSAR, and the predicted toxicity values were correlated with the measured values. The predicted values of the single QSAR model gave better correlation with the measured values ($R^2 = 0.67$) than the predicted values of the mechanism-based QSARs ($R^2 = 0.43$), although the mechanism-based approach is conceptually sounder. The QSAR for soft electrophilic phenols had a low statistical quality ($R^2 = 0.43$), which contributed to the worse prediction of the toxicity by the mechanism-based QSAR approach.

In the work of Cronin et al. (2001), 203 substituted aromatic compounds containing a nitro- or cyano group were investigated for toxicity to *T. pyriformis* ($\log 1/IC_{50}$) values measured from 40-h population growth impairment assay). Toxicity was related to hydrophobicity, quantified by the octanol-water partition coefficient (logP) and electrophilic reactivity, quantified by the energy of the lowest unoccupied molecular orbital (LUMO) or maximum acceptor superdelocalisability (A_{max}). The two parameters, LUMO and A_{max}, were intercorrelated ($R = 0.87$) for the full data set.

The following QSARs were obtained:

$\log 1/IC_{50} = 0.40\ logP - 0.94\ LUMO - 1.27$ (I.3.15)

$n = 203$, $R^2 = 0.60$, $s = 0.49$, $F = 151$

$\log 1/IC_{50} = 0.37\ logP + 13.1\ A_{max} - 4.30$ (I.3.16)

$n = 203$, $R^2 = 0.70$, $s = 0.42$, $F = 237$

A_{max} gave a QSAR with better statistical parameters than LUMO. Although both A_{max} and LUMO describe electrophilic reactivity, there are differences in their meaning (Cronin et al.,

2001). The LUMO value describes the electron affinity of a molecule and characterises the susceptibility of the molecule to attack by nucleophiles. Superdelocalisabilities describe reactivity of occupied and unoccupied orbitals. Individual superdelocalisabilities describe the contribution of an atom to the stabilisation energy in the formation of a charge-transfer complex with a second molecule (Cronin et al., 2001). The acceptor superdelocalisability characterises reactions with the electrophilic centre (Karelson, 1996). The maximum value for this provides an indication of the ability of the centre to react as an electrophile. The better QSAR obtained with A_{max} than with LUMO might be due to the fact that A_{max} might be able to describe better transition states of reactions (Karelson, 1996). According to Cronin et al. (2001) it is possible that for the compounds investigated in the study A_{max} discriminates better between highly electrophilic and less reactive compounds than LUMO.

Another investigation of the toxicity to *T. pyriformis* by using the response surface approach is that of Dimitrov et al. (2003). The authors investigated inhibition of population growth of the ciliate by miscellaneous narcotic chemicals. The toxicity data were again presented as 50% inhibitory growth concentration (IC_{50} in units mol/l).

According to Dimirtov et al. (2003) the response-surface approach can be considered as generalisation of the traditional baseline model. It results in the development of a QSAR which models non-congeneric chemicals acting via non-specific, weak chemical interactions without covalent bond rearrangements (such as van der Waals interactions, H-bond formation, etc.). These week interactions are associated with the narcosis mode of action producing a reversible toxic effect (Dimitrov et al., 2003). The organism response to an environmental toxicant is considered as a result of the influence of two different processes: uptake of the chemical into the biophase (traditionally modelled by logP) and interaction with the site of action (Dimitrov et al., 2003). According to Dimitrov et al. (2003) interaction of the non-covalent acting chemicals with the site of action can be described by descriptors assessing the electrophilic character of molecules, like the energy of the lowest unoccupied molecular orbital (LUMO), electronegativity, average or maximum superdelocalisability, and largest charge at non-hydrogen atom. This statement differs in a way from the understanding of e.g. Cronin and Schultz (2001), Cronin et al. (2001), Cronin et al. (2002a), that LUMO and A_{max} describe electrophilic reactivity resulting in covalent change in biological systems (Shultz et al., 2003a), rather than non-covalent electrophilic interactions.

According to Dimitrov et al. (2003) chemicals not included in the surface-response (not modelled well by the response-surface approach) possess well-defined reactive groups, which could interact by covalent bond rearrangements (Karabunarliev et al., 1996). In this way, by eliminating chemicals possessing reactive substructures, the chemicals that populate the surface-response could be identified. These chemicals could act by different narcotic mechanisms, including baseline (non-polar), polar, ester, and amine narcosis.

The following QSARs were obtained for compounds acting by different narcotic mechanisms:

General neutral narcotics:

$\log 1/IC_{50} = 0.678 \log P - 0.123 \text{ LUMO} + 1.54$ (I.3.17)
$n = 318$, $R^2 = 0.900$, $s^2 = 0.109$, $F = 1420$, $R^2_{cv} = 0.898$.

In this equation LUMO was statistically significant at the 95% confidence level (presumably coresponding to a p value below 0.05). According to Dimitrov et al. (2003) this confirms the importance of LUMO when narcotic chemicals from different chemical groups with different potency levels at the same hydrophobic range are combined. In this way, according to Dimitrov et al. (2003) the addition of the LUMO as a second descriptor to the baseline model extends it to a more general base-surface model accounting for the global electrophilicity of molecules associated with their ability to interact by non-specific mechanisms.

The following QSAR was obtained for esters, which are usually identified to have narcotic mode of toxic action, which differs from the other types of narcosis (non-polar, polar, amine):

$\log 1/IC_{50} = 0.71 \log P - 0.25 \text{ LUMO} + 1.38$ (I.3.18)
$n = 93$, $R^2 = 0.878$, $s^2 = 0.119$, $F = 325$, $R^2_{cv} = 0.869$

The level of significance of LUMO in the ester model was 85% (presumably corresponding to p value of 0.15).

Amines, other than aniline derivatives, were investigated separately as they have been found to elicit an enhanced toxic effect in fish (Sinks et al., 1998). The following QSAR was

derived for aliphatic amines (LUMO was not included as the level of statistical significance of this parameter when added to the model was found to be only 60%, presumably corresponding to p value of 0.40):

$\log 1/IC_{50} = 0.68\ \log P + 1.80$ (I.3.19)
$n = 51,\ R^2 = 0.854,\ s^2 = 0.135,\ F = 288,\ R^2_{cv} = 0.841.$

Combining the neutral narcotics and esters resulted in the following model (the significance of LUMO was more than 95%):

$\log 1/IC_{50} = 0.68\ \log P - 0.13\ LUMO + 1.50$ (I.3.20)
$n = 411,\ R^2 = 0.890,\ s^2 = 0.116,\ F = 1652,\ R^2_{cv} = 0.888$

These models showed that LUMO could be a useful parameter in modelling narcotic compounds with a different nature of narcotic action.

In the work of Schultz et al. (2002), a large data set of 500 aliphatic chemicals was investigated for toxicity to *T. pyriformis* (toxicity expressed as IC_{50} values obtained in a two-day population growth inhibition assay). The investigated chemicals represented a number of structural classes possessing a variety of mechanisms of toxic action, including narcosis and electrophilic mechanisms. Again, the response-surface approach was applied to derive QSARs, using the octanol-water partition coefficient (logP) (encoding hydrophobicity) and the energy of the lowest unoccupied molecular orbital (LUMO) to quantify electrophilic reactivity. Compounds having logP values between -1 and 5 were only included in the study. These limits were chosen pragmatically and represent the range of hydrophobicity of compounds that may be assessed by the *T. pyriformis* test assay. Very hydrophilic compounds (i.e. $\log P < -1$) are frequently observed to be insufficiently toxic to provide an accurate IC_{50} value, and hence confidence in these biological data is low. Conversely, hydrophobic compounds (i.e. $\log P > 5$) have insufficient water solubility in the test system to bring about 50% lethality. These limits in logP define the domain of QSAR applicability in terms of hydrophobicity. Further, the upper limit of logP is consistent with the practical limitations on its measurement, e.g. by using the shake flask method (Dearden and Bresnen, 1988).

The following QSAR, which represents a baseline toxicity relationship, was derived for the group of nonpolar narcotics (saturated alcohols, ketones, nitriles, esters, and sulphur-containing compounds):

$$\log 1/IC_{50} = 0.723 \log P - 1.79 \qquad \text{(I.3.21)}$$

$n = 215$, $R_A^2 = 0.926$, $s = 0.274$, $R^2_{cv} = 0.925$

QSARs were developed separately for several chemical classes, showing toxicity in excess of baseline (toxicity greater than that determined by the logP relationship only). These included: diesters, sodium salts of carboxylic acids, simple aldehydes, lactones, R-unsaturated (triple bond) alcohols (i.e., proelectrophiles), selected electrophilic chemicals thought to be capable of undergoing an SN_2 reaction, and a miscellaneous set of reactive chemicals. The SN_2-acting chemicals included those with an activated leaving group (i.e., halogen). The miscellaneous reactive group included hydrazides, dinitriles, and diones. The best models obtained are presented in the Table I.3.1:

Table I.3.1. Regression models for toxicity to *T. pyriformis* of selected classes, based on hydrophobicity and/or electrophilicity (reproduced from Schultz et al., 2002).

Class	N	logP	LUMO	Intercept	R^2_A	S	R^2pred
diesters	24	0.608	0.394	-1.64	0.927	0.242	0.917
salts	6	0.404		-1.45	0.912	0.185	0.811
aldehydes	17	0.503	-3.386[c]	2.06	0.919	0.200	0.898
	17	0.528		-0.98	0.908	0.214	0.889
lactones	10	0.512		-1.11	0.760	0.366	0.637
	9[a]	0.619		-1.29	0.944	0.187	0.916
proelectrophiles	22	0.525	-0.073[c]	-0.81	0.615	0.446	0.302
	22	0.520		-0.81	0.634	0.435	0.586
S_N2	49	0.657	-0.088[c]	-1.06	0.790	0.285	0.765
	49	0.668		-1.14	0.789	0.285	0.770
	48[b]	0.687		-1.18	0.821	0.265	0.805
Miscellaneous	11	0.818	-0.243[c]	-0.073	0.866	0.292	0.812
	11	0.892		-0.90	0.870	0.288	0.845

[a] (±)-β-butyrolactone excluded; [b] 3-bromopropionic acid excluded; [c] LUMO is not significant at 95% level ($p > 0.05$).

The model for proelectrophiles had moderate statistical parameters (see Table I.3.1). It improved with the addition of the molecular connectivity cluster index of third order ($^3\chi_c$):

$\log 1/IC_{50} = 0.588\ \log P - 0.976\ LUMO + 0.917\ ^3\chi_c + 1.18$ (I.3.22)
$n = 20,\ R_A^2 = 0.842,\ s = 0.283,\ R^2_{cv} = 0.803$

Two compounds, 1-hexyn-3-ol and 3-hexyn-2-ol, had high prediction residuals and were excluded from the model.

Also, a general QSAR including all compounds was developed:

$\log 1/IC_{50} = 0.645\ \log P - 0.342\ LUMO - 1.11$ (I.3.23)
$n = 353,\ R_A^2 = 0.859,\ s = 0.353,\ R^2_{cv} = 0.857$

(±)-β-Butyrolactone was poorly predicted and excluded from this equation.

Compounds containing a carboxylic acid moiety were not modelled well by the general model. The following QSAR was derived for the carboxylic acids:

$\log 1/IC_{50} = 0.\ 273\ \log P - 0.116\ LUMO - 0.558$ (I.3.24)
$n = 35,\ R_A^2 = 0.873,\ s = 0.141,\ R^2_{cv} = 0.838$

Aliphatic amines were also poorly modelled by the equation for neutral narcosis or by the general model. The following QSAR was derived for the group of aliphatic amines only:

$\log 1/IC_{50} = 0.676\ \log P - 1.23$ (I.3.25)
$n = 30,\ R_A^2 = 0.873,\ s = 0.\ 336,\ R^2_{cv} = 0.848$

A comparison of the baseline equation (Equation I.3.21) and the equation (Equation I.3.25) for aliphatic amines derived by Schultz et al. (2002) showed a similar slope for the logP dependence but a greater intercept for amines. This is consistent with previous studies (Sinks et al., 1998) that amine toxicity is reversible and dependent only on partitioning, but in excess to that obtained by baseline model. At the pH of the test system, amines with an ionisation constant (p*K*a) greater than 9.0 are protonated. According to Austin, et al., (1995) protonated

amines have a higher affinity for membranes than nonionised compounds, resulting in increased toxicity of these compounds over that expected from nonpolar narcosis. According to Schultz et al. (2002) it is not clear whether a separate mechanism of action exists for the amines or whether their requirements for separate modelling is an artefact of poorly calculated logP values.

Another group of chemicals that were not predicted well by the general model was the group of isothiocyanates. The IC_{50} values of saturated aliphatic isothiocyanates were found to be essentially constant (0.0202 ± 0.0023 mmol/l).

R-haloactivated compounds (structures such as RC(X)C(=O)R or RC(X)C(#N)R, where X represents a halogen atom) were also not predicted well by the general model. A QSAR for these compounds was derived with three descriptors – LUMO, the maximum acceptor superdelocalisability (*A*max), and the ellipsoidal volume (ElipVol):

$$\log 1/IC_{50} = -2.14\ LUMO + 25.1\ A_{max} + 0.000942\ ElipVol - 8.03 \qquad (I.3.26)$$
$$n = 30,\ R_A^2 = 0.831,\ s = 0.342,\ R^2_{cv} = 0.792$$

1-Bromo-3,3-dimethyl-2-butanone and ethyl 2,3-dibromopropionate were excluded from the equation.

LUMO reflects the reactivity of the whole molecule and depends mainly on the number and type of the halogen atoms, as well as on the type of the polarised group. A_{max} depends more on the arrangement of the heteroatoms. The third descriptor, ElipVol, reflects molecular bulk and shape. The positive coefficient of this descriptor suggested higher toxicity for the longer and slimmer (more linear) compounds, and lower toxicity for the shorter and more bulky (less linear) compounds.

For the toxicity of the amino alcohols, a three-descriptor model was obtained, using the energy of the highest occupied molecular orbital (HOMO), the connectivity valence path index of third order $^3\chi^v_p$, and the logarithm of Henry's law constant (logH):

$$\log 1/IC_{50} = -1.13\ HOMO + 1.11\ {}^3\chi^v_p - 0.0827\ logH - 12.8 \qquad (I.3.27)$$
$$n = 16,\ R_A^2 = 0.841,\ s = 0.214,\ R^2_{cv} = 0.788$$

This QSAR suggested a different nature of toxicity of amino alcohols compared to other studied chemicals. According to Schultz et al. (2002) amino alcohols, especially β-hydroxylamines, are corrosive. Their toxicity was found to be independent of hydrophobicity. The appearance of HOMO in the QSAR suggests a nucleophilic rather than an electrophilic type of reactivity. The connectivity index $^{3}\chi^{v}_{p}$ is sensitive both to the size and branching of the molecules. It is higher for molecules which have substituents in an α,β-position to each other. The third descriptor, logH (air-water partition coefficient), accounts for volatility, predicting higher toxicity for less volatile chemicals.

Another study on toxicity to *T. pyriformis* was that of Schultz et al. (2003b), who investigated a strategy for selecting optimal training and test sets of compounds from a chemical space, with a view to deriving valid QSARs with optimised applicability. They used *T. pyriformis* toxicity data for a series of 385 benzenes. Their QSAR models had the following form:

$$TOX = a*logP + b*A_{max} - c \qquad (I.3.28)$$

where TOX = $\log 1/IC_{50}$, logP is the octanol-water partition coefficient, and A_{max} is the maximum acceptor superdelocalisability value.

According to the authors, the training set for QSAR development should be chemically diverse and should cover a broad descriptor space, to allow for a broad chemical applicability domain. The selection of the test set should be based on the representivity of the test chemicals, which should mimic the distribution of all the chemicals within the descriptor space. The authors used the design of optimal experiments (DOE) technique to select a chemically diverse training set, and the K nearest neighbour (KNN) technique to select a representative test set. They showed that with appropriate selection of chemicals it is possible to maximise the coverage of the descriptor space with a relatively small number of chemicals, and to develop and validate QSARs having high statistical fit. The authors concluded that a training to test set ratio of 1:3 appears adequate for the validation of QSARs. In the case of the 385 investigated compounds, 60 chemicals were enough to represent the chemical space, of which 15 were included in the training set and 45 in the test set.

Schultz and Seward (2000) developed QSARs for inhibition of *T. pyriformis* population growth (toxicity data presented as $\log 1/IC_{50}$ values) for 23 non-polar narcotics, polar narcotics, and esters. The QSARs derived included dimyristoyl phosphatidylcholine-water partition coefficient ($\log K_{DMPC}$), and the octanol-water partition coefficient (logP):

$$\log 1/IC_{50} = 0.73 \log K_{DMPC} - 1.62 \quad \text{(I.3.29)}$$
$$n = 23,\ R^2 = 0.926,\ s = 0.24,\ F = 263,\ p = 0.0001$$

The above model was statistically better than the model with logP, developed for the same compounds:

$$\log 1/IC_{50} = 0.71 \log P - 1.60 \quad \text{(I.3.30)}$$
$$n = 23,\ R^2 = 0.828,\ s = 0.39,\ F = 101,\ p = 0.0001$$

$\text{Log}K_{DMPC}$ appeared to model better the toxicity of compounds that act by different narcotic mechanisms of toxicity. However, the modeling depends on the availability of $\log K_{DMPC}$ values for chemical compounds. Outliers to the $\log K_{DMPC}$ model were primary aliphatic amines.

A different approach was adopted by Balaz and Lukacova (2002), who used subcellular pharmacokinetic (SP) theory to derive QSARs for the toxicity of phenolic compounds to *T. pyriformis*. SP theory can be used for modelling drug distribution in membranes, intracellular and extracellular aqueous phases. The distribution of a drug depends on various processes, including transport through (and accumulation in) membranes, covalent and non-covalent binding to proteins and other cell constituents, metabolic reactions, and excretion.

The toxicity data (expressed as the reciprocal concentration (mmol/l) of a phenolic compound causing 50% reduction in the growth of the protozoan after 96 h exposure) were collected from the literature (Schultz and Riggin, 1985; Schultz et al., 1986; Cajina-Quezada and Schultz, 1990; Schultz et al., 1990; Schultz et al., 1992; Jaworska and Schultz, 1993; Bryant and Schultz, 1994; Schultz et al., 1996).

On the basis of SP theory, the authors derived the following equation, relating toxicity T to lipophilicity (logP) and acidity (pK_a):

$$\log T = -\log(A_0 \times P^{\beta} + B_1 \times K_a + 1) - \frac{D_0 + D_1 \times K_a}{A_0 \times P^{\beta} + B_1 \times K_a + 1} - E \times pK_a + F \qquad \text{(I.3.31)}$$

The first term accounts for distribution of the compound in the different phases of the studied biosystem; the second term accounts for metabolism; and the third term reflects the binding to the receptors. β is an empirical coefficient. A, B, and D are coefficients that characterise individual processes the molecules undergo in the biological system: A, accumulation in the membranes and protein binding; B, distribution in aqueous phases; and D, hydrophobicity-independent elimination. Their values depend on the biological system under investigation. F is an intercept. The indices 0 and 1 account for non-ionised and first-degree ionised phenols. Originally four additional coefficients were present in the equation – A_1, accounting for the accumulation in the membranes and protein binding of the ionised phenols; B_0, representing distribution in aqueous phases of the non-ionised phenols; C_0 and C_1, accounting for hydrophobicity-independent elimination of non-ionised and ionised phenols respectively. Applying non-linear regression analysis to derive the values of the coefficients resulted in high standard deviations of these coefficients and no improvement of the quality of the fit. That is why these coefficients were fixed as: $B_0 = 1$, $A_1 = C_0 = C_1 = 0$, suggesting that the processes for which A_1, C_0, and C_1 account, namely accumulation in the membranes and protein binding of the ionised phenols and hydrophobicity-independent elimination, do not contribute sufficiently to the observed phenol toxicity. Non-linear regression analysis gave the following values for the remining coefficients in the equation: $A_0 = 0.1830 \pm 0.1033$, $\beta = 0.2665 \pm 0.0631$, $B_1 = 6.407 \pm 3.805$, $D_0 = 5.471 \pm 0.929$, $D_1 = 16018 \pm 8630$, $E = 0.1758 \pm 0.0143$, and $F = 5.393 \pm 0.330$. The statistical parameters of the fit for n = 122 compounds were: the correlation coefficient R = 0.953, the standard error of estimation s = 0.257, and the value of the Fisher test F = 191.

According to the coefficient values, the model was interpreted as follows: the differences in toxicities are determined mainly by the distribution of free non-ionised molecules, in addition to a small effect of receptor binding (as described by the coefficient E and decrease in toxicity with increasing pK_a values). Metabolism decreases significantly the toxicity of two groups of compounds: the first group has $pK_a > 6$ and logP < 4 and the second group has $pK_a < 2$. As $D_1 >> D_0$, the metabolism of ionised species is faster and causes a rapid decrease in the toxicity of very acidic derivatives ($pK_a < 2$), regardless their hydrophobicity.

The model was compared to two empirical models, obtained by non-linear regression analysis by fitting polynomials with cross-terms of logP and pK_a, and the inverse values of these descriptors, to the toxicity of the investigated phenols. These models had better statistical parameters than the SP-based model (n = 122, R = 0.963, s = 0.233 and F = 157; n = 122,R = 0.962, s = 0.235 and F = 156), and gave similar results for predictive quality as the SP-based model when leave-one-out and leave-group-out cross-validation procedures were performed. Another validation procedure was applied by excluding compounds with extreme values of logP and pKa, and deriving the models using the remaining compounds. The toxicity of the excluded compounds was predicted by the models, which was actually extrapolation outside the descriptor space used to derive the models, rather than the interpolation obtained with the standard cross-validation procedures. The SP-based model appeared to extrapolate much better to the toxicity values for the omitted compounds than the empirical models (predictive sum of squares of deviations between the calculated and experimental values of the omitted points (PRESS) = 6.055 for the SP-based model, and PRESS = 30.40 and 36.83 for the two empirical models), indicating the higher predictive quality of the SP-based model. The SP model, although more complex, was derived on the basis of theoretical considerations.

Castro et al. (2003) used the so-called optimisation of correlation weights of local graph invariants (OCWLGI) approach to model the toxicity of 66 aliphatic compounds to the ciliate *T. pyriformis* (50% growth inhibitory concentration). The descriptors used were calculated on the basis of the labelled hydrogen-filled graphs (LHFG) or the graphs of atomic orbital (GAO), using Morgan extended connectivity values (Morgan, 1965) and certain correlation weights for the graph vertexes and Morgan connectivity values. LHFG are molecular graphs, including hydrogen atoms, in which each vertex (atom) is labelled. GAO are built from LHFG by means of a certain computational scheme (Toropov and Toropova, 1998; Toropova et al., 2000). GAO consist of hydrogen-filled molecular graphs, in which each type of atomic orbital of each atom is considered as a vertex of the graph and is labelled separately. The Morgan extended connectivity value of zero order for a given vertex of the molecular graph is the number of vertexes to which this vertex is connected (the number of adjacent vertexes). The Morgan extended connectivity value of order k assigned to a given vertex is the sum of the Morgan connectivity values of order k-1 of its adjacent vertexes (Wong et al., 2003). The correlation weight for the graph vertexes and Morgan extended connectivity values are

obtained by means of the Monte Carlo optimisation method, in order to obtain correlation between the toxicity values and the corresponding calculated OCWLGI descriptor with highest correlation coefficient.

The data set of 66 compounds was taken from Cronin et al. (2000). It was divided to training set (33 compounds) and test set (33 compounds). OCWLGI descriptors were calculated based on LHFG and Morgan extended connectivity values of order zero to third, and based on GAO and Morgan extended connectivity values of zero and first orders. For each set of these descriptors the Monte Carlo optimisation procedure was run three times, in order to obtain correlation weights providing for correlation between the toxicity and the descriptor with high correlation coefficient for the training set.

The model that gave best fit in the training set (first run of the Monte Carlo procedure, R^2 = 0.810, s = 0.452, F = 132) and best prediction of the test set (R^2 = 0.796, s = 0.445, F = 121) included the OCWLGI descriptor based on GAO and the zero order Morgan extended connectivity values. Applied for the whole data set (n = 66) this model gave the following statistics: R^2 = 0.804, s = 0.445, F = 262.

Another QSAR toxicity study based on OCWLGI approach was this from Toropov and Schultz (2003). Toxicity data (50% growth inhibitory concentration) for substituted benzenes tested in the two-day *T. pyriformis* population growth impairment assay were used. OCWLGI descriptors were calculated based on LHFG only. The most appropriate OCWLGI model was based on third order of Morgan extended connectivity (n = 157, R^2 = 0.883, s = 0.274, F = 1170). The model was validated on a test set of 60 compounds and good correlation between the observed and predicted values with slope of 0.991 and intercept of 0.012 and R^2 = 0.863, s = 0.28, and F = 372 was obtained. Catechols and benzoquinones, which might react via series of mixed mechanisms, including electrophile reactivity (Schultz et al., 1997; Russom et al, 1997) were not included in the investigated data set by Toropov and Schultz (2003).

Toropov and Schultz (2003) investigated data set taken from Akers et al. (1999). The best OCWLGI model obtained was based on second order of Morgan extended connectivity (n = 39, R^2 = 0.901, s = 0.346, F = 337).

Ivanciuc (2000) investigated the toxicity of 47 nitrobenzenes to *T. pyriformis*. Toxicity (taken from Dearden et al., 1995; and Schüürmann et al., 1997) was expressed as the 50% inhibition concentration of growth of *T. pyriformis*, obtained in a 48-h exposure static assay (IC_{50} values in units mmol/l). Ivanciuc (2000) used the following descriptors to develop QSARs: molecular weight (MW); octanol-water partition coefficient corrected for ionisation at pH 7.35 (logD); several Kier and Hall's valence connectivity indices; 80 Wiener-type indices, based on molecular graphs.

The best QSAR obtained included logD and three Wiener-type indices and had $R^2 = 0.876$, s = 0.250, F = 73.7, $R^2_{cv} = 0.843$, $s_{cv} = 0.27$. During the statistical analysis two descriptors were not included in the same model if their intercorrelation coefficient was higher than 0.8.

The same data set had been previously investigated using comparative molecular field analysis (CoMFA) (Schüürmann et al., 1997). A 3D QSAR model with logD and the CoMFA steric and electrostatic fields was obtained, having good statistical parameters ($R^2 = 0.788$, $R^2_{cv} = 0.761$, and s = 0.314). An earlier investigation of the same data set by Dearden et al. (1995) gave a multiple linear regression QSAR model using logD, LUMO, and the absolute value of the change on the nitro oxygen upon substitution with better statistics ($R^2 = 0.858$, $R^2_{cv} = 0.826$, and s = 0.255).

The QSAR with logD and three Wiener-type indices had better statistics than both the CoMFA model (Schüürmann et al., 1997) and the QSAR with quantum-chemical descriptors (Dearden et al, 1995). Nevertheless Ivanciuc (2000) notes that graph indices (like the presented Wiener-type indices) are not intended to replace other descriptors in QSAR analysis. According to the author, each type of descriptors accounts for particular aspects of the molecular structure, and descriptors from all types must be used in order to fully explore the structural space.

I.4. QSARs for toxicity to algae

Schmitt et al. (2000) investigated 19 nitroaromatic compounds for their toxicity to the unicellular green alga *Scenedesmus vacuolatus* (formerly *Chlorella fusca*). The toxicity was measured as EC_{50} values, denoting the concentration that leads to 50% reduction of cellular reproduction after 24 h exposure.

A QSAR model was obtained with logP and LUMO:

$\log EC_{50} = -0.61 \log P + 1.60 \text{ LUMO} - 1.19$ (I.4.1)
$n = 18, R^2 = 0.89, R_A^2 = 0.88, s = 0.42, F = 61$

Picric acid was poorly predicted and omitted from this model. It was the only compound that is more than 99.9% deprotonated under the pH of the test medium (pH = 6.7).

Adding the maximum net atomic charge at the nitro nitrogen ($q_{nitro\text{-}N}$) improved the model (picric acid was again omitted):

$\log EC_{50} = -0.54 \log P + 1.67 \text{ LUMO} - 35.5\ q_{nitro\text{-}N} + 19.1$ (I.4.2)
$n = 18, R^2 = 0.92, R_A^2 = 0.91, s = 0.36, F = 56,$

When logP was replaced with $\log D^u$, which quantifies the partitioning into octanol of the undissociated species available at pH = 6.7 (the pH of the test medium), slightly better models were obtained:

$\log EC_{50} = -0.53 \log D^u + 1.81 \text{ LUMO} - 1.15$ (I.4.3)
$n = 19, R^2 = 0.91, R_A^2 = 0.90, s = 0.39, F = 82$

$\log EC_{50} = -0.55 \log D^u + 1.69 \text{ LUMO} - 34.3\ q_{nitro\text{-}N} + 18.4$ (I.4.4)
$n = 19, R^2 = 0.95, R_A^2 = 0.93, s = 0.32, F = 84$

When logD, which accounts for partitioning into octanol of both the ionised and unionised species and ion-pair distribution, was used, worse model was obtained:

$\log EC_{50} = -0.79 \log D + 1.19 \text{ LUMO} - 1.28$ (I.4.5)
$n = 19, R^2 = 0.89, R_A^2 = 0.87, s = 0.43, F = 64$

As a third descriptor, $q_{nitro\text{-}N}$ was not significant. In this equation, logD accounts for distribution of both the unionised and ionised compound fraction, while LUMO refers to only the neutral species.

According to Schmitt et al. (2000), nitroaromatic compounds may exert oxidative stress by acting as redox cyclers. For this mode of action, the nitroaromatic radical anion formed by one-electron reduction is oxidised back to the parent compound while forming superoxide (O_2^{-}•), which then leads to generation of hydrogen peroxide (H_2O_2) and hydroxyl radical (OH•) as highly reactive oxidants. Although the variation in the net atomic charge of the nitro nitrogen was only small (0.034 au), its appearance in the QSAR models made mechanistic sense, as the nitro nitrogen is likely to be the site of attack of the additional negative charge during the one-electron reduction of the molecule.

From the explained mechanism of action, Schmitt et al. (2000) suggested that the energy of the singly occupied molecular orbital (SOMO) of the respective radical anion generated by one-electron reduction may be a suitable parameter for differentiating between redox-cycling compounds and those that are preferentially being reduced further to metabolites with closed-shell electronic structures (like nitroso derivatives and hydroxylamines in the case of nitroaromatic chemicals). Due to the high degree of intercorrelation between LUMO and SOMO ($R^2 = 0.86$), the influence of both properties on toxicity could not be discriminated statistically by means of multlinear regression. A similar model, with $logD^u$ and SOMO as descriptors, was obtained:

$$\log EC_{50} = -0.47 \log D^u + 1.58 \text{ SOMO} - 4.64 \qquad \text{(I.4.6)}$$

$n = 19$, $R^2 = 0.91$, $R_A^2 = 0.90$, $s = 0.39$, $F = 80$

Cronin et al. (2002b) developed QSARs for the toxicity of1 3 mono- and di-substituted nitrobenzenes to the alga *Chorella vulgaris* in a 15-min assay. Toxicity was measured as compound concentration that causes 50% decrease in algae fluorescence (EC_{50} values). The algal fluorescence is due to hydrolysis of fluorescein diacetate, which does not fluoresce, by esterase enzymes to fluorescein, which fluoresces. Fluorescein diacetate is added artificially to the cells, it is a relatively small molecule, which readily penetrates the algal cell and accumulates in the cell plasma membrane due to its hydrophobic nature (Worgan et al., 2003). There it is converted to fluorescein, causing cell fluorescence. The decrease of the fluorescence in cells subject to a toxicant relative to a control can be used as a measure of the decrease of cell viability (enzymatic function and number of cells) due to the toxicant.

Cronin et al. (2002b) applied response-surface approach, using descriptors for hydrophobicity (as quantified by logP) and reactivity (i.e. electrophilicity as quantified by LUMO). The following QSAR was obtained:

$\log 1/ED_{50} = 0.911\ \log P - 1.55\ LUMO - 3.88$ (I.4.7)
$n = 13,\ s = 0.442,\ R^2 = 0.767,\ F = 20.8,\ R^2_{cv} = 0.701$

4-Chloronitrobenzene was observed to be a significant outlier. It was the only nitrobenzene in the data set that was mono-substituted in the 4-position. Its removal resulted in the following QSAR:

$\log 1/ED_{50} = 0.952\ \log P - 1.68\ LUMO - 4.24$ (I.4.8)
$n = 12,\ s = 0.353,\ R^2 = 0.861,\ F = 35.2,\ R^2_{cv} = 0.813$

The same endpoint (toxicity to the green alga *C. vulgaris*, expressed as effective chemical concentration that causes 50% inhibition in the fluorescence at 15-min assay, EC_{50} value) was investigated by Worgan et al. (2003) for two series of chemicals: 10 non-polar narcotics and 10 polar narcotics.

The following QSARs with logP were obtained:

Non-polar narcotics:

$\log 1/EC_{50} = 1.04\ \log P - 3.28$ (I.4.9)
$n = 10,\ R^2 = 0.96,\ s = 0.27,\ F = 206,\ R^2_{cv} = 0.95$

Polar narcotics:

$\log 1/EC_{50} = 0.641\ \log P - 1.91$ (I.4.10)
$n = 10,\ R^2 = 0.88,\ s = 0.16,\ F = 69,\ R^2_{cv} = 0.84$

The QSAR for non-polar narcotics represents the ‘baseline’ toxicity. The QSAR for polar narcotics had a lower slope and greater intercept than that for non-polar narcotics. Also, its statistical parameters were worse.

Netzeva et al. (2004) developed QSARs for toxicity to the alga *C. vulgaris*, based on 65 aromatic compounds. The compounds tested included phenols, anilines, nitrobenzenes, benzaldehydes and other poly-substituted benzenes.

The following QSAR was obtained by using multiple linear regression:

$$\log 1/EC_{50} = 0.731\ \log P - 0.590\ LUMO - 1.91 \qquad \text{(I.4.11)}$$
$$n = 65,\ R^2 = 0.839,\ s = 0.429,\ F = 161,\ R^2_{cv} = 0.819$$

Statistically significant models (showing slightly worse fit) were also obtained by using maximum acceptor superdelocalisability (A_{max}) or electronegativity (EN) in addition to logP (R^2 = 0.822 and R^2 = 0.815, respectively). The results of the QSAR analysis support the importance of the hydrophobicity and electrophilicity for toxicity.

The following model was obtained by using PLS (this technique was used because of the intercorrelation between the descriptors):

$$\log 1/EC_{50} = 0.404\ \log P - 0.233\ LUMO + 9.842\ A_{max} + 0.204\ {}^0\chi^v - 5.40 \qquad \text{(I.4.12)}$$
$$n = 65,\ R^2(X) = 0.930,\ R^2(Y) = 0.858,\ RMSEE = 0.403,\ R^2_{cv}(Y) = 0.843$$

The model had two significant principal components. 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 leave-one-out (LOO) cross-validated R^2_{cv} (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.

LogP and zero order valence corrected molecular connectivity index (${}^0\chi^v$) encode factors influencing interaction with biological membranes. ${}^0\chi^v$ is believed to account for molecular size (Dearden et al., 1988) with a correction for the presence of π-electrons and lone pair electrons. Both LUMO and A_{max} account for electrophilicity.

Both the MLR and PLS models were validated (in addition to the LOO cross-validation procedure) by dividing the data set into two groups in order to simulate external prediction. The compounds were ordered by increasing toxicity and then subdivided into two equal complementary subsets by taking alternate chemicals. Separate models were derived for the two groups. Consequently, the toxicity of the compounds in the first group was predicted by the model derived for the second group and *vice versa*. The simulated external validation showed that the 2-descriptor MLR model and the 4-descriptor PLS model demonstrated stability and good predictivity (R^2 of the correlation between observed and predicted values 0.837 and 0.848 respectively, s = 0.428 and 0.413 respectively).

The same endpoint (toxicity to the alga *C. vulgaris* assessed in a 15-min assay) was measured by Cronin et al. (2004) using an extended data set of 91 compounds. The chemicals investigated possessed different mechanisms of toxic action, including narcosis, electrophilic mechanisms, uncoupling of oxidative phosphorylation, or toxicity via specific interactions with biological macromolecules. Two approaches were used to develop QSARs – multiple linear regression and k-nearest neighbour (KNN) analysis. LogP (accounting for hydrophobicity) and lowest unoccupied molecular orbital (LUMO) (accounting for reactivity) were used in the QSAR modeling. To improve the quality of the two-parameter models, a third descriptor was selected using both forward and backward stepwise regression analysis.

KNN analysis was performed to test a presence of nonlinearity in the relationship between the toxicity and the structural descriptors. In the KNN approach the three descriptors selected from the regression analysis were used. The descriptors were standardised in the range of 0 to 1. The Euclidean distances [d(a,b)] between the compounds were calculated using the following equation:

$$d(a,b) = \sqrt{\sum_{i=1}^{n} (x_{a,i} - x_{b,i})^2} \qquad \text{(I.4.13)}$$

where $x_{a,i}$ and $x_{b,i}$ are the values of descriptor i for two compounds a and b and n is the number of descriptors.

The nearest neighbour was identified as the compound with the shortest Euclidean distance to the chemical of interest. Up to 20 nearest neighbors were investigated to obtain best prediction of the toxicity (k = 1-10, 12, 15, and 20). The predicted toxicity was calculated as a mean value of the toxicities of the KNNs to the considered compound (principle of active analogues).

The following regression model was obtained including logP and LUMO (response-surface approach):

$$\log 1/EC_{50} = 0.829\ \log P - 0.405\ LUMO - 2.21 \qquad \text{(I.4.14)}$$
$$n = 91,\ R^2 = 0.868,\ s = 0.538,\ F = 290,\ R^2_{cv} = 0.857$$

The model was improved by adding the first-order Δ valence connectivity index ($\Delta^1\chi^v$):

$$\log 1/EC_{50} = 0.838\ \log P - 0.268\ LUMO - 0.278\ \Delta^1\chi^v - 2.76 \qquad \text{(I.4.15)}$$
$$n = 91,\ R^2 = 0.890,\ s = 0.494,\ F = 235,\ R^2_{cv} = 0.875$$

$\Delta^1\chi^v$ is the difference between the first-order valence connectivity index ($^1\chi^v$) for the compound in question and the straight chain alkane with the same molecular formula. According to Cronin et al. (2004) it accounts for molecular size, and encodes information about the electronic structure of the molecules.

In the KNN analysis using the three selected structural descriptors, each compound was omitted once and its toxicity was predicted using the mean toxicity of its KNNs, which were selected from the remaining 90 compounds (LOO procedure). The best prediction was obtained for k = 7, with the correlation between observed and predicted toxicity from LOO for the 91 compounds having $R^2 = 0.855$ and s = 0.560. These results were similar to the results obtained by the LOO procedure in the regression analysis. Two compounds, methidathion and pentabromophenol, were poorly predicted by the KNN analysis.

Subsequently, Cronin et al. (2004) divided the data set into training and test sets. The compounds in the two subsets were the same for both the MLR and the KNN studies.

The following QSAR was obtained for the training set using linear regression analysis:

$\log 1/EC_{50} = 0.900\ \log P - 0.156\ LUMO - 0.369\ \Delta^1\chi^v - 3.03$ (I.4.16)

$n = 73,\ R^2 = 0.892,\ s = 0.496,\ F = 189,\ R^2_{cv} = 0.878$

The statistical parameters of this model have similar values to the parameters of the model for the whole data set (Equation I.4.15).

The correlation between the observed and the predicted toxicities using this equation for the compounds of the test set had $R^2 = 0.901$ (n = 18).

The best prediction in the KNN approach was obtained using k = 5-7. The correlation between the mean toxicities, predicted by the LOO procedure, for k = 5-7 of the compounds in the training set and the observed toxicities had $R^2 = 0.824$ and s = 0.623 (n = 73). Again, methidathion and pentabromophenol, and additionally methanol, were outside the 95% prediction interval. Better fit between observed and predicted (mean k = 5-7 neighbors) toxicity was obtained for the compounds in the test set: $R^2 = 0.941$ (n = 18).

The use of a test set showed that regression analysis is more stable in this kind of validation, resulting in similar R^2 values of the model for the training set and the correlation between the observed and the predicted toxicities of the test set. The KNN analysis resulted in a large difference in R^2 values (0.824 for the training set vs 0.941 for the test set). Cronin et al. (2004) suggested that the higher R^2 of the observed vs predicted toxicity of the test set is due to the fact that, by chance, the 3 poorly predicted compounds (methidathion, pentabromophenol, and methanol) were included in the training set. Additionally, Cronin et al. (2004) noted that the final result of the KNN analysis depends on the correct choice of the numbers of neighbors to be used for prediction of the toxicity. As this cannot be defined before the modeling, it contributes to the ambiguity of the method.

Cronin et al. (2004) also correlated the measured toxicity to *C. vulgaris* with the acute toxicity to other aquatic species such as the bacterium *V. fischeri* (n = 50, $R^2 = 0.577$), ciliate *T. pyriformis* (n = 73, $R^2 = 0.729$), and fish *P. promelas* (n = 42, $R^2 = 0.745$).

Lu et al. (2001) developed QSARs for toxicity to the alga *Scenedesmus obliquus* based on 40 substituted benzenes (the toxicity was determined as 48h-EC_{50} values for inhibition of algae

growth in units mol/l). The QSARs were obtained by using the response – surface approach, and are given in Table I.4.1.

Table I.4.1. QSARs for substituted benzenes to *S. obliquus* (reproduced from Lu et al., 2001)

Model No	log 1/EC50 = a*logP + b*LUMO + c						
	A	b	c	N	R^2	s	F
1	0.272	-0.659	2.54	40	0.793	0.316	71.07
2	0.237	-1.156	1.84	25	0.800	0.247	43.88
3	0.315	-0.567	2.51	23	0.810	0.299	42.54
4	0.563	-0.030	1.93	15	0.861	0.286	37.20

All compounds were included in model (1), nitrobenzenes only were included in model (2), anilines and phenols were included in model (3), and the anilines and phenols not containing nitro group in model (4).

The QSAR equations show that the toxicity of substituted benzenes to algae is related to their hydrophobicity and electronic properties. The LUMO values for nitro-containing compounds are negative. In the QSAR for these compounds (Model 2) the coefficient of logP (0.237) is the lowest, and the negative value of the coefficient of LUMO (-1.156) is the highest among the obtained QSARs. This suggests that the toxicity depends more on LUMO (encoding electronic affinity of the compounds) than on logP for these compounds in comparison to the remaining benzenes. For the compounds included in Model 4 (anilines and phenols, not containing nitro group) LUMO values are positive, the QSAR coefficient of logP 0.563 is the highest, the coefficient of LUMO has the smallest absolute value (-0.030). This suggests that their toxicity depends more on hydrophobicity than on electronic factors in comparison to the remaining compounds. This result could be ralted to polar narcotis as a mechanism of toxic action of these compounds.

A different approach was used by Cui et al. (2003). They applied the holographic quantitative structure-activity relationship (HQSAR) technique to develop QSARs for toxicity of 83 benzene derivatives to *C. vulgaris*. Toxicity (taken from Kramer and Trumper, 1986) was presented as the concentration that inhibits 50% of the population growth in a 6h assay (IC_{50} values in units mmol/l). The HQSAR analysis uses a molecular hologram, which encodes the frequency of occurrence of various molecular fragment types in a molecule, and subsequently

partial least squares (PLS) regression analysis. The development of the molecular hologram includes generation of all possible structural fragments (including branched, cyclic and overlapping fragments) of a given molecule by selecting a user-defined minimum and maximum number of atoms. The fragments are distinguished by fragment distinction parameters: atoms (distinction by atom types), bonds (distinction by bond order), connections (distinction by atomic hybridisation states, i.e., how many connections are made to the atoms in the fragment, and the bond order of those connections), hydrogen (inclusion or exclusion of hydrogen atoms during the fragment generation), and chirality (fragment distinction based on atomic and bond stereo chemistry) (Lowis, 1997). Each fragment is assigned a unique integer in the range of $0 - 2^{31}$. The integer identifier is then related to an element in an integer array of length L, named molecular hologram (L generally in the range of 50 to 500, Lowis, 1997). The same fragment integer identifier will always correspond to the same element of the hologram, but several fragment integer identifiers could be related to the same element of the hologram. This reduces the size of the hologram to a smaller than the number of fragments (Lowis, 1997). The values for the elements of the hologram are equal to the total number of fragments whose integer identifier corresponds to that element of the hologram. These values are used as predictor variables in the PLS statistical procedure.

Cui et al. (2003) investigated models with different fragment distinction parameters, different size of the fragments and different hologram size. They used LOO cross-validation procedure (R^2_{cv} and s_{cv}) to select a best-performing model. The robustness and predictive ability of the model were validated by leave-n-out (LNO) procedure excluding 12 (15%) of the compounds.

The PLS HQSAR model which had best statistical parameters from the LOO cross-validation procedure was derived by using atoms, bonds and connections as fragment distinction parameters, fragment size of 2-5 atoms, and hologram size of 199. It had five principal components, $R^2 = 0.912$, s = 0.165, Ratio = 6.44 %, $R^2_{cv} = 0.783$, $s_{cv} = 0.262$, $R^2_{LNO} = 0.782$, $s_{LNO} = 0.264$, and $Ratio_{LNO} = 10.3$ %. Ratio represents the ratio between the standard error of prediction to the variation range of the bioactivity. For good models this ratio should be less than 10%. The presented results showed that the HQSAR model had good fit and robustness.

Gramatica et al. (2001) developed QSARs for inhibition of reproduction of *Chlorella*. Toxicity data ($log1/EC_{50}$ values in units mmol/l) for 15 phenylureas and 18 triazines were

taken from Manthey et al. (1993), and Grimme et al. (1998). According to Gramatica et al. (2001) phenylureas and triazines are inhibitors of photosynthesis, inhibiting electron transport in the photosystem II (PSII). Their molecular site of action has been identified to be a subunit of the D1 protein of the PSII reaction centre.

The chemical structures of the compounds were described by a set of 171 molecular descriptors: molecular weight (MW), logP, 38 descriptors for counts of different kinds of atoms, bonds, functional groups, H-bond acceptors and donors, 34 topological descriptors including connectivity indices and information indices, 33 global weighted holistic invariant molecular (WHIM) descriptors and 66 directional WHIM descriptors. WHIM descriptors encode three-dimensional structural information for molecular shape and electronic structure (Todeschini and Gramatica, 1997). Directional WHIM are descriptors calculated along molecular axes.

Leave-one-out cross-validation procedure was performed, reporting the cross-validated R^2 (R^2_{cv}). Leave-n-out procedure (repeating 5000 times a random object selection, with 30% of objects left out at each step) was also performed (R^2_{LNO}).

The best QSAR models for phenylureas included logP, one distance information index and two directional WHIM descriptors, and had n = 15, R^2 = 94.7, s = 0.147, R^2_{cv} = 90.0, and R^2_{LNO} = 87.6. The best QSAR models for triazines included the second order kappa shape index and two directional WHIM descriptors, and had n = 18, R^2 = 89.7, s = 0.127, R^2_{cv} = 84.2, and R^2_{LNO} = 79.5.

LogP did not appear in the QSAR for triazines. The authors suggested that for these 18 compounds the parameters describing specific molecular features are more important than simple hydrophobicity. The molecular descriptors that appeared in the QSARs for phenylureas and triazines were different. This could suggest that although the mode of action (i.e., the specific interaction with a molecular target site) is similar, the toxicological behaviour (i.e., the complex sequence from exposure to toxicokinetic and toxicodynamic) of the two classes of chemicals can be different.

I.5. QSARs for toxicity to *Hydrozoa*

Hydrozoa are widely spread and prolific groups of aquatic animals. Most of them are marine inhabitants, but a fresh-water form is the common *Hydra*. The life cycle of Hydrozoans includes polyp and medusa body forms. The polyp is sessile and usually colonial, sexual budding can take place in this form but the individuals formed remain attached to each other. The medusa is a free-swimming form and represents the sexually reproducing, dispersal phase. *Hydra* is an exceptional member of this class in that it only occurs as a polyp, which can reproduce sexually.

Devillers et al. (2002a) developed QSARs for adult and developmental toxicities to *Hydra attenuata* of 17 chemical compounds (data were taken from Johnson et al., 1984; and Johnson et al, 1986). Developmental toxicity is measured on dissociated *Hydra* cells, which, if undisturbed, will achieve the developmental events characteristic of an embryo and undergo total whole-body regeneration. The chemical structures were described by using the autocorrelation method. The autocorrelation descriptors are simple two-dimensional molecular descriptors, designed from the hydrogen-suppressed graphs of the molecules. Autocorrelation vectors can be derived for all physicochemical properties, which can be encoded by atomic contributions. The autocorrelation vectors consist of components (autocorrelation descriptors) corresponding to different interatomic distances (i.e. 0-n). Components of distance zero encode the whole molecule, the other components encode mainly branching information and chemical specificity. In the study of Devillers et al. (2002a) autocorrelation vectors encoding lipophilicity (H), molar refractivity (MR), H-bonding acceptor ability (HBA) and H-bonding donor ability (HBD) were calculated. The QSAR model was derived by PLS analysis. The best results were obtained with the following components of the abovementioned autocorrelation vectors: H_0, H_1, H_4, H_5, H_7, MR_0, MR_1, MR_2, MR_3, HBA_2 and HBD_0. A PLS model with two components was obtained, which explained approximately 96 % of the variance in the toxicity.

Attempts to model the ratio of adult toxicity to developmental toxicity using the autocorrelation method were unsuccessful.

The same group of authors extended their study on QSARs for adult and developmental toxicities to *H. attenuata* of 30 chemicals (Devillers et al., 2002b). The toxicity data were

taken from Johnson and Gabel (1983), Johnson et al. (1984), Johnson et al. (1986), Mayura et al. (1991), and Bowden et al. (1995). To describe the chemical structures they used again autocorrelation descriptors, encoding lipophilicity, molar refractivity, the H-bonding acceptor and donor ability, and also indicator variables. The data set was divided into a training set of 26 chemicals and a test set of 4 chemicals. The best PLS model was obtained with the autocorrelation components H_0, H_1, H_2, H_5, MR_0, MR_1,MR_2, MR_3, HBA_2 and HBD_0, and also nCl (the number of chlorine atoms). It had three components, which explained approximately 96 % of the variance in the two types of toxicity. No large outliers for the two endpoints were observed in the training and test sets (residuals less than 0.75 log-units).

I.6. QSARs for toxicity to *Daphnia*

Liu et al. (2003a) developed QSARs for acute toxicity to *D. magna* (the toxicity was determined as the 50% immobilisation concentrations in 48-h tests, EC_{50} values in μmol/l) of 20 α-substituted phenylsulphonyl acetates. A *D. magna* was considered immobile when it was unable to swim for more than a few strokes within 15 s after gentle shaking of the test vessel. The descriptors used to develop QSARs were octanol-water partition coefficient and water solubility, TLSER descriptors and Charge Model descriptors. The obtained QSARs are given in Table I.6.1.

Table I.6.1. Regression models of the toxicity of chemicals to *D. magna* (reproduced from Liu et al., 2003a)

No.	Equation (n = 20)	R_A^2	s	F
1	$\log EC_{50} = -0.193\ \log P + 2.384$	0.668	0.139	39.26
2	$\log EC_{50} = 0.185\ \log S_W + 2.684$	0.500	0.170	19.97
3	$\log EC_{50} = -0.494\ V_{mc} - 3.401\ \pi^* - 12.81\ \varepsilon_b + 7.260$	0.877	0.084	46.08
4	$\log EC_{50} = -0.004\ MW + 1.190\ Qo - 0.025\ Qs^2 + 3.467$	0.922	0.067	76.14

where:

logP is the logarithm of the octanol-water partition coefficient; $\log S_W$ is the logarithm of the aqueous solubility; MW is the molecular weight;

TLSERs based descriptors: V_{mc} is $V_m/100$; V_m is molecular volume; π^* is equal to α/V_m; covalent contribution to Lewis basicity ε_b (H-bonding term) is equal to the difference in energy between LUMO of water and HOMO of solute;
Charge Model descriptors: Qo is the square root of sum of the squared net charges on oxygen atoms; Qs^2 is the sum of the squared net charges on sulphur atoms.

The QSARs showed that hydrophobicity, combined with geometric, and electronic descriptors, could be used to model toxicity to *D. magna*.

Tao et al. (2002b) investigated the toxicity to *D. magna* (48-h 50% effective concentrations in mmol/l) for 217 chemicals by using a fragment constant method. The toxicity data were taken from Montgomery (1993), and Tomlin (1994). The 217 chemicals were divided into 10 classes: carbamates, pyrethrins, acetyl chlorides, amines, derivatives of urea, alcohol/ketones, heterocyclic nitrogen compounds, acids, organophosphorus pesticides, and others.

To derive QSARs, basic chemical fragments were defined on the basis of the chemicals in the dataset. According to Leo (1975) a fragment is an atom, or atoms, whose exterior bonds are to isolating carbon atoms. An isolating carbon is either one that has four single bonds, at least two of which are to non-hetero atoms, or is multiply bonded to other carbon atoms. A single atom fragment can only be an isolating carbon atom, a hydrogen, or a hetero atom, while a multiple-atom fragment can be formed by any combination of non-isolating carbon, hydrogen, and/or hetero atoms.

From the data set of 217 chemicals, Tao et al. (2002b) identified 103 different fragments, taking into account also attachment type of the fragment. Twenty-one structural correction factors were also introduced into the analysis. They accounted for structural features like molecular flexibility, unsaturation, multiple halogenation, branching (Leo, 1982). A linear multivariate regression analysis was then used to derive the fragment constant model. The numbers of fragments and their structural features that appear in a compound were taken as independent variables; the dependent variable was the logarithmically transformed reciprocal value of EC_{50}. A general equation with the following form was derived:

$$\log \frac{1}{EC_{50}} = \sum_{i=1}^{a} n_i f_i + \sum_{j=1}^{b} m_j F_j \qquad \text{(I.6.1)}$$

where a and b are the total number of fragments and structural features; n_i and m_j are the numbers for the ith fragment and the jth structural feature of the chemical; f_i is the fragment constant for the ith fragment, and F_j is the structural correction factor for the j-th structural feature.

The 217 chemicals in the dataset were randomly divided into a training set (200 chemicals) and a validation set (17 chemicals). The regression model based on the training set accounted for 96.9% of the variation in the experimental data for the 200 chemicals, with a mean residual for the model of 0.38 log-units. When the model was applied to the validation data set of 17 compounds, it showed good predictivity for 16 of them with mean residual of 0.37 log-units. Desmedipham was predicted with a large residual of -4.04. The reason for this was that desmedipham has a specific fragment that did not appear in any of the compounds from the training set. The fragment constant for desmedipham was therefore taken as zero during the model's validation.

A final model was derived from the whole data set of 217 chemicals. The coefficient of determination R^2 for the final model was 0.969, the mean residual was 0.4 log-units.

The robustness of the model was evaluated using three different types of a modified jackknife test: (1) deletion of one randomly selected individual chemical at a time for 30 trials; (2) deletion of subsets of 20 randomly selected chemicals at a time for 30 trials; (3) deletion of a class of chemicals each time for all 10 classes. For each trial, the model was developed based on the non-deleted compounds, and the coefficient of determination R^2 was calculated.

For the first type of the jackknife test (deletion of one chemical) the obtained R^2 values varied from 0.9682 to 0.9700, with a small range of 0.0018, suggesting that the model is generally robust in respect to deletion of individual chemicals. For the second type of the jackknife test (deletion of 20 chemicals) the R^2 values varied from 0.9679 to 0.9734, with a range of 0.0055, and for the jackknife test where deletion of a chemical class was performed, the R^2 values varied from 0.9668 to 0.9808, covering a range of 0.140. The average number of chemicals that were excluded from the models for this type of the jackknife test (the average

number of chemicals in the different chemical classes) was 21.7, which allowed direct comparison to the results of the second type of the jackknife procedure (exclusion of 20 compounds). The range of the obtained R^2 values with the chemical class-deletion procedure (0.140) was considered by Tao et al. (2002b) to be significantly higher than that derived from the procedure when 20 chemicals were deleted (0.0055). According to Tao et al. (2002b) this indicates that the model is more suitable for prediction of compound toxicity within the chemical classes, than among chemical classes not used to derive the model.

LogP values for 190 compounds of the data set investigated by Tao et al. (2002b) were available, however the authors did not attempt a QSAR model using this descriptor.

Faucon et al. (2001) investigated 96 compounds for acute toxicity to *D. magna* (effective concentration causing 50% immobilisation of *Daphnia* after 48-h exposure). The dataset was taken from French, German and English notification files of New Chemicals registered in 1994, 1995 and 1996. The dataset was divided into a training set of 61 compounds and a test set of 35 compounds.

The QSAR obtained contained logP and the ‘hardness' (Ha) equal to 1/2 (HOMO - LUMO):

$$\log 1/EC_{50} = 0.27\ \log P + 0.63\ Ha + 6.64 \qquad \text{(I.6.2)}$$

$n = 61,\ R^2 = 0.54,\ s = 0.71,\ R^2_{cv} = 0.49$

According to Pearson (1963) compounds with a high absolute value of `hardness' (which takes negative values) in general have small atomic radius, high effective nuclear charge and low polarisability. In the study of Faucon et al. (2001), the decrease (increase of polarisability) of the absolute value of `hardness' led to an increase in the toxicity.

When *Daphnia* toxicity of the compounds of the test set was predicted, $R^2 = 0.57$ and $s = 0.77$ were derived. This R^2 value is close to the R^2 value of the model.

The stability of the model was also checked by carrying out 10 times a random selection of compounds from the dataset in order to define 10 different training (61 compounds) and test (35 compounds) sets. Models were developed with the training sets, and toxicity values were

predicted for the respective test sets. The statistical results showed similarity between the models, which were therefore relative independent of the composition of the training set.

The quality of the QSAR equation was tested also with 30 compounds randomly chosen from the AQUIRE database (http://www.epa.gov/med/databases/aquire.html). The predicted *Daphnia* EC_{50} values were worse ($R^2 = 0.48$, s = 0.87) than those observed for the training and test sets. For this set, the authors observed a stronger correlation between the *Daphnia* acute toxicity and logP alone ($R^2 = 0.58$, s = 0.77).

In general, the above-mentioned models had moderate statistical fit. According to Lessigiarska et al. (2004) the data from the notification files of New Chemicals are not sufficiently homogeneous to develop high quality QSARs for toxicity to daphnids. Therefore it might be accepted that Faucon et al. (2001) obtained a realistic statistical fit for this type of data. For a subset of chemicals with particular structures with overpredicted *Daphnia* EC_{50} values (overestimated by approximately a factor of 10), the authors observed a better relationship between their *Daphnia* acute toxicities and the logP values alone:

$$\log 1/EC_{50} = 0.38 \log P + 2.25 \qquad \text{(I.6.3)}$$

$n = 16$, $R^2 = 0.80$, $s = 0.38$

I.7. QSARs for fish toxicity

A number of QSAR studies on chemical toxicity to different fish species were found in the literature. QSARs for fish toxicity can be used in the evaluation of chemical ecotoxicity and risk assessment.

Huuskonen (2003) investigated the use of atom-type electrotopological state (E-state) indices for predicting toxicity to fathead minnow (*Pimephales promelas*) (expressed as the concentration causing death of 50% of the fish after a 96h exposure period, LC_{50} values; the data were taken from Gao et al., 1992) for a diverse set of 140 organic chemicals. He applied a group contribution approach, in which a compound is divided into fragments, and the endpoint of interest is calculated by the summation of the contributions of each fragment. In this particular study, atom-type E-state indices for different groups were used as fragment contributions. According to Rose et al. (2002), the E-state value of an atom encodes

information about the electron accessibility at that atom. Thus, the E-state values of the groups in a molecule are thought to account for the ability of molecule to enter into non-covalent intermolecular interactions.

In the investigation of Huuskonen (2003), the data set was divided into a training set of 130 compounds for developing the QSAR models, using multiple linear regression and artificial neuron network approaches, and a test set of 10 compounds for evaluating the predictive ability of the models.

The multiple linear regression model obtained was the following:

$$-\log LC_{50} = \Sigma (a_i S_i) + 0.918 \qquad \text{(I.7.1)}$$

$n = 130$, $R^2 = 0.84$, $s = 0.36$, $F = 41.8$, $R^2_{cv} = 0.83$, $s_{press} = 0.38$

In this equation, a_i and S_i are the regression coefficients and corresponding structural parameters for a set of 14 atom-type E-state indices.

Huuskonen also applied an artificial neural network approach to detect the presence of non-linear dependencies of the toxicity on the E-state indices. The model obtained had a slightly higher statistical fit with $n = 130$, $R^2 = 0.88$, $s = 0.31$.

Toropov and Toropova (2002) investigated the acute toxicity (in terms of LC_{50} values) to fathead minnow (*P. promelas*) of 69 benzene derivatives (data were collected from Gute and Basak, 1997; and Basak et al., 2000). They used a similar approach to that of Castro et al. (2003) and Toropov and Schultz (2003) in the investigation of toxicity to *T. pyriformis* (see above), namely the optimisation of correlation weights of local graph invariants (OCWLGI) approach. The best model obtained was based on the Morgan extended connectivity of third order. The model statistics were the follows: $n = 69$, $R^2 = 0.900$, $s = 0.244$, $F = 605$.

Toropov and Benfenati (2004) used the same approach (OCWLCI) to investigate the toxicity of 51 aldehydes to *P. promelas* (expressed as compound concentrations causing death for 50% of the animals, LC_{50} values in units mmol/l, data were taken from the US EPA ECOTOX database [http://www.epa.gov/ecotox/], and the US EPA MED-Duluth Fatherad Minnow Database [Russom et al., 1997]). Toropov and Benfenati (2004) used again Morgan

extended connectivity, and the so-called nearest neighbouring codes (NNCs) as local graph invariants. The NNC on a given k-th vertex of the labelled hydrogen-filled graph (LHFG) (athom in the molecule) is defined as follows:

$$NNC_k = 100\ N_t + 10\ N_c + N_h \qquad \text{(I.7.2)}$$

where N_t is the total number of neighbours, N_c is the number of neighbours which are images of carbon atoms, and N_h is the number of neighbours which are images of hydrogen atoms.

The NNC values are functions of both the chemical composition of neighbouring vertexes of given vertex in the LHFG and the total number of neighbouring vertexes.

The best model obtained was based on the NNC_k values and had the following statistics: n = 51, R = 0.828, s = 0.465, F = 107.

The authors note that the QSAR includes aliphatic and aromatic saturated and unsaturated aldehydes. The model predicts well the toxicity of these different types of aldehydes, especially aliphatic and unsaturated ones, which are more reactive and toxic.

Romanelli et al. (2000) investigated the toxicity of alcohols on fathead minnows (*P. promelas*) (toxicity expressed as $logLC_{50}$ values). They developed QSARs by using logP, molecular surface area, molecular volume, molar refractivity, and polarisability as molecular descriptors. The following model was derived with logP:

$$logLC_{50} = -1.33\ logP + 2.60 \qquad \text{(I.7.3)}$$
$$n = 12,\ R^2 = 0.993,\ s = 0.0246$$

Adding more descriptors improved the model R^2 up to $R^2 = 0.995$ and s down to s = 0.022. The authors also report models that include quadratic and cubic terms of the molecular descriptors, which improved the model R^2 up to $R^2 = 0.999$, but s increased up to s = 0.026. Some of the models in the work of Romanelli et al. (2000) have insufficient data points (n = 12) in relation to the included variables (up to 5). Also, one could notice that logP alone is sufficient for modelling the toxicity of the data set, giving correlation coefficient $R^2 = 0.993$, and adding more variables in order to improve the R^2 with only 0.006 could be considered as

unnecessary. Additionally, having a statistca fit with R^2 of 0.995 and s of 0.022 is beyond the statistical error of the test.

Freidig and Hermens (2000) developed QSARs for toxicity to fish (14-day LC_{50} for guppy (*Poecilia reticulata*) and 4-day LC_{50} for fathead minnow (*P. promelas*) (data were taken from Deneer et al., 1987; DeBruijn and Hermens, 1993; and Freidig et al., 1999). The compounds were divided into narcotics and reactive compounds, and for each group a separate one-parameter QSAR was developed. The criterion for this dividing the compounds was based on the ratio of excess toxicity, Te:

Te = observed toxicity / predicted toxicity by a given baseline equation (I.7.4)

The Te values for compounds investigated by Freidig and Hermens (2000) were taken from the literature (Verhaar et al., 1992; Russom et al., 1997). Freidig and Hermens (2000) defined that if the Te value of a compound is < 5, it should be considered as a narcotic, if Te > 5 it was considered a reactive chemical.

The developed QSARs are given in Table I.7.1. For a set of organophosphate esters (OP-esters), nine out of 20 compounds were identified as possible narcotic compounds and their toxicity was described with a narcosis QSAR (logP relationship, R^2 = 0.91, Table I.7.1). As an electronic descriptor to describe the 11 remaining compounds, identified as reactive, the measured the *in vitro* 2nd order acetylcholinesterase inhibition rate of oxon-analogues (k_i) was used. A good correlation with acute toxicity was found (R^2=0.68). When the two descriptors were combined together in a single QSAR for the whole data set (n = 20) R^2 was found to be 0.49 (Table I.7.1). The use of two separate QSARs for the compounds acting by a given mechanism of toxicity gave better overall R^2 of 0.80 compared to the combined QSAR for the whole data set (R^2 = 0.49).

Another set of 26 nitrobenzenes was investigated by Freidig and Hermens (2000). As a descriptor for reactivity, the sum of Hammett constants (σ^-) for the substituents was used. The correlation with logP for the narcotic subset (9 chemicals) gave an R^2 of 0.97 (Table I.7.1). The correlation with σ- for the reactive compound subset (17 chemicals) gave an R^2 of 0.74. A two-variable QSAR for the whole data set (n = 26) had an R^2 of 0.75 (Table I.7.1). The use of the two separate QSARs gave a better overall R^2 of 0.83.

In an investigation of another data set of 15 α,β-unsaturated carboxylates the 2nd order reaction rate with glutathione (k_{GSH}) was used as the reactivity descriptor. The correlation with logP for the narcotic subset (5 chemicals) gave an R^2 of 0.93 (Table I.7.1). The correlation with k_{GSH} for the reactive subset (10 chemicals) gave an R^2 = 0.82. The two-variable QSAR for the whole data set (n = 15) had an R^2 of 0.82 (Table I.7.1). The use of the two separate QSARs again gave a better overall R^2 of 0.86.

Table I.7.1. QSAR models obtained by Freidig and Hermens (2000) and squared coefficients of determination obtained from predicted toxicity values using mixed and separate approaches for QSAR analysis.

	QSARs used to predict LC_{50}	R^2 for predicted vs. observed LC_{50}
OP-esters		
mixed	$\log LC_{50} = -0.28 \log P - 0.45 k_i + 3.66$	0.49
separate		0.80
narcosis	$\log LC_{50} = -0.66 \log P + 3.66$	
reactive	$\log LC_{50} = -0.62 k_i + 3.59$	
Nitrobensenes		
mixed	$\log LC_{50} = -1.07 \log P - 0.18 \sigma^- + 2.53$	0.75
separate		0.83
narcosis	$\log LC_{50} = -1.03 \log P + 4.72$	
reactive	$\log LC_{50} = -1.00 \sigma^- + 2.07$	
α,β-unsaturated carboxylates		
mixed	$\log LC_{50} = -0.30 \log P - 0.67 k_{GSH} + 2.67$	0.82
separate		0.86
narcosis	$\log LC_{50} = -1.25 \log P + 5.25$	
reactive	$\log LC_{50} = -0.68 k_{GSH} + 2.45$	

The authors concluded that using separate QSAR models for compounds acting by different mechanisms, each model including a descriptor that characterises the particular toxicity mechanism, gives better results than using a single model that combines all compounds and descriptors.

Rose and Hall (2003) used electrotopological state (E-state) indices to model toxicity to the guppy (*P. reticulata*). The negative logarithm of the lethal concentration that reduces the fish

population by 50% (pLC_{50} value) was used as the response variable for 25 phenols, 31 anilines and 36 substituted aromatic hydrocarbons (benzenes, toluenes and xylenes). The data were taken from Verhaar et al. (1995b), who compiled them from different sources. The mechanism of toxic action of the compounds is mainly non-polar and polar narcosis.

According to Rose and Hall (2003), atom-type E-state indices encode information about the electron accessibility for atoms of the same type, the presence/absence of an atom type and the count of the atoms of a given atom type. The following QSARs were obtained:

Full set of compounds:

$$pLC_{50} = 0.559\ {}^{1}\chi^{v} + 0.0610\ S^{T}(\text{-Cl}) + 0.280\ H_{max} - 0.0726\ S^{T}(\text{-O-}) + 1.58 \qquad \text{(I.7.5)}$$

$n = 92,\ R^2 = 0.87,\ s = 0.25,\ F = 140,\ R^2_{cv} = 0.85,\ s_{cv} = 0.26$

Phenols:

$$pLC_{50} = 0.703\ {}^{1}\chi^{v} + 0.0840\ S^{T}(\text{-Cl}) - 0.0834\ S^{T}(\text{-O-}) + 1.78 \qquad \text{(I.7.6)}$$

$n = 25,\ R^2 = 0.91,\ s = 0.17,\ F = 74,\ R^2_{cv} = 0.87,\ s_{cv} = 0.21$

Substituted aromatic hydrocarbons:

$$pLC_{50} = 1.16\ {}^{1}\chi^{v} + 0.0733\ S^{T}(\text{-Cl}) + 0.385\ HS^{T}(\text{other}) + 1.81 \qquad \text{(I.7.7)}$$

$n = 36,\ R^2 = 0.92,\ s = 0.21,\ F = 118,\ R^2_{cv} = 0.90,\ s_{cv} = 0.25$

Anilines:

$$pLC_{50} = 0.177\ {}^{1}\chi^{v} + 0.0792\ S^{T}(\text{-Cl}) + 0.263\ {}^{3}\kappa_{\alpha} + 2.62 \qquad \text{(I.7.8)}$$

$n = 31,\ R^2 = 0.82,\ s = 0.26,\ F = 42,\ R^2_{cv} = 0.78,\ s_{cv} = 0.29$

${}^{1}\chi^{v}$ is the valence-corrected first-order molecular connectivity. It decreases with increasing skeletal branching, but increases with the number of skeletal atoms (Rose and Hall, 2003). According to the QSAR equations, increased branching and decreased general size decrease the toxicity. $S^{T}(\text{-Cl})$ is the atom-type E-state descriptor for chlorine atoms and encodes

information about the electron accessibility of chlorine atoms in each molecule. Electropositive atoms near to chlorine tend to increase its value. According to the QSAR equations, more chlorine atoms and more polar chlorines increase the toxicity. This result might be also related to the increased hydrophobicity of the aromatic compounds with increasing the number of the chlorine atoms. However, it should be noted that the coefficient for the S^T(-Cl) in the QSAR for anilines is not statistically significantly different from zero (its standard error had a value of 0.087). H_{max} is the largest hydrogen E-state atom value for the molecule, encoding the hydrogen atom accessibility for the most accessible hydrogen atom. It appears on the most polar X-H bond in the molecule. S^T(-O-) is the atom-type E-state descriptor for ether atoms, which encodes the electron accessibility for ether atoms. HS^T(other) is the atom-type descriptor for the hydrogen atoms. $^3\kappa_\alpha$ is the third-order kappa alpha shape index, encoding information about the shape of the molecule and the presence of heteroatoms.

The model based on the full data set was validated by two types of cross-validation procedures. The results from leave-one-out cross-validation are reported in the equation. The second cross-validation procedure was a 10% leave-out method: at each run, 10% of the compounds (nine compounds) were excluded and a model with the same predictor variables was generated from the remaining 83 chemicals. The toxicities of the nine compounds excluded were predicted by the model. This procedure was repeated ten times on ten unique subsets of nine compounds, until each compound in the data set had been predicted once. The whole process was repeated three times in a random manner, resulting in three predictions for each observation. The quality of prediction was assessed using the mean absolute error (mean of the absolute values of the prediction errors, MAE) value of the prediction in log-units. The MAE from the cross-validation procedure was found to be 0.19. According to Rose and Hall (2003) this level of MAE for a 10% leave-out cross-validation procedure is a good confirmation of the predictive quality of the model. In addition to the cross-validation procedures, the model based on the full data set was validated using two external test sets. The first one consisted of 5 compounds, consisting of toxicity data for 4 compounds to the crustacean *D. magna* and for 1 compound to the pond snail *Lymnaea stagnalis*. The MAE obtained after predicting the toxicity of these compounds with the model based on the full data set, was 0.45. The second data set had toxicity data to the fish *P. promelas* for 12 compounds, being present in the training set for the full set model. The MAE obtained was

0.19. When toxicity to the fish *P. promelas* of other 5 compounds not found in the training set of the full set model were examined, the MAE was 0.44. On this basis, it was concluded that the quality of the model is reasonable.

Roberts and Costello (2003) developed QSARs for toxicity of 18 non-polar and polar narcotics to *P. reticulata* (guppy) using logP (octanol-water) and $\log K_{MW}$ (membrane-water) partition coefficients. Toxicity was expressed as the negative logarithm of LC_{50} values (pLC_{50}). The aim of the study was to examine the validity of the statement of Vaes et al. (1998) that there is no real difference between non-polar and polar narcosis mechanism of toxicity, and the apparent distinction disappears when membrane-water partition coefficients K_{MW} are used for QSAR modelling instead of octanol-water partition coefficients. The difference between non-polar and polar narcosis is observed because the octanol-water partition coefficient is not an adequate descriptor of partitioning into a lipid membrane (Vaes et al., 1998).

Roberts and Costello (2003) obtained the following QSARs:

Non-polar narcotics:

$$pLC_{50} = 0.84 \log P + 1.12 \qquad \text{(I.7.9)}$$
$$n = 8,\ R^2 = 0.971,\ s = 0.24,\ F = 199$$

$$pLC_{50} = 0.80 \log K_{MW} + 1.37 \qquad \text{(I.7.10)}$$
$$n = 8,\ R^2 = 0.975,\ s = 0.22,\ F = 238$$

LogP and $\log K_{MW}$ intercorrelated with $R^2 = 0.996$.

Polar narcotics:

$$pLC_{50} = 0.76 \log P + 1.98 \qquad \text{(I.7.11)}$$
$$n = 10,\ R^2 = 0.890,\ s = 0.29,\ F = 64$$

$$pLC_{50} = 0.88 \log K_{MW} + 1.36 \qquad \text{(I.7.12)}$$
$$n = 10,\ R^2 = 0.967,\ s = 0.16,\ F = 234$$

LogP and $logK_{MW}$ intercorrelate with $R^2 = 0.936$.

LogP based equations for non-polar and polar narcotics were different. $LogK_{MW}$ based QSARs for non-polar and polar narcotics were similar. That is why Roberts and Costello (2003) derived a general equation for polar and non-polar narcotics, based on $logK_{MW}$:

$$pLC_{50} = 0.84\ logK_{MW} + 1.38 \qquad \text{(I.7.13)}$$
$$n = 18,\ R^2 = 0.963,\ s = 0.21,\ F = 419$$

Roberts and Costello (2003) added an indicator variable G to this equation, with a value 1 being assigned for the non-polar narcotics and 0 for the polar narcotics:

$$pLC_{50} = 0.82\ logK_{MW} - 0.22\ G + 1.54 \qquad \text{(I.7.14)}$$
$$n = 18,\ R^2 = 0.973,\ s = 0.19,\ F = 274$$

According to the authors the equation improved slightly after adding of the indicator variable, although an increase of the square of the coefficient of determination R^2 of 0.01, and a decrease of the standard error of estimate s of 0.02 could be insufficient for stating of an improvement in the model, and could be a result only from the adding one more parameter to the regression model.

The indicator variable G was significant at 3% level (presumably coresponding to p value of 0.03), and the authors concluded that with 97% confidence the differences between polar-and non-polar narcotics are not completely explained by differences in $logK_{MW}$ values. The authors propose that this difference could be based on a difference in mechanism of interaction of the non-polar and polar narcotics with the membrane. According to Roberts and Costello (2003), non-polar narcotics act via 3-D partitioning (they are able to move in all directions in the hydrocarbon-like interior of the membrane), while polar narcotics act via 2-D partitioning (involving binding between a functional group on the narcotic and the polar phosphatidyl choline head groups at the membrane surface, which allows for moving of the polar narcotic only in two directions).

Tao et al. (2002a) used the same approach as Tao et al. (2002b) (see QSARs for *Daphnia* toxicity), to developed a fragment constant method for the prediction of toxicity to rainbow trout (*Oncorhynchus mykiss*), based on the experimental LC_{50} (96 h, mmol/l) values of 258 compounds. The compounds were divided into 11 classes: formates (37 compounds), pyrethroids (16 compounds), acetyl chlorides (11 compounds), amines (24 compounds), ureas (17 compounds), alcohols, phenols and ethers (8 compounds), ketones (24 compounds), heterocyclic nitrogen and triazines (44 compounds), acids (23 compounds), organophosphorus pesticides (38 compounds), and others (16 compounds). The toxicty data were collected from Montgomery (1993) and Tomlin (1994). Again, a multivariate regression analysis was applied to derive a model with the following form (Equation I.7.15):

$$\log \frac{1}{LC_{50}} = \sum_{i=1}^{a} n_i f_i + \sum_{j=1}^{b} m_j F_j,$$

where a and b are the total numbers of the atomic or group fragments and structural features defined by the model; n_i and m_j are the numbers of the i^{th} fragment and the j^{th} structural factor of the chemical; f_i is the fragment constant for the ith fragment and F_j is the structural correction factor for the j^{th} structural feature.

The square of the coefficient of determination of the final model was $R^2 = 0.902$. The authors note that the ratio of the number of cases to the number of parameters in the models is low because of the limited experimental data (258 data points, and approximately 90 parameters in the equation).

The mean residual of the differences between the measured and calculated LC_{50} values was 0.42 log-units. The residuals were, in general, symmetrically distributed around zero. They were dependent on the hydrophobicity of the chemicals. All chemicals with relatively high residuals also had high logP values, suggesting that the model is more applicable for more hydrophilic compounds. Two-thirds of the 24 chemicals with residuals greater than 1.0 log-unit belonged to the organophosphorus pesticides (7 compounds), heterocyclic nitrogen compounds (5 compounds), and acids (4 compounds). The model appeared to be more suitable for the remaining classes. The authors suggested that one source of the error of prediction could be the errors in the experimental LC_{50} values, because differences have been observed in reported LC_{50} values for the same chemical in the literature.

The robustness of the model was evaluated using three different jackknife tests: (1) deletion of subsets of 20 randomly selected chemicals at a time for 30 trials; (2) deletion of individual chemicals having a higher residual than 1.0 log-unit, one at a time for the 24 chemicals; and (3) deletion of a class of chemicals each time for all 11 classes. For each trial, the model was developed based on the non-deleted compounds, and the coefficient of determination R^2 was calculated. On the basis of these calculated coefficients of determination the sensitivity of the model on the excluded compounds was judged. The model appeared to be generally robust in respect to deletion of individual chemicals. Deletion of the chemicals with residuals higher that 1.0 log-units gave better models, suggesting the negative impact of these chemicals on the prediction model. With the exception of pyrethroids and ureas, deletions of chemicals belonging to a particular chemical class generally improved the jackknifed R^2 values, suggesting that the model is more suitable for prediction of chemical toxicity within the 11 studied chemical classes than for extrapolation to chemicals belonging to other classes.

Di Marzio et al. (2001) investigated the toxicity of nine aromatic hydrocarbons (non-polar narcotics) to the fish *Cnesterodon decemmaculatus*. Toxicity was measured as LC_{50} values (lethal concentration to 50% of the organisms) in a 96 h assay. Additionally, LC_{50} values for toxicity to fish species *O. mykiss* (rainbow trout) and *P. reticulata* (guppy) (Galassi et al., 1988), and for toxicity to fish *P. promelas* (fathead minnow) (Geiger et al., 1990) were used to derive QSARs.

The QSARs were derived using logP and molecular weight (MW) as descriptors, together with 113 weighted holistic invariant molecular (WHIM) descriptors, calculated by using the WHIM package (Todeschini, 1996). Thirty seven global and 66 directional descriptors were calculated, defined on molecular size, molecular shape, molecular symmetry, atom distribution and density and molecular globe size.

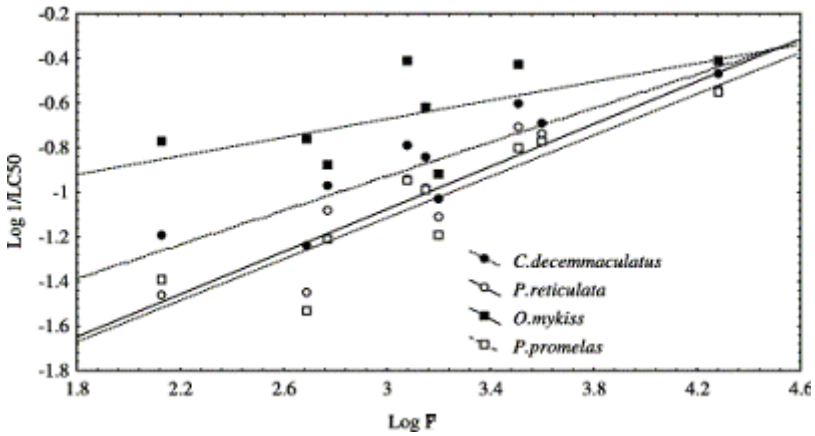
Good correlations between toxicity and logP were found for all species except *O. mykiss*. (*P. reticulata*: $R^2 = 0.858$, $R^2_{cv} = 0.775$; *P. promelas*: $R^2 = 0.812$, $R^2_{cv} = 0.691$; *C. decemmaculatus*: $R^2 = 0.804$, $R^2_{cv} = 0.705$; *O. mykiss*: $R^2 = 0.405$, $R^2_{cv} = 0.089$).

Better two-variable models were obtained using WHIM descriptors. The models had the following statistics: *P. reticulata*: $R^2 = 0.961$, $R^2_{cv} = 0.928$; *P. promelas*: $R^2 = 0.913$, $R^2_{cv} = 0.838$; *C. decemmaculatus*: $R^2 = 0.942$, $R^2_{cv} = 0.907$; *O. mykiss*: $R^2 = 0.881$, $R^2_{cv} = 0.784$.

The authors note that the two-variable models had better statistics, but models with logP were easier to interpret compared with those based on WHIM descriptors.

The figure shows the relationships between toxicity and logP for the four fish species considered in this study. It can be seen that the greatest differences between species are observed at logP values lower than 4. According to Di Marzio et al. (2001), this could be due to differences in the ability of chemicals to reach the site of action, since the mechanism of action for the investigated chemicals is similar (non-polar narcosis) in the four species.

Figure I.7.1. Relation between logP and log1/LC_{50} for fish species (taken from Di Marzio et al., 2001)



Ownby and Newman (2003) investigated toxicity data for four fish, carp (*Cyprinus carpio*), mummichog (*Fundulus heteroclitus*), bluegill (*Lepomis macrochirus*) and fathead minnow (*P. promelas*), taken from the US EPA ECOTOX database system (http://www.epa.gov/ecotox/). The toxicity data (96 h LC_{50} values in units mol/l) for eight metal chloride salts (Ag, Cd, Co, Cu, Hg, Ni, Pb, and Zn) were used.

The structural descriptors and statistics of the developed QSARs (with logLC_{50} used as dependent variable) are presented in Table I.7.2 (equations were not given):

Table I.7.2. Structural descriptors and statistical parameters of the QSARs developed by Ownby and Newman (2003)

Species	Metal ions	Descriptor	n	R^2	Dev [a]	F	p
C. carpio (carp)	Cd, Co, Cu, Hg, Ni, Zn	σ_p	6	0.56	18.6	5.01	0.089
L. macrochirus (bluegill)	Cd, Cu, Fe, Hg, Ni, Pb, Zn	σ_p	7	0.72	9.1	12.6	0.016
P. promelas (fathead minnow)	Ag, Cd, Co, Fe, Hg, Ni, Pb, Zn	σ_p	8	0.79	7.6	22.8	0.003
F. heteroclitus (mummichog)	Cd, Co, Cu, Hg, Ni, Zn	σ_p	6	0.64	15.0	7.07	0.056
F. heteroclitus (mummichog)	Cd, Co, Cu, Hg, Ni, Zn	$\|\log K_{OH}\|$	6	0.78	14.6	14.5	0.019

[a] Dev is the mean relative deviation from perfect fit expressed as a percentage, [(observed-fitted)/observed]*100.

where σ_p is the softness index; $|\log K_{OH}|$ is the logarithm of the first hydrolysis constant.

The softness index (σ_p) reflects metal ion softness, or the tendency for the outer electron shell to deform (i.e., polarisability), and the ion's tendency to share electrons with ligand donor atoms. The absolute value of the logarithm of the first hydrolysis constant ($|\log K_{OH}|$) reflects the metal affinity for ligands such as those with O donor atoms. Softness produced the best model for the freshwater fish (carp, bluegill, and fathead minnow), and $|\log K_{OH}|$ produced the best model for the saltwater fish (mummichog).

Xu et al. (2002) developed CoMFA and CoMSIA models for toxicity of 35 nitroaromatic compounds to *golden orfe fish* (data taken from Lang et al., 1998). The CoMFA and advanced CoMFA (using also H-bond field) models derived are given in the Table I.7.3.a. In advanced CoMFA the H-bond acceptor field is usually labelled as a "steric" field type and the H-bond donor field – as an "electrostatic" field type. A summary of the obtained CoMSIA models is given in Table I.7.3.b.

Table I.7.3.a. Summary of CoMFA models: CoMFA and advanced CoMFA (reproduced from Xu et al., 2002)

	R^2_{cv}	s_{press}	N [a]	R^2	s	F	Steric	Elec [b]	H (steric) [c]	H (elec) [d]
CoMFA (1)	0.533	0.543	4	0.870	0.293	50.2	0.496	0.504	-	-
CoMFA (2)	0.701	0.468	7	0.955	0.181	82.7	-	-	0.814	0.186
CoMFA (3)	0.727	0.432	5	0.944	0.195	97.9	0.188	0.286	0.415	0.111

[a] number of principal components in the PLS analysis; [b] relative contribution of electrostatic fields; [c] relative contribution of H-bond steric (acceptor) fields; [d] relative contribution of H-bond electrostatic (donor) fields.

Table I.7.3.b. Summary of CoMSIA models (reproduced from Xu et al., 2002)

	R^2_{cv}	s_{cv}	N [a]	R^2	s	F	Steric	Elec [b]	Hydro [c]	H-don [d]	H-acc [e]
CoMSIA (1)	0.827	0.344	5	0.930	0.218	77.4	0.053	0.228	0.322	0.167	0.231
CoMSIA (2)	0.830	0.341	5	0.931	0.217	78.5	0.030	0.252	0.338	0.172	0.208

[a] number of principal components of the PLS analysis; [b] relative contribution of electrostatic fields; [c] relative contribution of hydrophobic fields; [d] relative contribution of H-bond donor fields; [e] relative contribution of H-bond acceptor fields.

The column filtering of CoMSIA(1) model was set to 1.0 kcal/mol and that of CoMSIA(2) was 0.0 kcal/mol. Higher column filtering results in rejecting more variables as noisy in the PLS analysis of the CoMFA and CoMSIA approaches.

On the basis of the highest R^2_{cv} and the lowest s_{cv}, the CoMFA(3) model (optimal number of components 5) was chosen by Xu et al. (2002) as the best CoMFA model. It was obtained on the basis of steric, electrostatic, and H-bond acceptor and donor fields.

The CoMFA contour maps of Model 3 showed that positive charge above the benzene ring would increase the toxicity. Thus, the more numerous the nitrate groups, the more positive the charge of the phenyl ring and the more toxic the compound. Electron-withdrawing substituents at the *para* and *ortho* positions of nitrate group are also associated with enhanced toxicity.

The CoMSIA(2) model had slightly better statistical parameters than the CoMSIA(1), and was better than the CoMFA models (The CoMSIA(1) and the CoMSIA(2) models differed in the value of the used column filtering, see above). The contour maps of the CoMSIA(2) model showed that positive charge in the regions above the phenyl ring corresponds to high toxicity (consistent with the CoMFA results). On the contour map of the acceptor field, an area surrounding the benzene ring favourable for hydrogen-bond acceptor substituents for increasing toxicity was shown.

I.8. QSARs for toxicity to amphibians

Wang et al. (2001) developed QSARs for the acute toxicity of 31 substituted phenols to *Rana japonica* tadpoles. Toxicity was measured as the median lethal concentrations after 12 h and 24 h exposure, and converted to inverse logarithm form, expressed as 12 h-LC_{50} and 24 h-LC_{50} (mmol/l) respectively. The two types of LC_{50} value were highly intercorrelated (R^2 = 0.997), and the 24 h-LC_{50} values were used to derive QSARs. A mechanistically relevant QSAR model was derived for 31 phenols containing descriptors reflecting hydrophobicity, an electronic property, and steric effects:

$$24\text{ h-log1/LC}_{50} = 0.65\ \log P - 0.39\ \text{LUMO} + 0.38\ \text{HOF} + 0.21\ {}^{1}\chi_p + 1.76 \qquad \text{(I.8.1)}$$

$n = 31,\ R^2 = 0.85,\ s = 0.30,\ F = 37.7$

where LUMO is the energy of the lowest unoccupied orbital; HOF is the heat of formation (kcal); ${}^{1}\chi_p$ is the first-order simple molecular connectivity index.

The same organism was investigated by Huang et al. (2003), who determined the acute lethal toxicity (50% lethality) after a 12 h exposure (12 h-log1/LC_{50}) of 51 benzene derivatives. A general QSAR model was developed based on variables reflecting hydrophobicity, an electronic property, and molecular size respectively:

$$12\text{h-log1/LC}_{50} = 0.399\ \log D - 0.453\ \text{LUMO} + 0.0109\ \text{Vol.} + 1.39 \qquad \text{(I.8.2)}$$

$n = 51,\ R^2 = 0.914\ s = 0.175,\ F = 167,\ R^2_{cv} = 0.785$

I.9. Comparative studies on different species

Some authors developed and compared QSARs for toxicity to a number of species by applying the same methodology.

Selassie et al. (2002) derived QSARs for toxicity to a wide variety of biological systems, in which free radical formation is thought to take place as a mechanism of toxicity. QSARs for aromatic nitro compounds (chemical structure $X\text{-}C_6H_4NO_2$) were developed for the following endpoints: inhibition of luminescence (EC_{50} values) of *V. fischeri*; LC_{50} to fathead minnow (96 h assay); LC_{50} to the fish *Cyprinus carpio*; inhibition of growth (IC_{50} values) of *T. pyriformis* after 48 h. QSARs for phenols were also developed for toxicity to *T. pyriformis* (IC_{50} values); 8 day old flagfish larva (LC_{20} values); *C. vulgaris* (IC_{50} values).

The descriptors used were Hammett σ, σ-, and σ+ constants, LUMO, logP, Verloop's sterimol parameters for substituents (B5, and L), and indicator variables.

Parkerton and Konkel (2000) used toxicity data of phthalate esters (Pes) (data were taken from Staples et al., 1997; and Call et al., 1998) to different freshwater and marine species. On the basis of relationships derived between the aquatic toxicity (BE) and logP, the critical body residues (CBRs) for PE in fish were estimated according to the equation:

$$\log BE = \log CBR - b \log P \qquad \text{(I.9.1)}$$

The BE can refer to an acute LC_{50} or EC_{50} value or a chronic effect concentration. This equation is based on the assumption that adverse effects are elicited when the molar concentration in organism tissues exceeds a critical threshold.

The QSARs obtained for different species/endpoints were used to calculate predicted no-effect concentrations (PNECs) by using the statistical extrapolation procedure developed by Verhaar et al. (1994):

$$\log PNEC = \mu_a \cdot \log K_{ow} + \mu_b - K_z \sqrt{\sigma_a^2 \cdot \log K_{ow} + \sigma_b^2 + 2 \cdot \sigma_{ab}^2 \cdot \log K_{ow}} \qquad \text{(I.9.2)}$$

where PNEC is the aqueous concentration intended to protect 95% of the species (mmol/l), μ_a is the mean of QSAR slopes (i.e. b estimates), μ_b is mean of QSAR intercepts (i.e. logCBR

estimates), σ^2_a is the variance of QSAR slopes, σ^2_b is the variance of QSAR intercepts, σ^2_{ab} is covariance between QSAR slopes and intercepts, and K_z is the 95% confidence extrapolation factor (Alderberg and Slob, 1993).

For the toxicity of low-molecular-weight Pes with logP < 6, a simple logP-dependent QSARs for different test species/end points were obtained. All toxicity values were converted into mmol/l and log-transformed. These models are presented in Table I.9.1.

Table I.9.1 Aquatic toxicity QSARs for phthalate esters to various species (reproduced from Parkerton and Konkel, 2000).

Species	Type [a]	Endpoint	Slope	Intercept	n [b]	R^2
Algae						
Selenastrum capricornitum	F	96 h chl a EC_{50}	-0.89	1.18	9	0.93
Gymnodinium treve	M	96 h growth EC_{50}	-0.81	0.62	7	0.92
Skeletenema cornatum	M	96 h chl a EC_{50}	-0.81	0.90	3	0.66
Scenedesmus subspicatus	F	48 h growth	-0.65	0.82	3	0.70
Chlorella pyrenoidosa	F	96 h growth	-0.63	1.14	2	-
Protozoans						
Tetrahymena pyroformes	F	96 h growth EC_{50}	-0.65	1.40	3	0.99
Annelids						
Lumbriculus variegates	F	10 days LC_{50}	-0.51	0.34	4	0.97
Insects						
Paratanytarsus parthenogenica	F	96 h LC_{50}	-0.62	1.26	3	0.99
Chironomus tentans	F	10 days LC_{50}	-0.51	0.34	3	0.99
Crustacea						
Daphnia magna	F	48 h EC_{50}	-0.53	0.48	9	0.82
Daphnia magna	F	21 days surv and repro NOEC	-0.59	0.05	10	0.83
Artemia	M	24 h hatchability NOEC	-0.48	0.21	3	0.99
Nitroca spinipes	M	96 h LC_{50}	-0.54	0.30	3	0.97
Mysidopsis bahia	M	96 h LC_{50}	-0.66	0.43	6	0.96
Hyalella azteca	F	10 days LC_{50}	-0.57	-0.10	4	0.98
Fish						
Lepomis macrochirus	F	96 h LC_{50}	-0.60	0.31	5	0.97
Pimephales promelas	F	96 h LC_{50}	-0.63	0.51	14	0.95
Oncorhynchus mykiss	F	96 h LC_{50}	-0.51	0.14	6	0.96
Cyprinodon variegatus	M	96 h LC_{50}	-0.41	0.06	7	0.91

[a] F – freshwater, M- marine; [b] number of toxicity tests included in regression.

Two sample *t* tests indicated that the mean logCBR (intercepts of the QSARs) for algae (number of data points n = 5, mean value $\mu = 0.93$, standard deviation $\sigma = 0.23$) is statistically different than that for fish (n = 4, $\mu = 0.25$, $\sigma = 0.20$) (degrees of freedom df = 7; p = 0.002). Due to the wide variability in logCBRs for invertebrates (n = 9, $\mu = 0.63$, $\sigma = 0.56$), the mean logCBR for invertebrates was not statistically different from either fish (df = 11, p = 0.22) or algae (df = 12, p = 0.28). No statistical differences in logCBRs were observed between freshwater (n =12, $\mu = 0.73$, $\sigma = 0.52$) and marine (n = 6, $\mu = 0.42$, $\sigma = 0.30$) test species (df = 16, p = 0.19).

Algae possessed relatively steep slopes from QSAR analyses, while fish displayed more shallow slopes. Slope estimates for the invertebrates varied. Protozoans and annelids exhibit values similar to algae while Insects and crustaceans are characterised by values similar to fish. The Pearson correlation coefficient indicates a negative correlation between logCBR and b (R = -0.63, p = 0.005). This value was used to calculate the covariance estimate required in the equation for PNECs.

Results for high-molecular-weight Pes (logP > 6) indicated that these chemicals are not acutely or chronically toxic to freshwater or marine organisms. According to Parkerton and Konkel (2000), this might be due to the combined role of low water solubility and limited bioconcentration potential which precludes attainment of internal concentrations that are required to elicit adverse effects.

Dearden et al. (2000) investigated the toxicity of 33 non-polar narcotics and 15 polar narcotics (IC_{50} values in units mol/l) to *D. magna*, *V. fischeri* and the fathead minnow (*Pimephales promelas*).

For these compounds the measured octanol-water partition coefficient $logP_{meas}$ was well correlated with the polarisability (α) and the free energy H-bond acceptor factor (Ca), calculated with the software HYBOT98/HYBOT-PLUS.

$$logP_{meas} = 0.266\alpha - 1.06\ Ca + 0.144 \qquad \text{(I.9.3)}$$

n = 48, $R^2 = 0.981$, s = 0.172, F = 1156

The QSARs obtained for non-polar and polar narcotics are presented in Table I.9.2.a and Table I.9.2.b respectively:

Table I.9.2. QSARs for toxicity to *D. magna*, *V. fischeri* and *Pimephales promelas*, obtained by Dearden et al. (2000).

a. non-polar narcotics:

Species	QSAR equation	n	R^2	s	F
P. promelas	0.890 $logP_{meas}$ + 1.30	23	0.959	[a]	[a]
P. promelas	0.247 α – 0.900 Ca + 1.22	23	0.965	0.239	278
D. magna	0.828 $logP_{meas}$ + 1.61	23	0.838	[a]	[a]
D. magna	0.247 α + 0.910	23	0.848	303	117
V. fischeri	0.828 $logP_{meas}$ + 1.31	33	0.831	[a]	[a]
V. fischeri	0.300 α – 0.326 Ca + 0.026	33	0.925	0.334	186

b. polar narcotics:

Species	QSAR equation	n	R^2	s	F
P. promelas	0.583 $logP_{meas}$ + 2.63	10	0.799	[a]	[a]
P. promelas	0.118 α – 0.742 Ca + 3.38	10	0.877	0.189	24.9
D. magna	0.419 $logP_{meas}$ + 3.04	12	0.678	[a]	[a]
D. magna	0.154 α + 1.85	12	0.538	300	11.6
V. fischeri	0.558 $logP_{meas}$ + 2.67	15	0.790	[a]	[a]
V. fischeri	0.137 α – 0.587 Ca + 2.88	15	0.791	0.262	22.8

[a] not given by Dearden et al. (2000).

The QSARs showed differences in the way non-polar and polar narcotics correlate with logP and with α and Ca, indicating that the two classes of compound might exert their toxicity by mechanisms different to some extent. The QSARs were worse for the polar narcotics with respect to the non-polar narcotics. Species differences were also observed. Ca was not a statistically significant parameter in the equation with the polarisability for *D. magna*. The QSARs obtained for toxicity to *D. magna* had the worst statistical fit.

Dimitrov et al. (2000) developed QSARs for toxicity of narcotic compounds to three different species – the ciliate *T. pyriformis* (the toxicity data taken from Schultz, 1997), the water flea

D. magna (the toxicity data taken from Zhao et al., 1998a), and the fish fathead minnow (*P. promelas*) (the toxicity data taken from Nendza and Russom, 1991). They applied the "response- surface" approach, using logP as a descriptor of hydrophobicity, accounting for the compound biouptake, and LUMO as descriptor of electrophilicity, important for the interactions of the non-covalent acting chemicals with the site of action. Similarly to Dimitrov et al. (2003), Dimitrov et al. (2000) suggest these descriptors as relevant to model non-congeneric narcosis. The toxicity to *T. pyriformis* was expressed as 50% inhibitory growth concentrations (IC_{50}) in a 48-h assay. The toxicity to *D. magna* was evaluated as 50% immobilisation concentrations in a 24-h assay (IC_{50} values). Concentrations that cause 50% lethality of fish after 96 h exposure (LC_{50}) were used as estimate of the toxicity towards *P. promelas*. All toxicity values were expressed in mol/l. An interspecies QSAR was derived, and it was validated on three other data sets – toxicity to mosquito (*Cilex tarsalis*) larva (LC_{50} values of 24 h exposure assay), pond snail (*Lymnaea stagnalis*) (96-h LC_{50} values), and time-independent minimum narcosis concentration data for talpole (*Rana temporaria*) ($NC_{minimum}$).

Dimitrov et al. (2000) developed an interspecies QSAR, represented by the following equation:

$$\log(1/\text{endpoints}) = b(\text{organism, endpoint}) + 0.77\ \log P - 0.07\ \text{LUMO} \qquad \text{(I.9.4)}$$
$$n = 202,\ R^2 = 0.96,\ R^2_{cv} = 0.95,\ F = 1154$$

where log(1/endpoints) is the negative logarithm of the toxicity values for the three species, b(organism, endpoint) represents the values for the intercepts, which are different for the three species. The numbers datapoints for a single species (n), the values of b (organism, endpoint), and the standard deviations for the different endpoints, derived from the interspecies equation (σ_1), and from an regression equation with logP and LUMO as a descriptors, different for each endpoint (σ_2) are presented in Table I.9.3.

Table I.9.3. Numbers of data points (n), the values of the intercept b(organism, endpoint) for the different species, and standard deviations, derived from the interspecies equation (σ_1, Equation I.9.4), and from a separate regression equation with logP and LUMO for each endpoint (σ_2) (reproduced from Dimitrov et al., 2000)

Organism	Endpoint	n	Intercept (b)	Statistical comparisons	
				σ_1	σ_2
Training data					
T. pyriformis	IC_{50} (48-h)	125	1.12 ± 0.06	0.189	0.201
P. promelas	LC_{50} (96-h)	43	1.49 ± 0.04	0.237	0.290
D. magna	IC_{50} (24-h)	34	1.86 ± 0.06	0.283	0.309
Validation data					
C. tarsalis	LC_{50} (24-h)	13	0.93 ± 0.15	0.239	0.239
L. stagnalis	LC_{50} (96-h)	12	1.57 ± 0.26	0.282	0.389
R. temporaria	$NC_{minimum}$	9	1.02 ± 0.04	0.038	0.051

From the table it can be seen that the standard deviations for the different endpoints, obtained from the single interspecies model (Equation I.9.4) are similar to those, obtained by predicting the toxicities from separate QSARs for each endpoint. The correlation between the observed and predicted toxicities for all six endpoints, based on the interspecies equation (Equation I.9.4) gave R^2 of 0.97.

According to Dimitrov et al. (2000) the biological membranes are the sites of action of the investigated narcotic compounds, where they alter lipid properties and fatty acid composition. Due to this common mechanism of toxic action, the descriptors of the chemical structures that encode information relevant for this mechanism are similar for all endpoints. The different intercepts in the QSARs account for the differences in the biological complexities of the target organisms, and in the testing protocols.

I.10. Conclusions

A large number QSAR studies of acute toxicity have been published in the literature. The presented review focuses mainly on recent works, in which QSARs were developed using traditional statistical methods (regression analysis, partial least squares analysis), rather than neural network approaches, for example.

The toxicities to a wide range of aquatic species have been investigated by QSAR analysis, including bacteria, protozoa, algae, *Daphnia*, fish, amphibians. Very different groups of chemicals were investigated and different QSAR approaches have been applied.

A dependence of acute (narcotic) toxicity on a partition coefficent, especially the octanol-water partition coefficients, has been shown by many authors (Dearden et al., 2000; Freidig and Hermens, 2000; Parkerton and Konkel, 2000; Romanelli et al., 2000; Gramatica et al., 2001; Yu et al., 2001; Ren and Frymer, 2002; Worgan et al., 2003). In addition to octanol-water, partition coefficients in other phases were also used to develop QSARs for toxicity. For example Schultz and Seward (2000) concluded that dimyristoyl phosphatidylcholine-water partition coefficient gave better statistical fit than octanol-water partition coefficients. Roberts and Costello (2003) developed QSARs using octanol-water and membrane-water partition coefficients.

Some authors have used the so-called "response-surface approach" to model the toxicity of compounds based on hydrophobicity and electrophilicity. The QSARs obtained were based on octanol–water partition or distribution coefficients (logP or logD) and the energy of the lowest unoccupied molecular orbital (LUMO), which parameterise biouptake and tissue distribution, and electrophilic reactivity, respectively. This approach has been applied to different species, including the bacterium *V. fischeri* (Cronin et al., 2000); the protozoan *T. pyriformis* (Cronin et al., 2002a; Cronin and Schultz, 2001; Schultz et al., 2002); algae *Scenedesmus obliquus* (Lu et al., 2001), and *Chlorella vulgaris* (Cronin et al., 2002b). In some studies maximum acceptor superdelocalisability (A_{max}) has been used instead of LUMO as a measure of electrophilicity (Cronin et al., 2001; Schultz et al., 2003b). The response-surface approach is simple and mechanistically interpretable.

While e.g. Cronin and Schultz (2001), Cronin et al. (2001), Cronin et al. (2002a) used LUMO or A_{max} as descriptor of electrophilic reactivity resulting in covalent change in biological systems, according to Dimitrov et al. (2000) and Dimitrov et al. (2003) LUMO can be also used to describe electrophilic interaction of the non-covalent narcotic acting chemicals with the site of action.

Some authors extended the response-surface approach by adding additional indicator variables and other parameters to improve the statistical fit of the models (Schmitt et al., 2000; Wang et al., 2001; Huang et al., 2003; Cronin et al., 2004; Netzeva et al., 2004). Examples of such variables include charges of a certain atom in the compound (Schmitt et al., 2000), the heat of formation (Wang et al., 2001), the molecular volume (Huang et al., 2003),

molecular connectivity indices (Wang et al., 2001; Netzeva et al., 2004). These might help to model outliers with toxicity under- or overestimated by the response surface.

In general, outliers are thought to be compounds that act by different mechanisms of toxic action from the majority of the data set (Lipnick, 1991). Factors other than those accounted for in the model, such as pharmacokinetic properties, might account for the toxicity of the outliers. Another reason for the presence of outliers might be incorrectly determined or reported values of toxicity.

Ren (2003) developed separate QSARs for compounds acting by different mechanisms of toxic action (polar narcosis, respiratory uncoupling, pro-electrophilicity, and soft electrophilicity), and compared the predicition results with the predicition from a general QSAR including all compounds. He obtained better correlation between the predicted and measured toxicities using the prediction from the general QSAR model, rather than the predicition based on the mechanism-based QSARs. The worse predicition of the mechanism-based QSARs was attributed to the bad QSAR model for the soft electrophilies, obtained by Ren (2003). According to Schultz et al. (1998) QSAR modelling for electrophilic compounds is difficult because of data and descriptor limitation compared to QSAR modelling of compounds, acting by other toxic mechanisms.

QSARs for toxicity based on topological and/or electrotopological indices have been derived by many authors (Ivanciuc, 2000; Burden, 2001; Gramatica et al., 2001; Wang et al., 2001; Agrawal and Khadikar, 2002; Cronin et al., 2002a; Khadikar et al., 2002; Huuskonen, 2003; Ren, 2003; Rose and Hall, 2003). Descriptors accounting for the size and shape of molecules (molecular volume, molecular weight, Verloop's width parameters, charged molecular surface areas) have been used by Cronin et al. (2002a); Selassie et al., (2002); Huang et al. (2003); Ren (2003).

Another approach is based on the TLSER model descriptors, which represent cavity, dipolarity/polarisability, and H-bonding terms (Liu et al., 2001; Liu et al., 2003a). Dearden et al. (2000) also used polarisability and H-bond acceptor factor to develop QSARs.

Some QSARs were developed by using the energy of the highest unoccupied molecular orbital (HOMO) (Schultz et al., 2002; Ren, 2003) Another descriptor in some QSARs is

'hardness' equal to ½ (HOMO – LUMO), which accounts for atomic radius, nuclear charge and polarisability (Faucon et al., 2001).

3D QSAR (CoMFA and CoMSIA) have also been applied to investigate toxicity (Xu et al., 2002; Liu et al., 2003b; Liu et al., 2003c). Important interaction fields and regions around the molecules for the compound toxicity were identified. There is, however, implicit difficulty in applying these models, because they are usually developed for a particular series of compounds, for which the toxic effect is due to specific interactions with biological macromolecules.

Also, several other QSAR approaches have been explored in investigating acute toxicity. For example Balaz and Lukacova (2002) used subcellular pharmacokinetic theory to derive QSARs. Toropov and Toropova (2002), Castro et al. (2003), Toropov and Schultz (2003), and Toropov and Benfenati (2004) used optimisation of correlation weights of local graph invariants (OCWLGI). Cui et al. (2003) applied holographic quantitative structure-activity relationship (HQSAR) approach. Weighted holistic invariant molecular (WHIM) descriptors encoding three-dimensional structural information for molecular shape and electronic structure were used by Gramatica et al. (2001), and Di Marzio et al. (2001). Several studies using autocorrelation vectors encoding lipophilicity, molar refractivity, H-bonding acceptor ability and H-bonding donor ability were carried out by Devillers et al. (2002a); Devillers et al., (2002b). Tao et al. (2002a) and Tao et al. (2002b) applied a fragment constant method to develop toxicity QSARs. Gonzales et al. (2004) applied the TOPological Substructural Molecular Design (TOPS-MODE) approach, based on the hydrogen-depleted molecular graphs and bond adjacency matrix.

Interesting was the conclusion of Freidig and Hermens (2000) that using separate QSAR models for compounds acting by different mechanisms, each of the model including a descriptor that characterises the particular toxicity mechanism, gives better results than using a single model that combines all compounds and descriptors.

QSARs for metal toxicity were also developed using ion-specific physicochemical parameters (Ownby and Newman, 2003).

Development of QSARs for toxicity of mixtures of chemical compounds is an interesting research area, which is growing in the recent years. The environmental poutants are usually released as a chemical mixtures rather than single chemicals. Development of models for mixture toxicity will assist the evaluation of the chemical impact on the environment. Some authors used mixture partition coefficients and H-bond mixture properties to develop QSARs for toxicity of mixtures compoused from non-polar and polar narcotics (Yu et al., 2001; Lin et al., 2002; Lin et al., 2003b). Other investigations include specific characteristics of the chemical compounds in the mixtures (Lin et al., 2003a).

In Table I.10.1 a summary of the applied QSAR approaches and/or structural descriptors and the corresponding references is given.

Table I.10.1 Summary of the applied QSAR approaches and/or structural descriptors and the corresponding references.

QSAR approach/structural descriptorts used	References
Dependence of narcotic toxicity on a parttion coefficent	Dearden et al. (2000); Freidig and Hermens (2000); Parkerton and Konkel (2000); Romanelli et al. (2000); Schultz and Seward (2000); Gramatica et al. (2001); Yu et al. (2001); Ren and Frymer (2002); Roberts and Costello (2003); Worgan et al. (2003).
Response-surface approach	Cronin et al. (2000); Dimitrov et al. (2000); Cronin and Schultz (2001); Cronin et al. (2001); Lu et al. (2001); Cronin et al. (2002a); Cronin et al. (2002b); Dimitrov et al. (2003); Schultz et al. (2002); Schultz et al. (2003b)
Response-surface approach, extended by adding indicator variables and other parameters	Schmitt et al. (2000); Wang et al. (2001); Huang et al. (2003); Cronin et al. (2004); Netzeva et al. (2004)
Topological and/or electrotopological indices	Ivanciuc (2000); Burden (2001); Gramatica et al. (2001); Wang et al. (2001); Agrawal and Khadikar (2002); Cronin et al. (2002a); Rose and Hall (2003)
Descriptors accounting for the size and shape of molecules (molecular volume, molecular weight, Verloop's width parameters, Taft's steric parameter, charged molecular surface areas)	Cronin et al. (2002a); Selassie et al. (2002); Huang et al. (2003); Ren (2003).
HOMO, and 'hardness'	Faucon et al. (2001); Schultz et al. (2002); Ren (2003);
Theoretical linear solvation energy relationship (TLSER) descriptors	Liu et al. (2001); Liu et al. (2003a)
3D QSAR (CoMFA and CoMSIA)	Xu et al. (2002); Liu et al. (2003b)

Table I.10.1 (continued)

QSAR approach/structural descriptorts used	References
Subcellular pharmacokinetic theory	Balaz and Lukacova (2002)
Optimisation of correlation weights of local graph invariants (OCWLGI)	Toropov and Toropova (2002); Castro et al. (2003); Toropov and Schultz (2003); Toropov and Benfenati (2004)
Holographic quantitative structure-activity relationship (HQSAR) approach	Cui et al. (2003)
Weighted holistic invariant molecular (WHIM)	Di Marzio et al. (2001); Gramatica et al. (2001);
TOPological Substructural Molecular Design (TOPS-MODE)	Gonzales et al. (2004)
Autocorrelation vectors encoding lipophilicity, molar refractivity, H-bonding acceptor ability and H-bonding donor ability	Devillers et al. (2002a); Devillers et al., (2002b)
Fragment constant method	Tao et al. (2002a); Tao et al. (2002b)
QSARs for metal toxicity	Ownby and Newman (2003)
QSARs for mixture toxicity	Yu et al. (2001); Lin et al. (2002); Lin et al. (2003a); Lin et al. (2003b)

References

Agrawal, V.K., and Khadikar P.V. (2002). QSAR study on narcotic mechanism of action and toxicity: a molecular connectivity approach to *V. fischeri* toxicity testing. *Bioorganic and Medicinal Chemistry*, **10**, 3517-3522.

Akers K.S., Sinks G.D., Schultz T.W. (1999). Structure–toxicity relationships for selected halogenated aliphatic chemicals. *Environmental Toxicology and Pharmacology*, **7**, 33-39.

Alderberg, T., and Slob, W. (1993). Confience limits for hazardous concentrations based on logistically distributed NOEC toxicity data. *Ecotoxicology and Environmental Safety*, **25**, 48-63.

Arnold, L.M., Lin D.T., Schultz T.W. (1990). QSAR for methyl- and/or chloro-substituted anilines and the polar narcosis mechanism of toxicity. *Chemosphere*, **21**, 183-191.

Austin, R.P., Davis, A.M., Manners, C.N. (1995). Partitioning of ionizing molecules between aqueous buffers and phospholipid vesicles. *Journal of Pharmacological Sciences*, **84**, 1180-1183.

Balaz S., and Lukacova V. (2002). Subcellular pharmacokinetics and its potential for library focusing. *Journal of Molecular Graphics and Modelling*, **20**, 479-490.

Basak, S.C., Grunwald, G.D., Gute, B.D., Balasubramanian, K., Opitz, D. (2000). Use of statistical and neural net approaches in predicting toxicity of chemicals. *Journal of Chemical Information and Computer Sciences*, **40**, 885-890.

Bowden, H.C., Wilby, O.K., Botham, C.A., Adam, P.J., Ross, F.W. (1995). Assessment of the toxic and potential tearatogenic effects of four glycol ethers and two derivatives using the *Hydra* regeneration assay and rat whole embryo culture. *Toxicology in Vitro*, **9**, 773-781.

Boyd, E.M., Killham K., Meharg, A.A. (2001). Toxicity of mono-, di- and tri-chlorophenols to *lux* marked terrestrial bacteria, *Burkholderia* species *Rasc c*2 and *Pseudomonas fluorescens*. *Chemosphere*, **43**, 157-166.

Bryant, S.E., and Schultz, T.W. (1994). Toxicological assessment of biotransformation products of pentachlorophenol—*T.* population growth impairment. *Archives of Environmental Contamination and Toxicology*, **26**, 299–303.

Burden, F.R. (1996). Using artificial neural networks to predict biological activity from simple molecular structural considerations. *Quantitative Structure-Activity Relationships*, **15**, 7-11.

Burden, F.R. (2001). Quantitative structure-activity relationship studies using Gaussian processes. *Journal of Chemical Information and Computer Sciences,* **41,** 830-835.

Cajina-Quezada, M., and Schultz T.W. (1990). Structure-toxicity relationships for selected weak acid respiratory uncouplers. *Aquatic Toxicology*, **17**, 239-252.

Call, D. J., Brooke, L. T., Markee, T. P., Cox, D. A., Geiger, D. L., Varde Venter, F. A., Polkinghorne, C. N., Genisot, K. I. (1998). Sediment Toxicity Testing Program for Phthalate Esters. Lake Superior Research Institute, Univ. Wisconsin-Superior, Draft Report to the Chemical Manufacturers Assoc., Washington, DC, pp 318.

Castro, E.A., Toropov, A.A., Nesterova A.I., Nazarov, A.U. (2003). QSAR study of the toxic action of aliphatic compounds to the bacteria *V. fisheri* based on correlation weighting of local graph invariants. *Journal of Molecular Structure(Theochem)*, **639**, 129-135.

Chen, C.Y., and Lu, C.L. (2002). An analysis of the combined effects of organic toxicants. *Science of the Total Environment,* **289**, 123-132.

Chen, J.W., and Wang, L.S. (1996). Acute toxicity of alkyl (1-phenylsulfonyl) cycloalkane-carboxylates to V. fischeri and quantitative structure–activity relationship study based on the AM1 Hamiltonian. *Toxicological and Environmental Chemistry*, **57**, 17–26.

Cronin M.T.D., and Schultz T.W. (1996). Structure-toxicity relationships for phenols to *T. pyriformis*. *Chemosphere*, **32**, 1453-1468.

Cronin M.T.D., and Schultz T.W. (1998). Structure-toxicity relationships for three mechanisms of action of toxicity to *V. fischeri*. *Ecotoxicology and Environmental Safety*, **39**, 65-69.

Cronin, M.T.D, and Schultz, T.W. (2001). Development of quantitative structure-activity relationships for the toxicity of aromatic compounds to *T. pyriformis*: comparative assessment of the methodologies. *Chemical Research in Toxicology,* **14**, 1284-1295.

Cronin M.T.D., Aptula A.O., Duffy J.C., Netzeva T.I., Rowe P.H., Valkova I.V. Schultz T.W. (2002a). Comparative assessment of methods to develop QSARs for the prediction of the toxicity of phenols to *T. pyriformis. Chemosphere*, **49**, 1201-1221.

Cronin, M.T.D.; Bowers, G.S., Sinks, G. D., Schultz, T.W. (2000). Structure-toxicity relationships for aliphatic compounds encompassing a variety of mechanisms of toxic action to *V. fischeri*. *SAR and QSAR in Environmental Research*, **11**, 301-312.

Cronin, M.T.D., Bryant, S.E., Dearden, J.C., Schultz, T.W. (1995). Quantitative structure-activity study of the toxicity of benzonitriles to the ciliate *T. pyriformis*. *SAR and QSAR in Environmental Research*, **3**, 1-13.

Cronin, M.T.D., Dearden, J.C., Duffy, J.C., Edwards, R., Manga, N., Worth, A.P., Worgan, A.D.P. (2002b). The importance of hydrophobicity and electrophilicity descriptors in mechanistically-based QSARs for toxicological endpoints. *SAR and QSAR in Environmental Research,* **13**, 167-176.

Cronin, M.T.D., Gregory, B.W., Schultz, T.W. (1998). Quantitative structure-activity analyses of nitrobenzenes to *T. pyriformis*. *Chemical Research in Toxicology*, **11**, 902-908.

Cronin, M.T.D, Manga, N., Seward, J.R., Sinks, G.D., Schultz, T.W. (2001). Parametrization of electrophilicity for the prediction of the toxicity of aromatic compounds. *Chemical Research in Toxicology***,** 14, 1498-1505.

Cronin, M.T.D., Netzeva, T.I, Dearden, J.C., Edwards, R., Worgan, A.D.P. (2004). Assessment and Modeling of the Toxicity of Organic Chemicals to *Chlorella vulgaris*: Development of a Novel Database. *Chemical Research in Toxicology*, **17**, 545-554.

Cui, S., Wang, X., Liu, S., Wang, L. (2003). Predicting toxicity of benzene derivatives by molecular hologram derived quantitative structure-activity relationships (QSARs). *SAR and QSAR in Environmental Research,* 14, 223-231.

Dearden, J.C., and Bresnen, G.M. (1988). The measurement of partition coefficient. *Quantitative Structure-Activity Relationships*, **7**, 133-144.

Dearden, J.C.; Bradburne, S.J.A.; Cronin, M.T.D.; Solanki, P. (1988). The physical significance of molecular connectivity. In: Turner, J.E., England, M.W., Schultz, T.W., Kwaak, N.J. (Eds.), *Proceedings of the Third International Workshop on Quantitative Structure-Activity Relationships in Environmental Chemistry*. US DOE, Oak Ridge, TN, pp 43-50.

Dearden, J.C., Cronin, M.T.D., Schultz, T.W., Lin, D.T. (1995). QSAR study of the toxicity of nitrobenzenes to *T. pyriformis*. *Quantitative Structure-Activity Relationships*, **14**, 427-432.

Dearden, J.C., Cronin, M.T.D., Zhao, Y.-H., Raevsky, O.A. (2000). QSAR studies of compounds acting by polar and non-polar narcosis: an examination of the role of polarisability and hydrogen bonding. *Quantitative Structure-Activity Relationships*, **19**, 3-9.

DeBruijn, J., and Hermens, J. (1993). Inhibition of acetylcholinesterase and acute toxicity of organophosphorous compounds to fish: a preliminary structure-activity analysis. *Aquatic Toxicology*, **24**, 257-274.

Deneer, J.W., Sinnigie, T.L., Seinen, W., Hermens, J.L.M. (1987). Quantitative structure-activity relationship for the toxicity and bioconcentration factor of nitrobenzene derivatives towards the guppy (*Poecilia reticulata*). *Aquatic Toxicology*, **10**, 115-129.

Devillers, J., Chezeau, A., Thybaud, E., Rahmani, R. (2002a). QSAR modelling of the adult and developmental toxicity of glycols, glycol ethers and xylenes to *Hydra attenuata*. *SAR and QSAR in Environmental Research***, 13**, 555-566.

Devillers, J., Chezeau, A., Thybaud, E. (2002b). PLS-QSAR of the adult and developmental toxicity of chemicals to *Hydra attenuata*. *SAR and QSAR in Environmental Research***, 13**, 705-712.

Dimitrov, S.D., Mekenyan, O.G., Sinks, G.D., Schultz, T.W. (2003). Global modeling of narcotic chemicals: ciliate and fish toxicity. *Journal of Molecular Strucrure (Theochem),* **622**, 63-70.

Dimitrov, S.D., Mekenyan, O.G., Schultz, T.W. (2000). Interspecies modeling of narcotic toxicity to aquatic animals. *Bulletin of Environmental Contamination and Toxicology*, **65**, 399-406.

Di Marzio, W., Galassi, S., Todeschini, R., Consolaro, F. (2001). Traditional versus WHIM molecular descriptors in QSAR approaches applied to fish toxicity studies. *Chemosphere,* **44**, 401-406.

EC (2001). White Paper on a Strategy for a Future Chemicals Policy. Brussels: Commission of the European Communities. Available on the internet: http://europa.eu.int/comm/environment/chemicals/whitepaper.htm

EC (1986). Council Directive *86/609/EEC* of 24 November 1986 on the approximation of laws, regulations and administrative provisions of the Member States regarding the protection of animals used for experimental and other scientific purposes. *Official Journal of the European Community* L358, 18.12.1986.

EC (2003). Commission's proposal COM 2003 0644 (03) of 29 October 2003 for a Regulation of the European Parlament and of the Council concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH). Available on the internet: http://europa.eu.int/eur-lex/en/com/pdf/2003/com2003_0644en.html

Estrada, E., and Gutierrez, Y. (1999). Modeling chromatographic parameters by a novel graph theoretical sub-structural approach. *Journal of Chromatography A*, **858**, 187–199.

Faucon, J.C., Bureau, R., Faisant, J., Briens, F., Rault, S. (2001). Prediction of the Daphnia acute toxicity from heterogeneous data. *Chemosphere,* **44**, 407-422.

Feng, L., Han, S.K., Zhao, Y.H., Wang, L.S. (1996). Quantification and application of Lewis acidity and basicity of benzene and its derivatives. *Toxicological and Environmental Chemistry,* **53**, 119–126.

Freidig, A.P., Verhaar, H.J.M., Hermens, J.L.M. (1999). Combining the potency of chemicals with multiple models of action in aquatic toxicology: acute toxicity due to narcosis versus reactive toxicity of acrylic compounds. *Environmental Science and Technology*, **33**, 3038-3043.

Freidig, A.P., and Hermens, J.L.M. (2000). Narcosis and chemical reactivity QSARs for acute fish toxicity. *Quantitative Structure-Activity Relationships,* **19**, 547-553.

Galassi, S., Mingazzini, M., Viganò, L., Cesareo, D., Tosato, M. (1988). Approaches to modeling toxic responses of aquatic organisms to aromatic hydrocarbons. *Ecotoxicology and Environmental Safety,* **16**, 158–169.

Gao, C., Govind, R., Tabak, H. (1992). Application of the group contribution method for predicting the toxicity of organic chemicals. *Toxicological and Environmental Chemistry*, **11**, 631–636.

Geiger, D., Brooke, L., Call, D. (1990). Acute toxicities of organic chemicals to fathead minnows *Pimephales promelas* vol. 5. Center for Lake Superior Environmental Studies, University of Wisconsin Superior, Wisconsin.

Gonzales, M.P., Diaz, H.G., Cabrera, M.A., Ruiz, R.M. (2004). A novel approach to predict a toxicological property of aromatic compounds in the *T. pyriformis*. *Bioorganic and Medicinal Chemistry*, **12**, 735–744.

Gramatica, P., Vighi, M., Consolaro, F., Todeschini, R., Finizio, A., Faust. M. (2001). QSAR approach for the selection of congeneric compounds with a similar toxicological mode of action. *Chemoshere*, **42**, 873-883

Grimme, L.H., Altenburger, R., Backhaus, T., Bödeker, W., Faust, M., Scholze, M. (1998). Voresagbarkeit und Beurteilung der aquatischen Toxizität von Stoffgemischen – Multiple Kombinationen von unähnlich wirkenden Substanzen in niedrigen Konzentrazionen, UFZ_Bericht, Umweltforschungszentrum Leipzig-Halle, Leipzig, p 319.

Gute, B.D., and Basak, S.C. (1997). Predicting acute toxicity (LC_{50}) of benzene derivatives using theoretical molecular descriptors: a hierarchical QSAR approach. *SAR and QSAR in Environmental Research*, **7**, 117-131.

Hong, H., Han, S.K., Wang, L.S., Zhao, Y.H., Zou, G.W., Zhang, Z. (1995). Analysis of aromatic sulfur-containing compounds based on Free-Wilson method. *Chinese Science Bulletin*, **4**, 1621–1624.

Huang, H., Wang, X., Ou, W., Zhao, J., Shao, Y., Wang, L. (2003). Acute toxicity of benzene derivatives to the tadpoles (*Rana japonica*) and QSAR analyses. *Chemosphere,* **53**, 963-970.

Huuskonen J. (2003). QSAR modeling with the electrotopological state indices: predicting the toxicity of organic chemicals. *Chemosphere*, **20**, 949-953.

HYBOT98/HYBOT-PLUS, MolPro Project, Russia (http://www.ibmh.msk.su/qsar/molpro)

Ivanciuc, O. (2000). QSAR comparative study of Wiener descriptors for weighted molecular graphs. *Journal of Chemical Information and Computer Sciences***, 40**, 1412-1422.

Jaworska, J.S., and Schultz, T.W. (1993). Quantitative relationships of structureactivity and volume fraction for selected non-polar and polar narcotic chemicals. *SAR and QSAR in Environmental Research*, **1**, 3–19.

Johnson, E.M., and Gabel, B.E.G. (1983). The role of an artificial embryo in detecting potential teratogenic hazards. In: Johnson, E.M., and Kochhar, D.M. (Eds.), *Handbook of Experimental Pharmacology*. Springer-Verlag, Berlin, **65**, pp 335-348.

Johnson, E.M., Gabel, B.E.G., Christian, M.S., Sica, E. (1986). The developmental toxicity of xylene and xylene isomers in the *Hydra* assay. *Toxicology and Applied Pharmacology*, **82**, 323-328.

Johnson, E.M., Gabel, B.E.G., Larson, J. (1984). Developmental toxicity and structure/activity correlates of glycol and glycol ethers. *Environmental Health Perspectives*, **57**, 135-139.

Kaiser, K.L.E. (2003a). The use of neural networks in QSARs for acute aquatic toxicological endpoints. *Journal of Molecular Structure (Theochem)*, **622**, 85–95.

Kaiser, K.L.E. (2003b). Neural networks for effect prediction in environmental and health issues using large datasets. *QSAR and Combinatorial Science*, **22**, 185-190.

Karabunarliev, S., Mekenyan O.G., Karcher W., Russom C.L., Bradbury S.P. (1996). Quantum-chemical descriptors for estimating the acute toxicity of electrophiles to the fathead minnow (*Pimephales promelas*). An analysis based on molecular mechanism. *Quantitative Structure-Activity Relationships*, **15**, 302–310.

Karelson, M., Lobanov, V. S., Katritzky, A. R. (1996). Quantum-chemical descriptors in QSAR/QSPR studies. *Chemical Reviews,* **96**, 1027-1043.

Khadikar P.V., Phadnis A., Shrivastava A. (2002). QSAR study on toxicity to aqueous organisms using the PI index. *Bioorganic and Medicinal Chemistry*, **10**, 1181-1188.

Kier, L.B., and Hall, L. H. (1995). The molecular connectivity chi indexes and kappa shape indexes in structure-property modelling. In: Lipkowitz K.B., and Boyd, D.B., (Eds.), *Reviews in Computational Chemistry*. VCH Publishers: New York, Vol. 2, pp 367-422.

Kramer, C.R., and Trumper, L. (1986). Quantitative strukur-wirkyngs-beziehungen für die wachstumshemmung von autotrophen *Chlorella-vulgaris*-suspensionen durch monosubstituierte benzene, toluene, halogenbenzene and methoxybenzene", *Biochemie und Physiologie der Pflanzen*, **181**, 645-657.

Lang, P.Z. et al., 1998. Reaction and toxicity of organic pollutants in the Songhua river. JiLin Technology Publishing Company, JiLin.

Leo, A.J. (1975). In: Veith, G.D., and Korasevich, D.E. (Eds.), *Proceedings of Symposium on Structure–Activity Correlations in Studies of Toxicity and Bio-concentration with Aquatic Organisms*. Great Lakes Research Advisory Board, Burlington, Ontario, p 151.

Leo, A.J. (1982). Octanol/water partition coefficient. In: Lyman, W.J., and Rosenblatt, D.H. (Eds.), *Handbook of Chemical Property Estimation Methods*. McGraw-Hill, New York, pp 1–54.

Lessigiarska, I., Worth, A.P., Sokull-Klüttgen, B., Jeram, S., Dearden, J.C., Netzeva, T.I., Cronin, M.T.D. (2004). QSAR Investigation of a large data set for fish, algae and *Daphnia* toxicity. *SAR and QSAR in Environmental Research*, **15**, 413-431.

Lin, Z., Yin, K., Shi, P., Wang, L., Yu, H. (2003a). Development of QSARs for predicting the joint effects between cyanogenic toxicants and aldehydes. *Chemical Research in Toxicology,* **16,** 1365-1371.

Lin, Z., Yu, H,, Wei, D., Wang, G., Feng, J., Wang, L. (2002). Prediction of mixture toxicity with its total hydrophobicity. *Chemosphere,* **6**, 305-310.

Lin, Z., Zhong, P., Yin, K.,Wang, L.,Yu, H. (2003b). Quantification of joint effect for hydrogen bond and development of QSARs for predicting mixture toxicity. *Chemosphere,* **52**, 1199-1208.

Lipnick R.L. (1991). Outliers: their origin and use in the classification of molecular mechanisms of toxicity. *Science of the Total Environment*, **109/110**, 131–153.

Liu, X., Wang, B., Huang, Z., Han, S., Wang, L (2003a). Acute toxicity and quantitative structure-activity relationships of alpha-branched phenylsulfonyl acetates to *D. magna*. *Chemosphere,* **50**, 403-408.

Liu, X., Wu, C., Han, S., Wang, L., Zhang, Z. (2001). The acute toxicity of alpha-branched phenylsulfonyl acetates in *V. fischeri* test. *Ecotoxicology and Environmental Safety,* **49**, 240-244.

Liu, X., Yang, Z., Wang, L. (2003b). Three-dimensional quantitative structure–activity relationship study for phenylsulfonyl carboxylates using CoMFA and CoMSIA. *Chemosphere*, **53**, 945-952.

Liu, X, Yang, Z., Wang, L. (2003c). CoMFA of the acute toxicity of phenylsulfonyl carboxylates to *V. fischeri*. *SAR and QSAR in Environmental Research,* **14**, 183-190.

Lowis, D.R. (1997). HQSAR. A New, Highly Predictive QSAR Technique. *Tripos Technical Notes*, **Vol. 1, No 5**, Tripos, Inc., http://www.tripos.com/sciTech/inSilicoDisc/media/LITCTR/HQSAR_AP.PDF

Lu, G.H., Yuan, X., Zhao, Y.H. (2001). QSAR study on the toxicity of substituted benzenes to the algae (Scenedesmus obliquus). *Chemosphere,* **44**, 437-440.

Manthey, M., Faust, M., Smolka, S., Grimme, L.H. (1993). Herbicide bioconcentration in algae studies on lipophilicity-sorption-activity relationships (LSAR) with *Chlorella fusca*. *Science of the Total Environment*, Suppl. 1993, 453-460.

Mayura, K., Smith, E.E., Clement, B.A., Phillips, T.D. (1991). Evaluation of the developmental toxicity of chlorinated phenols utilizig *Hydra attenuata* and postimplantation rat embryos in culture. *Toxicology and Applied Pharmacology*, **108**, 253-266.

Montgomery, J.H. (1993). Agrochemicals Desk Reference: Environmental Data. Lewis Publishers, Boca Raton, FL.

Morgan, H.L. (1965). The generation of a unique machine description for chemical structures – a technique developed at chemical abstracts service. *Journal of Chemical Documentation,* **5**, 107-113.

Nendza, M., and Russom, C.L. (1991). QSAR modelling of the ERL-D fathead minnow acute toxcity database. *Xenobiotica*, **21**, 147-170.

Netzeva, T.I., Dearden, J.C., Edwards, R., Worgan, A.D.P., Cronin, M.T.D. (2004). QSAR analysis of the toxicity of aromatic compounds to *Chlorella vulgaris* in a novel short-term assay. *Journal of Chemical Information and Computer Sciences*, **44**, 258-265.

Ownby, D.R., and Newman, M.C. (2003). Advances in quantitative ion character-activity relationships (QICARs): using metal-ligand binding characteristics to predict metal toxicity. *QSAR and Combinatorial Science*, **22**, 241-246.

Parkerton, T.F., and Konkel, W.J. (2000). Application of quantitative structure-activity relationships for assessing the aquatic toxicity of phthalate esters. *Ecotoxicology and Environmental Safety,* **45**, 61-78.

Preston, S., Coad, N., Townend, J., Killham K., Paton, G.I. (2000). Biosensing the acute toxicity or metal interactions: Are they additive, synergistic, or antagonistic? *Environmental Toxicology and Chemistry,* **19,** 775–780.

Randic, M. (1975). On characterization of molecular branching. *Journal of American Chemical Society,* **97,** 6609-6615.

Ren, S. (2003). Ecotoxicity prediction using mechanism- and non-mechanism-based QSARs: a preliminary study. *Chemosphere,* **53**, 1053-1065.

Ren S., and Frymier P.D. (2002). Estimating the toxicities of organic chemicals to bioluminescent bacteria and activated sludge. *Water Research*, **36**, 4406-4414.

Roberts, D.W., and Costello, J.F. (2003). Mechanisms of action for general and polar narcosis: A difference in dimension. *QSAR and Combinatorial Science*, **22**, 226 – 233.

Romanelli, G.P., Cafferata, L.F.R., Castro, E.A. (2000). An improved QSAR study of toxicity of saturated alcohols. *Journal of Molecular Structure (Theochem)*, **504**, 261–265.

Rose K., and Hall L.H. (2003). E-state modelling of fish toxicity independent of 3D structure information. *SAR and QSAR in Environmental Research*, **14**, 113-129.

Rose K., Hall L.H., and Kier L.B. (2002). Modeling blood-brain barrier partitioning using the electrotopological state. *Journal of Chemical Information and Computer Sciences***, 42**, 651-666.

Russom, C.L., Bradbury, S.P., Broderius, S.J., Hammermeister, D.E., Drummond, R.A. (1997). Predicting modes of action from chemical structure: Acute toxicity in the fathead minnow *(Pimephales promelas)*. *Environmental Toxicology and Chemistry,* **16,** 948-967.

Schmitt, H., Altenburger, R., Jastorff, B., Schüürmann, G. (2000). Quantitative structure-activity analysis of the algae toxicity of nitroaromatic compounds. *Chemical Research in Toxicology,* **13**, 441-450.

Schultz, T.W. (1997). TETRATOX: The *T. pyriformis* population growth impairment endpoint – a surrogate for fish lethality. *Toxicology Methods*, **7**, 289-309.

Schultz, T.W., and Riggin, G.W. (1985). Predictive correlations for the toxicity of alkyl- and halogen-substituted phenols. *Toxicology Letters*, **25**, 47–54.

Schultz T.W., and Seward, J.R. (2000) Dimyristoyl phosphatidylcholine/water partitioning-dependent modeling of narcotic toxicity to *T. pyriformis*. *Quantitative Structure-Activity Relationships,* **19**, 239-344

Schultz, T.W., Arnold, L.M., Wilke, T.S., Moulton, M.P. (1990). Relationships of quantitative structure-activity for normal aliphatic alcohols. *Ecotoxicology and Environmental Safety*, **19**, 243–253.

Schultz, T.W., Bearden, A.P., Jaworska, J.S. (1996). A novel QSAR approach for estimating toxicity of phenols. *SAR and QSAR in Environmental Reseach,* **5**, 99–112.

Schultz, T.W., Cronin, M.T.D., Netzeva, T.I., Aptula, A.O. (2002). Structure-toxicity relationships for aliphatic chemicals evaluated with *T. pyriformis*. *Chemical Research in Toxicology,* **15***,* 1602-1609.

Schultz T.W., Cronin M.T.D., Walker J.D., Aptula, A.O. (2003a). Quantitative structure–activity relationships (QSARs) in toxicology: a historical perspective. *Journal of Molecular Structure (Theochem),* **622**, 1–22.

Schultz, T.W., Holcombe, G.W., Phipps, G.L. (1986), Relationships of quantitative structure-activity to comparative toxicity of selected phenols in the Pimephales promelas and *T. pyriformis* test systems. *Ecotoxicology and Environmental Safety*, **12**, 146–153.

Schultz, T.W., Lin, D.T., Wesley, S.K. (1992). QSARs for monosubstituted phenols and the polar narcosis mechanism of toxicity. *Quality Assurance, Good Practice, Regululation and Law*, **1**, 132–143.

Schultz T.W., Netzeva T.I., Cronin M.T.D. (2003b). Selection of data sets for QSARs: analyses of *T.* toxicity from aromatic compounds. *SAR and QSAR in Environmental Research*, **14**, 59-81.

Schultz, T.W., Sinks, G.D., Bearden, A.P., (1998). QSAR in aquatic toxicology: a mechanism of action approach comparing toxic potency to *Pimephales promelas, T. pyriformis,* and *V. fischeri*. In: Devillers, J. (Ed.), *Comparative QSAR*. Taylor & Francis, New York, pp 51–109.

Schultz, T.W., Sinks, G.D., Cronin, M.T.D. (1997). Identification of mechanism of toxic action of phenols to *T. pyriformis* from molecular structure. In: Fredenslund, F.C., and Schuurmann, G. (Eds.), *Quantitative Structure-Activity Relationships in Environmental Sciences*. Vol. VII, SETAC Press, Pensacola, FL, pp 329–342.

Schüürmann, G., Flemmig, B., Dearden, J. C. (1997). CoMFA study of acute toxicity of nitrobenzenes to *T. pyriformis*. In: Chen, F., and Schüürmann, G., (Eds.), *QuantitativeStructure-Activity Relationships in Environmental Sciences*. Vol. VII, SETAC Press: Pensacola, FL, pp 315-327.

Selassie, C.D., Garg, R., Kapur, S., Kurup, A., Verma, R.P., Mekapati, S.B., Hansch, C. (2002). Comparative QSAR and the radical toxicity of various functional groups. *Chemical Reviews,* **102**, 2585-2605.

Sinks, G.D., Carver, T.A., Schultz, T.W. (1998). Structure-toxicity relationships for aminoalcohols: A comparison with alkanols and alkanamines. *SAR and QSAR in Environmental Research*, **9**, 217–228.

Staples, C. A., Adams, W. J., Parkerton, T. F., Gorsuch, J. W., Biddinger, G. R., Reinert, K. (1997). Aquatic toxicity of eighteen phthalate esters. *Environmental Toxicology and Chemistry,* **16**, 875-891.

Tao S., Xi X., Xu F., Dawson R. (2002a). A QSAR model for predicting toxicity (LC_{50}) to rainbow trout. *Water Research*, **36**, 2926-2930.

Tao, S., Xi, X., Xu, F., Li, B., Cao, J., Dawson, R. (2002b). A fragment constant QSAR model for evaluating the EC_{50} values of organic chemicals to *D. magna. Environmental Pollution,* **116**, 57-64.

Terada, H. (1990). Uncouplers of oxidative phosphorylation. *Environmental Health Perspectives,* **87**, 213–218.

Todeschini, R. (1996). *WHIM-3D/QSAR*-Rel. 3.5 for Windows, Talete srl, Milan (Italy). Free download from the web page: http://www.disat.unimi.it/chm.

Todeschini, R., Gramatica, P. (1997). 3D-modelling and prediction by WHIM descriptors. Part 5. Theory development and chemical meaning of the WHIM descriptors. *Quantitative Structure-Activity Relationships*, **16**, 113-119.

Tomlin, C. (1994). The Pesticide Manual: Incorporating the Agrochemicals Handbook, 10th Edition. British Crop Protection Council, Farnham, Surrey.

Toropov, A.A., and Benfenati, E. (2004). QSAR modelling of aldehyde toxicity by means of optimization of correlation weights of nearest neighbouring codes. *Journal of Molecular Structure (Theochem)*, **676**, 165–169.

Toropov, A.A., and Schultz, T.W. (2003). Prediction of aquatic toxicity: use of optimization of correlation weights of local graph invariants. *Journal of Chemical Information and Computer Sciences,* **43**, 560-567.

Toropov, A.A., and Toropova, A.P. (1998). Optimization of correlation weights of the local graph invariants: use of the enthalpies of formation of complex compounds for the QSPR modelling. *Russian Journal of Coordination Chemistry*, **24**, 81–85.

Toropov A.A., and Toropova, A.P. (2002). QSAR modeling of toxicity on optimization of correlation weights of Morgan extended connectivity. *Journal of Molecular Structure (Theochem),* **578**, 129-134.

Toropova, A.P., Toropov, A.A., Ishankhodzhaeva, M.M., Parpiev, N.A. (2000). QSPR modeling of stability constants of coordination compounds by optimization of correlation weights of local graph invariants. *Russ. J. Inorg. Chem.*, **45**, 1057–1059.

Vaes, W.H.J., Ramos, E.U., Verhaar, H.J.M., Hermens, J.L.M. (1998). Acute toxicity of non-polar versus polar narcosis: Is there a difference?. *Environmental Toxicology and Chemistry*, **17**, 1380-1384.

Verhaar, H.J.M., Busser, F.J.M. and Hermens, J.L.M. (1995a). A surrogate parameter for the baseline toxicity content of contaminated water. *Environ. Sci. Technol.,* **29,** 726–734.

Verhaar, H.J.M., Mulder, W., Hermens, J.L.M., Rorije, E., Langenberg, J.H., Peijnenberg, W.J.G.M., Sabljic, A., Gusten, H., Eriksson, L., Sjostrom, M., Hansen, B., Nouwen, J., Karcher, W. (1995b). Overview of structure-activity relationships for environmental endpoints. Part 1: general outline and procedure. Report of the EU-DG-XII project QSAR for predicting fate and effects of chemicals in the environment (contract No EV3V-CT92-0211).

Verhaar, H.J., Van Leeuwen, C.J., Bol, J., Hermens, L.M. (1994). Application of QSARs in risk management of existing chemicals. *SAR and QSAR in Environmental Research*, **2**, 39-58.

Verhaar, H.J.M., Van Leeuwen, C.J., Hermens, J.L.M. (1992). Classifying environmental pollutants. 1: Strucuture activity relationships for predicting aquatic toxicity. *Chemosphere***, 25,** 471-491.

Wang, X., Dong, Y., Wang, L., Han, S. (2001). Acute toxicity of substituted phenols to *Rana japonica* tadpoles and mechanism-based quantitative structure-activity relationship (QSAR) study. *Chemosphere*, **44**, 447-455.

Wei, D.B., Zhai, L.H., Hu, H.Y. (2004). QSAR-based toxicity classification and prediction for single and mixed aromatic compounds. *SAR and QSAR in Environmental Research***, 15**, 207-216.

Wilson, L.Y., and Famini, G.R. (1991). Using theoretical descriptors in quantitative structure-activity relationships: Some toxicological indices. *Journal of Medicinal Chemistry*, **34**, 1668-1674.

Wong, H-W., Li, X., Swihart, M.T., Broadbelt, L.J. (2003). Encoding of Polycyclic Si-Containing Molecules for Determining Species Uniqueness in Automated Mechanism Generation. *Journal of Chemical Information and Computer Sciences,* **43***,* 735-742.

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

Xu, M., Zhang, A., Han, S., Wang, L. (2002). Studies of 3D-quantitative structure–activity relationships on a set of nitroaromatic compounds: CoMFA, advanced CoMFA and CoMSIA. *Chemosphere*, **48**, 707–715.

Yu, H.X., Lin, Z.F., Feng, J.F., Xu, T.L., Wang, L.S. (2001). Development of quantitative structure activity relationships in toxicity prediction of complex mixtures. *Acta Pharmacologica Sinica*, **22**, 45-49.

Zhao, Y.H., Cronin, M.T.D., Dearden, J.C. (1998a). Quantitative Structure-Activity Relationships of Chemicals Acting by Non-polar Narcosis - Theoretical Considerations. *Quantitative Structure-Activity Relationships*, **17**, 131-138.

Zhao, Y.H., Ji, G.D., Cronin, M.T.D., Dearden, J.C. (1998b). QSAR study of the toxicity of benzoic acids to *V. fischeri*, *D. magna*, and carp. *Science of the Total Environment,* **216**, 205-215.

PART II. QSARS FOR TOXICITY TO TERRESTRIAL ORGANISMS

Abstract

A large number of QSARs for acute toxicity have been developed and published. Toxicity to a wide variety of biological systems was investigated, including bacteria, algae, *Daphnia*, fish, fungi, plants, and mammals. The aim of this paper is to review and compare some of these QSARs. The review focuses on recently-published QSAR models (since 2000), with the exception of neural network models. In the first part of the review QSARs for toxicity to aquatic organisms were presented. The second part includes QSARs for toxicity to terrestrial organisms – bacterial species, fungi, worms, insects, plants, mammalian cells, and mammals, including humans.

II.1. Introduction

The aim of this paper is to review and compare some of these QSARs. The review focuses on recently-published QSAR models (since 2000), with the exception of neural network models. Recent reviews on the application of neural networks for developing QSARs for acute aquatic and health toxicological endpoints are given in Kaiser (2003a) and Kaiser (2003b). In the first part of the review QSARs for toxicity to aquatic organisms were summarised. The second part includes published QSARs for toxicity to terrestrial organisms – bacterial species, fungi, worms, insects, plants, mammalian cells, and mammals, including humans.

II.2. QSARs for toxicity to terrestrial bacteria

Less QSAR studies were found in the literature compared to these on toxicity to aquatic bacteria. Toxicity to two Gram-negative bacteria *Burkholderia* and *Pseudomonas fluorescens* was investigated by means of the QSAR analysis.

Boyd et al. (2001) used Theoretical Linear Solvation Energy Relationship (TLSER) descriptors (Wilson and Famini, 1991) to develop QSARs for toxicity to *Burkholderia RASC c*2 and *P. fluorescens*. TLSER descriptors represented cavity, dipolarity/polarisability, and H-bonding terms. Cavity was represented by V_{mc}, equal to $V_m/100$, where V_m is the molecular solvent-excluded volume. Dipolarity/polarisability was represented by polarisability index

(π^*) defining the ease with which an electron could be moved or polarised, equal to α/V_m, where α is molecular polarisability. Hydrogen-bond acceptor ability was represented by two terms - the covalent contribution to Lewis basicity, ε_b, equal to the difference in energy between LUMO (the energy of the lowest unoccupied molecular orbital) of water and HOMO (the energy of the highest occupied molecular orbital) of solute, and the electrostatic basicity contribution (q^-) representing the largest negative atomic charge in the solute molecule. The most negatively charged atom interactmost strongly with a proton from a neighbouring molecule (Boyd et al., 2001). Analogously, hydrogen bond-donating ability was presented by ε_a – the energy difference between HOMO of water and LUMO of solute, and qH^+ - the largest positive charge of a hydrogen atom in the solute molecule.

The two bacterial species, *Burkholderia RASC c*2 and *Pseudomonas fluorescens*, were marked with *lux* genes, encoding for bioluminescence, and used to investigate toxicity of mono-, di- and tri-chlorophenols. Toxicity was measured as the concentration of compound that caused a 50% decline in bioluminescence following exposure to the compounds in aqueous solution (50% effective concentration value, EC_{50}).

The QSAR equation for the toxicity of chlorophenols to *RASC c*2 was the following:

$$\log EC_{50} = 349\ \varepsilon_a - 40\ q^- + 32\ qH^+ - 40.8 \qquad \text{(II.2.1)}$$
$$n = 15,\ R_A^2 = 0.885,\ s = 0.188,\ F = 37.0,\ p = 0.000$$

The QSAR derived for the toxicity of chlorophenols to *P. fluorescens* was the following:

$$\log EC_{50} = -1.35\ \log P - 511\ \varepsilon_a - 315\ \pi^* + 105 \qquad \text{(II.2.2)}$$
$$n = 15,\ R_A^2 = 0.880,\ s = 0.207,\ F = 43.8,\ p = 0.000$$

In the QSARs for both *RASC c*2 and *P. fluorescens*, the H-bond acidic term (ε_a) was used. The derived QSARs indicated that molecular parameters describing the H-bonding nature of a chlorophenol provided a better fit than regressions between toxicity data and logP alone.

The $\log EC_{50}$ values for the two luminescent bacteria were correlated with each other and with toxicity values for the bacterium *Vibrio fischeri* (30 min exposure inhibition concentrations EC_{50} in mmol/l), the fish *Pimephales promelas* (96-h exposure lethal concentration LC_{50} in

mol/l) and the protozoan *Tetrahymena pyriformis* (48-h exposure lethal concentration causing a 50% decline in population growth, mmol/l). The correlation between the $logEC_{50}$ data for *RASC c*2 and *P. fluorescens* gave a Pearson coefficient of R = 0.71. Stronger correlations between *RASC c*2 and *V. fischeri*, *P. promelas* and *T. pyriformis* were observed, with R values of 0.89, 0.92 and 0.93, respectively. Poorer correlations were observed between the *P. fluorescens* data and the *V. fischeri*, *P. promelas* and *T. pyriformis* data (R values of 0.81, 0.65 and 0.65, respectively). This suggested that *lux*-marked *RASC c*2 could provide an environmentally relevant surrogate for determining the toxicity of organic xenobiotic compounds to higher freshwater organisms.

Bundy et al. (2001) investigated the toxicity of one-, two-, and three-ring compounds to *P. fluorescens*. The EC values for 20% inhibition in the bacterial bioluminescence (EC_{80}) were used. Compounds were classified in two classes: class I – non-polar narcotics; class II – basic (pyridine-ring containing) heterocycles (polar narcotics).

The obtained QSARs were the following:

Class I compounds:

$$logEC_{80} = -1.31\ logP + 6.32 \qquad \text{(II.2.3)}$$

$n = 10$, $R_A^2 = 0.95$, $s = 0.37$, $p < 0.001$, $R^2_{cv} = 0.88$

Class II compounds: (II.2.4)

$$logEC_{80} = -0.80\ logP + 4.59$$

$n = 5$, $R_A^2 = 0.79$, $s = 0.36$, $p = 0.028$, $R^2_{cv} = 0.57$

Class I and class II compounds:

$$logEC_{80} = -1.14\ logP + 5.76 \qquad \text{(II.2.5)}$$

$n = 15$, $R_A^2 = 0.87$, $s = 0.49$, $p < 0.001$, $R^2_{cv} = 0.82$

It should be noted the high slope of logP in these equations.

The authors tried to introduce LUMO, HOMO, or the energy difference (LUMO – HOMO) in the QSARs, but no statistical improvement was shown.

II.3. QSARs for toxicity to fungi

Several investigations on toxicity to yeasts and moulds were found in the literature. These organisms are widely distributed in nature, and play important roles in many ecosystems (Wang et al., 2002b).

In a study by Trohalaki et al. (2000), toxicity data for 52 halogenated alkanes (HA) to the mould *Aspergillus nidulans* were used to compare QSARs developed by using semi-empirical and *ab initio* molecular orbital calculations. Toxicity was expressed as the lowest concentration inducing lethality in 63% of the cells of the mould (marked as D_{37} because 37% of the cells survive). The data were taken from Crebelli et al. (1995). The original data set consisted of 55 compounds (Crebelli et al., 1995). Trohalaki et al. (2000) excluded two compounds from their study, for which the toxicity data were based on mixtures of isomers, the reason for the exclusion of the third compound was not given by Trohalaki et al. (2000).

The geometries of the neutral species of each compound were optimised by using the semi-empirical quantum mechanical method AM1 and different *ab initio* methods. LUMO, polarisability (α), and molecular volume (V) were calculated on the basis of the optimised geometries.

LUMO energies calculated from HA anions generally showed correlations with the toxicity about two times higher than those obtained with the LUMO of the neutral species. This suggests that HA anions are more important for the toxic response than the neutral species.

Two-parameter QSARs were derived using LUMO of the neutral species and one of α or *V*. QSARs with LUMO and α were better than these with LUMO and V. In Table II.3.1. the QSARs derived with LUMO and polarisability (α) are given. Three-parameter QSARs were derived using polarisability (α), LUMO of the neutral species, and molecular volume (V) (presented in Table II.3.2).

Table II.3.1. QSARs derived with polarisability (α) and LUMO (reproduced from Trohalaki et al., 2000).

Theory [a]	Intercept	α [b] x 10^2	LUMO (eV^{-1})	R^2	F
AM1	-2.722 ± 0.241	5.312 ± 0.609	-0.505 ± 0.063	0.708	59.5
HF/STO-3G	-1.120 ± 0.383	7.319 ± 0.988	-0.209 ± 0.038	0.615	39.2
HF/6-31G**	-1.589 ± 0.274	3.986 ± 0.379	-0.355 ± 0.044	0.787	90.3
B3LYP/6-31G**	-3.229 ± 0.217	3.829 ± 0.365	-0.368 ± 0.049	0.775	84.3
MP2/6-31G**	-1.647 ± 0.274	3.893 ± 0.387	-0.331 ± 0.043	0.770	82.0

[a] AM1 - semi-empirical quantum mechanical method for molecular optimisation, HF/STO-3G, HF/6-31G**, B3LYP/6-31G**, and MP2/6-31G** - *ab initio* quantum mechanical methods; [b] polarisability was calculated in atomic units.

Table II.3.2. QSARs derived using polarisability (α), LUMO, and molecular volume (V) (reproduced from Trohalaki et al., 2000).

Theory [a]	Compounds	Intercept	α [b] x 10^2	LUMO (eV^{-1})	V x 10^2 ($Å^{-3}$)	R^2	F
AM1	all	-2.568 ± 0.306	7.474 ± 2.706	-0.498 ± 0.064	-1.080 ± 1.317	0.712	39.6
AM1	1-35	-2.408 ± 0.198	6.967 ± 1.689	-0.311 ± 0.052	-1.160 ± 0.802	0.829	46.7
HF/STO-3G	all	-1.030 ± 0.412	10.295 ± 4.986	-1.972 ± 0.043	-1.043 ± 1.712	0.618	25.9
HF/STO-3G	1-35	-1.584 ± 0.237	9.003 ± 2.504	-0.129 ± 0.026	-0.821 ± 0.846	0.852	55.7
HF/6-31G**	all	-1.594 ± 0.277	4.442 ± 1.560	-0.343 ± 0.060	-0.318 ± 1.054	0.787	59.1
HF/6-31G**	1-35	-1.495 ± 0.246	8.202 ± 2.396	-0.235 ± 0.048	-2.978 ± 1.514	0.844	52.1
B3LYP/6-31G**	all	-2.943 ± 0.279	6.053 ± 1.443	-0.301 ± 0.064	-1.640 ± 1.031	0.786	58.8
B3LYP/6-31G**	1-35	-2.307 ± 0.246	10.032 ± 2.284	-0.169 ± 0.052	-4.572 ± 1.532	0.836	49.3
MP2/6-31G**	all	-1.646 ± 0.277	4.197 ± 1.719	-0.323 ± 0.063	-0.215 ± 1.186	0.770	53.6
MP2/6-31G**	1-35	-1.283 ± 0.276	10.974 ± 3.692	-0.215 ± 0.053	-4.896 ± 2.392	0.833	48.3

[a] AM1 - semi-empirical quantum mechanical method for molecular optimisation, HF/STO-3G, HF/6-31G**, B3LYP/6-31G**, and MP2/6-31G** - *ab initio* quantum mechanical methods; [b] polarisability was calculated in atomic units.

These there-parameter QSARs had only slightly better statistics than the two-parameter models. The negative coefficients obtained for LUMO (i.e. toxicity increases as LUMO decreases) suggested that electrophilicity is an important factor for the toxicity by determining the relative ease with which HAs can undergo reductive metabolic transformation and free radical generation (Crebelli et al., 1995). For all quantum mechanical methods, coefficients for α are positive indicate that toxicity increases with increasing

polarisability. All coefficients for V are negative, although in some cases large errors were observed, which made the signs questionable.

QSARs were also derived with a subset of chlorinated compounds only (compounds 1–35). The models had similar signs for the coefficients, and were slightly better than the models for the whole data set (Table II.3.2).

The atomic-charge-weighted partial positive surface area, PPSA-3, was also calculated and used to derive QSARs. PPSA-3 is defined as:

$$PPSA\text{-}3 = \Sigma(P_i Q_i^+) \qquad (II.3.1)$$

where P_i are the surface area contributions of the i-th positive atoms in the molecule and Q_i are the partial atomic charges for the i-th positive atoms.

In Table II.3.3. the QSARs derived using polarisability (α), LUMO, and PPSA-3 are given.

Table II.3.3. QSARs derived with polarisability (α), LUMO, and PPSA-3 (reproduced from Trohalaki et al., 2000).

Theory [a]	Compounds	Intercept	α [b] x 10^2	LUMO (eV^{-1})	PPSA-3 x 10^2 ($Å^{-2}$)	R^2	F
AM1	all	-2.838 ± 0.197	4.094 ± 0.551	-0.659 ± 0.059	4.870 ± 0.959	0.810	68.3
AM1	1-35	-2.579 ± 0.177	4.338 ± 0.546	-0.410 ± 0.127	1.421 ± 1.760	0.820	44.1
HF/STO-3G	all	-1.690 ± 0.415	7.914 ± 0.953	-0.093 ± 0.055	-7.926 ± 2.874	0.668	32.1
HF/STO-3G	1-35	-1.770 ± 0.299	6.716 ± 0.566	-0.104 ± 0.049	-2.026 ± 2.428	0.851	55.2
HF/6-31G**	all	-1.274 ± 0.359	3.443 ± 0.551	-0.423 ± 0.067	0.829 ± 0.615	0.794	61.8
HF/6-31G**	1-35	-1.600 ± 0.514	3.502 ± 0.054	-0.279 ± 0.118	0.052 ± 0.798	0.823	44.8
B3LYP/6-31G**	all	-3.274 ± 0.214	3.084 ± 0.535	-0.494 ± 0.083	1.235 ± 0.662	0.790	60.2
B3LYP/6-31G**	1-35	-2.776 ± 0.209	3.597 ± 0.554	-0.145 ± 0.136	-0.628 ± 0.856	0.790	36.3
MP2/6-31G**	all	-1.303 ± 0.360	3.201 ± 0.613	-0.406 ± 0.067	1.047 ± 0.724	0.780	56.6
MP2/6-31G**	1-35	-1.299 ± 0.454	3.085 ± 0.575	-0.346 ± 0.103	0.635 ± 0.830	0.813	42.0

[a] AM1 - semi-empirical quantum mechanical method for molecular optimisation, HF/STO-3G, HF/6-31G**, B3LYP/6-31G**, and MP2/6-31G** - *ab initio* quantum mechanical methods; [b] polarisability was calculated in atomic units.

These QSARs had better statistical parameters than those derived using α, LUMO, and V.

The study demonstrated that QSARs based on the semi-empirical (AM1) MO calculations are worse than those based on *ab initio* (HF, B3LYP, and MP2) methods.

The same data set as Trohalaki et al. (2000) was investigated by Cronin et al. (2002b), who used response-surface approach to develop QSARs for the toxicity of the 55 halogenated alkanes to the mould *Aspergillus nidulans* (the toxicity data were presented as the lowest concentration indusing lethality in 63 % (i.e. survival in 37 %) of the cells of the mould, D_{37} values, reported as mmol/l). The following QSAR was obtained:

$$\log 1/D_{37} = 0.598 \log P - 0.331 \text{ LUMO} - 2.17 \qquad \text{(II.3.2)}$$
$$n = 55,\ R^2 = 0.615,\ s = 0.413,\ F = 44.2,\ R^2_{cv} = 0.567$$

This equation has worse statistical fit than the ones obtained by Trohalaki et al. (2000) using polarisability and LUMO to predict the mould toxicity.

Wang et al. (2002b) developed QSARs based on 24 nitroaromatic compounds for inhibition of growth of the yeast *Saccharomyces cerevisiae*. The minimum concentration that produced a clear inhibition zone within 12 h (C_{miz}) was used as the toxicity parameter. The toxicity data were taken from Liao et al. (1997). The inhibition zone was considered to be clear when single yeast colonies were distinguished in it. Wang et al. (2002b) used the response-surface approach to develop QSARs (Cronin and Schultz, 2001), based on the octanol/water partition coefficients (logP) expressing hydrophobicity, and LUMO values, expressing electrophilicity (affinity of a molecule to accept electrons).

According to Wang et al. (2002b), the nitro group is a strong π-electron acceptor, lowering the electron density of the aromatic ring. Inside the nitro group, excess electronic charges are mainly localised at the oxygen atoms, while the nitrogen atom is typically electron-deficient. That is why nitroaromatic compounds show enhanced reactivity for the attack of nucleophilies at aromatic ring carbons, and for reactions with reducing agents. According to Wang et al. (2002b) the exact mechanisms of toxic action for nitrobenzenes are complex and unclear due to the variety of substituents on the nitrobenzenes.

Wang et al. (2002b) separated the investigated nitrobenzenes into two groups: mono-nitrobenzenes and di-nitrobenzenes. All di-nitrobenzenes had greater toxicity than the mono-nitrobenzenes with similar logP values.

For the group of mono-nitrobenzenes, a model based on logP was derived.

$C_{miz} = 0.83 \log P + 0.28$ (II.3.3)
$n = 16, R^2 = 0.68, s = 0.22, F = 29.9$

For the toxicity of the di-nitrobenzenes, a significant correlation with logP was not found.

The following relationship was derived for the whole data set, including both mono-nitrobenzenes and di-nitrobenzenes with LUMO:

$C_{miz} = -0.84 \text{ LUMO} + 0.62$ (II.3.4)
$n = 24, R^2 = 0.71, s = 0.32, F = 52.8$

The following QSARs with LUMO were obtained for the separated groups of mono-nitrobenzenes (n = 16) and di-nitrobenzenes (n = 8):

$C_{miz} = -0.75 \text{ LUMO} - 1.66$ (II.3.5)
$n = 16, R^2 = 0.56, s = 0.24, F = 18.8$

$C_{miz} = -0.91 \text{ LUMO} + 1.03$ (II.3.6)
$n = 8, R^2 = 0.82, s = 0.28, F = 26.5$

The QSARs showed that there is a difference between the relative reactivity of mono-nitrobenzenes and di-nitrobenzenes: for mono-nitro derivatives, both logP and LUMO gave significant correlations, while for di-nitrobenzenes, a significant relationship was observed only with LUMO, with the correlation being better than that between toxicity and LUMO for mono-nitrobenzenes.

The response-surface approach gave the following QSAR for all investigated compounds:

$C_{miz} = 0.41 \log P - 0.89 \text{ LUMO} - 0.46$ (II.3.7)

$n = 24$, $R^2 = 0.87$, $s = 0.22$, $F = 72.0$, $R^2_{cv} = 0.86$

According to Wang et al (2002b) the derived QSARs have satisfactory statistical parameters and sound mechanistic interpretation.

II.4. QSARs for toxicity to worms

An investigation of the acute toxicity of mixtures to the worm *Tubifex tubifex* was performed by Tichý et al. (2002). The toxicity was measured as the compound concentration that inhibits 50% of the movement of the worms after a 3-min exposure (EC_{50} values). The authors investigated the dependence of toxicity on the mixture composition (molar ratios of mixture components) and developed quantitative composition-activity relationships (QCAR). Two types of mixtures were investigated - benzene with nitrobenzene, and aniline with ethanol.

The mixture toxicity was expressed as normalised EC_{50} values ($EC50_{norm}$):

$$EC50_{norm} = EC50(A+B) / (R_{A,\,norm} * EC50_A + R_{B,\,norm} * EC50_B) \quad \text{(II.4.1)}$$

where $R_{A,\,norm} = (R_A/EC50_A) / (R_A/EC50_A + R_B/EC50_B)$ is the normalised molar ratio for component A (similarly $R_{B,\,norm}$ is the normalised molar ratio of compound B in the mixture);
$R_A = c_A / (c_A + c_B)$ is the molar ratio of compound A in the mixture (similarly R_B is the molar ratio of compound B in the mixture);
c_A and c_B are the concentrations of compounds A and B in the mixture.

$EC50_{norm}$ is used as a measure of additivity of the mixture components. The compounds A and B are additive in their potency if a replacement of some part of one compound in a mixture by an equipotent part of other compound does not change the potency of the mixture ("concentration addition"). $EC50_{norm}$ could also be presented by the following equation:

$$EC50_{norm} = EC50(A)_{mix} / EC50_A + EC50(B)_{mix} / EC50_B \quad \text{(II.4.2)}$$

where $EC50(A)_{mix}$ and $EC50(B)_{mix}$ are the individual contributions of the two compounds to .the toxicity EC50(A+B) of the mixture (equal to R_A*EC50(A+B) and R_B*EC50(A+B) respectively). From this equation it can be seen that additivity is defined by $EC50_{norm} = 1$.

The changes in mixture toxicity with changing mixture composition were visualised as R-plots (plotting mixture toxicity $EC50_{norm}$ against R_{norm} of one of the compounds).

The results confirmed the additivity of the toxicity in the binary mixture of benzene with nitrobenzene ($EC50_{norm} = 1$ in the whole range of composition).

The binary mixture of aniline with ethanol was found to exhibit both potentiation and inhibition (synergism, $EC50_{norm} < 1$, and antagonism, $EC50_{norm} > 1$) depending on mixture composition. The following non-linear QCAR was found:

$$EC50_{norm} = 1 + 0.42\ R_{norm} - 7.67\ R_{norm}^2 + 28.74\ R_{norm}^3 - 35.08\ R_{norm}^4 + 13.68\ R_{norm}^5 \quad \text{(II.4.3)}$$

where R_{norm} is the normalised molar ratio for ethanol.

Potentiation of the mixture occurred when $R_{norm,\ ethanol}$ was between 0.1 and 0.35, and inhibition occurred for $R_{norm,\ ethanol}$ between 0.5 and 0.9. For $R_{norm,\ ethanol}$ between 0.35 and 0.5 $EC50_{norm}$ was approximately equal to 1 within the experimental error.

The gaseous phase composition of the mixtures was also analysed. It was found that if the concentrations of the components in the gaseous phase behave non-ideally (i.e. the gaseous concentration of a mixture component depends on the concentration of the other component in the mixture), acute toxicity of the same mixture is not additive.

II.5. QSARs for toxicity to insects

Grodnitzky and Coats (2002) investigated the toxicity of 30 monoterpenoids to the house fly by using the GEometry, Topology and Atomic Weights AssemblY (GETAWAY) descriptors. LD_{50} values were obtained after 24 h exposure. GETAWAY descriptors were used to encode information about the three-dimensional structures, size and shape of the molecules (Consonni et al., 2002). The descriptors assign higher values to atoms that are at a greather

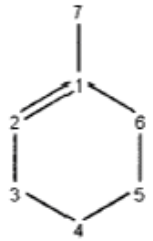
distance from the molecule's geometric centre. Higher values can be interpreted as the atoms that are more accessible to external interactions. In addition, Mulliken populations (encoding information for the electron density) and an electrotopological state descriptor were used to represent electron density around certain atoms in the molecule.

A QSAR model obtained for 20 alicyclic monoterpenoids was the following:

$\log 1/LD_{50} = 15.1$ [E-state] + 214 [GETAWAY] – 106 [interaction] – 30.7 (II.5.1)
$n = 20$, $R^2 = 0.86$, $s = 0.11$, $F = 32.6$, $R^2_{cv} = 0.72$

In this QSAR model, the GETAWAY descriptor accounts for molecular size and shape. The E-state descriptor on atom 2 (Figure II.5.1) represents electronic properties. In this model, an interaction effect between the two descriptors was also introduced. The negative coefficient of the interaction term indicated that toxicity decreases if both of the descriptor values are too large.

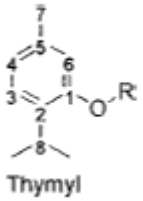
Figure II.5.1. Methyl cyclohexane carbon skeleton with a double bond present between atoms 1 and 2 (taken from Grodnitzky and Coats, 2002).



Another QSAR was developed for thymol and nine derivatives. Two of the derivatives were ethers and the other seven were esters. A linear relationship was obtained between toxicity of the thymol derivatives and the Mulliken population (electron density) around atom 1 (Figure II.5.2):

$\log 1/LD_{50} = -11.6$ [Mulliken population] + 65.3 (II.5.2)
$n = 10$, $R^2 = 0.90$, $s = 0.08$, $F = 68.5$, $R^2_{cv} = 0.84$

Figure II.5.2. Numbering of carbon atoms in thymol compounds (taken from Grodnitzky and Coats, 2002).



The model provides insight into the important regions of the molecules responsible for their anti-insecticide activity.

Sverdrup et al. (2002) investigated toxicity of 16 polycyclic aromatic hydrocarbons (PAHs) to soil-dwelling springtail *Folsomia fimetaria*. Toxicity was determined as compound pore-water concentrations estimated to give a 50% reduction in survival (LC_{50} values in units µmol/l), and compound pore-water concentrations estimated to give 10% reduction in reproductive output (EC_{10} values in units µmol/l).

LC_{50} and EC_{10} values were significantly negatively correlated with logP for PAHs having logP in the range 3.3 – 5.2:

$\log LC_{50} = -1.2 \log P + 5.6$ (II.5.3)
$n = 8, R^2 = 0.88, p = 0.0005$

$\log EC_{10} = -0.97 \log P + 4.0$ (II.5.4)
$n = 7, R^2 = 0.80, p = 0.006$

Anthracene was more than 6 times more toxic than predicted by its logP, and was excluded from this equation.

According to the authors, the observed absence of toxicity for highly lipophilic PAHs (logP > 5.6) could, in most cases, be explained by limited water solubility, resulting in insufficient compound concentrations in the pore-water.

II.6. QSARs for toxicity to plants

Wang et al. (2002a) investigated the effects of 27 substituted phenol derivatives on the root elongation of terrestrial macrophytes *Cucumis sativus*. Seeds of the plans were incubated for 48 h to a solution of the toxicant, and the root lengths were measured after that. The toxicity was measured as RC_{50} values – compound concentrations (mg/l) for which the average root length was 50% of those in the corresponding control. The RC_{50} data were converted to the logarithmic form of the inverse molar concentration (IRC_{50}).

Again, the response-surface approach was applied, using the combination of hydrophobicity and eletrophilicity parameters (logP and LUMO) to derive QSARs:

$$IRC_{50} = 0.65\ logP - 0.72\ LUMO + 2.40 \qquad (II.6.1)$$
$$n = 27,\ R^2 = 0.91,\ p < 0.000001$$

Adding the largest negative net atomic charge (q^-), used to describe the H-bonding acceptor capacity (Boyd et al., 2001), resulted in a slightly better QSAR:

$$IRC_{50} = 0.64\ logP - 0.72\ LUMO + 0.23\ q^- + 2.40 \qquad (II.6.2)$$
$$n = 27,\ R^2 = 0.94,\ p < 0.000001$$

Another investigation from the same group of authors (Wang et al., 2002c) included development of QSARs for the toxic effect of 42 phenols on the germination rate of *Cucumis sativus*. The toxicity was quantified by the GC_{50} value, the concentration for which the average germination rate was 50% of that in the corresponding control. It was converted to the logarithmic form of the inverse molar concentration prior to QSAR analysis.

Again, QSARs were developed by using the response-surface approach:

$$log1/GC_{50} = 0.56\ logP - 0.65\ LUMO + 2.26 \qquad (II.6.3)$$
$$n = 42,\ R^2 = 0.74,\ s = 0.22,\ F = 52.3$$

Four benzoic acid derivatives were poorly predicted by this equation, with their toxicity significantly overestimated. According to the authors, a possible reason could be the strong dissociation of carboxyl group, which makes it incorrect to model their toxicity with logP dependent QSARs. After removing the four benzoic acids, the following relationship was obtained:

$\log 1/GC_{50} = 0.60\ \log P - 0.71\ LUMO + 2.24$ (II.6.4)
$n = 38,\ R^2 = 0.84,\ s = 0.18,\ F = 89.3$

Another two poorly predicted compounds (catechol and 2,4-dinitrophenol) were identified from this equation, with their toxicity significantly underestimated. The removal of the two chemicals resulted in the following QSAR:

$\log 1/GC_{50} = 0.70\ \log P - 0.66\ LUMO + 2.17$ (II.6.5)
$n = 36,\ R^2 = 0.89,\ s = 0.14,\ F = 134$

2,4-Dinitrophenol was the only dinitrophenol investigated in this study. It could elicit its high toxicity by a different toxic mechanism, such as respiratory uncoupling (Russom et al., 1997) or electrophilic reactivity. Additionally, it might be strongly dissociated in the test medium. According to Schultz et al. (1997) catechols are capable of undergoing tautomeric transforamtion to more electrophilic semiquinones, which could possess a more potent toxic mechanism of action via futile redox recycling (O'Brien, 1991).

Enache et al. (2000) investigated *in vivo* toxicities of 12 metal ions to cabbage plants (*Brassica oleracea L var capitata cv Soshu*). Toxicity data were measured and expressed as the negative logarithm of the heavy metal content in the outer leaves, which produced a 50% reduction of the relative growth of the cabbage ($pRrg_{50}$ values in units mol/l). QSARs were developed using ion-specific physicochemical parameters:

$pRrg_{50} = -8.54\ X_{AR} - 1.59\ \Delta E_0 - 3.04X + 12.2$ (II.6.6)
$n = 11$ (Cu^{2+} excluded), $R_A = 0.90,\ s = 0.49,\ F = 15.1$

$pRrg_{50} = 0.0187\ AW - 1.10\ \Delta E_0 + 0.264\ \Delta IP - 1.02$ (II.6.7)
$n = 11$ (Cu^{2+} excluded), $R_A = 0.85,\ s = 0.59,\ F = 10.0$

where X_{AR} is Allred-Rochow electronegativity; X is the Pauling's electronegativity; ΔE_0 is the absolute value of the electrochemical potential between the ion and its first stable reduced state; AW is the atomic weight; ΔIP ionisation potential differential.

The same group of authors (Enache et al., 2003), investigated metal toxicity to "Sunspot" of 12 metal ions: Ca(II), Cd(II), Co(II), Cu(II), K(I), Li(I), Mg(II), Mn(II), Na(I), Ni(II), Zn(II) and La(III). "Sunspot" is a new ornamental dwarf variety of sunflower that produces giant single yellow blooms. Toxicity data in sunflower callus (undifferentiated mass of dividing cells) cultures were determined for a 14 day exposure. *In vitro* EC_{50} concentrations (metal concentration causing a 50% decrease in the rate of callus development) in the applied solution and determined in the callus tissue were used to develop QSARs. The correlations between the toxicity data in callus tissue and solution were modest (R approximately 0.65).

The following QSARs were obtained for the ED_{50} values determined in solution:

$$-\log EC_{50} = -0.502\ |\log K_{OH}| + 7.93 \qquad \text{(II.6.8)}$$
$n = 12,\ R_A^2 = 0.77,\ s = 0.61,\ F = 38.1$

$$-\log EC_{50} = 0.165\ \log K\ (EDTA) + 0.430 \qquad \text{(II.6.9)}$$
$n = 12,\ R_A^2 = 0.74,\ s = 0.64,\ F = 32.8$

where $|\log K_{OH}|$ is the first hydrolysis constant of the metal, logK (EDTA) is the stability constant of metal ion complexes with EDTA. The first hydrolysis constant is a measure of the ability of a metal ion to form a metal hydroxide. The parameter is considered to account for metal ion affinity to intermediate ligands such as those with oxygen donor atoms.

The following QSARs obtained for the ED_{50} values determined in solution (excluding Lanthanum (La(III)), whose toxic concentration was not determined in callus tissues):

$$-\log EC_{50} = 1.48\ X_m^2 r - 0.149 \qquad \text{(II.6.10)}$$
$n = 11,\ R_A^2 = 0.88,\ s = 0.43,\ F = 75.3$

$$-\log EC_{50} = 2.91\ X - 1.74 \qquad \text{(II.6.11)}$$

n = 11, R_A^2 = 0.86, s = 0.45, F = 66.1

$-\log EC_{50} = 0.881E^o + 3.58$ (II.6.12)
n = 11, R_A^2 = 0.83, s = 0.52, F = 48.22

where X_m^2r is the covalent index, X is the electronegativity of the element, E^o is the standard reduction potential.

The following QSARs were obtained for the EC_{50} values determined in callus tissue (n = 11 because the concentration for La(III) was not determined in callus tissue):

$-\log EC_{50} = -3.64\ IR + 2.01$ (II.6.13)
n = 11, R_A^2 = 0.81, s = 0.38, F = 44.30

$-\log EC_{50} = -0.0222\ \text{At radius} + 1.87$ (II.6.14)
n = 11, R_A^2 = 0.80, s = 0.40, F = 40.91

where IR is the ionic radius, At radius is the atomic radius

In these QSARs different types of properties appeared to correlate with toxicity, for example atomic properties like atomic radii and ionisation energies, chemical properties of metals like electronegativity, and other properties considered to reflect interaction with ligands such as the chemical softness parameter, the covalent index, the first hydrolysis constant, the stability constant of a metal complex in solution. According to Enache et al. (2003) it is difficult to identify the potential significance of these descriptors. Despite the limitations of the data and their interpretation, the study showed that it is possible to model metal toxicity in plant systems using a variety of properties of the metal ions.

II.7. QSARs for toxicity to mammalian cells

Kapur et al. (2000) measured the cytotoxicities of a series of para-substituted benzyl alcohols in a rapidly dividing cancer cell line, L1210 (murine leukemia cells), after a 48 h exposure.

The QSAR obtained was:

$\log 1/C = 0.51 \log P + 2.09$ (II.7.1)

$n = 11$, $R^2 = 0.870$, $s = 0.158$, $R^2_{cv} = 0.804$

omitted compounds: 4-$C(CH_3)_3$-benzyl alcohol; 4-NH_2-benzyl alcohol.

C represents the molar concentration producing 50% inhibition of cell growth in 48 h (IC_{50}).

The authors suggested that the observed toxicity of 4-$C(CH_3)_3$-benzyl alcohol is lower than that determined by logP due to the branching of the substituent. The authors suggested that the observed toxicity of the 4-NH_2 derivative is higher than predicted from logP because of free radical formation by hydrogen abstraction from the amino group.

On the basis of the derived model, the authors concluded that most para-substituted benzyl alcohols exert their cytotoxicity in rapidly growing murine leukemia cells via a polar narcosis mechanism that is determined by the hydrophobicity alone.

The authors tried to improve the statistical parameters of the model by adding the Brown variation of the Hammett constant σ^+, but they did not obtain better results. The Brown variation of the Hammett constant σ^+ takes into account the contribution of substituents through conjugation to electron-deficient reactivity sites attached to a benzene ring. π-Electron-releasing and π-electron-withdrawing groups are associated with negative and positive values of σ^+ respectively (Morao and Hillier, 2001).

Hansch and Gao (1997) reviewed the literature on free radical reactions. They found that in 27 QSAR equations describing the formation of free radicals by abstracting $H^{\cdot}$ from the phenolic hydroxy group, 25 equations were correlated with – negative values of σ^+, where negative values indicate that substituents are electron-releasing, i.e. increasing electron density in the benzene ring.

Hansch et al. (2000) investigated 15 phenol derivatives substituted at the 2-, 3- and 4-positions for their cytotoxic action on rat liver epithelial cells. The toxicity was determined by measuring the molar concentration (C) of phenol that increases the extracellular

concentration of lactate dehydrogenase (LDH) by 50%, 1 h after initiation of chemical exposure.

The QSAR obtained was the following:

$$\log 1/C = -0.98\ \sigma^{+} + 0.77\ \log P + 0.23 \qquad \text{(II.7.2)}$$

$n = 12,\ R^2 = 0.872,\ s = 0.213,\ R^2_{cv} = 0.764$

Three compounds were excluded as outliers from the derivation of this equation: 4-OH, 3-OH, 4-$CH_2CH_2NH_2$ phenols. According to the authors, the 4-$CH_2CH_2NH_2$ phenol is partially ionised under experimental conditions and therefore its logP value is uncertain. The higher activity of the 4-OH-phenol (hydroquinone) is possibly due to its ease of oxidation to quinone. The reason why 3-OH phenol is an outlier was not explained.

An equation with the bond dissociation energy (BDE) was also obtained (Equation II.7.4). BDE is the energy associated with the abstraction of a hydrogen atom from the hydroxy moiety of the substituted phenols by a phenoxy radical according to the following reaction (Hansch et al., 2000):

$$X\text{-}C_6H_4OH + C_6H_5O^{\cdot} \rightarrow X\text{-}C_6H_4O^{\cdot} + C_6H_5OH \qquad \text{(II.7.3)}$$

$$\log 1/C = -0.11\ BDE + 0.76\ \log P + 0.21 \qquad \text{(II.7.4)}$$

$n = 12,\ R^2 = 0.925,\ s = 0.163,\ R^2_{cv} = 0.827$

These models suggested a free radical mechanism of toxicity.

Moridani et al. (2002) investigated the toxicity of a set of 22 polyphenols to isolated rat hepatocytes. As a measure of toxicity, the concentration (μmol/l) at which a compound caused 50% cell death after incubation for 2 h was used. The toxicity to isolated rat hepatocytes was compared with that to HeLa tumor cells (the polyphenol concentration (μmol/l) required to inhibit HeLa tumor cell proliferation after 3 days). Values for toxicity to HeLa tumor cells were available for 13 of the investigated compounds. Experimental logP values, measured by the authors, were applied in deriving QSARs.

QSAR Equations II.7.5. and II.7.6. were derived for the toxicity to isolated rat hepatocytes:

$\log C_{hepatocyte} = -0.653 \log P + 4.12$ (II.7.5)
$n = 19, R^2 = 0.800, s = 0.326$
outliers excluded: nordihydroguaiaretic acid, kaempferol and genistein.

Nordihydroguaiaretic acid was 10 times more toxic than the calculated toxicity from the equation, whereas genistein and kaempferol were several times less toxic. The authors offered no reason for the anomalous behavior of nordihydroguaiaretic acid, kaempferol and genistein.

The authors added the polyphenol redox potential ($E_{p/2}$, V) to the QSAR, to account for the ease of oxidation of the compounds. Values of $E_{p/2}$ were available for 16 of the investigated compounds.

$\log C_{hepatocyte} = -0.736 \log P + 0.383 E_{p/2} + 4.16$ (II.7.6)
$n = 15, R^2 = 0.847, s = 0.318$
outlier excluded: kaempferol

Nordihydroguaiaretic acid and genistein were not included in this equation because of missing $E_{p/2}$ values.

The QSARs for the toxicity to HeLa tumor cells are given by Equations II.7.7. and II.7.8:

$\log C_{tumor} = -0.337 \log P + 2.40$ (II.7.7)
$n = 11, R^2 = 0.899, s = 0.131$
outliers excluded: chrysin and luteolin

Chrysin was seven times less toxic than the calculated value and luteolin was four times more toxic than its calculated value.

$\log C_{tumor} = -0.353 \log P + 0.223 E_{p/2} + 2.39$ (II.7.8)
$n = 10, R^2 = 0.939, s = 0.114$

outliers excluded: chrysin and luteolin

The data for the two toxicity endpoints were correlated by the following equation:

$\log C_{tumor} = 0.529 \log C_{hepatocyte} + 0.235$ (II.7.9)
$n = 10$, $R^2 = 0.900$, $s = 0.130$, $F = 72$
outliers excluded: kaempferol, chrysin and luteolin

The equation suggests that the mechanisms of toxicity for kaempferol, chrysin and luteolin to isolated rat hepatocytes may be different from those of HeLa tumor cells, because inclusion of these chemicals in the regression equation significantly reduces the correlation.

The authors argued that the polyphenols act via a mitochondrial mechanism of cytotoxicity, on the basis of a previously shown correlation between the cytotoxicity of polyphenols to hepatocytes and the ability of these chemicals to collapse the hepatocyte mitochondrial membrane potential (Galati et al., 2000). Another piece of evidence for mitochondrial toxicity is the observation by Moridani et al. (2002) of an increase in polyphenol toxicity to hepatocytes isolated from fasted rats compared to the toxicity to hepatocytes prepared from fed rats. As fasting depletes glycogen stores in hepatocytes, placing a greater demand on mitochondrial function, the increased susceptibility of hepatocytes isolated from fasted rats supports the suggestion that the mitochondrion is the site of toxic action for polyphenols.

According to QSAR Equations II.7.5-II.7.8., higher lipophilicity favours higher toxicity. According to the proposed mechanism, higher toxicity is ensured by high intracellular concentration. Lipophilicity plays role by determining the penetration of the cell membrane in order to reach the necessary concentration in the cell.

Moridani et al. (2003b) obtained QSARs for cytotoxicity to isolated rat hepatocytes of 31 phenols, using $\log LC_{50}$ (concentration inducing 50% cytotoxicity in isolated rat hepatocytes, μmol/l) as the response variable and logP as the predictor variable:

$\log LC_{50} = -0.588 \log P + 4.65$ (II.7.10)
$n = 27$, $R^2 = 0.801$, $s = 0.261$

outliers excluded: hydroquinone, catechol, 4-nitrophenol, and 2,4-dinitrophenol

According to Moridani et al. (2003b), the incorrect predictions of the toxicities of hydroquinone, catechol, 4-nitrophenol and 2,4-dinitrophenol indicate that although logP contributes to the toxicity, the mechanisms of toxicity of these outliers were different from the mechanism of the remaining phenols, which agrees with previous work on other systems (for example Cronin et al., 2002a).

The following QSAR was obtained by including the dissociation constant (pK_a) as an additional descriptor:

$$\log LD_{50} = -0.595 \log P + 0.197\ pK_a + 2.67 \qquad \text{(II.7.11)}$$

$n = 28$, $R^2 = 0.859$, $s = 0.218$

outliers excluded: hydroquinone, catechol, 2-nitrophenol

The equation suggests that cytotoxicity increases with increasing lipophilicity and ionisation at pH 7.4. At this pH, 2,4-dinitrophenol and 4-nitrophenol are 99.9% and 64% ionised, respectively.

The dissociation constant pK_a was replaced with the Brown variation of the Hammett constant σ^+, resulting in the QSARs given by Equations II.7.12 and II.7.13:

$$\log LD_{50} = -0.594 \log P - 0.552\ \sigma^+ + 4.54 \qquad \text{(II.7.12)}$$

$n = 28$, $R^2 = 0.853$, $s = 0.223$

outliers excluded: hydroquinone, catechol, 2-nitrophenol

According to this equation, phenols with electron-withdrawing groups (positive σ^+ values) are more toxic, whereas phenols with electron-releasing groups (more negative σ^+ values) are less toxic to isolated rat hepatocytes.

An equation was also derived for only phenols with electron-releasing groups, which had slightly better statistical parameters:

$\log LD_{50} = -0.597 \log P - 0.424 \sigma^{+} + 4.59$ (II.7.13)

$n = 21$, $R^2 = 0.892$, $s = 0.206$

outliers excluded: hydroquinone and catechol

Replacing σ^{+} with the bond dissociation energy (BDE) or redox potential $E_{p/2}$ resulted in two more QSAR equations (II.7.14 and II.7.15):

$\log LD_{50} = -0.601 \log P - 0.040 BDE + 4.61$ (II.7.14)

$n = 23$, $R^2 = 0.827$, $s = 0.223$

outliers excluded: hydroquinone, catechol, 2-nitrophenol, 4-nitrophenol

$\log LD_{50} = -0.529 \log P + 2.077 E_{p/2} + 2.806$ (II.7.15)

$n = 15$, $R^2 = 0.561$, $s = 0.383$

outlier excluded: 4-nitrophenol

Miyoshi et al. (1987) reported the mitochondrial uncoupling activities of phenols such as 2,4-dinitrophenol, 3-nitrophenol, 2,4-dichlorophenol, 4-chlorophenol, 4-*t*-butylphenol, 4-methylphenol and phenol.

σ^{+} and BDE have negative coefficients in the presented QSARs, which means that compounds with numerically smaller σ^{+} (possessing electron-releasing groups) and BDE values are less toxic. As smaller values of σ^{+} and BDE favour toxicity by forming free radicals (Hansch et al. (2000), this result implies that phenoxy radical formation does not play a major role in the toxicity of simple phenols to hepatocytes, and mitochondrial uncoupling could be the more likely toxicity mechanism. However, Moridani et al. (2003b) noted that most of the phenols in their study contained electron-releasing groups, which could contribute to phenoxy radical formation. Equation II.7.15, which suggests that phenols with numerically smaller redox potentials are more toxic, is evidence for the possible involvement of phenoxy radicals in phenol toxicity to hepatocytes.

Catechol and hydroquinone were outliers from the QSARs. They can undergo oxidation and form quinones, which are reactive intermediate species that can react with protein thiols, resulting in toxic effects.

Another investigation of the same group of authors was that of Moridani et al. (2004), who determined the toxocty of catechols to isolated rat hepatocites (toxicity expressed as the concentration at which a compound causes 50% cell death, LC_{50} in μmol/l).

Moridani et al. (2004) used the following rules to identify the outliers from the QSAR equations: (1) when an individual data point falls outside of 95% confidence limits around the regression line (certainty limit rule), (2) when the removal of a data point improves R^2 at least with 0.05 (correlation coefficient rule), and (3) when there is a scientific explanation why the outlier behaves differently from the rest of data points, either by a previously published mechanism of toxic action or a new explanation (scientific reasoning rule). Moridani et al. (2004) note that an outlier should satisfy all three rules.

The following QSARs were obtained:

$\log LD_{50} = -0.464 \log P + 3.72$ (II.7.16)
$n = 20$, $R^2 = 0.740$, $s = 0.372$
outliers excluded: 4-methoxycatechol, 3-methoxycatechol, L-dopa

When the ionisation constant (pKa) was included in the equation as a second descriptor, the toxicity predicition for the outliers inproved:

$\log LD_{50} = -0.343 \log P - 0.116 pKa + 4.39$ (II.7.17)
$n = 22$, $R^2 = 0.738$, $s = 0.375$
outlier excluded: 4-methoxycatechol

The authors note that both hydroxy groups of the catechols are electron-releasing groups, which could contribute to the toxicity via semiquinone radical formation. However, the aouthors state that the descriptors, included in the QSARs might account for the fraction of a catechol that is in its unionised form and is available to cross the biological membrane and to enter the hepatocyte or intracellular organelles where is the supposed site of toxic action. Unpublished data of the authors have shown that catechols induced a collapse of the hepatocyte mitochondrial membrane potential before cytotoxicity ensued.

According to Moridani et al. (2004), if a compound satisfies the first two rules to be identified as an outlier, (certainty limit and correlation coefficient rules), it could be automatically considered that there is a scientific reason for it to be an outlier (difference in the mechanism of toxic action of this compound). The authors note that sometimes it is difficult to explain the outliers scientifically by an alternative toxicity mechanism because the mechanism is not known or has not been investigated yet. Moridani et al. (2004) suggested all investigated catechols share a common non-specific mechanism, however some may have additional specific mechanisms either through metabolism or other mechanisms such as decoupling. For instance, to explain the outliers based on the scientific reasoning, the authors had found that the cytotoxicity induced by 3-hydroxyanisole and its metabolism to 4-methoxycatechol was prevented by cytochrome P450 inhibitors and the cytotoxicity was attributed to the quinone metabolite (Moridani et al., 2003a).

Etzenhouser et al. (2001) investigated ten esters of caffeic acid and dihydrocaffeic acids for their cytotoxicity to L1210 leukaemia and MCF-7 breast cancer cells in culture (expressed as IC_{50}, compound concentration that induces 50% inhibition of growth).

The following QSAR was derived for the inhibition of growth of L1210 cells by caffeic acid esters:

$\log 1/IC_{50} = 0.46\ \log P + 3.84$ (II.7.18)
$n = 9,\ R^2 = 0.915,\ s = 0.165,\ R^2_{cv} = 0.881$
outlier: C_6H_{13} ester (hexyl ester)

The following QSAR was derived for the inhibition of growth of L1210 cells by esters of dihydrocaffeic acid:

$\log 1/IC_{50} = 0.15\ \log P + 0.14\ E_S + 4.62$ (II.7.19)
$n = 9,\ R^2 = 0.906,\ s = 0.067,\ R^2_{cv} = 0.793$
outlier: $CH_2C_6H_5$ ester (benzylic ester)

where logP is the octanol/water partition coefficient of each compound; E_S is the Taft's steric parameter for each alkyl group of the alkoxy moiety of each ester.

The positive coefficient of E_S indicates that cytotoxicity is enhanced by a decrease in the size of the substituents. However, the low value of this coefficient and its high 5% confidence interval (comparable to the regression coefficient) suggests that it does not play an important role in toxicity.

The following QSARs (Equations II.7.20 and II.7.21) for the inhibition of growth of MCF-7 cells by caffeic esters were derived:

$$\log 1/C = 0.37 \log P + 2.64 \qquad \text{(II.7.20)}$$

$n = 9$; $R^2 = 0.956$; $s = 0.075$; $R^2_{cv} = 0.931$

outlier: $CH_2(CH_2)_6CH3$ ester

$$\log 1/C = 0.64 \log P - 0.06 \log P^2 + 2.34 \qquad \text{(II.7.21)}$$

$n = 10$, $R^2 = 0.958$, $s = 0.081$, $R^2_{cv} = 0.882$

The parabolic model was a marginal improvement over the linear model, and was used to establish an optimal hydrophobicity for high inhibitory activity in the breast cancer line. The optimum inhibition was found to be at a logP of 5.75.

The following QSAR was obtained for the inhibition of growth of MCF-7 cells by esters of dihydrocaffeic acid:

$$\log 1/C = 0.36 B_1 + 3.40 \qquad \text{(II.7.22)}$$

$n = 9$; $R^2 = 0.885$; $s = 0.050$, $R^2_{cv} = 0.367$

outlier: $CH_2(CH_2)_6CH_3$ ester

where B_1 is Verloop's width parameter. The presence of B_1 in the QSAR suggests that shape is important for toxicity. The *n*-octyl ester was a statistical outlier from this equation. According to Etzenhouser et al. (2001), it is possible that the ease of flexibility of the octyl chain combined with its ability to fold on itself is not adequately represented by Verloop's B_1 descriptor. The authors note that there is a lack of diversity in the biological activity of the compounds included in this equation, with the average activity of the compounds falling within the following range: 4.02 ± 0.15. It should also be noted that the compound values of

B_1 lack diversity as well, with 8 out of 10 compounds having the same value of 1.52. Hence this QSAR has not been developed according to good QSAR development practice.

Verma et al. (2003) obtained the cytotoxicities of a series of substituted thiophenols to rapidly growing mouse leukemia L1210 cells *in vitro*. The ID_{50} values (the concentration of the substituted thiophenol that induces 50% inhibition of growth in the L1210 cells after 48 h) were correlated to the Brown variation of the Hammett substituent constant σ^+:

$\log 1/ID_{50} = -0.93\ \sigma^+ + 0.86\ I_H + 3.99$ (II.7.23)
$n = 23$, $R^2 = 0.852$, $s = 0.168$, $R^2_{cv} = 0.793$
outliers: 4-NO_2-thiophenol; 4-$CH(CH_3)_2$-thiophenol; 4-OC_6H_5-thiophenol

An indicator variable (I_H) was used, having a value of 1 if the substituent was a halogen (e.g., Cl, Br) or a pseudohalogen (e.g., CF3), and 0 for all other substituents.

The QSAR suggested that radical formation may be involved in cellular toxicity. The negative coefficient of σ^+ indicates that compounds with electron-releasing substituents (such as 4-amino thiophenol and the 4-alkoxy thiophenols) are more toxic. The radical formation goes by abstraction of $H^{\cdot}$ from the –SH group of the substituted thiophenols.

The 4-isopropyl analogue was twice as active as predicted by the QSAR. According to the authors, the isopropyl group easily undergoes hydrogen abstraction and subsequent radical formation. Thus, 4-isopropyl thiophenol could afford competing sites for radical formation, which may result in enhanced cytotoxicity. The nitro analogue was nine times more active than predicted. According to the authors, this could be due to the formation of the active nitro anion radical and its subsequent reduction to a nitrosobenzene and/or phenylhydroxylamine. The 4-phenoxy analogue was overpredicted by a factor of 10. The authors suggested that its low activity could be due to its bulk and/or geometry.

Calculated homolytic ΔBDEs were also used as electronic descriptors in deriving a QSAR:

$\log 1/ID_{50} = -0.20\ \Delta BDE + 0.59\ I_H + 4.02$ (II.7.24)
$n = 23$, $R^2 = 0.794$, $s = 0.199$, $R^2_{cv} = 0.704$
outliers: 4-NO_2-thiophenol; 4-$CH(CH_3)_2$-thiophenol; 4-OC_6H_5-thiophenol

ΔBDE represents the bond energy of the SH bond in a substituted thiophenol as compared to the same bond in thiophenol itself, i.e. ΔBDE is the energy difference for the following reaction:

$$X\text{-}C_6H_4SH + C_6H_5S\bullet \rightarrow X\text{-}C6H4S\bullet + C_6H_5SH \qquad \text{(II.7.25)}$$

QSAR equation II.7.24 also suggest toxicity via radical formation. It has worse statistical parameters than the QSAR based on σ^+ (Equation II.7.23).

Geiss and Frazier (2001) investigated 10 volatile brominated and chlorinated methanes for their potential to induce oxidative stress in primary rat hepatocytes. According to Halliwell and Chirico (1993), halogenated aliphatic (HA) compounds induce oxidative stress *in vitro* and *in vivo*, mediated by bioactivation, primarily mediated by cytochrome P450 metabolism (Waller and McKinney, 1996). Reactive (radical) species are produced by reductive dehalogenation of the parent HA compounds (Waller and McKinney, 1996). In addition, oxidative metabolism of HAs may result in the formation of haloacetic acids. Haloacetic acids take part in a variety of oxidative stress-related effects, including DNA damage and lipid peroxidation (Parrish et al., 1996). Cytochrome P450s are active and inducible in liver cells, which is why Geiss and Frazier (2001) selected primary rat hepatocytes for studying the metabolism of HAs.

The thiazolyl blue (MTT) dye reduction assay, which is an indicator of mitochondrial function (Al Casey et al., 1995), was used as cytotoxicity endpoint. $EC50_{MTT}$ values denote the HA concentration causing a 50% decrease in the cellular reduction of thiazolyl blue dye, measured spectrophotometrically by the change in optical density.

Another endpoint investigated by Geiss and Frazier (2001) was obtained by performing the thiobarbituric acid reactive species (TBARS) assay as an indicator of lipid peroxidation (Yokoyama et al., 1995). The lowest effective chemical concentrations to cause a statistically significant increase in the level of TBARS in hepatocytes greater than that measured in control cells (LEC_{LP} values) were used.

Another endpoint was described by the lowest effective chemical concentrations to cause an increase in the level of reactive oxygen species (ROS) than that measured in the control cells (LEC_{ROS} values).

Descriptors used to derive QSARs were: LUMO, HOMO, logP, molar refractivity (MR), difference between LUMO and HOMO energies (E_{DIFF}), and the longest carbon–halogen bond length (DIST). In the QSAR models the negative logarithm form of the toxicity parameters were used. The statistical significance was $p < 0.05$.

The one-descriptor QSARs are presented in Table II.7.1.

Table II.7.1. One-descriptor QSAR models for each endpoint (reproduced from Geiss and Frazier, 2001)

Descriptor	–log (LEC_{LP})			–log(LEC_{ROS})			–log($EC50_{MTT}$)		
	X [a]	R^2	F	X [a]	R^2	F	X [a]	R^2	F
LUMO	-1.431	0.84	40.76	-1.140	0.72	20.91	-1.181	0.77	27.33
logP	1.192	0.76	25.49	1.039	0.79	29.58	1.007	0.74	22.28
MR	0.136	0.67	15.90	0.127	0.79	30.01	1.248	0.76	24.70
E_{DIFF}	-1.050	0.42	6.99	-0.976	0.49	7.76	-1.022	0.54	9.34
DIST	*	0.05	3.20	*	0.04	2.40	*	0.07	1.21

[a] descriptor coefficient from regression model; * no coefficient reported due to lack of statistical correlation in a one-descriptor model.

LUMO, logP, and MR were individually highly correlated ($R^2 > 0.67$) with each of the response values. LUMO was negatively correlated. A decrease in LUMO energy results in a greater affinity to accept an electron during reductive dehalogenation, involved in the production of the free radical metabolite. The QSARs with logP suggest that an increase in the lipophilicity results in an increase in toxicity. This effect may be related to uptake or transport of the chemical, to interaction with mitochnodrial membrane, or to the interaction with the cytochrome P450 enzyme. MR may reflect steric and electronic factors involved in the interaction of the HA molecule with the enzyme. It should be noted that these descriptors (LUMO, logP, MR and E_{DIFF}) were highly intercorrelated (R values > 0.70), with the

intercorrelation between logP and LUMO having R of 0.90. Therefore it is not possible to differentiate between the contributions of the different descriptors.

The best two-descriptor QSARs are presented in Table II.7.2 (only descriptors with low inter-correlation (R < 0.50) were considered for use in the same model).

Table II.7.2. Best two descriptor QSAR models for each endpoint (reproduced from Geiss and Frazier, 2001)

Endpoint	Descriptor	Coefficient	Intercept	R^2	F	$s^{2\,a}$	R^2_{cv}
–log (LEC_{LP})	LUMO	-1.60 ± 0.23	-4.37 ± 3.32	0.88	25.8	0.169	0.76
	DIST	-2.86 ± 1.78					
–log(LEC_{ROS})	MR	0.15 ± 0.02	2.79 ± 2.21	0.89	27.3	0.116	0.75
	DIST	-3.79 ± 1.56					
–log(LEC_{ROS})	LUMO	-1.28 ± 0.28	3.37 ± 3.29	0.77	11.5	0.244	0.59
	DIST	-2.40 ± 2.12					
–log($EC50_{MTT}$)	MR	0.14 ± 0.03	1.14 ± 3.41	0.81	15.4	0.194	0.58
	DIST	-2.78 ± 1.99					

[a] square of the standard error.

The authors note that two-descriptor models have only slightly better statistical parameters than the one-descriptor models. In all of the best two-parameter models the DIST descriptor appeared.

Schmidt and Heilmann (2002) developed a QSAR model for cytotoxicity to the human cervical carcinoma cell line of a set of 37 sesquiterpene lactones (STLs). The toxicity was determinited as IC_{50} values. Predictive QSAR models were obtained by using PCR (principal component regression) and PLS analysis with a descriptor set representing fractional accessible molecular surface areas (Q_frASAs). To calculate this set of descriptors the molecular surface accessible to a spherical atom probe of radius 1.4 Å was used. The parts of the surface attributable to atoms possessing charges within 14 charge intervals from –0.3 to 0.3 e (elementary charge) were calculated, resulting in 14 Q_frASA descriptors. The model obtained by PCR had the following statistical parameters (number of compounds n = 37): number of significant principal components N = 6, $R^2 = 0.831$, s = 0.206, $R^2_{cv} = 0.724$, s_{cv} = 0.266. The model obtained by PLS had the following statistical parameters (again n = 37):

number of significant latent variables (PLS components) k = 2, R^2 = 0.828, s = 0.216, F = 82.1, R^2_{cv} = 0.743, s_{cv} = 0.265. The QSARs showed that variance in STL cytotoxicity data can be explained to a high degree by electronic and surface properties.

II.8. Toxicity estimation using enzyme activity and submitochondrial particles

Cronin et al. (2002b) developed QSARs investigated the toxic and metabolic effects of 23 aliphatic alcohols on the perfused rat liver. The biological data were obtained form Strubelt et al. (1999). Each alcohol was tested at a constant concentration, 65.1 mmol/l. The activities of the following enzymes were assessed: glutamate-pyruvate-transaminase (GPT); lactate dehydrogenase (LDH); and glutamate dehydrogenaze (GLDH). The extracellular release of the enzymes was used as the toxicological endpoint. The reduction in the amount of intracellular ATP was also determined as a toxicological endpoint. Mechanistically based QSARs were developed by using descriptors for hydrophobicity (as quantified by logP) and reactivity (i.e. electrophilicity as quantified by LUMO). A parameter describing molecular branching was also used (third order path-cluster molecular connectivity index, $^3\chi_{PC}$). The following QSARs were obtained:

The activity of GPT:

$$\log GPT = 0.576 \log P - 0.193 \, LUMO - 0.494 \, {}^3\chi_{PC} + 1.19 \qquad (II.8.1)$$

n = 23, R^2 = 0.836, s = 0.183, F = 38.5, R^2_{cv} = 0.810

The activity of LDH (3-methyl-2-buten-1-ol was an outlier):

$$\log LDH = 0.561 \log P - 0.297 \, LUMO - 0.487 \, {}^3\chi_{PC} + 1.57 \qquad (II.8.2)$$

n = 22, R^2 = 0.848, s = 0.184, F = 40.1, R^2_{cv} = 0.813

The activity of GLDH (four outliers, 2-methyl-1-butanol, 2-propyn-1-ol, 1-buten-3-ol, 2-methyl-2-propen-1-ol were removed from the equation):

$$\log GLDH = 0.399 \log P - 0.037 \, LUMO - 0.384 \, {}^3\chi_{PC} + 0.579 \qquad (II.8.3)$$

n = 19, R^2 = 0.846, s = 0.132, F = 34.1, R^2_{cv} = 0.797

ATP concentration (three outliers, 1-butanol, 2-methyl-1-butanol, 3-methyl-2-buten-1-ol were removed):

$$\log 1/ATP = 0.393\ \log P - 0.362\ LUMO - 0.263\ {}^{3}\chi_{PC} + 1.48 \qquad \text{(II.8.4)}$$

$n = 20$, $R^2 = 0.857$, $s = 0.159$, $F = 38.9$, $R^2_{cv} = 0.789$

The QSARs for GPT and LDH were similar in terms of slope and intercept, possibly indicating a similar mechanism of action. The QSAR for the GLDH had a significantly different slope and intercept, with the LUMO term much less significant. This was in accordance with the founding that increased leakage of GPT and LDH is associated with cell membrane damage, whereas leakage of GLDH is associated with mitochondrial membrane damage (Strubelt et al, 1999). The QSAR for the concentration of ATP was also different from the other models.

Argese et al. (2002) investigated the toxicity of eighteen substituted anilines by means of a short-term *in vitro* assay, using submitochondrial particles (SMP) as biosensors. The assay determines the effects of toxicants that act specifically on mitochondrial respiratory functions, such as uncouplers and enzyme inhibitors, or non-specifically, by disturbing the structure and functioning of the inner mitochondrial membrane. The test with phosphorylating SMPs was carried out by determining the effect of toxicants on the process of reverse electron transfer, where exogenous NAD^+ is reduced to NADH, which strongly absorbs light at 340 nm. The toxicant concentration at which the rate of NADH production was diminished by 50% (EC_{50} values in mol/l) was used to develop QSARs.

The anilines investigated had substituents with a wide range of the electron donor/withdrawing capabilities, whereas compound hydrophobicity varied in a narrow range (logP values varied from 0.04 to 1.89). This study aimed to characterise the contribution to the toxicity due to the electronic properties of the substituents only.

The investigated anilines with comparable logP values exhibited different toxic effects. Good correlation between logP and compound toxicity was not found ($R^2 = 0.22$), probably due to the low variation in compound hydrophobicity.

The EC_{50} values were correlated with the Hammett sigma constants (σ), LUMO, HOMO, qH^+ (the largest positive partial charge on any hydrogen atom) and q^- (the largest negative partial charge on any atom). The Hammett σ constant was used to describe electronic effects. For strong electron-withdrawing substituents ($COCH_3$, CN, NO_2), the nucleophilic σ^-_p was used, which takes into account the resonance effect, present when these groups are conjugated with an electron-donating group, such as NH_2 in anilines. For the disubstituted anilines, a summation of the single substituent constants was used.

According to Urrestarazu Ramos et al. (1999) and Cramer (1993), HOMO and q^- could be associated with H-bonding acceptor capacity, which increases with increasing HOMO and decreasing q^-. The H-bonding donor capacity increases with a decrease in LUMO and an increase in qH^+.

The QSARs derived by Argese et al. (2002) are given in the Table II.8.1.

Table II.8.1. Results of regression analysis between EC_{50} values and QSAR parameters (reproduced from Argese et al., 2002)*

Descriptor	n	R^2	R^2_{adj}	R^2_{cv}	A	b	F	s
σ	18	0.81	0.80	0.76	3.20 ± 0.08	0.57 ± 0.07	69	0.25
σ (excluding outliers)	16	0.91	0.90	0.88	3.16 ± 0.05	0.53 ± 0.05	136	0.16
HOMO	18	0.71	0.69	0.65	-0.83 ± 0.71	-13.4 ± 2.1	39	0.31
HOMO (excluding outliers)	15	0.84	0.82	0.74	-0.30 ± 0.47	-11.7 ± 1.4	66	0.20
LUMO	18	0.82	0.81	0.78	4.62 ± 0.13	-11.0 ± 1.3	72	0.25
LUMO (excluding outliers)	17	0.86	0.85	0.83	4.58 ± 0.12	-11.0 ± 1.2	90	0.22
qH^+	18	0.67	0.65	0.59	-12.2 ± 2.8	44.0 ± 7.7	33	0.39
qH^+ (excluding outliers)	17	0.77	0.75	0.70	-13.1 ± 2.4	46.2 ± 6.5	50	0.28

* a and b are the regression coefficients for the equation: $\log 1/EC_{50} = a + b*X$, where X represents the molecular descriptor.

The electronic Hammett parameter σ gave the best QSAR model. The QSAR equation showed that toxicity increases by increasing the electron-withdrawing effects of the substituents. Two compounds, 3,5-dinitroaniline and 4'-aminoacetophenone, were found to be outliers and were excluded from the regression. According to Argese et al. (2002) one

explanation of the excess toxicity exhibited by these compounds could be related to the possibility of the NO_2 and $COCH_3$ groups to form further H-bonding by acting as acceptors.

Significant correlations were also found with LUMO and qH^+ (descriptors encoding H-bonding donor capacity). In both cases removing 4'-aminoacetophenone (as a statistical outlier) improved the results. The coefficients for LUMO and qH^+ in the QSARs indicate that toxicity increases with an increase in the H-bonding donor capacity.

A QSAR with HOMO was also obtained. Three compounds were statistical outliers: 3,5-dinitroaniline, 4'-aminoacetophenone and 2-aminophenol. No correlation of toxicity with q^- was found ($R^2 = 0.07$).

The following two-variable QSAR was obtained:

$$\log 1/EC_{50} = 0.68\ \sigma - 8.4\ q^- - 4.0 \qquad \text{(II.8.5)}$$
$$n = 18,\ R^2 = 0.89,\ R_A^2 = 0.87,\ s = 0.20,\ F = 59$$

The exclusion of the identified statistical outliers 3,5-dinitroaniline and 4'-aminoacetophenone resulted in the equation:

$$\log 1/EC_{50} = 0.61\ \sigma - 5.9\ q^- - 1.9 \qquad \text{(II.8.6)}$$
$$n = 16,\ R^2 = 0.95,\ R_A^2 = 0.94,\ s = 0.12,\ F = 130$$

According to the authors, the QSARs indicate the importance of electronic interactions and H-bonding donor capacity in the toxicity of anilines. These findings support a mechanism of toxic action based on H-bonding between the NH_2 group of substituted anilines and polar groups at the membrane/water interface of the submitochondrial particles, leading to a disorder in the membrane structure and disturbance of its functioning.

II.9. QSARs for toxicity to rodents and humans

Cronin et al. (2002b) developed QSAR for toxicity of 21 pyridines to mice (LD_{50} values from a single i.p. ingections, in units mol/kg) using response-surface approach. The toxicity data were taken from Richard et al. (1985); and Richard et al. (1990).

$\log 1/LD_{50} = 0.380\ \log P - 0.660\ LUMO + 1.81$ (II.9.1)

$n = 21$, $s = 0.224$, $R^2 = 0.808$, $F = 43.1$, $R^2_{cv} = 0.766$

Hansch and Kurup (2003) used the number of valence electrons (NVE), obtained by adding up the number of valence electrons of the elements in a compound (H = 1, C =1 4, O = 6, N = 5, halogens = 7, S = 6, P = 5) to develop QSARs for toxicity to different biological systems (data were taken from different literature reports). According to Hansch et al. (2003), compounds containing chemical elements of lower atomic number than chlorine have NVE values proportional to their polarisability. Therefore Hansch et al. (2003) used NVE as a descriptor in QSARs for nerve toxicity. The authors concluded that when correlation with NVE occurs, the toxic effect could occur at the synapse of the nerve axon, which appears to be a polar region, in which compound polarisability might play role in its toxic effect.

In Hansch and Kurup (2003) the following endpoints were investigated and the following QSARs were developed:

LD_{50} for female white mice. Toxicity data for barbiturates, taken from Cope and Hancock (1939):

$\log 1/C = -1.44\ \log P + 0.16\ NVE - 8.70$ (II.9.2)

$n = 11$, $R^2 = 0.924$, $s = 0.077$, $R^2_{cv} = 0.879$

outliers: 5-C_2H_5, 5-C(Me)=CHC_2H_5; 5-$CH(Me)_2$, 5-CH(Me)=CHC_2H_5, 5-Me, 5-C(Me)=CHC_4H_9

According to the authors, the negative logP term corresponds to the suggestion that the synapse is polar in character.

LD_{100} of miscellaneous drugs to humans. Toxicity data for various drugs, taken from King (1985). The data come from studies on individuals, who commit suicide or dies from an overdose of a drug, and the drug concentration in their blood is immediately determined. By taking the average values of many individuals, King (1985) established log1/C values.

$\log 1/C = 0.83\ \log P + 0.13\ NVE + 1.26$ (II.9.3)

n = 10, R^2 = 0.970, s = 0.270, R^2_{cv} = 0.922
outlier: phenobarbital

The model based on logP alone had an R^2 value of 0.876.

The following QSAR was obtained with a larger data set (data taken from King, 1987):

$$\log 1/C = 0.61 \log P + 0.017 NVE + 1.44 \quad \text{(II.9.4)}$$

n = 36, R^2 = 0.850, s = 0.438, R^2_{cv} = 0.817
outliers: morphine; theophylline; $CF_2(Cl)_2$; halothane; paraldehyde

II.10. Comparative studies on different species

Some authors developed and compared QSARs for toxicity to a number of species by applying the same methodology.

Selassie et al. (2002) derived QSARs for toxicity to a wide variety of biological systems, in which free radical formation is thought to take place as a mechanism of toxicity. QSARs for aromatic nitro compounds (chemical structure X-$C_6H_4NO_2$) were developed for the following endpoints: chromosomal aberrations (EC_{50} values) induced in human peripheral lymphocytes; inhibition of growth (IC_{50} values) of pollen tubes in tobacco plants; minimum inhibitory zone concentration (MIC) of growth of *S. cerevisiae* (yeast). A QSAR for inhibition of growth of *E. coli* by chloramphenicol analogues was also developed. QSARs for phenols were also developed for toxicity (IC_{50} values) to L1210 cells; *P. aeruginosa* (LC_{50}); uncoupling (EC_{50} values) of phosphorylation in *Ascaris* muscle mitochondria. A QSAR for toxicity (IC_{50} values) to L1210 leukemia cells was also developed for thiophenols (chemical formula X-C6H4SH), and for butylated hydroxy toluene and butylated hydroxy anisole analogues. Benzylalcohols (chemical formula X-$C_6H_4CH_2OH$) were investigated by QSAR analysis for growth inhibition of Gram-negative bacteria (*Proteus vulgaris*, *E. coli*, and *Pseudomonas*); growth inhibition of Gram-positive bacteria (*S. faecalis*, *S. albus*, *S. aureus*); growth inhibition of moulds *(A. niger, penicillum, cladosporium,* and *mucor)*. A QSAR was derived for toxicity (IC_{50} values) of methyl-(E)-3-methoxy-2-2(2-(E,E)-(4-X-phenylbuta-1,3-dienyl)phenyl)acrylates) to chloroquine-resistant *P. falciparum*.

The descriptors used were Hammett σ- and σ+ constants, BDE (homolytic bond dissociation energy), Otsu's radical parameter E_R (based on radical abstraction of •H from X-C_6H_4-$CH(CH_3)$), logP, Verloop's sterimol parameters for substituents (B1, B5, and L), Taft's steric constant E_S, and indicator variables.

II.11. Conclusions

A large number QSAR studies of acute toxicity to terrestrial organisms have been published in the literature. In the presented review mainly recent works are summarised, in which traditional statistical methods (regression analysis, partial least squares analysis) are applied to develop QSARs, rather than neural network approaches, for example.

The first part of this review included recently published QSARs for aquatic organisms. The present second part focuses on QSARs for toxicity to terrestrial organisms. The toxicities to a wide range of species have been investigated, including bacteria, fungi, insects, plants, mammalian cells and mammals, and humans. Again, different groups of chemicals were investigated and different QSAR approaches have been applied.

Some authors have shown a dependence of acute (narcotic) toxicity on the octanol/water partition coefficent (Kapur et al., 2000; Bundy et al., 2001; Sverdrup et al., 2002).

The so-called "response-surface approach" was also used to model the toxicity of compounds based on hydrophobicity and electrophilicity. The QSARs obtained were based on logP and LUMO, which parameterise biouptake and tissue distribution, and electrophilic reactivity, respectively. This approach has been applied to the yeast *Saccharomyces cerevisiae* (Wang et al., 2002b); the mould *Aspergillus nidulans* (Cronin et al., 2002b); plants *Cucumis sativus* (Wang et al., 2002a; Wang et al., 2002c); mice (Cronin et al., 2002b); toxic and metabolic effects on perfused rat liver (Cronin et al., 2002b). According to Croninn et al., (2002a) the response-surface approach is simple and mechanistically interpretable.

Another approach has been to use the Brown variation of the Hammett constant (σ^+) to model acute toxicity of aromatic compounds forming free radicals (Hansch et al., 2000; Selassie et al., 2002; Moridani et al., 2003b; Verma et al., 2003).

Descriptors accounting for the size and shape of molecules (molecular volume, Verloop's width parameters, Taft's steric parameter, charged molecular surface areas) have been used by Trohalaki et al. (2000); Etzenhouser et al. (2001); Selassie et al., (2002). Polarisability and molar refractivity have been used by Geiss and Frazier (2001); Hansch and Kurup (2003); Trohalaki et al. (2000). Interesting was the study of Trohalaki et al. (2000), who compared QSARs developed with LUMO, polarisability, molecular volume, and atomic-charge-weighted partial positive surface area, calculated by using semi-empirical and *ab initio* molecular orbital methods for geometry optimisation. It should be noted, however, that QSARs based on these parameters are often tought to be less mechanistically interpretable.

Other descriptors encoding electronic effects were also used (HOMO, Hammett constants σ, atomic charges, Argese et al., 2002).

Also, several other QSAR approaches have been explored in investigating acute toxicity. Boyd et al. (2001) applied the TLSER model descriptors, which represent cavity, dipolarity/polarisability, and H-bonding terms. Grodnitzky and Coats (2002) used geometry, topology and atomic weights assembly descriptors (GETAWAY), which encode information about the three-dimensional structures, size and shape of the molecules. QSARs for metal toxicity were also developed using ion-specific physicochemical parameters (Enache et al., 2000; Enache et al., 2003).

Development of QSARs for toxicity of mixtures of chemical compounds is an interesting research area, which is growing in the recent years. The environmental poutants are usually released as a chemical mixtures rather than single chemicals. Development of models for mixture toxicity will assist the evaluation of the chemical impact on the environment. Tichý et al. (2002) investigated dependence of toxicity on the mixture composition (molar ratios of mixture compounds).

In Table II.11.1 a summary of the applied QSAR approaches and/or structural descriptors and corresponding references is given.

Table II.11.1 Summary of the applied QSAR approaches and/or structural descriptors and corresponding references.

QSAR approach/structural descriptorts used	References
Dependence of narcotic toxicity on a parttion coefficent	Kapur et al., (2000); Bundy et al. (2001); Sverdrup et al. (2002)
Response-surface approach	Cronin et al. (2002b); Wang et al. (2002a); Wang et al. (2002b); Wang et al. (2002c)
Model of aromatic compounds forming free radicals by the Brown variation of the Hammett constant (σ^+)	Hansch et al. (2000); Selassie et al. (2002); Moridani et al. (2003b); Verma et al. (2003)
Descriptors accounting for the size and shape of molecules (molecular volume, Verloop's width parameters, Taft's steric parameter, charged molecular surface areas)	Trohalaki et al. (2000); Etzenhouser et al. (2001); Selassie et al. (2002);
HOMO, atomic charges, polarisability and molar refractivity	Trohalaki et al. (2000); Geiss and Frazier (2001); Hansch and Kurup (2003); Argese et al. (2002)
Theoretical linear solvation energy relationship (TLSER) descriptors	Boyd et al. (2001)
Geometry, topology and atomic weights assembly descriptors (GETAWAY)	Grodnitzky and Coats (2002)
QSARs for metal toxicity	Enache et al. (2000); Enache et al. (2003)
QSARs for mixture toxicity	Tichý et al. (2002)

References

Al Casey, S., Brewster, D., Viau C., Acosta, D. (1995). Effect of glutathione depletion and oxidative stress on the in vitro cytotoxicity of velnacrine maleate. *Toxicology Letters,* **76,** 257–265.

Argese, E., Bettiol, C., Fasolo, M., Zambon, A., Agnoli, F. (2002). Substituted aniline interaction with submitochondrial particles and quantitative structure-activity relationships. *Biochimica et Biophysica Acta,* **1558**, 151-160.

Boyd, E.M., Killham K., Meharg, A.A. (2001). Toxicity of mono-, di- and tri-chlorophenols to *lux* marked terrestrial bacteria, *Burkholderia* species *Rasc c*2 and *Pseudomonas fluorescens*. *Chemosphere*, **43**, 157-166.

Bundy, J.G., Morriss, A.W.J., Durham, D.G., Campbell, C.D., Paton, G.I. (2001). Development of QSARs to investigate the bacterial toxicity and biotransformation potential of aromatic heterocylic compounds. *Chemosphere*, **42**, 885-892.

Cope, A.C., and Hancock, E.M. (1939). Substituted vinyl barbituric acids III. Derivatives containing a dialkylvinyl group having five or more carbon atoms. *Journal of American Chemical Society,* **61**, 776-779.

Cramer, C.J., Famini, G.R., Lowrey, A.H. (1993). Use of Quantum Chemical Properties as Analogs for Solvatochromic Parameters in Structure-Activity Relationships. *Accounts of Chemical Research*, **26**, 599-605.

Crebelli, R., Andreoli, C., Carere, A., Conti, L., Crochi B., Cotta-Ramusino M., Benigni, R. (1995). Toxicology of halogenated aliphatic hydrocarbons: structural and molecular determinants for the disturbance of chromosome segregation and the induction of lipid peroxidation. *Chemico-Biological Interactions,* **98**, 113-129.

Consonni, V., Todeschini, T., Pavan, R. (2002). Structure/response correlations and similarity/diversity analysis by GETAWAY descriptors: 1. Theory of the novel 3D molecular descriptors. *Journal of Chemical Information and Computer Sciences,* **42**, 682-692.

Cronin, M.T.D, and Schultz, T.W. (2001). Development of quantitative structure-activity relationships for the toxicity of aromatic compounds to *T. pyriformis*: comparative assessment of the methodologies. *Chemical Research in Toxicology,* **14**, 1284-1295.

Cronin M.T.D., Aptula A.O., Duffy J.C., Netzeva T.I., Rowe P.H., Valkova I.V. Schultz T.W. (2002a). Comparative assessment of methods to develop QSARs for the prediction of the toxicity of phenols to *T. pyriformis. Chemosphere*, **49**, 1201-1221.

Cronin, M.T.D., Dearden, J.C., Duffy, J.C., Edwards, R., Manga, N., Worth, A.P., Worgan, A.D.P. (2002b). The importance of hydrophobicity and electrophilicity descriptors in mechanistically-based QSARs for toxicological endpoints. *SAR and QSAR in Environmental Research,* **13**, 167-176.

Enache, M., Palit, P., Dearden, J.C., Lepp, N.W. (2000). Correlation of physico-chemical parameters with toxicity of metal ions to plants. *Pest Management Science*, **56**, 821-824.

Enache, M., Dearden, J.C., Walker, J.D. (2003). QSAR analysis of metal ion toxicity data in sunflower callus cultures (*Helianthus annuus* "Sunspot"). *QSAR and Combinatorial Science*, **22**, 234 – 240.

Etzenhouser, B., Hansch, C., Kapur, S., Selassie, C.D. (2001). Mechanism of toxicity of esters of caffeic and dihydrocaffeic acids. *Bioorganic and Medicinal Chemistry,* **9**, 199-209.

Galati, G., Teng S., Moridani M.Y., Chan T.S., O'Brien P.J. (2000). Cancer chemoprevention and apoptosis mechanisms induced by dietary polyphenolics. *Drug Metabolism and Drug Interactions*, **17**, 311–349.

Geiss, K.T., Frazier, J.M. (2001). QSAR modeling of oxidative stress in vitro following hepatocyte exposures to halogenated methanes. *Toxicology in Vitro,* **15**, 557-563.

Grodnitzky, J.A., and Coats, J.R. (2002). QSAR evaluation of monoterpenoids' insecticidal activity. *Journal of Agricultural and Food Chemistry,* **50,** 4576-4580.

Halliwell, B., and Chirico, S. (1993). Lipid peroxidation: its mechanism, measurement, and significance. *American Journal of Clinical Nutrition,* **57,** 715S–725S

Hansch C., and Gao H. (1997). Comparative QSAR: radical reactions of benzene derivatives in chemistry and biology. *Chemical Reviews*, **97**, 2995-3059.

Hansch, C., and Kurup, A. (2003). QSAR of chemical polarisability and nerve toxicity. 2. *Journal of Chemical Information and Computer Sciences,* **43**, 1647-1651.

Hansch, C., McKarns S.C., Smith C.J., Doolittle D.J. (2000). Comparative QSAR evidence for a free-radical mechanism of phenol-induced toxicity. *Chemico-Biological Interactions*, **127**, 61–72.

Hansch, C., Steinmetz, W.E., Leo, A.J., Mekapati, S.B., Kurup, A., Hoekman, D. (2003). On the role of polarisability in chemical-biological interactions. *Journal of Chemical Information and Computer Sciences,* **43**, 120-125.

Kaiser, K.L.E. (2003a). The use of neural networks in QSARs for acute aquatic toxicological endpoints. *Journal of Molecular Structure (Theochem)*, **622**, 85–95.

Kaiser, K.L.E. (2003b). Neural networks for effect prediction in environmental and health issues using large datasets. *QSAR and Combinatorial Science*, **22**, 185-190.

Kapur S., Shusterman A., Verma R.P., Hansch C., Selassie C.D. (2000). Toxicology of benzyl alcohols: a QSAR analysis. *Chemosphere*, **41**, 1643-1649.

King, L.A. (1985). Fergusons principle and the prediction of fatal drug levels in blood. *Human Toxicol.,* **4,** 273-278.

King, L.A. (1987). The Use of Drug Solubility Data in Forensic Toxicology. Meeting of the International Association of Forensic Toxicologists. Rigi Kalbad, Switzerland. R. Bandenberger.

Liao, Y.Y., Zhou, D., Liu, Z.T., Han, S.K., Wang, L.S. (1997). QSARs of the toxicity to yeast of substituted aromatic compounds. In: Wang, L.S., Han, S.K. (Eds.), *Progress in Organic Pollutants Chemistry*. Nanjing University Press, Nanjing, pp 181–186 (in Chinese).

Miyoshi, H., Nishioka T., Fujita T. (1987). Quantitative relationship between protonophoric and uncoupling activities of substituted phenols. *Biochimica et Biophysica Acta,* **891**, 194–204.

Morao, I., and Hillier I.H. (2001). Magnetic analysis (NICS) of monoarylic cations. Linear relationship between aromaticity and Hammett constants (σ_p^+). *Tetrahedron Letters*, **42**, 4429–4431.

Moridani, M.Y., Cheon, SS., Khan, S., O'Brien, P.J., (2003a) Metabolic activation of 3-hydroxyanisole by isolated rat hepatocytes, *Chemico-Biological Interactions*, **142**, 317–333.

Moridani M.Y., Galati G., O'Brien P.J. (2002). Comparative quantitative structure toxicity relationships for flavonoids evaluated in isolated rat hepatocytes and HeLa tumor cells. *Chemico-Biological Interactions*, **139**, 251-264.

Moridani, M.Y., Siraki, A., Chevaldina, T., Scobie, H., O'Brien, P.J (2004). Quantitative structure toxicity relationships for catechols in isolated rat hepatocytes. *Chemico-Biological Interactions*, **147**, 297–307.

Moridani M.Y., Siraki A., O'Brien P.J. (2003b). Quantitative structure toxicity relationships for phenols in isolated rat hepatocytes. *Chemico-Biological Interactions*, **145**, 213-223.

O'Brien, J. (1991). Molecular mechanisms of quinone cytotoxicity. *Chemico-Biological Interactions,* **80,** 1–41.

Parrish, J., Austin, E., Stevens, D., Kinder D., Bull, R. (1996).Haloacetate-induced oxidative damage to DNA in the liver of male B6C31 mice. *Toxicology,* **110,** 103–111.

Richard, J.M., Taillandier, G., Benoit-Guyod,J.-L. (1985). A quantitative structure-activity relationship study on substituted pyridines as a contribution to the knowledge of the toxic effects of orellanine, a toxin from the mushroom *Crotinarius orellanus*. Toxicon, **23**, 815-824.

Richard, J.M., Taillandier, G., Ramos, M.N., Galdino, S.L., Da Rocha Pitta, I. (1990). QSAR analysis of series of pyridines related to the mushroom orellanine: comparison of the results obtained with physico-chemical, sterical and quantal parameters. *Pharmazie*, **45**, 771-774.

Russom, C.L., Bradbury, S.P., Broderius, S.J., Hammermeister, D.E., Drummond, R.A. (1997). Predicting modes of action from chemical structure: Acute toxicity in the fathead minnow *(Pimephales promelas)*. *Environmental Toxicology and Chemistry,* **16,** 948-967.

Schmidt, T.L., and Heilmann, J. (2002). Quantitative structure-cytotoxicity relationships of sesquiterpene lactones derived from partial charge (Q)-based fractional Accessible Surface Area Descriptors (Q_frASAs). *Quantitative Structure-Activity Relationships*, **21**, 276-287.

Schultz, T.W., Sinks, G.D., Cronin, M.T.D. (1997). Identification of mechanism of toxic action of phenols to *T. pyriformis* from molecular structure. In: Fredenslund, F.C., and Schüürmann, G. (Eds.), Quantitative Structure–Activity Relationships in Environmental Sciences, vol. VII. SETAC Press, Pensacola, FL, pp 329–342.

Selassie, C.D., Garg, R., Kapur, S., Kurup, A., Verma, R.P., Mekapati, S.B., Hansch, C. (2002). Comparative QSAR and the radical toxicity of various functional groups. *Chemical Reviews,* **102**, 2585-2605.

Strubelt, O., Deters, M., Pentz, R., Siegers, C.P. Younes, M. (1999). The toxic and metabolic effects of 23 aliphatic alcohols in the isolated perfused rat liver. *Toxicological Sciences*, **49**, 133-142.

Sverdrup, L.E., Nielsen, T., Krogh, P.H. (2002). Soil ecotoxicity of polycyclic aromatic hydrocarbons in relation to soil sorption, lipophilicity, and water solubility. *Environmental Science and Technology,* **36,** 2429-2435.

Tichý, M., Borek-Dohalský, V., Matousová, D., Rucki, M., Feltl, L., Roth, Z. (2002). Prediction of acute toxicity of chemicals in mixtures: worms Tubifex tubifex and gas/liquid distribution. *SAR and QSAR in Environmental Research***, 13**, 261-269.

Trohalaki, S., Gifford, E., Pachter, R. (2000). Improved QSARs for predictive toxicology of halogenated hydrocarbons. *Computers and Chemistry*, **24**, 421-427.

Urrestarazu Ramos, E., Vaes, W.H.J., Mayer, P., Hermens, J.L.M. (1999). *Aquatic Toxicology*, **46**, 1-10.

Verma, R.P., Kapur, S., Barberena, O., Shusterman, A., Hansch, C., Selassie, C.D. (2003). Synthesis, cytotoxicity, and QSAR analysis of X-thiophenols in rapidly dividing cells. *Chemical Research in Toxicology,* 16, 276-284.

Waller, C., Evans, M., McKinney, J. (1996). Modeling the cytochrome P450-mediated metabolism of chlorinated volatile organic compounds. *Drug Metabolism and Disposition,* **24,** 203–210.

Wang, X., Sun, C., Wang, Y., Wang, L. (2002a). Quantitative structure-activity relationships for the inhibition toxicity to root elongation of *Cucumis sativus* of selected phenols and interspecies correlation with *T. pyriformis*. *Chemosphere***, 46**, 153-161

Wang, X., Yin, C., Wang, L., (2002b). Structure-activity relationships and response-surface analysis of nitroaromatics toxicity to the yeast (*Saccharomyces cerevisiae*). *Chemosphere,* **46**, 1045-1051.

Wang, X., Yu, J., Wang, Y., Wang, L. (2002c). Mechanism-based quantitative structure-activity relationships for the inhibition of substituted phenols on germination rate of *Cucumis sativus*. *Chemosphere,* **46**, 241-250.

Wilson, L.Y., and Famini, G.R. (1991). Using theoretical descriptors in quantitative structure-activity relationships: Some toxicological indices. *Journal of Medicinal Chemistry*, **34**, 1668-1674.

Yokoyama, H., Horie, T., Awazu, S. (1995). Oxidative stress in isolated rat hepatocytes during naproxen metabolism. *Biochemical Pharmacology,* **49,** 991–996.