# Structural Alerts—A New Classification Model to Discriminate Excess Toxicity from Narcotic Effect Levels of Organic Compounds in the Acute Daphnid Assay

Peter C. von der Ohe, Ralph Kühne, Ralf-Uwe Ebert, Rolf Altenburger, Matthias Liess, and Gerrit Schüürmann*

*Department of Chemical Ecotoxicology, UFZ Centre for Environmental Research, Permoserstrasse 15, D-04318 Leipzig, Germany*

*Received July 27, 2004*

Quantitative and qualitative structure−activity relationships (QSARs) have a great potential to support the risk assessment of chemicals, provided there are tools available that allow evaluation of the suitability of QSARs for the compounds of interest. In this context, a pragmatic approach is to discriminate excess toxicity from narcotic effect levels, because the latter can be estimated from QSARs and thus have a low priority for experimental testing. To develop a respective scheme for the acute daphnid toxicity as one of the primary ecotoxicological endpoints, 1067 acute toxicity data entries for 380 chemicals involving the daphnid species *Daphnia magna* were taken from the on-line literature, and quality checks such as water solubility were employed to eliminate apparently odd data entries. For 36 known narcotics with $LC_{50}$ values referring to *D. magna*, a reference baseline QSAR is derived. Compounds with $LC_{50}$ values above a certain threshold defined relative to their predicted baseline toxicity are classified as exerting excess toxicity. Three simple discrimination schemes are presented that enable the identification of excess toxicity from structural alerts based on the presence or absence of certain heteroatoms and their chemical functionality. Moreover, a two-step classification approach is introduced that enables a prioritization of organic compounds with respect to their need for experimental testing. The discussion includes reaction mechanisms that may explain the association of structural alerts with excess toxicity, a comparison with predictions derived from mode of action-based classification schemes, and a statistical analysis of the discrimination performance in terms of detailed contingency table statistics.

## Introduction

According to the upcoming European chemical policy (*1*), it is likely that until 2012, approximately 30 000 chemical substances with production volumes of more than one ton per year will require data for their toxicological and ecotoxicological evaluation. At present, however, we are far from having a complete picture with respect to the chemicals in use (*2*). It was estimated recently that for less than 1% of the chemicals in commerce in the United States, experimental results for their aquatic toxicity are available (*3*).

To enable an efficient use of the available testing resources, a promising way forward would be to include low cost screening methods that allow the identification of those chemicals where according to present knowledge, substantial toxicity for the endpoints of interest is to be expected. In this context, theoretical methods such as quantitative or qualitative structure−activity relationships (QSARs) may form a particularly efficient component besides experimental in vitro methods and high-throughput techniques. Recent QSAR research includes a variety of endpoints such as skin sensitization (*4*), androgen receptor binding (*5*), aryl-hydrocarbon receptor binding (*6*), mixture toxicity prediction (*7*, *8*), acute aquatic toxicity (*9*), and rat chronic toxicity (*10*). Besides models to predict effect levels, there are also classification schemes that relate modes of toxic action to structural features (*11*, *12*) or to property profiles calculated from molecular structure (*13*−*15*).

For the regulation of industrial chemicals, the American and European legislations still differ significantly with regard to the use of QSARs. While the U.S. Environmental Protection Agency (EPA) as well as the Interagency Testing Committee of the Toxic Substances Control Act use QSARs to estimate the toxicity of existing and new chemicals, the current European regulation is based on experimental data, employing QSARs only for specific and limited purposes such as data evaluation or the provision of additional evidence for conducting long-term tests (*16*, *17*). However, it is likely that under the new REACH (Registration, Evaluation, and Authorization of Chemicals) system to be introduced in Europe (*1*), more efforts will be given to include nonanimal methods such as in vitro techniques and QSAR models in the regulatory decision making process, provided that these have been validated appropriately.

Among the few QSAR packages with a history in U.S. regulatory programs, ECOSAR (*18*) and ASTER (*12*, *19*) are mainly based on linear regression equations relating toxicity to the octanol/water partition coefficient, $K_{ow}$ (*3*). For a given chemical, the selection of the appropriate QSAR proceeds with ECOSAR by chemical class, which

* To whom correspondence should be addressed. Tel: +49-341-235-2309. Fax: +49-341-235-2401. E-mail: gerrit.schuurmann@ufz.de.

10.1021/tx0497954 CCC: $30.25 
Published on Web 02/22/2005

may be problematic for multifunctional compounds as well as for compounds where the log $K_{ow}$ is not a good descriptor for the effect level of interest (*3*). ASTER first allocates the chemical to a likely mode of toxic action through analysis of substructural features. If narcotic type toxicity or oxidative uncoupling was predicted as a mode of action (MOA), the associated effect level is then predicted by accordingly selected QSARs, again employing log $K_{ow}$ as the only molecular descriptor.

Among the commercial packages that are used by the U.S. EPA is TOPKAT (*20*), which employs both classification schemes and multiple linear regression relationships to predict toxicity from molecular structure. According to a recent independent evaluation of the prediction power of TOPKAT for chronic oral rat toxicity, however, TOPKAT could not yield predictions for 108 (34%) out of 315 organic compounds (*10*). Among the remaining 208 chemicals, the predicted lowest observed adverse effect level (LOAEL) was within a factor of 2, 5, and 10 of the experimental value for 43, 66, and 80% of the test set, respectively (*10*). For a different test set of 313 compounds, TOPKAT could not yield LOAEL predictions for 84 chemicals (27%), and for the remainder, the prediction error was similar to the one achieved with the other test set (*10*).

A recent comparative analysis of the performance of six QSAR packages to estimate the acute fish toxicity showed some variation between the different methods, also with regard to the application range (*3*). Although the test set of 130 compounds consisted mainly of simple organics with low functionality, with more than 60% being nonpolar narcotics, the overall performances of the six QSAR packages were only low to moderate: The mean absolute prediction error of the $LC_{50}$ ranged from 0.9 to 1.3 log units, and the squared correlation coefficient ($r^2$) ranged from 0.154 to 0.279. The proportion of chemicals with prediction errors above a factor of 10 was between 32 and 45%, and prediction errors above factors of 100 and 1000 were observed for test set proportions between 13 and 22% and between 5 and 9%, respectively (*3*).

Here, the performance of TOPKAT was particularly interesting: Only 37% of the compounds were in the so-called optimum prediction space, and for these compounds, the model performance was significantly better than when considering all formally applicable chemicals. With regard to the latter, there were also substantial differences observed; only two of the six packages (including ECOSAR) could handle all compounds of the test set, while the ASTER methodology could be applied to only 92 of the 130 compounds.

With regard to MOA classification schemes in the area of aquatic toxicology (*11*, *12*, *22*), we are not aware of a corresponding comparative study about the statistical performance with a test set of compounds that had not been included in the training sets. However, some experience has been reported with individual packages. In a recent evaluation of ASTER, MOA-based QSAR selection and subsequent application lead to an $r^2$ of 0.90 for a validation set of 97 chemicals (*21*), which is somewhat surprising when considering the much inferior statistics reported in the above-mentioned comparative study (*3*). Note further that with a descriptor-based classification of chemicals according to prevalent MOAs in fish, systematic errors were reported for epoxides, fluorinated hydrocarbons, thiols, and α,β-unsaturated carbonyl compounds (*14*).

According to REACH (*1*), the acute daphnid toxicity will become the primary trigger for the aquatic toxicity evaluation in the European legislation. So far, there is no independent analysis available about the performance of QSAR models for this endpoint, and there is also no classification scheme available that allocates chemical structures to MOAs prevalent in the daphnid bioassay. In the present investigation, three classification models (CMs) are introduced that allow discrimination between compounds exerting narcotic effect levels and those exerting excess toxicity in the acute 48 h daphnid test, employing only simple structural features as molecular descriptors. To this end, 1067 $LC_{50}$ (lethal concentration 50%) values have been collected from the on-line literature (*23*) and critically evaluated, resulting in a final set of 300 organic compounds.

For the a priori identification of the narcotic effect range, the excess toxicity ($T_e$) concept (*24*, *25*) is employed in connection with a reference baseline QSAR developed for a subset of 36 compounds that had already been classified as narcotic (*26*). Considering factors of 10 to account for both the data uncertainty and the difference between nonpolar and polar narcotics when referring to the octanol/water partition coefficient ($K_{ow}$) as a hydrophobicity parameter (*11*), $LC_{50}$ values within a factor of 100 from baseline toxicity are classified as belonging to the narcotic effect range, and the remainder indicate significant excess toxicity. For the three CMs, the concordance (overall agreement between experimental and predicted toxicity categories) ranges from 0.63 to 0.92, with predictivities (for each category, the ratio of the numbers of correctly predicted over predicted compounds) up to 0.98.

Moreover, three existing CMs (*11*, *12*, *22*) that make use of substructural features to predict MOAs of organic compounds in the acute fish toxicity test are evaluated for their suitability to separate excess toxicity from the narcotic effect range. The results show that more than 60% of the chemicals with reactive or specific MOAs in the fish test have acute daphnid toxicity values in the narcotic effect range. It suggests that for setting test priorities in tiered chemical hazard evaluation schemes, $T_e$-based CMs are superior to MOA-based CMs.

## Materials and Methods

From the U.S. EPA database AQUIRE (*23*), 1067 acute toxicity values (48 h $LC_{50}$, lethal concentration 50%) for the cladoceran *D. magna* were collected for a total of 380 compounds. The query was conducted for the endpoint mortality as recorded in AQUIRE. Note, however, that some studies use mortality ($LC_{50}$) and immobilization ($EC_{50}$, effective concentration 50%) as identical endpoints in the context of daphnid toxicity, as is, for example, reported in the toxicity analysis of parathion (*27*) that is also included in the presently selected AQUIRE data set.

When multiple test values were found for one substance, these values were checked for consistency. If values differed by more than a factor of 30 from the closest one in a group of at least three other references, the aberrant value was discarded so as to remove outliers from the data set. Of all the remaining values for a given substance, the arithmetic mean was taken as the valid experimental value.

**Training Set.** From the initial set of 1067 $LC_{50}$ data, 77 values were excluded as outliers as described above, which led to a set of 349 organic chemicals with at least one $LC_{50}$ value per substance. Subsequently, 49 chemicals were excluded because their $LC_{50}$ values exceeded the predicted water solubil-

ity (*28*) or because they contained metal atoms or were inorganic, leading to the final set of 300 organic compounds that cover a log $K_{ow}$ (octanol/water partition coefficient) range from −2 to 8. All log $K_{ow}$ were predicted using the KOWWIN software (*28*).

As regards the chemical domain, the data set includes hydrocarbons, aliphatic alcohols, phenols, ethers, and esters; anilines, amines, nitriles, nitroaromatics, amides, and carbamates; urea and thiourea derivatives; isothiocyanates; thioles; phosphorothionate and phosphate esters; and halogenated derivatives.

For 36 compounds of the data set, the prevalent mode of toxic action had been reported as narcosis (*26*). Consequently, these 36 compounds were used to derive a baseline QSAR for the acute toxicity toward *D. magna*. The remaining 264 compounds formed the training set for the derivation of CMs to discriminate between narcotic effect levels and excess toxicity.

**Excess Toxicity.** The ratio between the QSAR-predicted baseline toxicity and the experimental toxicity was evaluated in terms of the excess toxicity $T_e$ (*24*, *25*):

$$T_e = \frac{LC_{50}\ (\text{baseline})}{LC_{50}\ (\text{exp})} \quad (1)$$

Polar narcotic compounds are known to be on the average 10 times more toxic than nonpolar narcotics (*11*), which corresponds to a log $T_e$ of one. Moreover, data uncertainty is estimated to cover 1 order of magnitude, keeping in mind that the $LC_{50}$ values collected in AQUIRE come from different laboratories. Therefore, it was decided that the narcotic toxicity range should comprise compounds whose toxicity values are within a factor of 100 of the baseline toxicity. A log $T_e$ greater than two, corresponding to toxicity 100 times above baseline toxicity, was defined to indicate excess toxicity. In this way, excess toxicity is most likely associated with a reactive or specific MOA, while it is not unlikely to find some nonnarcotic chemicals with $T_e$ values below 100 and even below 10, which is indeed the case and will be discussed in more detail below. Following this approach, all 264 compounds of the training set were classified as narcotic level or excess toxic, which served as experimental categories for the subsequent classification modeling.

**Derivation of CMs.** Visual inspection of the chemical structures of the compounds exerting excess toxicity, followed by cross-checking the structural patterns of narcotic effect level compounds, led to the empirical identification of structural alerts (SAs) as indicators of excess toxicity in terms of the presence or absence of certain heteroatoms and chemical functionalities. This approach was applied on two levels of complexity: First, only the presence or absence of certain heteroatoms was analyzed in comparison with log $T_e$. Second, substructural units encoding specific functional groups were identified as indicators of excess toxicity. To this end, all compounds were allocated to major chemical classes, keeping in mind that a unique classification is not feasible for multifunctional compounds. The distribution of chemical classes and subclasses with regard to the observed $T_e$ values resulted in initial hypotheses about functional groups that might discriminate, as good as possible, between narcosis level and excess toxicity. Detailed visual analyses of the chemical structures in relation to their log $T_e$ values finally led to the derivation of three CMs, which are comparatively analyzed for their overall prediction performance (concordance) as well as for their sensitivity (recognition power) and predictivity (prediction power) using contingency table statistics (see below).

For comparative purposes, the following three existing MOA-based classification schemes (CMs) were included in the analysis: a 4-MOA CM (*11*), a 7-MOA CM (*12*), and a 2-MOA CM (*22*). For convenience, all classification rules have been implemented in our ChemProp software system (*29*), allowing automated classification runs according to all six schemes.

**Contingency Table Statistics.** The performance of the different classification schemes was evaluated in terms of detailed contingency table statistics as outlined elsewhere (*15*), considering the two categories narcotic effect range and excess toxicity. Here, the concordance is calculated as the sum of all compounds that are correctly predicted to belong to one of the two categories ($n_{cpred}$), divided by the total number of compounds ($N$). It represents the proportion of compounds where the predicted and the experimental classification agree:

$$\text{concordance} = \frac{1}{N}\sum_{i=1}^{2} n_{cpred}(i) \quad (2)$$

To evaluate the category specific performances, the following two parameters have been employed (*15*). For a given category, the sensitivity

$$\text{sensitivity} = \frac{n_{cpred}}{n_{exp}} \quad (3)$$

is calculated as the number of compounds correctly predicted to belong to this category ($n_{cpred}$), divided by the number of compounds that actually belong to this category ($n_{exp}$). As such, the sensitivity may also be referred to as the recognition power of the CM (*15*). Correspondingly, the predictivity

$$\text{predictivity} = \frac{n_{cpred}}{n_{pred}} \quad (4)$$

is defined as the number of compounds correctly predicted to belong to one category ($n_{cpred}$), divided by the total number of compounds predicted to belong to this class ($n_{pred}$).

**QSAR Modeling.** For subsets of narcotic effect level compounds, linear regression of log $LC_{50}$ on calculated log $K_{ow}$ was performed and compared with baseline toxicity QSARs. The statistical tests and graphs have been carried out with SPSS (*30*) and the SigmaPlot graphics package (*31*).

## Results

**Baseline Toxicity.** For the 36 compounds reported as narcotics (*26*), linear regression of log $LC_{50}$ on log $K_{ow}$ yields

$$\log LC_{50}\ (\text{mol/L}) = -0.857\ (\pm 0.049)\ \log K_{ow} - 1.281\ (\pm 0.125) \quad (5)$$

where $n = 36$, $r^2 = 0.90$, SE = 0.44, and $F_{1,34} = 311$. Equation 5 thus represents a baseline QSAR for the acute toxicity toward *D. magna* and was used to evaluate the excess toxicity in terms of $T_e$ values according to eq 1.

Interestingly, the subset of 33 compounds classified as polar narcotics according to the 4-MOA scheme trained on acute fish toxicity (*11*) yields

$$\log LC_{50}\ (\text{mol/L}) = -0.802\ (\pm 0.085)\ \log K_{ow} - 2.206\ (\pm 0.244) \quad (6)$$

for the acute daphnid toxicity ($n = 33$, $r^2 = 0.74$, SE = 0.45, and $F_{1,31} = 89.9$), with a quite similar slope (−0.802 vs −0.857) and an intercept one log unit lower (−2.206 vs −1.281) as compared to eq 5 that represents daphnid baseline narcosis. For the derivation of eq 6, all anilines were left out, because there is evidence that anilines show substantial excess toxicity toward daphnids despite their polar narcotic effect level in fish (*32*).

A similar result is achieved when applying the 7-MOA scheme that was also derived from acute fish toxicity data (*12*), and the respective data distributions are shown in

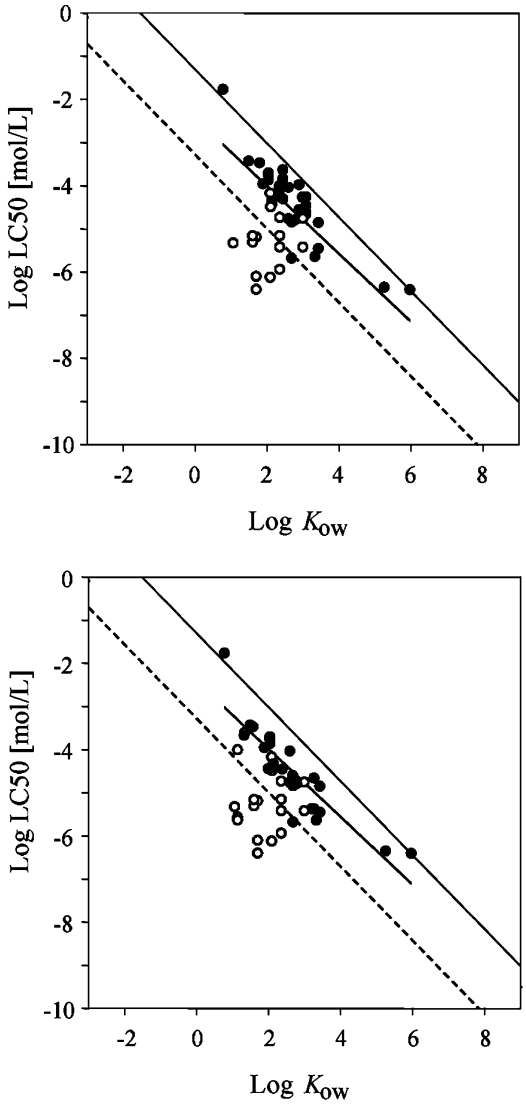



**Figure 1.** Log $LC_{50}$ (mol/L) vs log $K_{ow}$ for compounds classified as polar narcotics except anilines (filled circles) according to the 4-MOA CM (33 compounds, top plot) and 7-MOA CM (25 compounds, bottom plot), respectively, together with their associated linear regression lines. The anilines classified as polar narcotics according to these two schemes are shown as open circles (top, 15 compounds; bottom, 20 compounds). Moreover, the lines representing baseline toxicity (eq 5, solid line) and the threshold of excess toxicity (dashed line) are included for comparison.

the top and bottom part of Figure 1, respectively. It follows that in contrast to findings with polar narcosis toward fish (*33*), a convergence of the log $LC_{50}$ vs log $K_{ow}$ relationships for baseline and polar narcosis with increasing log $K_{ow}$ is not observed for the acute daphnid toxicity according to the presently available AQUIRE data. Note also that the log $K_{ow}$-based regression relationships for nonpolar and polar narcosis toward the fish species *Pimephales promelas* (slope, −0.85 vs −0.73; intercept, −1.39 vs −2.16) as recommended by the European Union (*17*) intersect each other only at the high log $K_{ow}$ of 6.42.

**Discrimination between Narcotic Effect Levels and Excess Toxicity.** For the training set of 264 compounds, baseline daphnid toxicity was predicted through eq 5, and their actual experimental $LC_{50}$ values were classified through eq 1 as narcotic effect level or excess toxic. Here, log $T_e$ > 2 indicated excess toxicity, while $LC_{50}$ values with log $T_e \leq 2$ were classified as belonging to the narcotic effect range (see Materials and Methods). In this way, 78 of the 264 compounds were classified as exerting excess toxicity, and the remaining 186 compounds were classified as yielding narcotic effect levels. The associated $LC_{50}$ data cover more than 9 orders of magnitude, ranging from $8.04 \times 10^{-10}$ mol/L (deltamethrin) to 0.34 mol/L (triethylene glycol). Note further that for the subset of 78 compounds exerting excess toxicity, the $LC_{50}$ variation is 7 orders of magnitude.

In Table 1, all 300 compounds are listed together with their mean experimental daphnid toxicities in terms of log $LC_{50}$, calculated log $K_{ow}$ values, and the logarithmic excess toxicities (log $T_e$). In addition, the toxicity category (narcotic effect range vs excess toxicity) predicted by CM1, CM2, and CM3 as well as by the three literature schemes 4-MOA CM (*11*), 7-MOA CM (*12*), and 2-MOA CM (*22*) are summarized and will be explained in more detail below.

**CM1.** According to Könemann (*34*), baseline toxicity is expected for simple organic compounds such as aliphatic and aromatic hydrocarbons, aliphatic alcohols, aldehydes and ketones, and halogenated derivatives. A generalization of this approach leads to the following simple condition for a chemical to exert a narcotic level toxicity: The compound may contain any combination of C, H, O, and halogen, excluding α,β-unsaturated carbonyl groups as electrophilic functionalities. As will be shown below, this simple approach is particularly powerful in predicting chemicals as exerting narcotic effect levels.

According to CM1, all compounds containing heteroatoms other than O and halogen as well as all α,β-unsaturated carbonyl derivatives would be classified as excess toxic in the 48 h *D. magna* test. Interestingly, this prediction of excess toxicity is much less reliable as compared to the complementary prediction of narcotic effect levels, because CM1 tends to predict too many compounds as excess toxic. Despite this disadvantage, the predictive mode of CM1 for narcotic effect levels appears to be highly reliable, suggesting that in a tiered approach, CM1 could serve as a first-tier component for the predictive identification of compounds that are highly probable to exert only narcotic effect levels (see below). A more detailed evaluation of the statistical performance of CM1 is given below.

**CM2 and CM3—SAs.** CM2 and CM3 are based on SAs as indicators of excess toxicity. CM2 is based on nine SAs (SA1−SA9) each of which is defined through a single substructural unit (see below), forming the primary rules to identify excess toxic chemicals. The statistical selection criterion was that for each correspondingly defined SA, at least three compounds with $T_e$ values above 100 are present in the data set.

Inspection of the remaining classification errors led to the identification of a further eight substructural units, each of which belonged to one or two compounds exerting excess toxicity (of course, without counterexamples). Because these structural features form variants of the primary rules of CM2, they were considered as additional tentative alerts and led to the definition of a correspondingly augmented classification scheme 3, CM3, consisting of nine partially extended SAs (SA1−SA2, SA3*, SA4, SA5*−SA7*, and SA8−SA9; see below).

In the respective list of SAs given below, open valences indicate attachment to any of the elements C, H, N, O, P, S, F, Cl, Br, or I. For CM2, only the first substructural

**Table 1. Compounds with 48 h Daphnia Toxicity in Terms of Log $LC_{50}$, Log $K_{ow}$, Log $T_e$, and the Prediction Results of Six CMs[a]**

| no. | CAS | name | log $LC_{50}$ (mol/L) | log $K_{ow}$ | log $T_e$ | CM1 | CM2 | CM3 | 4-MOA CM | 7-MOA CM | 2-MOA CM |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | training set | | | | | | | |
| 1 | 50293 | DDT | −7.89 | 6.79 | 0.79 | 0 | 0 | 0 | 4 | 7 | 0 |
| 2 | 51285 | 2,4-dinitrophenol | −4.62 | 1.73 | 1.86 | 1 | 0 | 0 | 3 | 4 | 0 |
| 3 | 52686 | trichlorofon | −6.31 | −0.28 | 5.27 | 1 | 0 | 3.2 | NA | 5/6 | 0 |
| 4 | 55389 | fenthion | −6.79 | 4.08 | 2.01 | 1 | 3.1 | 3.1 | 4 | 6 | 0 |
| 5 | 55630 | nitroglycerine | −3.85 | 1.51 | 1.28 | 1 | 0 | 0 | NA | 1 | 0 |
| 6 | 56382 | parathion | −8.17 | 3.73 | 3.70 | 1 | 3.1 | 3.1 | 4 | 6 | 0 |
| 7 | 58140 | pyrimethamine | −4.63 | 2.41 | 1.29 | 1 | 0 | 0 | NA | 1 | 0 |
| 8 | 58899 | lindane | −5.39 | 4.26 | 0.46 | 0 | 0 | 0 | 1 | 7 | 0 |
| 9 | 58902 | 2,3,4,6-tetrachlorophenol | −6.12 | 4.09 | 1.34 | 0 | 0 | 0 | NA | 4 | 0 |
| 10 | 59063 | ethopabate | −3.07 | 1.90 | 0.17 | 1 | 0 | 0 | 3 | 3 | 0 |
| 11 | 59507 | 4-chloro-3-methylphenol | −4.85 | 2.70 | 1.26 | 0 | 0 | 0 | 2 | 2 | 0 |
| 12 | 60515 | dimethoate | −4.94 | 0.28 | 3.42 | 1 | 3.1 | 3.1 | 4 | 6 | 0 |
| 13 | 60571 | dieldrin | −6.28 | 5.45 | 0.33 | 0 | 0 | 0 | 3 | 5/7 | 1 |
| 14 | 62533 | aniline | −5.33 | 1.08 | 3.13 | 1 | 8.1 | 8.1 | 2 | 2 | 0 |
| 15 | 62555 | thioacetamide | −3.64 | −0.83 | 3.07 | 1 | 0 | 0 | NA | 1 | 0 |
| 16 | 62566 | thiourea | −3.84 | −1.31 | 3.68 | 1 | 7.1 | 7.1 | NA | 1 | 0 |
| 17 | 62737 | dichlorvos | −9.10 | 0.60 | 7.30 | 1 | 2.1 | 2.1 | NA | 6 | 1 |
| 18 | 63252 | carbaryl | −7.33 | 2.35 | 4.04 | 1 | 6.1 | 6.1 | NA | 1 | 0 |
| 19 | 68122 | *N*,*N*-dimethylformamide | −0.70 | −0.93 | 0.22 | 1 | 0 | 0 | NA | 5 | 0 |
| 20 | 72208 | endrin | −6.38 | 5.45 | 0.43 | 0 | 0 | 0 | 3 | 5/7 | 1 |
| 21 | 74839 | methyl bromide | −4.63 | 1.18 | 2.34 | 0 | 0 | 0 | 1 | 1 | 0 |
| 22 | 75058 | acetonitrile | −1.06 | −0.15 | −0.10 | 1 | 0 | 0 | NA | 1 | 0 |
| 23 | 75070 | acetaldehyde | −0.55 | −0.17 | −0.59 | 0 | 0 | 0 | 3 | 5 | 1 |
| 24 | 75081 | ethyl mercaptan | −5.56 | 1.27 | 3.19 | 1 | 4.1 | 4.1 | NA | 1 | 1 |
| 25 | 75150 | carbon disulfide | −4.56 | 1.94 | 1.62 | 1 | 0 | 0 | NA | 1 | 0 |
| 26 | 75218 | ethylene oxide | −2.32 | −0.05 | 1.08 | 0 | 0 | 0 | 3 | 5 | 1 |
| 27 | 75252 | bromoform | −3.74 | 1.79 | 0.92 | 0 | 0 | 0 | 1 | 1 | 0 |
| 28 | 75354 | 1,1-dichloroethene | −3.28 | 2.12 | 0.18 | 0 | 0 | 0 | 1 | 1 | 0 |
| 29 | 77474 | hexachlorocyclopentadiene | −6.72 | 4.63 | 1.47 | 0 | 0 | 0 | 1/3 | 7 | 0 |
| 30 | 78591 | isophorone | −3.06 | 2.62 | −0.47 | 1 | 1.1 | 1.1 | 3 | 5 | 1 |
| 31 | 78999 | 1,1-dichloropropane | −3.57 | 2.25 | 0.36 | 0 | 0 | 0 | 1 | 1 | 0 |
| 32 | 79061 | acrylamide | −2.65 | −0.81 | 2.06 | 1 | 1.1 | 1.1 | NA | 5 | 1 |
| 33 | 79094 | propionic acid | −3.17 | 0.58 | 1.39 | 0 | 0 | 0 | NA | 1 | 0 |
| 34 | 83410 | 1,2-dimethyl-3-nitrobenzene | −4.56 | 2.91 | 0.78 | 1 | 0 | 0 | 2 | 1 | 0 |
| 35 | 83421 | 2-chloro-6-nitrotoluene | −4.61 | 3.00 | 0.76 | 1 | 0 | 0 | 2 | 1 | 0 |
| 36 | 84662 | diethyl phthalate | −3.61 | 2.65 | 0.06 | 0 | 0 | 0 | 3 | 3 | 0 |
| 37 | 84742 | dibutyl phthalate | −4.88 | 4.61 | −0.36 | 0 | 0 | 0 | 3 | 3 | 0 |
| 38 | 85018 | phenanthrene | −5.36 | 4.35 | 0.35 | 0 | 0 | 0 | 1 | 1 | 0 |
| 39 | 85687 | butyl benzyl phthalate | −5.19 | 4.84 | −0.24 | 0 | 0 | 0 | 3 | 3 | 0 |
| 40 | 86306 | *N*-nitrosodiphenylamine | −4.40 | 3.16 | 0.42 | 1 | 0 | 0 | 3 | 5 | 0 |
| 41 | 86748 | carbazole | −4.70 | 3.23 | 0.65 | 1 | 0 | 0 | NA | 1 | 0 |
| 42 | 87865 | pentachlorophenol | −5.64 | 4.74 | 0.30 | 0 | 0 | 0 | NA | 4 | 0 |
| 43 | 88722 | 1-methyl-2-nitrobenzene | −4.14 | 2.36 | 0.84 | 1 | 0 | 0 | 2 | 1 | 0 |
| 44 | 88733 | 1-chloro-2-nitrobenzene | −3.64 | 2.46 | 0.25 | 1 | 0 | 0 | 2 | 1 | 1 |
| 45 | 88857 | 2-(1-methylpropyl)-4,6-dinitrophenol | −6.00 | 3.67 | 1.57 | 1 | 0 | 0 | 3 | 4 | 0 |
| 46 | 88891 | 2,4,6-trinitrophenol | −3.43 | 1.54 | 0.83 | 1 | 0 | 0 | 3 | 4 | 0 |
| 47 | 89598 | 4-chloro-2-nitrotoluene | −4.27 | 3.00 | 0.41 | 1 | 0 | 0 | 2 | 1 | 0 |
| 48 | 89612 | 1,4-dichloro-2-nitrobenzene | −4.26 | 3.10 | 0.32 | 1 | 0 | 0 | 2 | 1 | 1 |
| 49 | 90028 | salicylaldehyde | −4.45 | 2.01 | 1.44 | 0 | 0 | 0 | 3 | 2 | 0 |
| 50 | 90040 | o-aminoanisole | −4.01 | 1.16 | 1.74 | 1 | 0 | 0 | NA | 2 | 0 |
| 51 | 90051 | 2-methoxyphenol | −3.68 | 1.34 | 1.25 | 0 | 0 | 0 | NA | 2 | 0 |
| 52 | 90131 | 1-chloronaphthalene | −5.01 | 3.81 | 0.46 | 0 | 0 | 0 | NA | 1 | 0 |
| 53 | 90437 | 2-phenylphenol | −5.38 | 3.28 | 1.29 | 0 | 0 | 0 | NA | 2 | 0 |
| 54 | 91225 | quinoline | −3.53 | 2.14 | 0.41 | 1 | 0 | 0 | NA | 5 | 0 |
| 55 | 91645 | coumarin | −4.03 | 1.51 | 1.46 | 0 | 0 | 0 | NA | 5 | 0 |
| 56 | 91941 | 3,3′-dichlorobenzidine | −5.38 | 3.21 | 1.35 | 1 | 0 | 0 | NA | 2 | 0 |
| 57 | 92524 | biphenyl | −4.66 | 3.76 | 0.15 | 0 | 0 | 0 | 1 | 1 | 0 |
| 58 | 92693 | 4-phenylphenol | −4.67 | 3.28 | 0.58 | 0 | 0 | 0 | NA | 2 | 0 |
| 59 | 94757 | 2,4-dichlorophenoxyacetic acid | −3.17 | 2.62 | −0.35 | 0 | 0 | 0 | NA | 5 | 0 |
| 60 | 95158 | benzo[*b*]thiophene | −3.36 | 2.99 | −0.49 | 1 | 0 | 0 | NA | 1 | 0 |
| 61 | 95487 | *o*-cresol | −3.87 | 2.06 | 0.82 | 0 | 0 | 0 | 2 | 2 | 0 |
| 62 | 95512 | 2-chloroaniline | −5.19 | 1.72 | 2.44 | 1 | 0 | 0 | 2 | 2 | 0 |
| 63 | 95534 | ortho-toluidine | −5.31 | 1.62 | 2.64 | 1 | 0 | 0 | 2 | 2 | 0 |
| 64 | 95578 | 2-chlorophenol | −4.34 | 2.16 | 1.21 | 0 | 0 | 0 | 2 | 2 | 0 |
| 65 | 95761 | 3,4-dichlorobenzenamine | −5.95 | 2.37 | 2.64 | 1 | 8.1 | 8.1 | 2 | 2 | 0 |
| 66 | 95829 | 2,5-dichloroaniline | −4.74 | 2.37 | 1.43 | 1 | 0 | 0 | 2 | 2 | 0 |
| 67 | 95954 | 2,4,5-trichlorophenol | −4.86 | 3.45 | 0.63 | 0 | 0 | 0 | 2 | 2 | 0 |
| 68 | 96093 | 1,2-epoxyethylbenzene | −4.02 | 1.59 | 1.37 | 0 | 0 | 0 | 3 | 5 | 1 |
| 69 | 96457 | ethylene thiourea | −3.59 | −0.49 | 2.73 | 1 | 0 | 7.3 | NA | 1 | 0 |
| 70 | 97007 | 1-chloro-2,4-dinitrobenzene | −5.40 | 2.27 | 2.18 | 1 | 0 | 0 | 3 | 5 | 1 |
| 71 | 97745 | bis(dimethylthiocarbamyl)sulfide | −4.86 | 0.75 | 2.93 | 1 | 0 | 6.3 | NA | 1 | 0 |

**Table 1 (Continued)**

| no. | CAS | name | log $LC_{50}$ (mol/L) | log $K_{ow}$ | log $T_e$ | CM1 | CM2 | CM3 | 4-MOA CM | 7-MOA CM | 2-MOA CM |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | training set | | | | | | | |
| 72 | 97778 | bis(diethylthiocarbamoyl)disulfide | −5.56 | 3.67 | 1.13 | 1 | 0 | 0 | 3 | 5 | 0 |
| 73 | 98828 | cumene | −3.64 | 3.45 | −0.60 | 0 | 0 | 0 | 1 | 1 | 0 |
| 74 | 98953 | nitrobenzene | −3.48 | 1.81 | 0.65 | 1 | 0 | 0 | 2 | 1 | 0 |
| 75 | 99081 | 1-methyl-3-nitrobenzene | −4.04 | 2.36 | 0.74 | 1 | 0 | 0 | 2 | 1 | 0 |
| 76 | 99514 | 1,2-dimethyl-4-nitrobenzene | −3.98 | 2.91 | 0.20 | 1 | 0 | 0 | 2 | 1 | 0 |
| 77 | 99650 | 1,3-dinitrobenzene | −3.59 | 1.63 | 0.91 | 1 | 0 | 0 | 3 | 5 | 0 |
| 78 | 99876 | cymene | −4.32 | 4.00 | −0.39 | 0 | 0 | 0 | 1 | 1 | 0 |
| 79 | 99990 | 4-methylnitrobenzene | −4.01 | 2.36 | 0.71 | 1 | 0 | 0 | 2 | 1 | 0 |
| 80 | 100005 | 4-chloronitrobenzene | −4.31 | 2.46 | 0.92 | 1 | 0 | 0 | 2 | 1 | 1 |
| 81 | 100027 | 4-nitrophenol | −3.96 | 1.91 | 1.04 | 1 | 0 | 0 | 2 | 2 | 0 |
| 82 | 100414 | ethyl benzene | −3.54 | 3.03 | −0.34 | 0 | 0 | 0 | 1 | 1 | 0 |
| 83 | 100425 | styrene | −3.41 | 2.89 | −0.35 | 0 | 0 | 0 | 1 | 5 | 0 |
| 84 | 100618 | *N*-methylaniline | −5.79 | 1.62 | 3.12 | 1 | 8.1 | 8.1 | NA | 1 | 0 |
| 85 | 102089 | diphenylthiourea | −3.53 | 3.21 | −0.50 | 1 | 0 | 0 | NA | 1 | 0 |
| 86 | 103695 | ethylaniline | −5.46 | 2.11 | 2.37 | 1 | 8.1 | 8.1 | NA | 1 | 0 |
| 87 | 103720 | isothiocyanatobenzene | −6.13 | 3.33 | 2.01 | 1 | 5.1 | 5.1 | NA | 5 | 0 |
| 88 | 103855 | phenylthiourea | −3.54 | 0.95 | 1.44 | 1 | 0 | 0 | NA | 1 | 0 |
| 89 | 104949 | 4-methoxybenzenamine | −5.57 | 1.16 | 3.30 | 1 | 8.1 | 8.1 | NA | 2 | 0 |
| 90 | 105373 | ethyl propionate | −2.78 | 1.36 | 0.34 | 0 | 0 | 0 | NA | 3 | 0 |
| 91 | 105555 | 1,3-diethylthiourea | −2.84 | 0.60 | 1.04 | 1 | 0 | 0 | NA | 1 | 0 |
| 92 | 105679 | 2,4-dimethylphenol | −4.77 | 2.61 | 1.25 | 0 | 0 | 0 | 2 | 2 | 0 |
| 93 | 106412 | 4-bromophenol | −4.46 | 2.40 | 1.13 | 0 | 0 | 0 | NA | 2 | 0 |
| 94 | 106445 | *p*-cresol | −3.71 | 2.06 | 0.66 | 0 | 0 | 0 | 2 | 2 | 0 |
| 95 | 106478 | 4-chloroaniline | −6.41 | 1.72 | 3.65 | 1 | 8.1 | 8.1 | 2 | 2 | 0 |
| 96 | 106489 | 4-chlorophenol | −4.42 | 2.16 | 1.28 | 0 | 0 | 0 | 2 | 2 | 0 |
| 97 | 106898 | epichlorohydrin | −3.58 | 0.63 | 1.76 | 0 | 0 | 0 | 3 | 5 | 1 |
| 98 | 107028 | acrolein | −6.00 | 0.19 | 4.55 | 1 | 1.1 | 1.1 | 3 | 5 | 1 |
| 99 | 107039 | 1-propanethiol | −6.10 | 1.76 | 3.31 | 1 | 4.1 | 4.1 | NA | 1 | 1 |
| 100 | 107073 | 2-chloroethanol | −2.61 | 0.11 | 1.24 | 0 | 0 | 0 | NA | 5 | 0 |
| 101 | 107119 | allylamine | −3.15 | 0.21 | 1.69 | 1 | 0 | 0 | NA | 1 | 0 |
| 102 | 107131 | acrylonitrile | −3.78 | 0.21 | 2.32 | 1 | 1.1 | 1.1 | 3 | 5 | 1 |
| 103 | 107153 | ethylenediamine | −3.36 | −1.62 | 3.46 | 1 | 0 | 0 | NA | 1 | 0 |
| 104 | 107926 | *n*-butyric acid | −3.16 | 1.07 | 0.96 | 0 | 0 | 0 | NA | 1 | 0 |
| 105 | 108189 | bis(isopropyl)amine | −2.35 | 1.64 | −0.33 | 1 | 0 | 0 | 1 | 1 | 0 |
| 106 | 108394 | *m*-cresol | −3.76 | 2.06 | 0.71 | 0 | 0 | 0 | 2 | 2 | 0 |
| 107 | 108429 | 3-chloroaniline | −6.11 | 1.72 | 3.35 | 1 | 8.1 | 8.1 | 2 | 2 | 0 |
| 108 | 108441 | *m*-toluidine | −5.17 | 1.62 | 2.50 | 1 | 8.1 | 8.1 | 2 | 2 | 0 |
| 109 | 108850 | bromocyclohexane | −3.89 | 3.45 | −0.35 | 0 | 0 | 0 | 1 | 1 | 0 |
| 110 | 108952 | phenol | −3.44 | 1.51 | 0.87 | 0 | 0 | 0 | 2 | 2 | 0 |
| 111 | 109466 | dibutylthiourea | −3.52 | 2.57 | 0.04 | 1 | 0 | 0 | NA | 1 | 0 |
| 112 | 109524 | pentanoic acid | −3.36 | 1.56 | 0.74 | 0 | 0 | 0 | NA | 1 | 0 |
| 113 | 109897 | diethylamine | −3.12 | 0.81 | 1.14 | 1 | 0 | 0 | 1 | 1 | 0 |
| 114 | 110021 | thiophene | −2.42 | 1.81 | −0.41 | 1 | 0 | 0 | NA | 1 | 0 |
| 115 | 110838 | cyclohexene | −3.94 | 2.96 | 0.12 | 0 | 0 | 0 | 1 | 1 | 0 |
| 116 | 110861 | pyridine | −1.77 | 0.80 | −0.19 | 1 | 0 | 0 | 2 | 2 | 0 |
| 117 | 111422 | 2,2′-iminobisethanol | −2.93 | −1.71 | 3.11 | 1 | 0 | 0 | NA | 1 | 0 |
| 118 | 111444 | 2,2′-dichlorodiethyl ether | −2.78 | 1.56 | 0.16 | 0 | 0 | 0 | 3 | 5 | 1 |
| 119 | 111911 | propoxur | −2.94 | 1.30 | 0.54 | 0 | 0 | 0 | 3 | 5 | 0 |
| 120 | 114261 | 2-(1-methylethoxy)phenol, methyl carbamate | −4.91 | 1.90 | 2.02 | 1 | 6.1 | 6.1 | NA | 6 | 0 |
| 121 | 115297 | endosulfan | −6.14 | 3.50 | 1.86 | 1 | 0 | 0 | 3 | 7 | 0 |
| 122 | 115311 | isobornyl thiocyanatoacetate | −6.50 | 3.75 | 2.01 | 1 | 0 | 5.2 | 3 | 5 | 1 |
| 123 | 115866 | phosphoric acid | −5.51 | 4.70 | 0.20 | 1 | 0 | 0 | NA | 6 | 0 |
| 124 | 116063 | aldicarb | −5.61 | 1.36 | 3.16 | 1 | 6.1 | 6.1 | NA | 6 | 0 |
| 125 | 118967 | 2,4,6-trinitrotoluene | −4.39 | 1.99 | 1.40 | 1 | 0 | 0 | 3 | 5 | 0 |
| 126 | 119653 | isoquinoline | −3.71 | 2.14 | 0.60 | 1 | 0 | 0 | NA | 5 | 0 |
| 127 | 120832 | 2,4-dichlorophenol | −4.80 | 2.80 | 1.12 | 0 | 0 | 0 | 2 | 2 | 0 |
| 128 | 120934 | ethyleneurea | −1.19 | −0.74 | 0.54 | 1 | 0 | 0 | NA | 1 | 0 |
| 129 | 121142 | 2,4-dinitrotoluene | −3.72 | 2.18 | 0.57 | 1 | 0 | 0 | 3 | 5 | 0 |
| 130 | 121299 | pyrethrine II | −7.40 | 5.33 | 1.55 | 0 | 0 | 0 | 4 | 5/7 | 1 |
| 131 | 121733 | 3-nitrochlorobenzene | −3.84 | 2.46 | 0.45 | 1 | 0 | 0 | 2 | 1 | 0 |
| 132 | 121755 | malathion | −7.36 | 2.29 | 4.11 | 1 | 3.1 | 3.1 | 4 | 5/6 | 0 |
| 133 | 121879 | 2-chloro-4-nitroaniline | −4.49 | 2.12 | 1.39 | 1 | 0 | 0 | 2 | 2 | 0 |
| 134 | 122145 | fenitrothion | −6.68 | 3.30 | 2.57 | 1 | 3.1 | 3.1 | 4 | 6 | 0 |
| 135 | 122349 | simazine | −3.33 | 2.40 | −0.01 | 1 | 0 | 0 | NA | 1 | 0 |
| 136 | 122667 | 1,2-diphenylhydrazine | −4.65 | 3.06 | 0.75 | 1 | 0 | 0 | 3 | 5 | 0 |
| 137 | 123546 | 2,4-pentanedione | −3.32 | 0.05 | 1.99 | 0 | 0 | 0 | NA | 1 | 0 |
| 138 | 124403 | dimethylamine | −2.96 | −0.17 | 1.82 | 1 | 0 | 0 | 1 | 1 | 0 |
| 139 | 126738 | tributyl phosphate | −4.86 | 3.82 | 0.31 | 1 | 0 | 0 | NA | 6 | 0 |
| 140 | 131113 | dimethyl phthalate | −3.77 | 1.66 | 1.07 | 0 | 0 | 0 | 3 | 3 | 0 |
| 141 | 132650 | dibenzothiophene | −5.06 | 4.17 | 0.20 | 1 | 0 | 0 | NA | 5 | 0 |
| 142 | 135193 | 2-naphthol | −4.61 | 2.69 | 1.02 | 0 | 0 | 0 | NA | 2 | 0 |
| 143 | 137268 | thiram | −6.06 | 1.70 | 3.32 | 1 | 0 | 6.3 | 3 | 5 | 0 |
| 144 | 140669 | 4-tert-octylphenol | −6.36 | 5.28 | 0.55 | 0 | 0 | 0 | 2 | 2 | 0 |

**Table 1 (Continued)**

| no. | CAS | name | log $LC_{50}$ (mol/L) | log $K_{ow}$ | log $T_e$ | CM1 | CM2 | CM3 | 4-MOA CM | 7-MOA CM | 2-MOA CM |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | training set | | | | | | | |
| 145 | 141786 | ethyl acetate | −2.09 | 0.86 | 0.07 | 0 | 0 | 0 | NA | 3 | 0 |
| 146 | 141902 | thiouracil | −4.22 | 0.90 | 2.17 | 1 | 9.1 | 9.1 | NA | 1 | 0 |
| 147 | 142961 | butyl ether | −3.70 | 3.01 | −0.16 | 0 | 0 | 0 | 1 | 1 | 0 |
| 148 | 148016 | dinitolmide | −3.14 | 0.19 | 1.69 | 1 | 0 | 0 | 3 | 5 | 0 |
| 149 | 149315 | 2-methyl-1,3-pentanediol | −1.22 | 0.62 | −0.59 | 0 | 0 | 0 | 1 | 1 | 0 |
| 150 | 150196 | 3-methoxyphenol | −3.48 | 1.59 | 0.84 | 0 | 0 | 0 | NA | 2 | 0 |
| 151 | 156605 | *trans*-1,2-dichloroethylene | −2.64 | 1.98 | −0.33 | 0 | 0 | 0 | 1 | 1 | 0 |
| 152 | 206440 | fluoranthene | −6.28 | 4.93 | 0.78 | 0 | 0 | 0 | 1 | 1 | 0 |
| 153 | 260946 | acridine | −4.81 | 3.32 | 0.68 | 1 | 0 | 0 | NA | 5 | 0 |
| 154 | 298000 | methyl parathion | −7.34 | 2.75 | 3.70 | 1 | 3.1 | 3.1 | 4 | 6 | 0 |
| 155 | 298022 | phorate | −7.13 | 3.37 | 2.96 | 1 | 3.1 | 3.1 | 4 | 6 | 0 |
| 156 | 311455 | diethyl *p*-nitrophenyl phosphate | −9.14 | 1.97 | 6.17 | 1 | 0 | 0 | NA | 6 | 0 |
| 157 | 333415 | diazinon | −8.45 | 3.86 | 3.86 | 1 | 3.1 | 3.1 | 4 | 6 | 0 |
| 158 | 470906 | chlorfenvinfos | −6.56 | 4.15 | 1.72 | 1 | 0 | 0 | NA | 5/6 | 0 |
| 159 | 503877 | 2-thioxo-4-imidazolinone | −3.77 | −0.43 | 2.86 | 1 | 0 | 7.3 | NA | 1 | 0 |
| 160 | 532558 | benzoyl isothiocyanate | −4.93 | 1.65 | 2.24 | 1 | 5.1 | 5.1 | NA | 5 | 1 |
| 161 | 534134 | *N*,*N*′-dimethylthiourea | −3.85 | −0.38 | 2.90 | 1 | 7.1 | 7.1 | NA | 1 | 0 |
| 162 | 534521 | dinitro-*o*-cresol | −4.79 | 2.27 | 1.57 | 1 | 0 | 0 | 3 | 4 | 0 |
| 163 | 536903 | 3-methoxybenzeneamine | −5.64 | 1.16 | 3.37 | 1 | 8.1 | 8.1 | NA | 2 | 0 |
| 164 | 542756 | 1,3-dichloropropene | −4.25 | 2.29 | 1.01 | 0 | 0 | 0 | 3 | 5 | 0 |
| 165 | 542858 | isothiocyanatoethane | −5.31 | 1.79 | 2.49 | 1 | 5.1 | 5.1 | NA | 5 | 1 |
| 166 | 554007 | 2,4-dichloroaniline | −5.43 | 2.37 | 2.12 | 1 | 0 | 0 | 2 | 2 | 0 |
| 167 | 556616 | isothiocyanatomethane | −5.42 | 1.30 | 3.02 | 1 | 5.1 | 5.1 | NA | 5 | 1 |
| 168 | 576261 | 2,6-dimethylphenol | −4.04 | 2.61 | 0.52 | 0 | 0 | 0 | 2 | 2 | 0 |
| 169 | 578541 | 2-ethylbenzenamine | −4.18 | 2.11 | 1.09 | 1 | 0 | 0 | 2 | 2 | 0 |
| 170 | 589162 | 4-ethylaniline | −6.13 | 2.11 | 3.04 | 1 | 8.1 | 8.1 | 2 | 2 | 0 |
| 171 | 592825 | 1-isothiocyanatobutane | −5.43 | 2.77 | 1.77 | 1 | 5.1 | 5.1 | NA | 5 | 1 |
| 172 | 598163 | tribromoethene | −4.33 | 2.18 | 1.18 | 0 | 0 | 0 | 1 | 1 | 0 |
| 173 | 598527 | methylthiourea | −3.98 | −0.84 | 3.42 | 1 | 7.1 | 7.1 | NA | 1 | 0 |
| 174 | 602017 | 2,3-dinitrotoluene | −5.44 | 2.18 | 2.29 | 1 | 0 | 0 | 3 | 5 | 0 |
| 175 | 609198 | 3,4,5-trichlorophenol | −5.46 | 3.45 | 1.23 | 0 | 0 | 0 | 2 | 2 | 0 |
| 176 | 611063 | 2,4-dichloro-1-nitrobenzene | −4.66 | 3.10 | 0.72 | 1 | 0 | 0 | 2 | 1 | 1 |
| 177 | 618622 | 1,3-dichloro-5-nitrobenzene | −4.46 | 3.10 | 0.52 | 1 | 0 | 0 | 2 | 1 | 0 |
| 178 | 622786 | benzylisothiocyanate | −6.54 | 3.01 | 2.68 | 1 | 5.1 | 5.1 | NA | 5 | 1 |
| 179 | 625536 | ethylthiourea | −4.00 | −0.35 | 3.01 | 1 | 0 | 7.2 | NA | 1 | 0 |
| 180 | 626437 | 3,5-dichloroaniline | −5.16 | 2.37 | 1.85 | 1 | 8.1 | 8.1 | 2 | 2 | 0 |
| 181 | 630206 | 1,1,1,2-tetrachloroethane | −3.84 | 2.93 | 0.05 | 0 | 0 | 0 | 1 | 1 | 0 |
| 182 | 632224 | 1,1,3,3-tetramethylurea | −1.60 | −0.20 | 0.49 | 1 | 0 | 0 | NA | 1 | 0 |
| 183 | 634673 | 2,3,4-trichloroaniline | −5.43 | 3.01 | 1.57 | 1 | 0 | 0 | 2 | 2 | 0 |
| 184 | 634833 | 2,3,4,5-tetrachloroaniline | −5.56 | 3.65 | 1.15 | 1 | 0 | 0 | NA | 2/4 | 0 |
| 185 | 636306 | 2,4,5-trichloroaniline | −4.76 | 3.01 | 0.90 | 1 | 0 | 0 | 2 | 2 | 0 |
| 186 | 680319 | hexamethyl phosphoramide | −1.43 | −0.22 | 0.34 | 1 | 0 | 0 | NA | 1 | 0 |
| 187 | 693210 | diethylene glycol dinitrate | −3.34 | 0.90 | 1.29 | 1 | 0 | 0 | NA | 1 | 0 |
| 188 | 732116 | phosmet | −5.60 | 2.48 | 2.19 | 1 | 3.1 | 3.1 | 4 | 6 | 0 |
| 189 | 759944 | dipropylcarbamothioic acid, S-ethyl ester | −4.61 | 3.02 | 0.74 | 1 | 0 | 0 | NA | 1 | 0 |
| 190 | 786196 | carbophenothion | −6.44 | 5.19 | 0.71 | 1 | 3.1 | 3.1 | 4 | 6 | 0 |
| 191 | 825445 | benzo[*b*]thiophene S,S-dioxide | −4.07 | 0.78 | 2.12 | 1 | 0 | 0 | NA | 5 | 0 |
| 192 | 877430 | 2,6-dimethylquinoline | −3.62 | 3.24 | −0.44 | 1 | 0 | 0 | NA | 1/5 | 0 |
| 193 | 935955 | 2,3,5,6-tetrachlorophenol | −5.61 | 4.09 | 0.82 | 0 | 0 | 0 | NA | 4 | 0 |
| 194 | 944229 | fonofos | −7.43 | 4.02 | 2.71 | 1 | 0 | 3.3 | NA | 1 | 0 |
| 195 | 1014706 | simetryn | −3.63 | 2.90 | −0.14 | 1 | 0 | 0 | NA | 1 | 0 |
| 196 | 1016053 | dibenzothiophene-5,5-dioxide | −4.57 | 2.61 | 1.05 | 1 | 0 | 0 | NA | 1 | 0 |
| 197 | 1024573 | heptachlor epoxide | −6.21 | 4.56 | 1.02 | 0 | 0 | 0 | 3 | 5/7 | 1 |
| 198 | 1516321 | butylthiourea | −3.85 | 0.63 | 2.03 | 1 | 0 | 7.2 | NA | 1 | 0 |
| 199 | 1563662 | carbofuran | −6.52 | 2.30 | 3.27 | 1 | 6.1 | 6.1 | NA | 6 | 0 |
| 200 | 1570645 | 4-chloro-*o*-cresol | −5.69 | 2.70 | 2.10 | 0 | 0 | 0 | 2 | 2 | 0 |
| 201 | 1570656 | 2,4-dichloro-6-methylphenol | −5.65 | 3.35 | 1.50 | 0 | 0 | 0 | 2 | 2 | 0 |
| 202 | 1582098 | trifluralin | −6.24 | 5.31 | 0.41 | 1 | 0 | 0 | 3 | 5 | 0 |
| 203 | 1825214 | pentachloroanisole | −7.01 | 5.30 | 1.19 | 0 | 0 | 0 | NA | 5 | 0 |
| 204 | 1836777 | chlornitrofen | −5.88 | 4.96 | 0.34 | 1 | 0 | 0 | 3 | 1 | 0 |
| 205 | 1897456 | chlorothalonil | −6.21 | 3.66 | 1.79 | 1 | 0 | 0 | 3 | 1 | 0 |
| 206 | 1912249 | atrazine | −3.60 | 2.82 | −0.10 | 1 | 0 | 0 | NA | 1 | 0 |
| 207 | 1918021 | picloram | −3.61 | 1.36 | 1.16 | 1 | 0 | 0 | 3 | 2 | 0 |
| 208 | 1982474 | chloroxuron | −4.99 | 4.08 | 0.22 | 1 | 0 | 0 | NA | 1 | 0 |
| 209 | 2008584 | 2,6-dichlorobenzamide | −2.35 | 0.90 | 0.29 | 1 | 0 | 0 | 3 | 1 | 0 |
| 210 | 2051607 | 2-chlorobiphenyl | −5.42 | 4.40 | 0.37 | 0 | 0 | 0 | NA | 1 | 0 |
| 211 | 2051618 | 3-chlorobiphenyl | −5.64 | 4.40 | 0.59 | 0 | 0 | 0 | NA | 1 | 0 |
| 212 | 2051629 | 4-chloro-1,1′-biphenyl | −5.60 | 4.40 | 0.55 | 0 | 0 | 0 | NA | 1 | 0 |
| 213 | 2257092 | (2-isothiocyanatoethyl)benzene | −6.10 | 3.50 | 1.82 | 1 | 5.1 | 5.1 | NA | 5 | 1 |
| 214 | 2437798 | 2,4,2′,4′-tetrachlorobiphenyl | −6.99 | 6.34 | 0.27 | 0 | 0 | 0 | NA | 1 | 0 |
| 215 | 2489772 | trimethylthiourea | −2.19 | −0.17 | 1.05 | 1 | 0 | 0 | NA | 1 | 0 |
| 216 | 2539175 | 2-methoxytetrachlorophenol | −6.08 | 3.92 | 1.44 | 0 | 0 | 0 | NA | 2/4/5 | 0 |

**Table 1 (Continued)**

| no. | CAS | name | log $LC_{50}$ (mol/L) | log $K_{ow}$ | log $T_e$ | CM1 | CM2 | CM3 | 4-MOA CM | 7-MOA CM | 2-MOA CM |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | training set | | | | | | | |
| 217 | 2556425 | tetrapropylthioperoxydicarbonic-diamide | −6.19 | 5.63 | 0.08 | 1 | 0 | 0 | 3 | 5 | 0 |
| 218 | 2668248 | 2-methoxy-4,5,6-trichlorophenol | −5.37 | 3.27 | 1.29 | 0 | 0 | 0 | NA | 2/5 | 0 |
| 219 | 2741062 | 1-phenyl-3-ethyl thiourea | −3.35 | 1.91 | 0.43 | 1 | 0 | 0 | NA | 1 | 0 |
| 220 | 2764729 | diquat | −5.01 | 2.36 | 1.71 | 1 | 0 | 0 | NA | 1/5 | 0 |
| 221 | 2782914 | tetramethyl thiourea | −2.23 | 0.04 | 0.92 | 1 | 0 | 0 | NA | 1 | 0 |
| 222 | 2921882 | clorpyrifos | −8.64 | 4.66 | 3.37 | 1 | 3.1 | 3.1 | 4 | 6 | 0 |
| 223 | 3209221 | 1,2-dichloro-3-nitrobenzene | −4.62 | 3.10 | 0.69 | 1 | 0 | 0 | 2 | 1 | 1 |
| 224 | 3483123 | dithiothreitol | −3.76 | −0.48 | 2.89 | 1 | 4.1 | 4.1 | NA | 1 | 1 |
| 225 | 3547044 | DDE | −6.86 | 5.44 | 0.91 | 0 | 0 | 0 | NA | 5 | 0 |
| 226 | 3689245 | TEDP | −9.15 | 3.98 | 4.45 | 1 | 3.1 | 3.1 | 4 | 1 | 0 |
| 227 | 3766812 | 2-(1-methylpropyl)phenol, methylcarbamate | −6.32 | 2.86 | 2.58 | 1 | 6.1 | 6.1 | NA | 6 | 0 |
| 228 | 4044659 | 1,4-diisothiocyanatobenzene | −6.40 | 4.67 | 1.11 | 1 | 5.1 | 5.1 | NA | 5 | 0 |
| 229 | 4104750 | *N*-methyl-*N*-phenylthiourea | −3.36 | 1.50 | 0.79 | 1 | 0 | 0 | NA | 1 | 0 |
| 230 | 6317186 | thiocyanic acid, methylene ester | −6.25 | 0.62 | 4.44 | 1 | 0 | 5.2 | 3 | 5 | 0 |
| 231 | 6972050 | *N*,*N*-dimethylthiourea | −3.39 | −0.63 | 2.65 | 1 | 7.1 | 7.1 | NA | 1 | 0 |
| 232 | 7012375 | 2,4,4′-PCB | −6.21 | 5.69 | 0.05 | 0 | 0 | 0 | NA | 1 | 0 |
| 233 | 8018017 | mancozeb | −5.21 | 0.62 | 3.40 | 1 | 0 | 0 | NA | 1 | 0 |
| 234 | 10605217 | carbendazim | −5.54 | 1.55 | 2.93 | 1 | 6.1 | 6.1 | NA | 1 | 0 |
| 235 | 12002481 | trichlorobenzene | −4.40 | 3.93 | −0.25 | 0 | 0 | 0 | 1 | 1 | 0 |
| 236 | 15245440 | 2,4,6-trinitro-1,3-benzenediol | −2.19 | 1.06 | 0.00 | 1 | 0 | 0 | 3 | 4 | 0 |
| 237 | 15263533 | dithiocarbamate | −7.38 | −0.95 | 6.91 | 1 | 0 | 6.2 | NA | 1 | 0 |
| 238 | 18259057 | 2,3,4,5,6-PCB | −7.61 | 6.98 | 0.35 | 0 | 0 | 0 | NA | 1 | 0 |
| 239 | 23564058 | thiophanate-methyl | −4.33 | 1.10 | 2.11 | 1 | 6.1 | 6.1 | NA | 1 | 0 |
| 240 | 25154523 | nonylphenol | −6.41 | 5.99 | 0.00 | 0 | 0 | 0 | 2 | 2 | 0 |
| 241 | 25167833 | 2,3,4,5-tetrachlorophenol | −5.76 | 4.09 | 0.97 | 0 | 0 | 0 | NA | 2/4 | 0 |
| 242 | 25875518 | robenidine | −6.65 | 3.75 | 2.15 | 1 | 9.1 | 9.1 | 3 | 1 | 0 |
| 243 | 28249776 | thiobencarb | −4.67 | 3.90 | 0.05 | 1 | 0 | 0 | NA | 1 | 0 |
| 244 | 29232937 | pirimiphos-methyl | −9.14 | 3.44 | 4.91 | 1 | 3.1 | 3.1 | 4 | 6 | 0 |
| 245 | 32598133 | 3,3′,4,4′-tetrachloro-1,1′-biphenyl | −8.16 | 6.34 | 1.45 | 0 | 0 | 0 | NA | 1 | 0 |
| 246 | 33813206 | 5,6-dihydro-3H-imidazo[2,1-c]-1,2,4-dithiazole-3-thione | −5.92 | 1.60 | 3.27 | 1 | 0 | 6.4 | 3 | 5 | 0 |
| 247 | 33820530 | isopropalin | −7.01 | 5.80 | 0.76 | 1 | 0 | 0 | 3 | 5 | 0 |
| 248 | 35065271 | 2,2′,4,4′,5,5′-hexachloro-1,1′-biphenyl | −8.44 | 7.62 | 0.63 | 0 | 0 | 0 | NA | 1 | 0 |
| 249 | 35367385 | diflubenzuron | −7.77 | 3.59 | 3.41 | 1 | 9.1 | 9.1 | 3 | 1 | 0 |
| 250 | 35693993 | 2,2′,5,5′-tetrachloro-1,1′-biphenyl | −6.99 | 6.34 | 0.27 | 0 | 0 | 0 | NA | 1 | 0 |
| 251 | 37680652 | 2,2′,5-trichloro-1,1′-biphenyl | −6.67 | 5.69 | 0.51 | 0 | 0 | 0 | NA | 1 | 0 |
| 252 | 37680732 | 2,4,5,2′,5′-PCB | −7.51 | 6.98 | 0.25 | 0 | 0 | 0 | NA | 1 | 0 |
| 253 | 38380073 | 2,2′,3,3′,4,4′-PCB | −8.78 | 7.62 | 0.97 | 0 | 0 | 0 | NA | 1 | 0 |
| 254 | 51630581 | fenvalerate | −8.13 | 6.76 | 1.05 | 1 | 0 | 0 | NA | 3 | 0 |
| 255 | 52315078 | cypermethrin | −9.06 | 6.38 | 2.31 | 1 | 2.1 | 2.1 | 3 | 3 | 0 |
| 256 | 52645531 | permethrin | −8.23 | 7.43 | 0.58 | 1 | 2.1 | 2.1 | 4 | 7 | 0 |
| 257 | 52918635 | deltamethrin | −10.09 | 6.18 | 3.52 | 1 | 2.1 | 2.1 | 3 | 3 | 0 |
| 258 | 55406536 | iodopropynyl butylcarbamate | −6.85 | 2.45 | 3.47 | 1 | 6.1 | 6.1 | NA | 1 | 0 |
| 259 | 57057837 | 3,4,5-trichloroguaiacol | −5.56 | 3.27 | 1.48 | 0 | 0 | 0 | NA | 2/5 | 0 |
| 260 | 59756604 | fluridone | −4.86 | 4.48 | −0.26 | 1 | 0 | 0 | NA | 1/5 | 0 |
| 261 | 66230044 | esfenvalerate | −9.19 | 6.76 | 2.12 | 1 | 0 | 0 | NA | 3 | 0 |
| 262 | 68359375 | cyfluthrin | −9.42 | 5.74 | 3.22 | 1 | 2.1 | 2.1 | 3 | 3 | 0 |
| 263 | 76738620 | paclobutrazol | −4.01 | 3.36 | −0.15 | 1 | 0 | 0 | 3 | 1 | 0 |
| 264 | 91465086 | cyhalothrin | −8.72 | 6.85 | 1.57 | 1 | 0 | 0 | NA | 3 | 0 |
| | | | | narcotic | | | | | | | |
| 265 | 56235 | tetrachloromethane | −3.64 | 2.44 | −0.27 | 0 | 0 | 0 | 1 | 1 | 0 |
| 266 | 64175 | ethanol | −0.59 | −0.14 | 0.57 | 0 | 0 | 0 | 1 | 1 | 0 |
| 267 | 67641 | acetone | −0.62 | −0.24 | 0.45 | 0 | 0 | 0 | 1 | 1 | 0 |
| 268 | 67663 | trichlormethane | −2.72 | 1.52 | −0.13 | 0 | 0 | 0 | 1 | 1 | 0 |
| 269 | 67721 | hexachloroethane | −4.83 | 4.03 | −0.10 | 0 | 0 | 0 | 1 | 1 | 0 |
| 270 | 71238 | 1-propanol | −0.93 | 0.35 | 0.65 | 0 | 0 | 0 | 1 | 1 | 0 |
| 271 | 71432 | benzene | −2.48 | 1.99 | 0.51 | 0 | 0 | 0 | 1 | 1 | 0 |
| 272 | 75092 | dichlormethane | −2.59 | 1.34 | −0.16 | 0 | 0 | 0 | 1 | 1 | 0 |
| 273 | 76017 | pentachloroethane | −3.97 | 3.11 | −0.03 | 0 | 0 | 0 | 1 | 1 | 0 |
| 274 | 78831 | 2-methyl-1-propanol | −1.82 | 0.77 | 0.12 | 0 | 0 | 0 | 1 | 1 | 0 |
| 275 | 78875 | 1,2-dichloropropane | −3.34 | 2.25 | −0.13 | 0 | 0 | 0 | 1 | 1 | 0 |
| 276 | 79005 | 1,1,2-trichloroethane | −3.09 | 2.01 | −0.09 | 0 | 0 | 0 | 1 | 1 | 0 |
| 277 | 79016 | trichloroethene | −3.35 | 2.47 | 0.05 | 0 | 0 | 0 | 1 | 1 | 0 |
| 278 | 79345 | 1,1,2,2-tetrachloroethane | −3.45 | 2.19 | −0.29 | 0 | 0 | 0 | 1 | 1 | 0 |
| 279 | 91203 | naphthalene | −4.12 | 3.17 | −0.12 | 0 | 0 | 0 | 1 | 1 | 0 |
| 280 | 95476 | *o*-xylene | −3.78 | 3.09 | 0.15 | 0 | 0 | 0 | 1 | 1 | 0 |
| 281 | 95501 | 1,2-dichlorobenzene | −4.81 | 3.28 | −0.71 | 0 | 0 | 0 | 1 | 1 | 0 |
| 282 | 96184 | 1,2,3-trichloropropane | −3.72 | 2.50 | −0.29 | 0 | 0 | 0 | 1 | 1 | 0 |
| 283 | 101553 | 4-bromophenyl-phenyl ether | −5.84 | 4.94 | −0.33 | 0 | 0 | 0 | NA | 1 | 0 |
| 284 | 101848 | diphenyl ether | −5.46 | 4.05 | −0.71 | 0 | 0 | 0 | NA | 1 | 0 |

**Table 1 (Continued)**

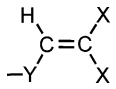
| no. | CAS | name | log $LC_{50}$ (mol/L) | log $K_{ow}$ | log $T_e$ | CM1 | CM2 | CM3 | 4-MOA CM | 7-MOA CM | 2-MOA CM |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | narcotic | | | | | | | |
| 285 | 106423 | *p*-xylene | −3.52 | 3.09 | 0.41 | 0 | 0 | 0 | 1 | 1 | 0 |
| 286 | 106467 | 1,4-dichlorobenzene | −4.17 | 3.28 | −0.08 | 0 | 0 | 0 | 1 | 1 | 0 |
| 287 | 107062 | 1,2-dichloroethane | −2.29 | 1.83 | 0.56 | 0 | 0 | 0 | 1 | 1 | 0 |
| 288 | 107211 | 1,2-ethanediol | −0.48 | −1.20 | −0.23 | 0 | 0 | 0 | NA | 1 | 0 |
| 289 | 107415 | 2-methyl-2,4-pentanediol | −1.22 | 0.58 | 0.55 | 0 | 0 | 0 | 1 | 1 | 0 |
| 290 | 108383 | *m*-xylene | −3.43 | 3.09 | 0.50 | 0 | 0 | 0 | 1 | 1 | 0 |
| 291 | 108883 | toluene | −2.80 | 2.54 | 0.66 | 0 | 0 | 0 | 1 | 1 | 0 |
| 292 | 108907 | monochlorobenzene | −3.77 | 2.64 | −0.22 | 0 | 0 | 0 | 1 | 1 | 0 |
| 293 | 111706 | 1-heptanol | −3.22 | 2.31 | 0.04 | 0 | 0 | 0 | 1 | 1 | 0 |
| 294 | 111900 | 2-(2-ethoxyethoxy)ethanol | −1.53 | −0.69 | −0.84 | 0 | 0 | 0 | 1 | 1 | 0 |
| 295 | 112276 | triethylene glycol | −0.46 | −1.75 | −0.68 | 0 | 0 | 0 | NA | 1 | 0 |
| 296 | 115208 | 2,2,2-trichloroethanol | −3.00 | 1.21 | −0.69 | 0 | 0 | 0 | NA | 5 | 0 |
| 297 | 120821 | 1,2,4-trichlorobenzene | −4.16 | 3.93 | 0.49 | 0 | 0 | 0 | 1 | 1 | 0 |
| 298 | 127184 | tetrachloroethene | −4.04 | 2.97 | −0.22 | 0 | 0 | 0 | 1 | 1 | 0 |
| 299 | 142289 | 1,3-dichloropropane | −2.61 | 2.32 | 0.66 | 0 | 0 | 0 | 1 | 1 | 0 |
| 300 | 541731 | 1,3-dichlorobenzene | −4.18 | 3.28 | −0.09 | 0 | 0 | 0 | 1 | 1 | 0 |

[a] $LC_{50}$ denotes the experimental lethal concentration of 50% toward *D. magna* within 48 h (*23*), $K_{ow}$ is the calculated octanol/water partition coefficient (*28*), and $T_e$ is the excess toxicity (eq 1). CM1, CM2, and CM3 are the presently developed CMs, and 4-MOA CM, 7-MOA CM, and 2-MOA CM denote three MOA-based CMs taken from the literature (*11*, *12*, *22*). NA, not assigned.

unit of each SA is relevant (e.g., in the case of SA3, 3.1: phosphorothionates with different heteroatoms), and for CM3, the additional tentative rules—as far as present—are marked through a star (e.g., in the case of SA3, 3.2*: thiophosphonates).

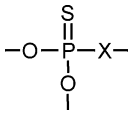
**SA1.** α,β-Unsaturated carbonyl and nitrile compounds

**SA2.** Carbon−carbon double bond activated by two halogens (X = halogen; Y = C, O, or N)

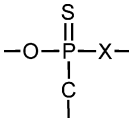


**SA3.** Organophosphorus compounds

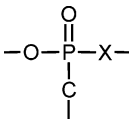
**SA3.1.** Phosphorothionates with different heteroatoms (X = O or S)



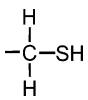
**SA3.2*.** Thiophosphonates (X = O or S)



**SA3.3*.** Phosphonates (X = O or S)

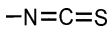


**SA4.** Aliphatic thioles

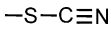


**SA5.** Isothiocyanates and thiocyanates

**SA5.1.** Isothiocyanates

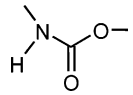


**SA5.2*.** Thiocyanates

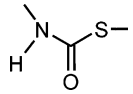


**SA6.** Carbamates

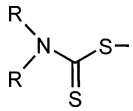
**SA6.1.** Simple carbamates



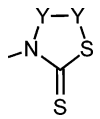
**SA6.2*.** Thiocarbamates



**SA6.3*.** Dithiocarbamates (R = H or $CH_3$)



**SA6.4*.** Rhodanin derivatives (Y = C, O, N, or S)



**SA7.** Thiourea derivatives

**SA7.1.** With at least two hydrogen atoms and a maximum of two methyl groups attached to the two nitrogens (R = H, $CH_3$)

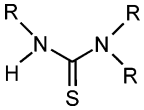
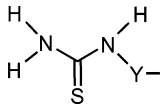


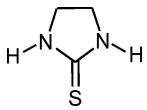
**Table 2. CMs Used for Discriminating between Narcotic Effect Levels and Excess Toxicity**

| | predicted category | |
|---|---|---|
| models | excess toxicity | narcotic effect level |
| CM1 | all organic chemicals other than narcosis | organic chemicals that contain any combination of C, H, O, and halogens, not showing SA1 |
| CM2 | organic chemicals containing SA1−SA9 | all other organic chemicals |
| CM3 | organic chemicals containing SA1, SA2, SA3*, SA4, SA5*−SA7*, SA8, SA9 | all other organic chemicals |
| 4-MOA CM (*11*) | organic chemicals classified as (3) reactive chemicals and (4) specifically acting chemicals | organic chemicals classified as (1) inert chemicals and (2) less inert chemicals |
| 7-MOA CM (*12*) | organic chemicals classified as (4) oxidative phosphorylation uncouplers, (5) electrophiles and proelectrophiles, (6) AChE inhibitors, and (7) CNS seizure agents | organic chemicals classified as (1) baseline narcotics, (2) polar narcotics, and (3) ester narcotics |
| 2-MOA CM (*22*) | organic chemicals classified as (1) electrophiles or (2) proelectrophiles | all other organic chemicals |

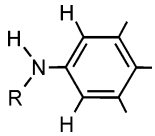
**SA7.2*.** With one aliphatic chain attached to nitrogen (Y ≠ H, $CH_3$)



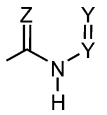
**SA7.3*.** Cyclic ethylene−thiourea derivatives



**SA8.** Primary or secondary anilines without ortho substituents (R = H, $CH_3$, or $C_2H_5$)



**SA9.** Imid derivatives with different heteroatoms (Y = C, O, or N; Z = O or N)



In Table 2, the relevant structural rules of CM1, CM2, and CM3 are summarized. Moreover, Table 2 contains the rules used to apply the MOA-based schemes 4-MOA CM (*11*), 7-MOA CM (*12*), and 2-MOA CM (*22*) for distinguishing between narcotic effect levels and excess toxicity as defined in the present context. To this end, for each MOA, all narcotic MOAs (e.g., "inert chemicals" and "less inert chemicals" in the case of the 4-MOA CM) were allocated to the narcotic toxicity range, and all reactive or specific MOAs (e.g., electrophilic and proelectrophilic toxicity in the case of the 2-MOA CM) were allocated to the excess toxicity. It should be noted that the original derivation of these three CMs did not consider the presently employed $T_e$ criterion. As a consequence, the statistical performances as presented below do not evaluate their validity in terms of MOA classifications but only their suitability to discriminate between the narcotic effect range and the excess toxicity.

Coming back to the prediction results of these six CMs for the present set of 300 compounds, the information summarized in Table 1 is illustrated with two examples: For DDT (compound 1 in Table 1), CM1, CM2, CM3, and the 2-MOA CM agree in their prediction of a narcotic toxicity range, while the 4-MOA CM classifies DDT as specifically acting compound, and the 7-MOA CM classifies DDT as a central nervous system (CNS) seizure agent. Note, however, that in the acute daphnid test, the log $T_e$ of DDT is 0.79 and thus well in the range expected already from narcotics, despite the well-known fact that DDT has the potential to affect specifically the CNS.

With carbaryl (compound 18 in Table 1), the 4-MOA CM does not offer structural rules to predict any MOA, while the 7-MOA CM classifies this compound as exerting nonpolar narcosis, which agrees with the respective implicit prediction when applying the 2-MOA CM (because carbaryl does not contain any of the electrophilic and proelectrophilic structural features as defined in this CM). According to CM1, CM2, and CM3, carbaryl is classified as excess toxic in agreement with the experimental log $T_e$ value of 4.04 toward *D. magna*. With CM1, the relevant structural criterion is that carbaryl contains nitrogen (in the form of −NH−), while the relevant SA of CM2 and CM3 is the carbamate functionality (SA5, −O−CO−NH−). Note that in both cases, fish-trained MOA predictions (if applicable) would suggest toxic effect levels in ranges significantly different from what is actually found in the daphnid bioassay.

**Sterical Influence.** Figure 2 contains examples of chemicals classified as thiourea derivatives by SA7. These compounds show similar molecular structures but differ in their log $T_e$. With increasingly bulky substitution at the thiourea nitrogen, the excess toxicity is lowered: The parent substance thiourea showed the highest toxicity with a log $T_e$ of 3.55, whereas 1,3-diethylthiourea with two ethyl side chains was not excess toxic.

**Classification Performances.** In Table 3, the performance of each CM is evaluated in terms of contingency tables. The last column shows the number of compounds that could actually be classified by the given CM. Taking CM2 as an example, 205 of the 264 compounds were predicted to exert a narcotic level toxicity, of which 179 actually show $LC_{50}$ values in the narcotic range as defined through log $T_e \leq 2$, and 26 compounds have $LC_{50}$ values exceeding that toxicity range. Moreover, with CM2, seven of the 186 narcotic level compounds are (wrongly) predicted to exert excess toxicity, and 52 of the 78 compounds with experimental excess toxicities are correctly recognized.

Besides the newly introduced CM1, CM2, and CM3, the 2-MOA CM (a scheme to identify electrophilic and proelectrophilic structures as indicators of excess toxicity)

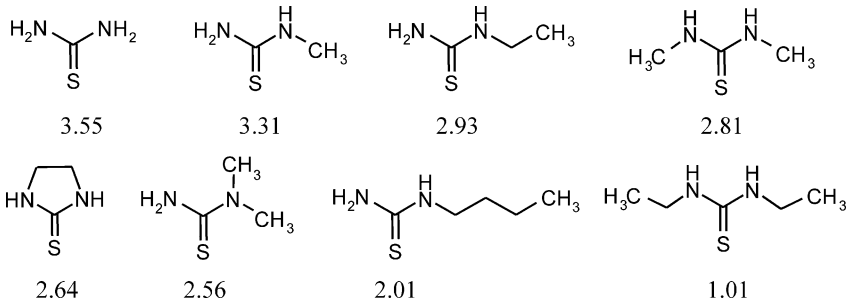



**Figure 2.** Thiourea derivatives that contain similar molecular structures, ranked according to their excess toxicity (log $T_e$).

**Table 3. Two-Dimensional 2 × 2 Contingency Table of Six CMs**

| | category type . | | | |
|---|---|---|---|---|
| | | experimental category | | |
| models | predicted category | narcotic effect level | excess toxicity | total |
| CM1 | narcotic effect level | 89 | 2 | 91 |
| | excess toxic | 97 | 76 | 173 |
| | total | 186 | 78 | 264 |
| CM2 | narcotic effect level | 179 | 26 | 205 |
| | excess toxic | 7 | 52 | 59 |
| | total | 186 | 78 | 264 |
| CM3 | narcotic effect level | 179 | 14 | 193 |
| | excess toxic | 7 | 64 | 71 |
| | total | 186 | 78 | 264 |
| 4-MOA CM | narcotic effect level | 59 | 11 | 70 |
| | excess toxic | 44 | 25 | 69 |
| | total | 103 | 36 | 139 |
| 7-MOA CM | narcotic effect level | 125 | 45 | 170 |
| | excess toxic | 61 | 33 | 94 |
| | total | 186 | 78 | 264 |
| 2-MOA CM | narcotic effect level | 169 | 65 | 234 |
| | excess toxic | 17 | 13 | 30 |
| | total | 186 | 78 | 264 |

**Table 4. Contingency Table Statistics of Six CMs**

| | | | statistical evaluation | |
|---|---|---|---|---|
| model | concordance | category | sensitivity | predictivity |
| CM1 | 0.625 | narcotic effect level | 0.478 | 0.978 |
| | | excess toxicity | 0.974 | 0.439 |
| CM2 | 0.875 | narcotic effect level | 0.962 | 0.873 |
| | | excess toxicity | 0.666 | 0.881 |
| CM3 | 0.920 | narcotic effect level | 0.962 | 0.927 |
| | | excess toxicity | 0.820 | 0.901 |
| 4-MOA CM | 0.604 | narcotic effect level | 0.573 | 0.843 |
| | | excess toxicity | 0.694 | 0.362 |
| 7-MOA CM | 0.598 | narcotic effect level | 0.672 | 0.735 |
| | | excess toxicity | 0.423 | 0.351 |
| 2-MOA CM | 0.689 | narcotic effect level | 0.909 | 0.722 |
| | | excess toxicity | 0.167 | 0.433 |
| | | without anilines | | |
| 4-MOA CM | 0.642 | narcotic effect level | 0.563 | 0.964 |
| | | excess toxicity | 0.926 | 0.373 |
| 7-MOA CM | 0.616 | narcotic effect level | 0.655 | 0.784 |
| | | excess toxicity | 0.508 | 0.351 |
| 2-MOA CM | 0.715 | narcotic effect level | 0.904 | 0.755 |
| | | excess toxicity | 0.200 | 0.433 |

and the 7-MOA CM are able to provide classifications for all 264 compounds of the training set. The latter, however, yields conflicting MOA allocations for 15 of the compounds (cf. Table 1). With the simpler 4-MOA CM, only 139 substances (52%) can be classified (which means that structural rules are missing for 125 compounds), and here, one compound is allocated to two different MOAs.

The associated performance statistics in terms of the concordance (eq 2), sensitivity (eq 3), and predictivity (eq 4) are summarized in Table 4. Taking CM1 as an example, a high predictivity for narcotic level toxicity (0.978) contrasts with a relatively low respective sensitivity (0.478). The former indicates that compounds predicted to exert a narcotic level $LC_{50}$ have a high probability that their acute daphnid toxicity is actually in the narcotic range. However, with the present training set, 52% of the narcotic level compounds are not recognized as such when applying CM1. By contrast, the CM1 recognition power for compounds exerting excess toxicity is very high (sensitivity 0.974), while its predictivity for such compounds is low (0.439). As a consequence, the overall agreement between predicted and actual toxicity category is only moderate for CM1 (concordance = 0.625).

CM2 provides a significant overall improvement as compared to CM1, except that the latter is particularly strong in predicting narcotic effect levels toward daphnids, which could be very helpful in identifying those compounds where narcotic effect level QSARs for the acute daphnid toxicity can be applied. The greatest concordance is achieved with CM3, keeping in mind that here several rules have been applied that at this stage are only tentative.

All three MOA-based CMs taken from the literature (*11*, *12*, *22*) show only moderate concordances in the present context of discriminating narcotic effect levels from excess toxicity. Interestingly, their prediction power for excess toxicity is particularly poor (0.35−0.43). Note further that when applying the 2-MOA CM SAs for electrophilic and proelectrophilic substructures, only 16.7% of the compounds exerting excess toxicity in the *D. magna* test are actually recognized.

**QSAR Modeling of Narcotic Level Toxicity.** Figure 3 shows the data distributions of log $LC_{50}$ vs log $K_{ow}$ for different subsets of compounds that are classified as exerting narcotic level toxicities toward *D. magna*. In the top left part of the figure, the respective plot with the subset of 36 compounds classified previously as narcotics (*26*) is shown, and the associated regression line of the baseline QSAR (eq 5) is included for comparison also in all other plots of the figure.

For the subset of 91 compounds predicted to exert narcotic level toxicities according to CM1, the respective data distribution in the top right of Figure 3 shows that most of the $LC_{50}$ values are slightly below the baseline regression line. Consequently, the respective regression equation for CM1 narcotic level toxicity:

$$\log LC_{50}\,(\text{mol/L}) = -0.801\ (\pm 0.038)\log K_{ow} - 2.139\ (\pm 0.144) \quad (7)$$

($n$ = 91, $r^2$ = 0.83, SE = 0.66, and $F_{1,89}$ = 448) has a similar slope as compared to eq 5 (−0.801 vs −0.857) but an intercept that is almost one log unit lower than the one of the baseline QSAR (−2.139 vs −1.281). The latter

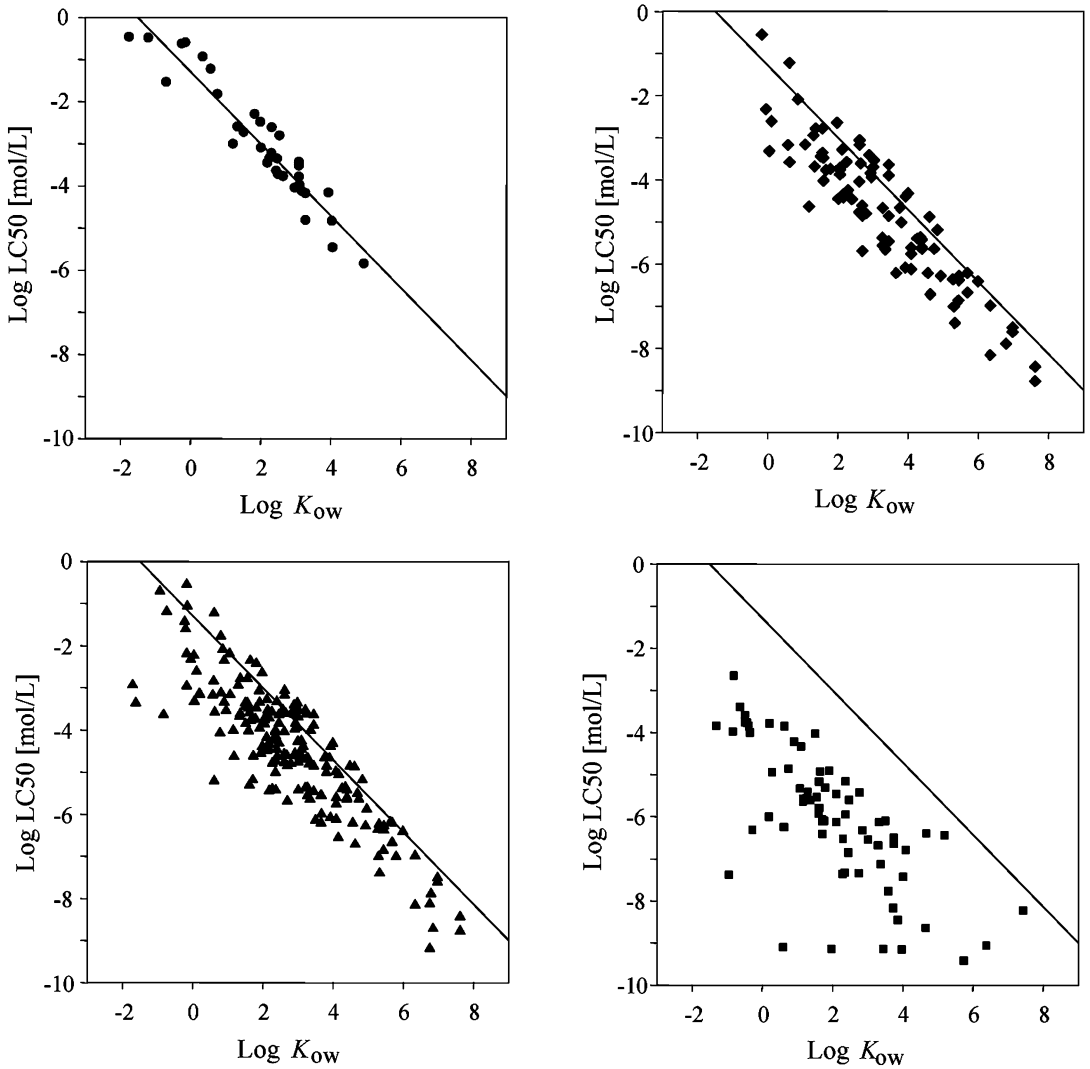



**Figure 3.** Log $LC_{50}$ (mol/L) vs log $K_{ow}$ for 36 known nonpolar narcotics (circles, top left), 91 CM1 narcotic level compounds (diamonds, top right), 193 CM3 narcotic level compounds (triangles, bottom left), and 71 chemicals predicted to exert excess toxicity according to CM3 (squares, bottom right). In all four plots, the regression line of eq 5 representing baseline toxicity is included for comparison.

has two causes: First, narcotic level compounds according to CM1 include, by definition, substances with excess toxicities ($T_e$) up to 100 to account for both data uncertainty and effect level differences between nonpolar and polar narcosis. Second, the 36 known baseline narcotics had—by intention—been excluded from the training set used for deriving CM1 (as well as CM2 and CM3) and so do not belong to the subset of 91 compounds used for the calibration of eq 7. Note also that eq 7 is very similar to eq 6. However, when applying the fish-trained 4-MOA scheme (*11*) to identify polar narcotics and leaving out anilines for the reason mentioned above, 2/3 of the compounds with polar narcotic effect levels toward daphnids are actually overlooked.

By construction, eq 7 yields an $LC_{50}$ estimate of compounds in the narcosis toxicity range, while application of eq 5 provides the minimum daphnid toxicity expected from baseline narcosis. Note further that eq 5 is in fact similar to a previously published baseline QSAR based on 17 narcotics (*26*) (baseline narcosis):

$$\log LC_{50}\,(\text{mol/L}) = -0.95 \log K_{ow} - 1.19 \qquad (8)$$

where $n = 17$, $r^2 = 0.99$, and SE = 0.21.

Because of the low CM1 sensitivity for narcotic level daphnid toxicity (cf. Tables 3 and 4), 97 compounds with experimental $LC_{50}$ values in the log $T_e$ range up to 2 are not included in eq 7. With CM2 and CM3, the respective sensitivity is significantly higher (0.962), with slightly reduced predictivities (0.873 and 0.927) as compared to CM1 (0.978).

For the subset of 193 compounds predicted as exerting narcotic level toxicities according to CM3, linear regression of log $LC_{50}$ on log $K_{ow}$ yields a CM3 narcotic level toxicity:

$$\log LC_{50}\,(\text{mol/L}) = -0.748\ (\pm 0.030) \log K_{ow} - 2.393\ (\pm 0.101) \qquad (9)$$

where $n = 193$, $r^2 = 0.76$, SE = 0.76, and $F_{1,191} = 614$. Again, the slope of the regression is similar to the baseline QSAR of eq 5, while the intercept is still lower as compared to eqs 5 and 7. As can be seen from the plot in the bottom left of Figure 3, the scatter of the data is somewhat larger due to the greater number of misclassified excess toxic chemicals, which is also reflected in the lower calibration $r^2$ and greater standard deviation of the regression statistics.

For comparison, all compounds predicted to exert excess toxicity according to CM3 are shown in the bottom right of Figure 3. Except for seven misclassified com-

**Table 5. Compounds with 48 h Toxicity to Six Daphnidae Species in Terms of Log $LC_{50}$, Log $K_{ow}$, Log $T_e$, and the Prediction Results of Three CMs[a]**

| no. | CAS | name | species | log $LC_{50}$ (mol/L) | log $K_{ow}$ | log $T_e$ | CM1 | CM2 | CM3 |
|---|---|---|---|---|---|---|---|---|---|
| | | | validation set 1 | | | | | | |
| 1 | 50293 | DDT | *D. pulex* | −8.58 | 6.79 | 1.2 | 0 | 0 | 0 |
| 4 | 55389 | fenthion | *D. pulex* | −8.35 | 4.08 | 3.5 | 1 | 3.1 | 3.1 |
| 6 | 56382 | parathion | *D. pulex* | −8.58 | 3.73 | 4.1 | 1 | 3.1 | 3.1 |
| 8 | 58899 | lindane | *D. pulex* | −5.14 | 4.26 | 0.1 | 0 | 0 | 0 |
| 14 | 62533 | aniline | *D. pulex* | −5.97 | 1.08 | 4.0 | 1 | 8.1 | 8.1 |
| 17 | 62737 | dichlorvos | *C. dubia* | −9.20 | 0.60 | 7.7 | 1 | 2.1 | 2.1 |
| 18 | 63252 | carbaryl | *D. pulex* | −7.53 | 2.35 | 4.3 | 1 | 6.1 | 6.1 |
| 20 | 72208 | endrin | *D. pulex* | −7.28 | 5.45 | 1.2 | 0 | 0 | 0 |
| 23 | 75070 | acetaldehyde | *C. dubia* | −0.88 | −0.17 | 0.1 | 0 | 0 | 0 |
| 38 | 85018 | phenanthrene | *D. pulex* | −5.20 | 4.34 | 0.1 | 0 | 0 | 0 |
| 42 | 87865 | pentachlorophenol | *D. pulex* | −5.37 | 4.74 | −0.1 | 0 | 0 | 0 |
| 49 | 90028 | salicylaldehyde | *D. pulex* | −4.36 | 2.01 | 1.5 | 0 | 0 | 0 |
| 59 | 94757 | 2,4-D | *D. pulex* | −4.84 | 2.62 | 1.4 | 0 | 0 | 0 |
| 61 | 95487 | *o*-cresol | *D. pulex* | −4.05 | 2.06 | 1.1 | 0 | 0 | 0 |
| 90 | 105373 | ethyl propionate | *D. pulex* | −3.17 | 1.36 | 0.9 | 0 | 0 | 0 |
| 92 | 105679 | 2,4-dimethylphenol | *C. dubia* | −4.43 | 2.61 | 1.0 | 0 | 0 | 0 |
| 94 | 106445 | *p*-cresol | *D. pulicaria* | −3.68 | 2.06 | 0.8 | 0 | 0 | 0 |
| 96 | 106489 | 4-chlorophenol | *C. dubia* | −4.15 | 2.16 | 1.2 | 0 | 0 | 0 |
| 100 | 107073 | 2-chloroethanol | *D. pulex* | −2.15 | 0.11 | 1.1 | 0 | 0 | 0 |
| 101 | 107119 | allylamine | *D. pulex* | −3.23 | 0.21 | 2.0 | 1 | 0 | 0 |
| 106 | 108394 | *m*-cresol | *D. pulicaria* | −3.04 | 2.06 | 0.1 | 0 | 0 | 0 |
| 110 | 108952 | phenol | *D. pulex* | −3.13 | 1.51 | 0.7 | 0 | 0 | 0 |
| 116 | 110861 | pyridine | *D. pulex* | −2.14 | 0.80 | 0.4 | 1 | 0 | 0 |
| 117 | 111422 | 2,2′-iminobisethanol | *D. pulex* | −4.64 | −1.71 | 5.3 | 1 | 0 | 0 |
| 121 | 115297 | endosulfan | *D. carinata* | −5.93 | 3.50 | 1.7 | 1 | 0 | 0 |
| 124 | 116063 | aldicarb | *D. laevis* | −6.13 | 1.36 | 3.9 | 1 | 6.1 | 6.1 |
| 132 | 121755 | malathion | *D. pulex* | −8.22 | 2.29 | 5.1 | 1 | 3.1 | 3.1 |
| 134 | 122145 | fenitrothion | *M. macrocopa* | −6.85 | 3.30 | 2.8 | 1 | 3.1 | 3.1 |
| 135 | 122349 | simazine | *D. pulex* | −3.00 | 2.40 | −0.2 | 1 | 0 | 0 |
| 137 | 123546 | 2,4-pentanedione | *D. pulex* | −3.30 | 0.05 | 2.3 | 0 | 0 | 0 |
| 145 | 141786 | ethyl acetate | *D. pulex* | −2.53 | 0.86 | 0.8 | 0 | 0 | 0 |
| 152 | 206440 | fluoranthene | *C. dubia* | −6.65 | 4.93 | 1.0 | 0 | 0 | 0 |
| 154 | 298000 | methyl parathion | *C. dubia* | −7.94 | 2.75 | 4.4 | 1 | 3.1 | 3.1 |
| 157 | 333415 | diazinon | *D. pulex* | −8.59 | 3.86 | 4.0 | 1 | 3.1 | 3.1 |
| 158 | 470906 | chlorfenvinfos | *C. dubia* | −8.95 | 4.15 | 4.1 | 1 | 0 | 0 |
| 175 | 609198 | 3,4,5-trichlorophenol | *C. dubia* | −5.70 | 3.45 | 1.5 | 0 | 0 | 0 |
| 195 | 1014706 | simetryn | *M. macrocopa* | −3.82 | 2.90 | 0.1 | 1 | 0 | 0 |
| 199 | 1563662 | carbofuran | *D. pulex* | −6.74 | 2.30 | 3.6 | 1 | 6.1 | 6.1 |
| 202 | 1582098 | trifluralin | *D. pulex* | −6.14 | 5.31 | 0.2 | 1 | 0 | 0 |
| 206 | 1912249 | atrazine | *D. pulex* | −3.72 | 2.82 | 0.1 | 1 | 0 | 0 |
| 218 | 2668248 | 2-methoxy-4,5,6-trichlorophenol | *C. dubia* | −5.10 | 3.27 | 1.0 | 0 | 0 | 0 |
| 222 | 2921882 | clorpyrifos | *D. pulex* | −8.48 | 4.66 | 3.1 | 1 | 3.1 | 3.1 |
| 227 | 3766812 | 2-(1-methylpropyl)phenol, methylcarbamate | *M. macrocopa* | −6.32 | 2.86 | 2.7 | 1 | 6.1 | 6.1 |
| 241 | 25167833 | tetrachlorophenol | *D. pulex* | −5.36 | 4.09 | 0.5 | 0 | 0 | 0 |
| 243 | 28249776 | thiobencarb | *C. dubia* | −5.70 | 3.90 | 1.1 | 1 | 0 | 0 |
| 256 | 52645531 | permethrin | *D. pulex* | −7.70 | 7.43 | −0.3 | 1 | 2.1 | 2.1 |
| 264 | 91465086 | cyhalothrin | *C. dubia* | −9.18 | 6.85 | 1.7 | 1 | 0 | 0 |
| | | | validation set 1 continued/validation set 2 | | | | | | |
| 301 | 51036 | piperonyl butoxide | *C. dubia* | −5.71 | 4.29 | 0.7 | 0 | 0 | 0 |
| 302 | 57556 | propylene glycol | *C. dubia* | −0.90 | −0.78 | 0.7 | 0 | 0 | 0 |
| 303 | 72435 | methoxychlor | *C. dubia* | −7.39 | 5.67 | 1.1 | 0 | 0 | 0 |
| 304 | 76448 | heptachlor | *D. pulex* | −6.95 | 5.86 | 0.4 | 0 | 0 | 0 |
| 305 | 88062 | dowicide 2S | *C. dubia* | −4.69 | 3.45 | 0.5 | 0 | 0 | 0 |
| 306 | 93721 | silvex | *D. pulex* | −5.05 | 3.68 | 0.6 | 0 | 0 | 0 |
| 307 | 108463 | resorcinol | *D. pulicaria* | −3.04 | 1.03 | 1.1 | 0 | 0 | 0 |
| 308 | 145733 | endothall | *C. dubia* | −3.59 | 1.89 | 0.8 | 0 | 0 | 0 |
| 309 | 330541 | diuron | *D. pulex* | −5.22 | 2.67 | 1.7 | 1 | 0 | 0 |
| 310 | 709988 | propanil | *C. dubia* | −4.75 | 2.88 | 1.1 | 1 | 0 | 0 |
| 311 | 959988 | alpha-endosulfan | *D. carinata* | −6.21 | 3.50 | 1.9 | 1 | 0 | 0 |
| 312 | 1031078 | endosulfan sulfate | *D. carinata* | −5.75 | 3.64 | 1.3 | 1 | 0 | 0 |
| 313 | 1194656 | dichlobenil | *D. pulex* | −4.67 | 2.83 | 1.0 | 1 | 0 | 0 |
| 314 | 1563388 | carbofuran phenol | *C. dubia* | −6.01 | 2.90 | 2.3 | 0 | 0 | 0 |
| 315 | 1646873 | 2-methyl-2-(methylsulfinyl)-propion-aldehyde, O-(methyl-carbamoyl)oxime | *D. laevis* | −6.34 | −0.78 | 6.1 | 1 | 6.1 | 6.1 |
| 316 | 1646884 | aldoxycarb | *D. laevis* | −5.34 | −0.67 | 5.0 | 1 | 6.1 | 6.1 |
| 317 | 1836755 | nitrofen | *C. dubia* | −6.12 | 4.32 | 1.1 | 1 | 0 | 0 |
| 318 | 2212671 | molinate | *C. dubia* | −4.83 | 2.91 | 1.1 | 1 | 0 | 0 |
| 319 | 7786347 | mevinphos | *C. dubia* | −8.37 | −0.24 | 7.6 | 1 | 1.1 | 1.1 |
| 320 | 8001352 | toxaphene | *D. pulex* | −7.44 | 6.79 | 0.1 | 0 | 0 | 0 |

**Table 5 (Continued)**

| no. | CAS | name | species | log $LC_{50}$ (mol/L) | log $K_{ow}$ | log $T_e$ | CM1 | CM2 | CM3 |
|---|---|---|---|---|---|---|---|---|---|
| validation set 1 continued/validation set 2 | | | | | | | | | |
| 321 | 8003347 | pyrethrum | *D. pulex* | −7.12 | 6.15 | 0.3 | 0 | 0 | 0 |
| 322 | 15972608 | lasso | *D. pulex* | −4.44 | 3.37 | 0.3 | 1 | 0 | 0 |
| 323 | 19666309 | oxadiazon | *M. macrocopa* | −5.80 | 4.81 | 0.3 | 1 | 0 | 0 |
| 324 | 21087649 | metribuzin | *C. dubia* | −3.78 | 1.49 | 1.4 | 1 | 0 | 0 |
| 325 | 33213659 | beta-endosulfan | *D. carinata* | −6.30 | 3.50 | 1.9 | 1 | 0 | 0 |
| 326 | 51218452 | metolachlor | *C. dubia* | −4.25 | 3.24 | 0.2 | 1 | 0 | 0 |
| 327 | 95737681 | pyriproxyfen | *D. carinata* | −6.60 | 5.55 | 0.4 | 1 | 0 | 0 |

[a] $LC_{50}$ denotes the experimental lethal concentration of 50% toward six daphnidae species within 48 h (*23*), $K_{ow}$ is the calculated octanol/water partition coefficient (*28*), and $T_e$ is the excess toxicity (eq 1). CM1, CM2, and CM3 are the presently developed CMs. NA, not assigned.

pounds (cf. Tables 1 and 3), all other substances have $T_e$ values above 100, corresponding to graphical locations in the lower left triangle of the log $LC_{50}$ vs log $K_{ow}$ plot.

**Validation.** Because the SAs have been derived through visual inspection of the chemical structures and $T_e$ values of all compounds of the training set, internal validation procedures such as cross-validation do not apply. Moreover, for the training set, all suitable *D. magna* $LC_{50}$ entries from the AQUIRE database had been included, such that the collection of an additional compound set with respective toxicity values does not appear to be feasible at this point in time.

To still perform some kind of validation, all AQUIRE entries referring to other species of the daphniidae family were collected and subjected to the same selection and quality criteria (only uniquely defined organic chemicals without metals, mean values without apparent outliers, and water solubility cutoff) as with *D. magna* (Table 5). In this way, an initial validation set of 74 compounds with $LC_{50}$ values covering six different daphnid species (*Daphnia pulex*, 36 values; *Ceriodaphnia dubia*, 23 values; *Daphnia carinata*, five values; *Moina macrocopa*, four values; *Daphnia laevis*, three values; and *Daphnia pulicaria*, three values) was constructed (validation set 1). However, 47 of the respective compounds belong also to the *D. magna* training set, leaving a final validation set of 27 compounds (validation set 2). Only the latter offers an external validation of the presently derived CMs, while the former allows one to evaluate the possibility of extrapolating the $T_e$-based classification concept to other daphnid species.

In the upper part of Table 6, the contingency table statistics for the extrapolation across species are summarized using validation set 1 (74 compounds). The statistical performances are similar to the ones achieved with the *D. magna* training set (Table 4), indicating that the $T_e$ criterion applied to discriminate narcotic effect levels from excess toxicity can well be extrapolated across different daphnid species. Because none of the CM3 specific tentative rules were applicable, the validation results of CM2 and CM3 are identical. With regard to the 27 compounds outside the *D. magna* training set (validation set 2), the CM1 prediction power for narcotic level toxicities is similar to the corresponding training result, while the CM1 validation performance for the excess toxicity is inferior. With CM2, the validation statistics are overall of similar quality as the training statistics.

**Table 6. External Validation of CM1, CM2, and CM3 Using $LC_{50}$ Values of Six Daphnidae Species Other than *D. magna*[a]**

| model | concordance | category | statistical evaluation: sensitivity | predictivity |
|---|---|---|---|---|
| validation set 1 (74 compounds) | | | | |
| CM1 | 0.662 | narcotic effect level | 0.574 | 0.940 |
| | | excess toxicity | 0.900 | 0.439 |
| CM2/CM3 | 0.932 | narcotic effect level | 0.981 | 0.930 |
| | | excess toxicity | 0.800 | 0.941 |
| validation set 2 (27 compounds) | | | | |
| CM1 | 0.481 | narcotic effect level | 0.435 | 0.909 |
| | | excess toxicity | 0.750 | 0.188 |
| CM2/CM3 | 0.963 | narcotic effect level | 1.000 | 0.958 |
| | | excess toxicity | 0.750 | 1.000 |

[a] Validation set 1 covers the following six daphnidae species (and the associated number of compounds): *Daphnia pulex* (36), *Ceriodaphnia dubia* (23), *Daphnia carinata* (5), *Moina macrocopa* (4), *Daphnia laevis* (3), and *Daphnia pulicaria* (3). However, only 27 of the 74 compounds do not belong to the *D. magna* training set and form validation set 2.

## Discussion

The intention of this work was to derive structural rules as a tool for a first-tier risk assessment that allows for discrimination of compounds with narcotic effect levels from those that are likely to exert excess toxicity in the acute daphnid test. Baseline toxicity QSARs allow one to estimate, with reasonable accuracy, the aquatic toxicity of narcotics. It follows that such compounds would have little priority for experimental testing. Thus, the ability to identify—directly from chemical structure—compounds that are likely to be toxic only in the narcotic range would offer a possibility to reduce the need for experimental testing and thus provide an attractive component of a tiered chemical hazard classification scheme.

**Excess Toxic and Narcotic Level Chemicals.** In the presently analyzed data set, 36 of the 78 compounds with log $T_e$ values above 2 are pesticides, and nine of the 10 compounds with a log $T_e$ above 4 are also pesticides. At the same time, 30 pesticides have log $T_e$ values below 2, and 17 have values even below 1. It demonstrates that the knowledge about whether a given compound is a pesticide would not be a reliable predictor of excess toxicity, although highly toxic chemicals have a high probability of being used as pesticides.

Among the 91 compounds with $T_e$ values in the range of 1−10 are 23 hydrocarbons, 11 phenols, one aniline, 17 nitroaromatics, six urea and thiourea derivatives, three esters, and three organophosphorus compounds. Note that in a previous MOA classification study, the fish toxicity of nitroaromatics was allocated to a reactive MOA with a likely enhanced toxicity (*14*), while with algae both narcotic type and excess toxicity were observed and related to electronic structure characteristics of the compounds (*35*). In the $T_e$ range 10−100 are four

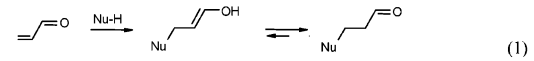
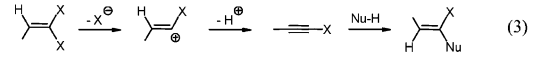
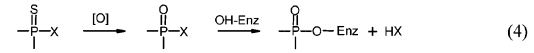
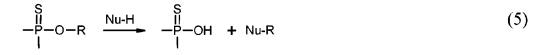
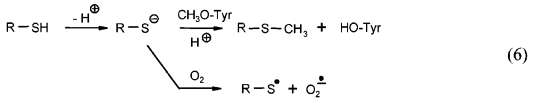
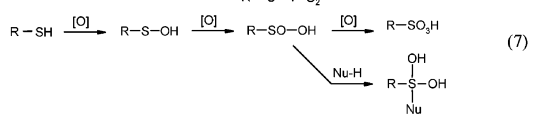
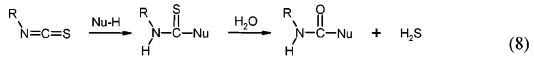
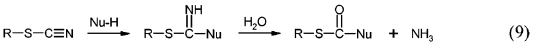
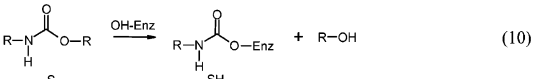
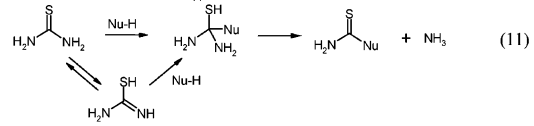
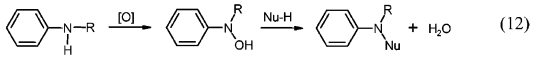
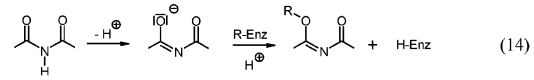
hydrocarbons, 12 phenols, seven anilines, three urea and thiourea compounds, and three isothiocyanates. Monofunctional compounds with $T_e$ values above 100 include 11 anilines but only one phenol. This distribution of chemical classes across effect levels shows that a classification scheme based on simple compound classes would not perform well in discriminating between narcotic level and excess toxicity toward daphnids.

Similarly, the focus on previously published lists of electrophilic structural features (*12*, *22*) would not yield a reliable identification of excess toxic chemicals with respect to daphnids. On one hand, well-known electrophiles such as the Michael type acceptors acrylamid (no. 10), acrolein (no. 98), and acrylonitrile (no. 102) as well as the $S_NAr$ sensitive 2,4-dinitro-1-chlorobenzene (no. 70) have $T_e$ values above 100. On the other hand, however, a variety of electrophiles classified according to the fish-trained 7-MOA scheme (*12*) or according to the 2-MOA scheme (*22*) are associated with the $T_e$ range of 1−100, and a still substantial number of electrophiles have $T_e$ values in the range of 1−10: When applying the 7-MOA scheme, 15 electrophiles have $T_e$ values below 10, and 13 electrophiles have $T_e$ values in the range of 10−100. With the 2-MOA scheme, 17 electrophiles have $T_e$ values below 100, and seven of them have values below 10. These findings show that existing schemes to identify electrophiles do not provide a good discrimination between excess toxic and narcotic level compounds.

Allylamine (no. 101) is an example of a total of 20 compounds with log $T_e$ values in the range of 1.5−2.0 that could be considered as a gray zone between narcotic level and excess toxicity. With mammals, allylamine is highly toxic to the heart, which is traced back to a metabolic activation by monoamine oxidase to yield acrolein (*36*). However, direct daphnid exposure to acrolein results in a much higher toxicity with a log $T_e$ value of 4.55. One possible reason for this discrepancy could be that the strong basicity of allylamine makes this compound significantly less bioavailable for the aqueous exposure pathway (through formation of the ammonium form prevalent under neutral pH conditions). In any case, allylamine belongs to the narcotic range as defined in this study and is correctly classified through CM2 and CM3 as well as through the 2-MOA scheme, while CM1 and the 7-MOA scheme predict this compound to be excess toxic, and the 4-MOA scheme does not offer a classification due to missing structural features. Interestingly, three of the seven compounds misclassified by CM2 and CM3 as excess toxic (cf. Table 3) have log $T_e$ values in this gray zone range of 1.5−2.0.

It should be noted, however, that there are also classes of electrophiles with only little representation in the presently analyzed set of 300 compounds. In the AQUIRE database, only two epoxides (nos. 26 and 68) were found that both have daphnid $LC_{50}$ values in the narcotic level (with log $T_e$ values of 1.08 and 1.37, respectively). Another example is given by the $S_N2$ reactants epichlorohydrin (no. 97) and 2-chloroethanol (no. 100), where again the daphnid toxicity is in the narcotic range for both compounds (log $T_e$ values of 1.76 and 1.24, respectively). It follows that a more comprehensive analysis of such underrepresented electrophilic structures cannot be undertaken at this stage but requires additional test data, as is similarly the case for existing fish-trained MOA classification schemes. A further aspect of interest in future studies may be how the presently selected $T_e$ value of 100 as a cutoff between the narcotic range (including both data uncertainty and the systematic difference between nonpolar and polar narcosis as outlined above) and the excess toxicity would perform as compared to alternative approaches.

**Table 7. Reaction Mechanisms Associated with the Nine SAs**

| Structural Alert | Reaction Mechanism | Reaction No. |
|---|---|---|
| SA1 |  | (1) |
| |  | (2) |
| SA2 |  | (3) |
| SA3 |  | (4) |
| |  | (5) |
| SA4 |  | (6) |
| |  | (7) |
| SA5 |  | (8) |
| |  | (9) |
| SA6 |  | (10) |
| SA7 |  | (11) |
| SA8 |  | (12) |
| SA9 |  | (13) |
| |  | (14) |

**Chemical Reactivity and Metabolic Potential of SAs.** The SAs of CM1 (SA1 and SA2), CM2 (SA1−SA9), and CM3 (SA1, SA2, SA3*, SA4, SA5*−SA7*, SA8, and SA9) are used as indicators for the potential of chemical structures to exert excess toxicity in the acute daphnid toxicity test. While these SAs have been identified empirically through visual inspection of the chemical structures and $T_e$ values of all training set compounds, mechanistic reasoning provides possible explanations for the correlation between the presence of individual SAs and the observed enhanced toxicity of the chemicals.

In Table 7, possible biotransformation reactions associated with individual SAs are listed that could explain why compounds containing these structural features are more toxic than narcotic level chemicals. The underlying reaction mechanisms are based on standard considerations about the reactivity associated with functional groups as discussed in organic chemistry textbooks. Some also include known metabolic processes such as the mechanism of acetylcholine esterase inhibition through certain classes of insecticides. Note further that the

reaction products listed in Table 7 may well undergo subsequent biotransformations, which are not considered further.

The chemical reactivity associated with SA1 is illustrated for the case of α,β-unsaturated carbonyl compounds. Because of the electronegative carbonyl oxygen, the β carbon is electron deficient and thus susceptible for an attack by endogeneous nucleophiles (Nu-H). The resultant conjugated 1,4-addition leads to an enol, which is likely to be tautomerized to the final carbonyl compound (reaction 1). A corresponding attack can also take place at the (somewhat less) electrophilic α carbon, yielding an allyl alcohol as a 1,2-adduct (reaction 2). Respective transformations may also occur for α,β-unsaturated nitrile compounds, which form a related class of xenobiotic electrophiles.

SA2 has been mainly found in pyrethroids that block the gate of the sodium channel protein and therefore affect the balance of nerve membranes (*37*). Apart from this toxicological route, it may also be hypothesized that due to the strong electron-attracting effect of halogen substituents in geminal alkene halogenides, two-step elimination via the vinyl cation as an intermediate may lead to alkynes, which are substantially more reactive toward the addition of endogeneous nucleophiles (reaction 3) than their alkene counterparts.

Organic phosphorothionates containing SA3 are known as potent acetylcholine esterase (AChE) inhibitors (*38*, *39*). Initial monooxygenase-mediated oxidation yields the bioactive phosphate, which is sufficiently reactive to attack the nucleophilic hydroxyl oxygen of the AChE serine group, resulting in a phosphorylated and thus deactivated AChE. In addition, an $S_N2$ attack of endogeneous nucleophiles (Nu-H) at alkoxy ester functions of correspondingly substituted phosphorothionates may take place, leading to alkylated derivatives (Nu-R; reaction 5 in Table 7).

Aliphatic thiols (SA4) are more acidic than corresponding alcohols by around 7 orders of magnitude, and the thiolate anions are among the strongest nucleophiles. Thus, a possible reaction path could be started by an initial deprotonation that yields the thiolate anion, which in turn is sufficiently reactive to dealkylate alkoxy functionalities such as the methoxy group of tyrosine ($CH_3O$-Tyr), one of the aromatic amino acid side chains (reaction 6). Here, a further toxicologically relevant sideway is the reaction of the thiolate anion with intracellular oxygen, leading to the formation of superoxoide anion and subsequently further reactive oxygen species. Another possible route is the stepwise oxidation via monooxygenases until the formation of sulfonic acid derivatives. Here, the intermediate sulfinic acid (R−SO−OH) should be quite reactive due to the electron deficient character of the sulfur atom and may thus interfere with endogeneous nucleophiles (reaction 7). Note that with regard to the acute toxicity toward fish, thiols have been classified as baseline narcotics (*12*, *14*), which contrasts with the clear excess toxicity in the acute daphnid test according to the present data set.

Isothiocyanates and thiocyanates (SA5) are electrophilic at the central $sp^2$ and $sp^3$ carbons, respectively, enabling the addition of endogeneous nucleophiles, possibly followed by hydrolysis (reactions 8 and 9). Carbamates (SA6) are another class of insecticides known as AChE inhibitors (*38*). In contrast to phosphorothionates, however, an oxidative activation step is not needed, and attack of the carbonyl carbon at the serine OH group yields the carbamylated enzyme (reaction 10). Thiourea derivatives (SA7) also contain an electrophilic carbonyl carbon atom that may attack electron-rich sites of biological macromolecules (Nu-H), leading to respective adducts that may further undergo desamination (reaction 11). Here, the tautomeric form, an isothiourea compound, would yield the same reaction product.

Aromatic amines (SA8) form a special case with regard to their acute toxicity toward daphnids, although the mechanistic origin has not been disclosed so far. While anilines act as polar narcotics toward fish under acute exposure regimes, certain substitution patterns of the aromatically bound amino group lead to a considerably enhanced toxicity toward *D. magna* and other waterfleas (*32*). Interestingly, prolonged exposure of rainbow trout (*Onchorhyncus mykiss*) and medaka (*Oryzia latipes*) toward aniline and 4-chloroaniline results in metabolic conversion and effects different from polar narcosis (*40*, *41*), which indicates that the MOA of aromatic amines in fish depends also on the duration of exposure. In line with previous findings that ortho substitution reduces the aniline excess toxicity toward daphnids (*32*), SA8 is confined to derivatives without ortho substituents. Possible metabolic transformations include the formation of the highly toxic hydroxylamine, which may attack endogeneous nucleophiles through an addition−elimination reaction (reaction 12).

Imides (SA9) are derivatives of carbonyl compounds. Although the carbonyl carbon atoms are somewhat reduced in their electrophilic reactivity as compared to the ones in aldehydes, ketones, and esters, the imide could still form adducts with electron-rich sites of biological macromolecules (Nu-H; reaction 13). Because imides are relatively acidic, another possible metabolic route could be given by an initial deprotonation, followed by a dealkylation of enzymes (R-Enz), membrane proteins, or DNA side chains through the strong nucleophile formed intermediately (reaction 14).

As mentioned above, the transformation reactions summarized in Table 7 provide possible pathways that may explain the enhanced toxicity of compounds containing the relevant structural features. In some cases such as with the thiols and imides, however, the proposed reactions are based only on principal reactivity considerations, without experimental biochemical or toxicological evidence except that such chemical structures are apparently associated with enhanced acute toxicities toward daphnids.

**Statistical Performance of CM1, CM2, and CM3.** Among our newly introduced classification schemes, CM1 is particularly strong in identifying excess toxicity (recognition power) as well as in predicting narcotic effect levels (prediction power). It follows that for a given compound, the CM1 prediction of a narcotic effect level in the acute daphnid test has a high probability to hold true. At the same time, the CM1 criterion for excess toxicity is likely to be less reliable, keeping in mind that a substantial portion of the respectively classified compounds may in fact turn out to exert effect levels in the range of narcosis.

The overall best-performing scheme is CM3, which however contains eight structural rules that at this stage are only tentative. These tentative rules require additional experimental data to prove or disprove their statistical significance. It follows that for the time being,

CM3 should not be used alone but only in the context of further information. Nonetheless, most of the tentative rules can be traced back to distinct biotransformation reactions as outlined above and thus can be considered as mechanistic hypotheses about the association of certain structural features with the occurrence of excess toxicity.

CM2 is confined to statistically significant rules and shows an overall performance between CM1 and CM3. While the CM2 prediction power with respect to narcotic effect levels is lower than the one of CM1 (0.873 vs 0.978; Table 4), the corresponding recognition power is much better than with CM1 (0.962 vs 0.478). In absolute numbers, with CM2, only seven of the 186 compounds with narcotic effect levels were overlooked to belong to this class.

With regard to excess toxicity, the CM2 prediction power is again twice as good as with CM1 (0.881 vs 0.439), but at the same time, the respective recognition power is significantly inferior to the CM1 result (0.666 vs 0.974). The results suggest that a properly combined use of CM1 (optimized to predict narcotic effect levels) and CM2 or CM3 (optimized to predict excess toxicity) is the current method of choice to screen organic compounds for their priority to undergo an experimental test of their acute daphnid toxicity.

**Combined Two-Step CM Approach.** Using the specific strengths of the CMs CM1 and CM2/CM3, a combined two-step approach can be applied that yields high prediction rates for both compounds with narcotic effect levels and compounds exerting excess toxicity. In a first step, CM1 is used to predict narcotic effect levels as well as excess toxicity. At this stage, only the category of narcotic effect levels is used, for which CM1 is relatively conservative and was shown to have a high prediction power. The correspondingly identified compounds have a high probability of not exceeding the narcotic range in the acute daphnid test and thus have a low priority for experimental testing. The latter results from the fact that baseline narcosis can be relatively well-predicted through $K_{ow}$-based QSARs.

Second, CM2 (or CM3) is applied to the prediction of both narcotic range compounds and compounds exerting excess toxicity. In general, there will be a subset of compounds that should exert a narcotic effect level according to CM2 (or CM3) but at the same time should be excess toxic according to CM1. This subset is allocated an intermediate priority for experimental testing.

Finally, the remaining subset where both CM1 and CM2 (or CM3) agree consists of compounds with the highest probability of exerting excess toxicity. For these compounds, reasonable QSAR-based predictions are generally not available (except possibly for structurally related groups based on specific knowledge), and as a consequence, the priority for undertaking an experimental test to determine the acute daphnid toxicity is high.

The approach is visualized in Figure 4, where the three corresponding subgroups are ordered according to their predicted priority for experimental testing. With this two-step procedure, 91 compounds are predicted (by both CM1 and CM2) to exert narcosis level toxicity, of which 89 actually belong to this class (predictivity = 0.978; see respective CM1 entry in Table 4). These 91 compounds would be allocated lowest test priority when applying the classification in a predictive mode.

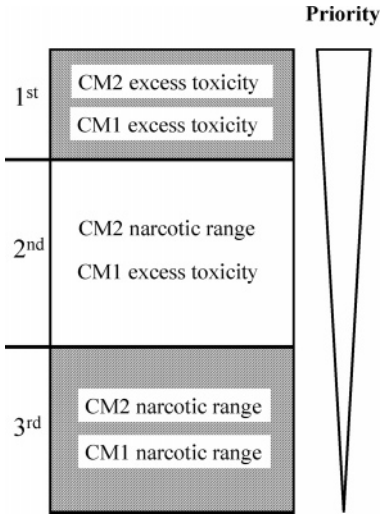



**Figure 4.** Two-step classification approach combining CM1 and CM2. Compounds predicted to exert narcotic effect levels according to CM1 are likely to have corresponding $LC_{50}$ values and, thus, have a low priority for experimental testing. By contrast, compounds predicted by both CM1 and CM2 to exert excess toxicity have a high priority for experimental testing, because their effect level cannot be estimated reasonably well from QSARs. Compounds with conflicting predictions according to CM1 and CM2 are allocated an intermediate priority for experimental testing. Note that in this two-step scheme, CM2 may also be replaced by CM3.

For 59 compounds, both CM1 and CM2 agree in predicting excess toxicity in the sense of $T_e > 100$, of which 52 compounds are actually excess toxic (predictivity = 0.881; see respective CM2 entry in Table 4). This subset of 59 compounds would thus be allocated with the highest test priority, because their toxicity is expected to exceed the baseline toxicity at least by a factor of 100 and is generally difficult to be predicted quantitatively from existing QSAR models.

For the remainder of 114 compounds, CM1 and CM2 yield conflicting prediction results. This is mainly caused by the low predictivity of CM1 for excess toxic compounds (0.439, see Table 4). As noted above, CM1 is biased toward a high predictivity for the narcotic level toxicity of organic compounds and, in this respect, is superior to CM2 and CM3 (cf. Tables 3 and 4). Ninety of the 114 compounds have $LC_{50}$ values in the narcosis range in agreement with the respective CM2 prediction, while 24 compounds are in fact excess toxic (as predicted by CM1). Within the two-step approach as outlined in Figure 4, this intermediate class would thus be allocated an intermediate priority for experimental testing.

**Statistical Performance of MOA-Based CMs.** For the present training set of 264 compounds, the 4-MOA CM (*11*) could be applied to only half of the compounds. It indicates that the applicability domain of this classification scheme is somewhat restricted. Moreover, the overall concordance is only moderate (0.604, Table 4) and in fact inferior to all other CMs except the 7-MOA CM (0.598), the latter of which refers, however, to a much greater number of compounds (264 vs 139, cf. Table 3).

In view of the above-mentioned systematic difference between the daphnid and the fish toxicity of anilines, one might suspect that the only moderate performance of the fish-trained CMs could be driven mainly by misclassifications for this compound class. However, leaving out the

22 anilines results in only moderately improved overall concordances of 0.642 (4-MOA CM), 0.616 (7-MOA CM), and 0.715 (2-MOA CM), with still very low predictivities for excess toxicity (0.373, 0.351, and 0.433; see Table 4). Only for the 4-MOA CM, the sensitivity for excess toxicity as well as the predictivity for narcotic effect levels are now much better than before (0.926 vs 0.694 and 0.964 vs 0.843, respectively; cf. Table 4), while the predictivity for excess toxicity and the sensitivity for narcotic level toxicity are essentially unchanged (0.563 vs 0.573 and 0.373 vs 0.362, respectively). Note further than when restricting the chemical domain to the subset of 123 compounds (without anilines) that can be classified by the 4-MOA scheme, the overall concordances of CM2 and CM3 increase from 0.875 (Table 4) to 0.911 and from 0.920 (Table 4) to 0.943, respectively.

Another group with an apparently systematic difference in the toxic level between daphnids and fish are the thiols. All three compounds (nos. 24, 99, and 224) are clearly excess toxic toward daphnids with $T_e$ values above 100, in contrast to their classification as narcotics according to the fish-based 7-MOA scheme (*12*) and a separate MOA classification study (*14*). Note, however, that the 4-MOA scheme has no classification rule for this functional group and that according to the 2-MOA scheme all three thiols would have been classified as excess toxic.

A further interesting case is given by the only two aldehydes present in the data set under investigation. Acetaldehyde (no. 23) is even less toxic than according to baseline narcosis (log $T_e$, −0.59), suggesting that the metabolic oxidation via aldehyde dehydrogenase to acetic acid acts as a quite efficient detoxification pathway. Although salicylaldehyde (no. 49) shows a moderately elevated toxicity (log $T_e$, 1.44), its daphnid $LC_{50}$ is still in the narcotic range, and both aldehydes are classified accordingly by CM1, CM2, and CM3 (cf. Table 1). By contrast, application of the 4-MOA scheme would classify both compounds as excess toxic, while according to both the 7-MOA scheme and the 2-MOA scheme only acetaldehyde would be predicted to be excess toxic. Here, the 7-MOA scheme would allocate salicylaldehyde to the group of polar narcotics (cf. Table 1).

The generally only moderate statistical performance of the MOA-based CMs for predicting the daphnid toxicity level has probably two causes: First, reactive and specific MOAs need not result in significant excess toxicities. This means, however, that MOA-based classification schemes are not necessarily suited to provide good testing priorities in the context of tiered chemical hazard evaluation schemes that include QSAR predictions for narcotic effect levels as potential or preliminary alternatives to experimental results.

Second, two of the three classification schemes were derived using acute fish toxicity data (actually $LC_{50}$ toward guppy, *Poecilia reticulata*, for the 4-MOA CM, and 96 h $LC_{50}$ data toward fathead minnow, *Pimephales promelas*, for the 7-MOA CM). It suggests that except for simple narcotics, the scope for extrapolating knowledge about prevalent MOAs across trophic levels appears to be limited.

The 2-MOA CM (*22*) was developed to identify electrophilic and proelectrophilic substructures as indicators for the likely occurrence of excess toxicity of the respective compounds toward fish and other aquatic organisms and has a strong basis on chemical reaction mechanisms. In the present context of acute daphnid toxicity, however, both its recognition and its prediction power for excess toxic chemicals are surprisingly poor (0.167 and 0.433; Table 4). Among the 30 compounds predicted to show excess toxicity, only 13 actually belong to this class, and 65 of the 78 chemicals with experimental excess toxicities would be (implicitly) classified as narcotic effect level compounds (Table 3). These results suggest that the applicability domain of the 2-MOA CM in terms of actually covered electrophilic and proelectrophilic substructures was too restricted for the presently analyzed set of 264 compounds. When comparing the 2-MOA CM with the presently derived CM2 and CM3, the former contains SAs identical or similar to SA1, SA4, and SA5.

**Excess Toxicity and Specific Modes of Toxic Action.** For the acute daphnid toxicity, the currently developed classification schemes allow discrimination of chemicals exerting excess toxicity from compounds with narcotic effect levels. The respective distinction is based on a $T_e$ criterion of 100: Only $LC_{50}$ values with a $T_e >$ 100 are classified as excess toxic, to account for both data uncertainties and systematic effect level differences between nonpolar and polar narcosis when employing the $K_{ow}$ scale of hydrophobicity. As a consequence, specifically acting compounds with $T_e$ values below 100 would also be classified as narcotic level, in line with the goal to sort out those compounds where the acute daphnid toxicity is within 2 orders of magnitude from baseline narcosis, the latter of which can be predicted pretty well from respective QSARs.

In the daphnid test, oxidative uncouplers are one group of specifically acting compounds with $T_e$ values below 100. Inspection of Table 1 reveals that according to the 7-MOA CM (*12*), the following nine compounds (eight phenols and one aniline) are classified as oxidative uncouplers: 2,4-dinitrophenol (no. 2), 2,3,4,6-tetrachlorophenol (no. 9), pentachlorophenol (no. 42), 2-(1-methylpropyl)-4,6-dinitrophenol (no. 45), 2,4,6-trinitrophenol (no. 46), dinitro-*o*-cresol (no. 162), 2,3,4,5-tetrachloroaniline (no. 184), 2,3,5,6-tetrachlorophenol (no. 193), and 2,4,6-trinitro-1,3-benzenediol (no. 236). Moreover, 2-methoxy-tetrachlorophenol (no. 216) and 2,3,4,5-tetrachlorophenol (no. 241) are predicted to exert different specific MOAs including oxidative uncoupling. Interestingly, all of these mostly phenolic uncouplers have $T_e$ values below 100 including 2,4-dinitrophenol and pentachlorophenol as well-known reference compounds for this MOA, and thus belong to the narcotic toxicity range.

For the five uncouplers pentachlorophenol, 2,4,6-trinitrophenol, 2,3,5,6-tetrachlorophenol, 2,3,4,5-tetrachlorophenol (that would exert both polar narcosis and oxidative uncoupling according to the 7-MOA scheme), and 2,4,6-trinitro-1,3-benzenediol, the $T_e$ values are even below 10 and thus in the effect level range of baseline and polar narcotics. It demonstrates that the $T_e$ criterion is clearly not strict in terms of distinguishing between narcosis and specific or reactive MOAs but straightforward in its focus on the effect level as compared to baseline toxicity.

For the subset of 11 oxidative uncouplers, the mean log $T_e$ is 1.08 with a standard deviation of 0.57, indicating that their toxicity is on the average a factor of 10 greater than baseline toxicity. Moreover, regression of log $LC_{50}$ on log $K_{ow}$ yields

$$\log LC_{50}\ (\text{mol/L}) = -0.879\ (\pm 0.149) \log K_{ow} - 2.287\ (\pm 0.505) \quad (10)$$

($n = 11$, $r^2 = 0.77$, SE = 0.60, and $F_{1,9} = 34.8$). As compared to the baseline QSAR (eq 5), the intercept is lowered by one log unit (−2.287 vs −1.281), and the slope is similar (−0.879 vs −0.857). The moderate statistics reflect the somewhat greater scatter within the range of two log units above baseline toxicity. It follows that eq 10 can be used as a reasonable estimate for the acute daphnid toxicity of (phenolic) oxidative uncouplers, with expected $LC_{50}$ values within two log units distant from baseline narcosis.

Interestingly, the log $K_{ow}$-based regression relationship for oxidative uncoupling in the fish *Pimephales promelas* (*12*) has a quite different slope (−0.67) and intercept (−2.95) as compared to eq 10 that refers to daphnids and is in fact almost parallel to the corresponding fish polar narcosis QSAR (*12*, *33*) with a slope of −0.65 and an intercept of −2.29. The latter indicates that with this fish species, $LC_{50}$ predictions for oxidative uncouplers would be consistently ca. 0.65 log units below the ones according to polar narcosis.

Nonetheless, the mean log $T_e$ of all 12 oxidative uncouplers of the Duluth database (*12*) is only 0.82 with a standard deviation of 0.46 and, thus, even slightly closer to fish baseline narcosis as was found for the comparison of uncouplers and narcotics in the daphnid toxicity test. It follows that even when employing a $T_e$ value of 10 as criterion to distinguish excess toxicity from the narcotic range, most of these uncouplers would have been classified as belonging to the group of narcotic type chemicals. Moreover, in the ciliate assay with *Tetrahymena pyriformis* (*15*), 19 oxidative uncouplers have a mean log $T_e$ of 1.29 with a standard deviation of 0.43. From these findings with three different aquatic organisms, it may be concluded that oxidative uncouplers appear to exert an only moderate excess toxicity toward aquatic organisms that is in fact still in the range expected for the combined group of nonpolar and polar narcotics.

## Conclusions

In the context of tiered chemical hazard assessment schemes, QSARs enable prediction of narcotic effect levels of organic compounds with reasonable accuracy, thus reducing the need for experimental testing. To this end, however, tools are required that allow identification of compounds likely to exert effect levels in the narcotic range. With the presently introduced $T_e$-based classification schemes, discrimination between narcotic level and excess toxic compounds referring to the acute daphnid test is feasible, employing only information about the chemical structure of the substances. As such, this approach is well-suited for the priority setting of organic compounds with regard to their need for experimental testing. Moreover, extrapolation to other invertebrate species and other taxonomic groups would allow one to extend the scope of QSAR-supported risk assessment schemes to be extended to a broader range of endpoints. The inferior performance of MOA-based classification schemes reflects the fact that toxic effects based on reactive and specific MOAs do not necessarily exceed the narcotic effect range, indicating that the knowledge about prevalent MOAs is in general not sufficient to predict the likely effect level.

**Acknowledgment.** P.C.v.d.O. thanks Matthias Grote for helpful discussions.

## References

(1) White paper, COM (2003) Directive of the European Parliament and of the Council: Amending council Directive 67/548/EEC in order to adapt it to Regulation (EC) of the European Parliament and of the Council concerning the registration, evaluation, authorisation and restriction of chemicals. 2003/0256/COD, 2003/0257/COD (29 October 2003).

(2) Betts, K. S. (1998) Chemical industry pressured to test high-production volume chemicals. *Environ. Sci. Technol. 32*, 251A.

(3) Moore, D. R. J., Breton, R. L., and MacDonald, D. B. (2003) A comparison of model performance for six quantitative structure−activity relationship packages that predict acute toxicity to fish. *Environ. Toxicol. Chem. 22*, 1799−1809.

(4) Estrada, E., Patlewicz, G., Chamberlain, M., Basketter, D., and Larbey, D. (2003) Computer-aided knowledge generation for understanding skin sensitization mechanisms: The TOPS-MODE approach. *Chem. Res. Toxicol. 16*, 1226−1235.

(5) Fang, H., Tong, W., Branham, W. S., Moland, C. L., Dial, S. L., Hong, H., Xie, Q., Perkins, R., Owens, W., and Sheehan, D. M. (2003) Study of 202 natural, synthetic, and environmental chemicals for binding to the androgen receptor. *Chem. Res. Toxicol. 16*, 1338−1358.

(6) Arulmohhiraja, S., and Morita, M. (2004) Structure−activity relationships for the toxicity of polychlorinated dibenzofurans: Approach through density functional theory-based descriptors. *Chem. Res. Toxicol. 17*, 348−356.

(7) Altenburger, R., Nendza, M., and Schüürmann, G. (2003) Mixture toxicity and its modeling by quantitative structure−activity relationships. *Environ. Toxicol. Chem. 22*, 1900−1915.

(8) Lin, Z., Yin, K., Shi, P., Wang, L., and Yu, H. (2003) Development of QSARs for predicting the joint effects between cyanogenic toxicants and aldehydes. *Chem. Res. Toxicol. 16*, 1365−1371.

(9) Serra, J. R., Jurs, P. C., and Kaiser, K. L. E. (2001) Linear regression and computational neural network prediction of *Tetrahymena pyriformis* acute toxicity for aromatic compounds from molecular structure. *Chem. Res. Toxicol. 14*, 1535−1545.

(10) Aptula, A. O., Kühne, R., Ebert, R.-U., Cronin, M. T. D., Netzeva, T. I., and Schüürmann, G. (2003) Modeling discrimination between antibacterial and nonantibacterial activity based on 3D molecular descriptors. *QSAR Comb. Sci. 22*, 113−128.

(11) Verhaar, H. J. M., van Leeuwen, C. J., and Hermens, J. L. M. (1992) Classifying environmental pollutants. 1: Structure−activity relationships for prediction of aquatic toxicity. *Chemosphere 25*, 471−491.

(12) Russom, C. L., Bradbury, S. P., Broderius, S. J., Hammermeister, D. E., and Drummond, R. A. (1997) Predicting modes of toxicity action from chemical structure: Acute toxicity in the fathead minnow (*Pimephales Promelas*). *Environ. Toxicol. Chem. 16*, 948−967.

(13) Venkatapathy, R., Moudgal, C. J., and Bruce, R. M. (2004) Assessment of the oral rat chronic lowest observed adverse effect level model in TOPKAT, a QSAR software package for toxicity prediction. *J. Chem. Inf. Comput. Sci. 44*, 1623−1629.

(14) Boxall, A. B. A., Watts, C. D., Dearden, J. C., Bresnen, G. M., and Scoffin, R. (1997) Classification of environmental pollutants into general mode of toxic action classes, based on molecular descriptors. In *QSAR in Environmental Sciences VII* (Chen, F., and Schüürmann, G., Eds.) pp 263−275, SETAC Press, Pensacola, FL.

(15) Schüürmann, G., Aptula, A. O., Kühne, R., and Ebert, R.-U. (2003) Stepwise discrimination between four modes of toxic action of phenols in the *Tetrahymena pyriformis* assay. *Chem. Res. Toxicol. 16*, 974−987.

(16) Comber, M. H. I., Walker, J. D., Watts, C., and Hermens, J. (2003) Quantitative structure−activity relationships for predicting potential ecological hazard of organic chemicals for use in regulatory risk assessment. *Environ. Toxicol. Chem. 22*, 1822−1828.

(17) Russom, C. L., Breton, R. L., Walker, J. D., and Bradbury, S. P. (2003) An overview of the use of quantitative structure−activity relationships for ranking and prioritizing large chemical inventories for environmental risk assessment. *Environ. Toxicol. Chem. 22*, 1810−1821.

(18) U.S. Environmental Protection Agency (1994) ECOSAR: A computer program for estimating the ecotoxicity of industrial chemicals based on structure−activity relationships. *EPA-748-*

*R-002*, Office of Pollution Prevention and Toxics, National Center for Environmental Publications and Information, Cincinnati, OH.
(19) Russom, C. L., Anderson, E. B., Greenwood, B. E., and Pilli, A. (1991) ASTER: An integration of the AQUIRE database and the QSAR system for use in ecological risk assessments. *Sci. Total Environ. 109/110*, 667−670.
(20) Gombar, V. K., and Enslein, K. (1995) Use of predictive toxicology in the design of new chemicals. *ACS Symp. Ser. 589*, 236−249.
(21) Bradbury, S. P., Russom, C. L., Ankley, G. T., Schultz, T. W., and Walker, J. D. (2003) Overview of data and conceptual approaches for derivation of quantitative structure−activity relationships for ecotoxicological effects of organic chemicals. *Environ. Toxicol. Chem. 22*, 1789−1798.
(22) Hermens, J. L. M. (1990) Electrophiles and acute toxicity to fish. *Environ. Health Perspect. 87*, 219−225.
(23) U.S. Environmental Protection Agency (2002) *AQUIRE (Aquatic Toxicity Information Retrieval Database)*, National Health and Environmental Effects Research Laboratory, Duluth, MN.
(24) Lipnick, R. L., Watson, K. R., and Strausz, A. K. (1987) A QSAR study of the acute toxicity of some industrial organic chemicals to goldfish. Narcosis, electrophile and proelectrophile mechanisms. *Xenobiotica 17*, 1011−1025.
(25) Lipnick, R. L. (1989) Base-line toxicity predicted by quantitative structure−activity relationships as a probe for molecular mechanism of toxicity. In *Probing Bioactive Mechanisms* (Magee, P. S., Henry, D. R., and Block, J. H., Eds.) Vol. 413, pp 366−389, American Chemical Society, Washington, DC.
(26) Van Leeuwen, C. J., Van der Zandt, P. T. J., Aldenberg, T., Verhaar, H. J. M., and Hermens, J. L. M. (1992) Application of QSARs, extrapolation and equilibrium partitioning in aquatic effects assessment. 1. Narcotic industrial pollutants. *Environ. Toxicol. Chem. 11*, 267−282.
(27) Sosak-Swiderska, B., Tyrawska, D., and Dzido, D. (1998) *Daphnia magna* ecotoxicity test with parathion. *Chemosphere 37*, 2989−3000.
(28) U.S. Environmental Protection Agency (2003) *EPI Suite for Windows*, v3.11, National Health and Environmental Effects Research Laboratory, Duluth, MN.
(29) Schüürmann, G., Kühne, R., Kleint, F., Ebert, R.-U., Rothenbacher, C., and Herth, P. (1997) A software system for automatic chemical property estimation from molecular structure. In *Quantitative Structure-Activity Relationships in Environmental Sciences - VII* (Chen, F., and Schüürmann, G., Eds.) pp 93−114, SETAC Press, Pensacola, FL.
(30) *SPSS for Windows*, Release 7.5.1 (1996) SPSS Inc., Chicago, IL.
(31) *Sigma Plot for Windows*, Release 4.0 (1997) SPSS Inc., Chicago, IL.
(32) Ramos, E. U., Vaal, M. A., and Hermens, J. L. M. (2001) Interspecies sensitivity in the aquatic toxicity of aromatic amines. *Environ. Toxicol. Pharmacol. 11*, 149−158.
(33) Veith, G. D., and Broderius, S. J. (1987) Structure-toxicity relationships for industrial chemicals causing type (II) narcosis syndrome. In *QSAR in Environmental Toxicology—II* (Kaiser, K. L. E., Ed.) pp 385−391, D. Reidel Publishing Company, Dordrecht, The Netherlands.
(34) Könemann, H. (1981) Quantitative structure−activity relationships in fish toxicity studies Part 1: Relationship for 50 industrial pollutants. *Toxicology 19*, 209−221.
(35) Schmitt, H., Altenburger, R., Jastorff, B., and Schüürmann, G. (2000) Quantitative structure−activity analysis of the algae toxicity of nitroaromatic compounds. *Chem. Res. Toxicol. 13*, 441−450.
(36) Timbrell, J. (2000) *Principles of Biochemical Toxicology*, 3rd ed., p 86, Taylor & Francis, London, United Kingdom.
(37) Krieger, R. (2001) *Handbook of Pesticide Toxicology—Volume II*, 2nd ed., Academic Press, San Diego, CA.
(38) Tomlin, C. D. S. (2000) *The Pesticide Manual, A World Compendium*, Crop Protection 23 Publications, Farnham.
(39) Schüürmann, G. (1992) Ecotoxicology and structure−activity studies of organophosphorus compounds. In *Rational Approaches to Structure, Activity and Ecotoxicology of Agrochemicals* (Fujita, T., and Draber, W., Eds.) pp 485−541, CRC Press, Boca Raton, FL.
(40) Dady, J. M., Bradbury, S. P., Hoffman, A. D., Voit, M. M., and Olson, D. L. (1991) Hepatic microsomal N-hydroxylation of aniline and 4-chloroaniline by rainbow trout (*Onchorhyncus mykiss*). *Xenobiotica 29*, 1605−1620.
(41) Bradbury, S. P., Dady, J. M., Fitzsimmons, P. N., Voit, M. M., Hammermeister, D. E., and Erickson, R. J. (1993) Toxicokinetics and metabolism of aniline and 4-chloroaniline in medaka (*Oryzias latipes*). *Toxicol. Appl. Pharmacol. 118*, 205−214.

TX0497954