Statistically Validated QSARs, Based on Theoretical Descriptors, for Modeling Aquatic Toxicity of Organic Chemicals in *Pimephales promelas* (Fathead Minnow)

Ester Papa, Fulvio Villa, and Paola Gramatica*

Department of Structural and Functional Biology, QSAR and Environmental Chemistry Research Unit, University of Insubria, via Dunant 3, 21100 Varese, Italy

Received May 24, 2005

The use of Quantitative Structure−Activity Relationships in assessing the potential negative effects of chemicals plays an important role in ecotoxicology. $(LC_{50})_{96h}$ in *Pimephales promelas* (Duluth database) is widely modeled as an aquatic toxicity end-point. The object of this study was to compare different molecular descriptors in the development of new statistically validated QSAR models to predict the aquatic toxicity of chemicals classified according to their MOA and in a unique general model. The applied multiple linear regression approach (ordinary least squares) is based on theoretical molecular descriptor variety (1D, 2D, and 3D, from DRAGON package, and some calculated logP). The best combination of modeling descriptors was selected by the Genetic Algorithm-Variable Subset Selection procedure. The robustness and the predictive performance of the proposed models was verified using both internal (cross-validation by LOO, bootstrap, Y-scrambling) and external statistical validations (by splitting the original data set into training and validation sets by Kohonen-artificial neural networks (K-ANN)). The model applicability domain (AD) was checked by the leverage approach to verify prediction reliability.

INTRODUCTION

The impact of the potential hazard of untested chemicals, a challenge confronting national and international regulatory agencies,[1−4] can be measured by experimental investigations, but this approach is both quite expensive and time-consuming.[5] An alternative is to rely on QSAR (Quantitative Structure−Activity Relationships) models that describe a mathematical relationship between the structural features of a set of chemicals and the particular activity associated with them.[6,7] According to recently stated OECD principles,[8] such models should, in order to be reliable and predictive, be checked by internal validation and assessed in terms of predictive power using data not employed in developing the model (external statistical validation); such models must be accompanied by a definition of the applicability domain (AD)[9,10] and could also possibly have a mechanistic interpretation.

In the initial research in the field of QSARs for ecological risk assessment it was assumed that chemicals of the same chemical class would behave in a toxicologically similar manner. However Russom et al.[11] illustrated that this kind of classification can be problematic as, in the same chemical class, there could be chemicals acting through different modes of action. Thus, QSAR development and application has evolved to a mode of action (MOA) perspective.[12,13]

On this topic, Verhaar et al. published an expert system that makes it possible to classify a large number of organic pollutants into four classes:[14,15] narcosis, polar-narcosis, reactive chemicals, and specific mechanism of action. The advantage of developing MOA-based QSARs is their capability to define high quality predictions.

* Corresponding author phone: +39-0332-421573; fax: +39-0332-421554; e-mail: paola.gramatica@uninsubria.it; http://dipbsf.uninsubria.it/qsar/.

However the difficulties in defining correct "a priori" MOAs can lead to QSARs under/overestimating toxicity. Considering that the general aim of a QSAR model should be to predict data for untested compounds, for which the mode of action is difficult to know with any certainty, it is considered useful to also provide global toxicity models independent of a chemical's MOA.[16]

Several QSAR models predicting acute chemical toxicity for aquatic environments have been published.[17−21] They are based mainly on the logarithm of the octanol−water coefficient (logP, also referred to as $\log K_{ow}$) as this hydrophobicity term reproduces the ability of a substance to enter cells through the lipid membranes and indicates both toxicant uptake and baseline toxicity. Nevertheless the experimental determination of logP can be a complex matter, and experimental values can differ greatly even when referred to the same compound.[22,23] Thus, several approaches have been developed for the theoretical calculation of logP,[24−27] but also in these calculations it is not uncommon to have differences of several orders of magnitude.[28] In 2001 Pontolillo and Eganhouse[22] raised the problem of the validity of this parameter for DDT, and the problem of logP reliability was later generalized by Renner.[23] For these reasons Kaiser, among others, recently proposed[28] a non-logP type neural network model of fathead minnow toxicity, based exclusively on other structural theoretical molecular descriptors.

Following this line, the present paper proposes new predictive multiple linear regression (MLR) QSAR models to assess the acute aquatic toxicity of industrial organic chemicals, classified according to their MOA, in terms of $(LC_{50})_{96h}$ in *Pimephales promelas* (fathead minnow): models with or without the use of logP as a molecular descriptor, together with other theoretical descriptors, have been developed and compared, and also, if possible, a mechanistic

10.1021/ci050212l CCC: $30.25
Published on Web 09/02/2005

interpretation is provided. In addition, global models, obtained for the whole data set, have been proposed as predictive tools, applicable even in the absence of "a priori" information about the MOA of a chemical.

The most important aspect of the proposed models is their development, taking some fundamental points required by OECD principles[8] for regulatory acceptability of QSARs into account: their validation for predictivity (both by internal and external statistical validation, by "a priori" splitting of available data by Kohonen Map-ANN in training and validation sets) and the possibility of verifying their chemical applicability domain by the leverage approach.

METHODS

Experimental Data. A selected set of experimental 96h LC_{50} data (from the original U.S.-E.P.A. Duluth Fathead Minnow Database[29]) was taken from Russom.[11] It consists of flow-through bioassays, conducted with juvenile fathead minnows, on chemicals selected from a cross section of the Toxic Substances Control Act Inventory of industrial organic chemicals. The compounds were "a priori" classified as narcotics (MOA 1), polar-narcotics (MOA 2), reactive chemicals (MOA 3), or specific acting compounds (MOA 4)[14,15] (salts, chemicals without the CAS number, undefined structure, or without an "a priori" defined mode of action were excluded from this study). The median lethal concentrations are reported as the logarithm of the inverse molar concentration: $\log(1/LC_{50})$. The final studied data set of 468 selected, structurally heterogeneous, chemicals is available as Supporting Information.

Molecular Descriptors. Molecular descriptors were computed using the current version of software DRAGON.[30] The input files for descriptors calculation, containing information on atom and bond types, connectivity, partial charges, and atomic spatial coordinates, relative to the minimum energy conformation of the molecule, were obtained by the molecular mechanics method of Allinger (MM+) using the package HYPERCHEM.[31]

A total of about 1200 molecular descriptors of different kinds was calculated to describe compound chemical diversity.[32] The descriptor typology is (a) 0D-constitutional (atom and group counts); (b) 1D-functional groups, 1D-atom centered fragments; (c) 2D-topological, 2D-BCUTs, 2D-walk and path counts, 2D-autocorrelations, 2D-connectivity indices, 2D- information indices, 2D-topological charge indices, and 2D-eigenvalue-based indices; and (d) 3D-Randic molecular profiles from the geometry matrix, 3D-geometrical, 3D-WHIMs, and 3D-GETAWAYs descriptors.[33,34]

To compare the modeling power of variables traditionally used to model toxicity end points with the above listed theoretical molecular descriptors, different kinds of logP (calculated values for AlogP,[30] MlogP,[30] and ClogP[11,26]) were always added and used for descriptor selection by genetic algorithm during model development.

Quantum-chemical descriptors such as HOMO (highest occupied molecular orbital), LUMO (lowest unoccupied molecular orbital), HOMO−LUMO gap (DHL), ionization potential (P ion), and heat of formation (H), calculated by the semiempirical PM3 Hamiltonian for the geometry optimization method available in the HYPERCHEM package,[31] were always added and used for descriptor selection during model development.

Constant values and descriptors found to be correlated pairwise were excluded in a prereduction step (when there was more than 98% pairwise correlation, one variable was deleted), and the genetic algorithm was applied for variables selection to a final set of about 400 descriptors for each class of toxicity.

Chemometric Methods. Multiple linear regression analysis and variable selection were performed by the software MOBY DIGS of Todeschini[35] using the Ordinary Least Square regression (OLS) method and GA-VSS (Genetic Algorithm-Variable Subset Selection).[36]

The outcome of the application of the genetic algorithms is a population of 100 regression models, ordered according to their decreasing internal predictive performance, verified by Q^2. The models with lower Q^2 are those with fewer descriptors. First of all, models with 1−2 variables were developed by the all-subset-method procedure in order to explore all the low dimension combinations. The number of descriptors was subsequently increased one by one, and new models were formed. The GA was stopped when increasing the model size did not increase the Q^2 value to any significant degree. Particular attention was paid to the collinearity of the selected molecular descriptors: in fact, to avoid multicollinearity without, or with, "apparent" prediction power (due to chance correlation), regression was calculated only for variable subsets with an acceptable multivariate correlation with response, by applying the *QUIK* rule (Q Under Influence of *K*).[37] Acceptable models are only those with a global correlation of [X+y] block (K_{XY}) greater than the global correlation of the X block (K_{XX}) variable, X being the molecular descriptors and *y* the response variable. The collinearity in the original set of molecular descriptors results in many similar models that more or less yield the same predictive power (in MOBY-DIGS software 100 models of different dimensionality). Therefore, when there were models of similar performance, those with higher ΔK ($K_{XY} - K_{XX}$) were selected and further verified.

The robustness of the models and their predictivity were evaluated by both Q^2_{LOO} and bootstrap. In this last procedure K *n*-dimensional groups are generated by a randomly repeated selection of *n*-objects from the original data set. The model obtained on the first selected objects is used to predict the values for the excluded sample, and then Q^2 is calculated for each model. The bootstrapping was repeated 5000 times for each validated model.

The proposed models were also checked for reliability and robustness by permutation testing: new models are recalculated for randomly reordered response (Y scrambling). Evidence that the proposed model is well founded, and not just the result of chance correlation, is provided by obtaining new models on the set with randomized responses that have significantly lower R^2 and Q^2 than the original model. When checked by the Y-scrambling procedure our suggested QSAR models verify this condition.

The real predictive capability of each model developed on the training set is verified on an external validation set and is calculated from $Q^2_{EXT} = 1 - \text{PRESS/SD}$, where PRESS is the sum of squared differences between the measured response and the predicted value for each molecule

in the validation set, and SD is the sum of squared deviations between the measured response for each molecule in the validation set and the mean measured value of the training set. A measure to define the accuracy of the proposed QSARs is the RMSE (root mean square of errors) that summarizes the overall error of the model. It is calculated as the root square of the sum of the squared errors in predictions divided by their total number:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_i (y_i - \hat{y}_i)^2}{n}}$$

To have compounds for external validation, the available sets of chemicals were split into a training set and an external validation set. The splitting of the data set, realized by Kohonen artificial neural network (K-ANN)[38] using the software KOALA,[39] takes advantage of the clustering capabilities of K-ANN.

The three most significant principal components, calculated from each group of DRAGON molecular descriptors, were used to synthesize the structural information of the chemicals. This structural information and the response were used as variables to organize the structure of a Kohonen map. After choosing the map dimensions, the number of epochs, and the splitting partition, all the objects of the studied toxicity class were presented to the map. For each class we developed more than one map so as to select the best splitting for each developed model: 300−500 epochs, $8 \times 8 - 12 \times 12$ dimension maps.

At the end of the net training, similar chemicals fell within the same neuron, i.e., they carried the same information. To select the training set of chemicals it was assumed that the compound closest to each neuron centroid was the most representative of all the chemicals within the same neuron. Thus, the training set chemicals were selected according to the minimal distance from the centroid of each cell in the top map. The remaining objects, close to the training set chemicals, were used for the validation set.

Standard deviation error in prediction (SDEP), standard deviation error in calculation (SDEC), and root mean square of errors (RMSE, calculated separately for the training and the test/validation sets) are reported for each model together with the coefficient of determination (R^2) and cross-validated explained variance by leave-one-out (Q^2_{LOO}), bootstrap (Q^2_{BOOT}), and the external validation parameter (Q^2_{EXT}).

In general the best predictive models were selected by maximizing the ΔK ($K_{\mathrm{XY}} - K_{\mathrm{XX}}$) values and then trying to find a trade off between external predictivity ($Q^2_{\mathrm{EXT}} > 0.7$) and internal predictivity and robustness ($Q^2_{\mathrm{LOO}} > 0.7$; $Q^2_{\mathrm{BOOT}} > 0.6$). In this study an average of about 50% of the models' population has a $Q^2_{\mathrm{EXT}} > 0.7$, and the best-selected model has the highest Q^2_{EXT}.

In the equations of the models, reported in this paper, the variables are listed in order of relative importance by their standardized regression coefficients. In fact, since molecular descriptors do not have equal variance (i.e. they are not autoscaled), their relative importance in the model is measured better by standardized regression coefficients (i.e. the coefficients multiplied by the standard deviation of the corresponding predictor). The errors of the regression coefficients have also been reported for each equation.

The presence of outliers (i.e. compounds with cross-validated standardized residuals greater than 2.5 standard deviation units) and chemicals very structurally influential in determining model parameters (i.e. compounds with a high *leverage* value (h)[40] greater than $3\,p'/n$ (h^*), where p' is the number of model variables plus one, and n is the number of the objects used to calculate the model) was verified by the Williams plot.[41] Also the reliability of the predicted data with regard to chemical domain was verified by the *leverage* approach: the predictions for chemicals of the validation set with a leverage value higher than h^* must be considered potentially unreliable, being out of the structural chemical domain of the training set.

RESULTS AND DISCUSSION

The majority of QSAR models predicting acute chemical toxicity for aquatic environments are based on "a priori" classification of chemicals according to their MOA and on the use of the logarithm of the octanol−water coefficient for narcotics, as this hydrophobicity term is regarded as the most mechanistically interpretable, being considered informative on a chemical's ability to enter cells through lipid membranes. However, it is important to remember that logP looks like a single descriptor but is, really, the condensation of all the structural information represented by fragments and correction factors; it is strongly dependent on the QSAR applied for its calculation. Moreover, as reported by Kaiser et al.[28] for many substances the logP value is generated by other QSARs that interpret molecular structure information in their own peculiar way. Therefore, it should be possible to omit logP as a variable when, included in the list of input variables, there is the type of information on which logP computation depends, or there are molecular descriptors able to furnish analogous information. In addition, it is important to note that the "a priori" determination of the mechanism of action of a compound is not always easy because toxicity is a complicated effect, and many steps involved in this effect are poorly understood or characterized: an approach for direct toxicity prediction of a global data set of heterogeneous chemicals can be developed and compared. Thus, in line with other studies on modeling toxicity,[16,28,42] we developed both QSAR models logP-based and logP-free, basing our models only on theoretical molecular descriptors, both for groups of chemicals classified according to the classical four MOAs and for the global data set (direct toxicity prediction, DTP model[16]).

In this study, we want mainly to verify the genetic algorithm efficiency in selecting the best modeling variables, even when a lot of different molecular descriptors are used as input, including also traditional variables considered "mechanistically interpretable", such as logP and quantum chemical descriptors.

Particular attention was paid to model predictivity, carrying out different kinds of internal and external validation,[9] and to chemical applicability domain.[9,10]

LogP as Toxicity Descriptor. As demonstrated in the literature[22,23] the octanol−water coefficient of a given compound could be subject to high variability due to the applied experimental procedure or the selected calculation

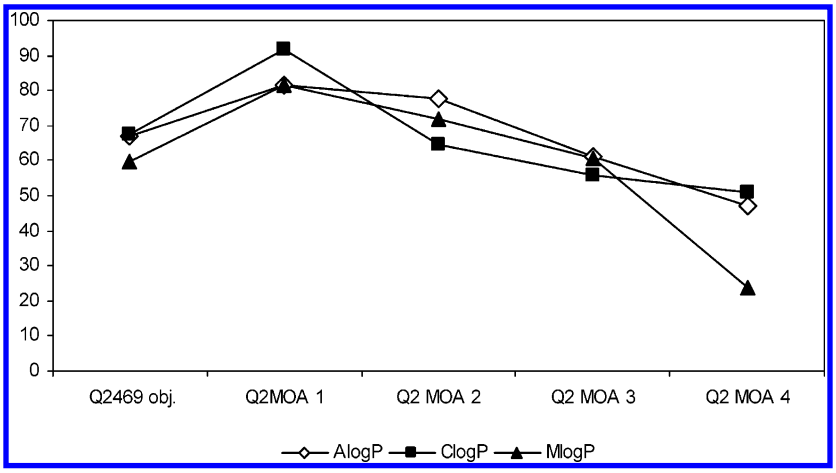

Figure 1. Comparison of the fitting power trends (LOO cross-validated) of the different logP values for modeling the complete data set and the subsets of the "a priori" defined MOAs.

Table 1. Comparison of Internal Predictive Power (Q^2_{LOO} Expressed as Percentage Value) for QSARs Based on Different logP (AlogP, MlogP, ClogP)

	Q^2				
	469 obj	MOA 1	MOA 2	MOA 3	MOA 4
AlogP	66.93	81.62	77.47	61.04	47.14
ClogP	67.42	91.68	64.72	55.8	50.94
MlogP	59.7	81.33	71.68	60.5	23.94
Δ^a Q^2 best-worst	7.72	10.35	12.75	5.24	27

[a] The Δ value was calculated as the difference between the best and the worst Q^2_{LOO} obtained for each group.

method. For this reason it cannot be considered a univocal descriptor since it could provide, for each studied compound, a range of predicted toxicity values strictly dependent on the logP used for the QSAR development, compromising model accuracy and quality.

A clear example of this aspect has been obtained by comparing performances in fitting and prediction for QSARs based exclusively on logP (AlogP, MlogP, and ClogP) and developed separately for the "a priori" selected MOA classes and for the complete data set (Table 1).

Table 1 clearly shows the differences in internal predictive power for the selected logP descriptors. The Δ value calculated as the difference between the best and worst Q^2_{LOO} obtained for each group gives an idea of the variability in these logP-based models and demonstrates that none of those logP can be defined as "the best" since a different one was selected by GA as the best modeler for each different MOA class: this demonstrates that it is erroneous to consider one specific logP (i.e. the most commonly used ClogP) as always the best, making it necessary to use the appropriate one for each modeled MOA.

The general fitting power trends (LOO cross-validated) of the different logP values for modeling the complete data set and the subsets of the a priori defined MOAs are shown in Figure 1.

As expected, logP importance and predictivity decreases in the model developed for the reactive chemicals and shows even poorer modeling performance for specific acting compounds, since the information about chemical partitioning (condensed by logP) in these cases is not sufficiently relevant to describe the chemical behavior producing their toxicity.

In general ClogP appears as the best one for MOA1, but its performance appears to be worse than AlogP and MlogP if applied to the other MOAs. However in the global model AlogP and ClogP have better modeling power than MlogP.

The general decreasing modeling trend shows that the need for structural information, unlike logP, increases from MOA1 to MOA4; this could justify the addition of different descriptors to a model or, eventually, total logP substitution by other theoretical molecular descriptors that are more able to describe chemical toxicity behavior.

Models according to the Mode of Action (MOA) Classification. Narcotics. The majority of industrial organic chemicals are deemed to exhibit a narcosis mode of action. Narcotic chemicals cause only noncovalent and reversible alterations at the site of action, which is considered to be the cellular membrane.

The best logP-based model obtained for narcotics is

$$\text{Log}\,(1/\text{LC}_{50})_{96h} = 2.6 + 0.72(\pm 0.02)\text{ClogP} - 0.13(\pm 0.02)\text{LUMO} - 1.03(\pm 0.2)\text{RARS} \quad (1)$$

$$n_{\text{training}} = 147;\ n_{\text{test}} = 116;\ R^2 = 0.95;$$
$$Q^2_{\text{LOO}} = 0.95;\ Q^2_{\text{BOOT}} = 0.94$$

$$Q^2_{\text{EXT}} = 0.93;\ \text{SDEC} = 0.283;\ \text{SDEP} = 0.293$$

$$\text{RMSE}_{\text{training}} = 0.29;\ \text{RMSE}_{\text{test}} = 0.34;$$
$$K_{\text{X}} = 49.53;\ K_{\text{XY}} = 61.4$$

The reported fitting and validation parameters have, as expected, very high values, indicating that the model has very good descriptive and predictive performance. The quality of Q^2_{EXT} (94.8%) and the small RMSE values (similar for training and test sets) confirm the high predictivity of this model. As expected, taking into account the high single correlations between logP and the studied end point commented on above, the best variables selected by the GAs for narcotic chemicals are ClogP (which is the most relevant descriptor in the equation), combined with LUMO (lowest unoccupied molecular orbital), related to the electron affinity, and RARS (R matrix average row sum-GETAWAY), inversely related to molecular dimension.

An analysis of the model applicability domain reveals the presence of only 3 chemicals with a slightly high leverage: triethylene glycol and 1-decanol in the training set and cyclohexane in the test set. Some outliers are also present: tolazoline and 4,4′-isopropylenebis(2,6-dichlorophenol) in the training set, while 2,6-dimethylmorpholine and 1,3-diethyl-2-thiobarbituric acid are outliers in the test set. Chemicals influential on the structural domain of the model, characterized by a hat value exceeding the cut off one, can be explained as compounds with peculiar features poorly represented in the training set, which could affect the variables selection for a better modeling of those chemicals. Outliers, of which the standardized residual values exceed the cut off value of 2.5 units, could be associated with errors in the experimental values, with peculiar structural features or with errors in the "a priori" assignment of the MOA. At this stage it is not possible to define which of these causes could explain the above-mentioned outliers.

For this class of chemicals we also developed a logP-free externally validated model, but the statistical quality is worse than model (1) ($R^2 = 0.83$; $Q^2_{LOO} = 0.81$; $Q^2_{EXT} = 0.78$), thus it is not useful to comment here on this model, it being clearly confirmed that logP is the best modeler of narcosis.

Polar Narcotics. Polar narcosis is elicited by molecules that possess strong electron-releasing substituents and that are able to form hydrogen bonds. Such compounds, often featuring an aromatic structure, exhibit an effect similar to nonpolar narcotics but at potency levels greater than that estimated by their hydrophobicity. A dipolarity and/or hydrogen bond acceptor and donor capacity was attributed to the difference in potency between narcotics and polar narcotics.[17,43−45]

Considering a training set made up of 57 objects we developed the following externally validated logP-based model:

$$\text{Log}\,(1/\text{LC}_{50})_{96h} = 2.6 + 0.34(\pm 0.06)\text{AlogP} + 0.82(\pm 0.19)\text{BEHv3} + 0.18(\pm 0.04)\text{nHDon} - 0.65(\pm 0.19)\text{C} - 029 \quad (2)$$

$$n_{training} = 57;\ n_{test} = 29;\ R^2 = 0.90;$$
$$Q^2_{LOO} = 0.88;\ Q^2_{BOOT} = 0.88$$

$$Q^2_{EXT} = 0.84;\ \text{SDEC} = 0.360;\ \text{SDEP} = 0.397$$

$$\text{RMSE}_{training} = 0.31;\ \text{RMSE}_{test} = 0.44;$$
$$K_X = 28.98;\ K_{XY} = 44.63$$

The reported fitting and validation parameters have very high values, indicating that the model has good descriptive and predictive performance. As expected, considering the single correlations between logP and the studied end point commented on above, the best logP selected by the GAs for polar narcotic chemicals is AlogP. The other variables are BEHv3 (highest eigenvalue n.3 of Burden matrix/weighted by van der waal volumes), related to the 2D-steric features of a molecule, nHDon (number of donor atoms for H-bonds), representing the donor capacity of hydrogens for H-bonds related to polar narcotics (phenols, amines, etc.), and C-029 (atom centered fragment counting for groups derived from carboxylic acids R−CX−X).

We also developed a population of externally validated logP-free models of which the following is the best:

$$\text{Log}\,(1/\text{LC}_{50})_{96h} = -0.12 + 0.22(\pm 0.03)\text{C} - 002 + 1.23(\pm 0.19)\text{BEHm3} + 0.64(\pm 0.11)\text{nBnz} - 0.22(\pm 0.03)\text{nN} \quad (3)$$

$$n_{training} = 57;\ n_{test} = 29;\ R^2 = 0.84;$$
$$Q^2_{LOO} = 0.81;\ Q^2_{BOOT} = 0.80$$

$$Q^2_{EXT} = 0.89;\ \text{SDEC} = 0.360;\ \text{SDEP} = 0.397$$

$$\text{RMSE}_{training} = 0.40;\ \text{RMSE}_{test} = 0.37;$$
$$K_X = 25.43;\ K_{XY} = 41.31$$

The model matches the high quality parameters that deal with fitting power and capability to assess the toxicity of external data. The very similar values of RMSE, for training and test chemicals, highlight the model generalizability.[46]

The capability to penetrate cell membranes, generally well described by the octanol−water coefficient, could be represented here by a combination of different descriptors: C-002 (number of R−CH_2−R fragments) and the 2D-BCUT descriptor BEHm3 (highest eigenvalue n.3 of Burden matrix/weighted by atomic masses). The first, giving information on the length of the aliphatic chains in the molecules, is related to their lipophilicity. The second takes into account the 2D-steric features of a molecule. It should also be pointed out that polar narcotics in the data set are mainly aromatic (e.g.: phenols, anilines...), so the nBnz (number of benzene-like rings) descriptor could represent this feature in the model; furthermore, in a multivariate approach the nBnz is selected in combination with the nN descriptor (number of nitrogen atoms), that could explain the potency to form hydrogen bonds with biomolecules.

On analyzing the model AD in the Williams plot of this last model (Figure 2) only one substance in the training set (*tert*-octylamine - 290) was identified as an outlier, while dodecylamine (311) and 3-amino-5,6-dimethyl-1,2,4-triazine (348) of the training set and tridecylamine (339) of the test set are chemicals with high leverage. Compared with the other chemicals of the data set, these molecules could have some structural anomalies that are not well modeled by the selected descriptors for the outliers or that are too particular for the influential chemicals (with leverage higher than the h^* value). It is important to note that none of the 29 chemicals in the validation set is an outlier for response, and only tridecylamine (339) is structurally high leverage. Thus, the predicted toxicity value for this chemical should be considered as extrapolated by the model and potentially unreliable.

Comparing the quality of Q^2_{EXT} for the logP-based (0.84) and the logP-free model (0.89), the slightly better performance of the second one can be explained by the residuals distribution and the hat values in the Williams plot of the logP-based model (Figure 3). This graph clearly shows that two compounds in the test set affect, negatively, the AD of the model (2): 6-chloro-2-pyridinol (347) is a strongly structural high leverage compound, of which the predicted value should not be considered reliable, while dibuthyl adipate (283) is an outlier for the response (3.5 standard

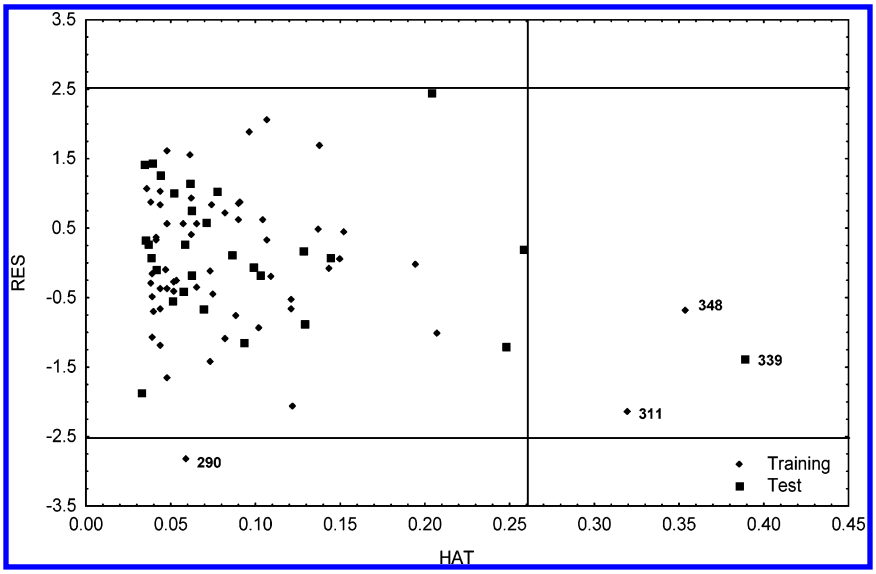

Figure 2. Williams plot for the externally validated logP-free model developed for polar narcotics (eq 3).

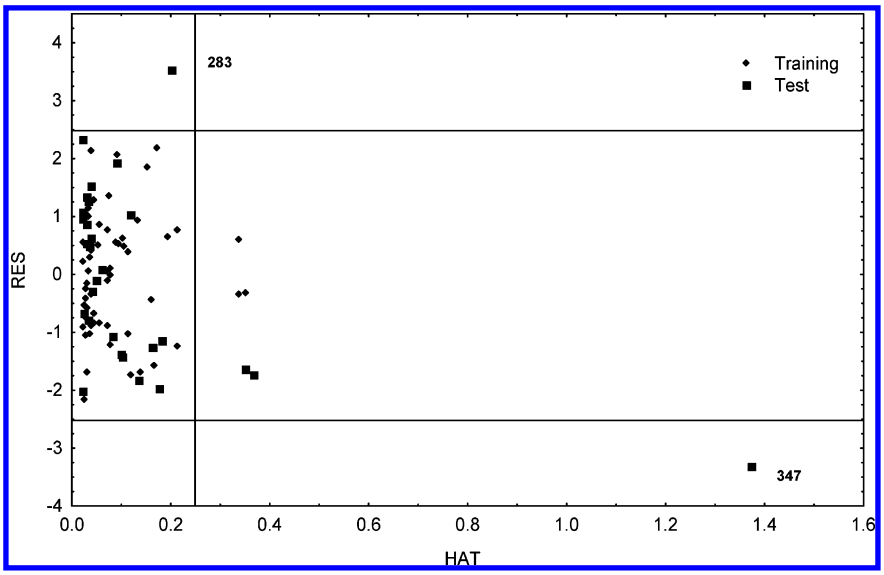

Figure 3. Williams plot for the externally validated logP-based model developed for polar narcotics (eq 2).

deviation units). The presence of those 2 anomalous chemicals in the test set justifies the worse external predictive power of the model (2) compared with the logP-free one (3) that, on the contrary, provides reliable predictions for both compounds. Moreover the same observation that the logP-based model appears better in its fitting power, while the logP-free model is the best one in prediction for external chemicals, can be verified also by the small RMSE values.

Reactive Chemicals. The toxicity of some compounds is considerably greater than that predicted for narcotics and polar-narcotics, and such compounds are generally defined as having "reactive toxicity". In this context the term "reactive" encompasses a wide spectrum of competitive electro- and nucleophilic, redox, and free radical processes. With respect to environmental pollutants, compounds that act as electrophiles are especially important as they can react with nucleophilic groups ($-NH_2$, $-OH$, $-SH$) in biological macromolecules such as DNA and proteins. Hermens[47] and Lipnick[48] correlated a particular mechanism of action with well-defined structural moieties. Such reactivity causes difficulty in QSAR development as the occurrence of more specific, irreversible toxic effects leads to complications in the modeling.[21]

A logP-based model was developed on a training set of 62 compounds with different typology of reactivity and was externally validated on a validation set of 19 chemicals, also in this case selected by Kohonen maps:

$$\begin{aligned}\text{Log}\,(1/\text{LC}_{50})_{96\text{h}} = {} & 3.22(\pm0.2) + 0.43(\pm0.07)\text{AlogP} - \\ & 18.34\,(\pm3.9)\text{H6v} + 0.14(\pm0.04)\text{L1m} + \\ & 0.27(\pm0.1)\text{BEHm7} - 2.49(\pm0.8)\text{R2u}+ \end{aligned} \quad (4)$$

$$n_{\text{training}} = 62;\ n_{\text{test}} = 19;\ R^2 = 0.76;$$

$$Q^2_{\text{LOO}} = 0.70;\ Q^2_{\text{BOOT}} = 0.69$$

$$Q^2_{\text{EXT}} = 0.75;\ \text{SDEC} = 0.483;\ \text{SDEP} = 0.542$$

$$\text{RMSE}_{\text{training}} = 0.54;\ \text{RMSE}_{\text{test}} = 0.56;$$

$$K_{\text{X}} = 48.96;\ K_{\text{XY}} = 50.05$$

This model is satisfactory as it allows the modeling of a relatively broad group of reactive chemicals that act by very

different mechanisms. In fact, electrophiles/proelectrophiles are generally subdivided into different classes[49] in order to focus on a particular mechanism of action. Veith and Mekenyan[50] developed a general QSAR equation that models not only the toxicity of different modes of toxic action but also the different modes of electrophilic reactivity. Their interesting model describes toxic potency by using hydrophobicity, quantified by the octanol−water partition coefficient (logP), to model the biouptake and the energy of the lowest unoccupied molecular orbital (E_{LUMO}) to model electrophilic interactions; but the quality of this QSAR was verified only by fitting power ($R^2 = 0.81$) and was not verified for predictivity, not even by internal cross-validation (LOO). Differently, the predictive power of our model was also checked by statistical external validation, with good results, highlighting the robustness and the real predictivity of the model.

Some additional structural information (mainly dimensional) is needed to implement the modeling power of logP in order to obtain a satisfactory model: H6v (H autocorrelation of lag6/weighted by atomic van der Waals volumes), L1m (1st component size directional WHIM index/weighted by atomic masses), BEHm7 (highest eigenvalue n.3 of Burden matrix/weighted by atomic masses), and R2u+ (R maximal autocorrelation of lag2/unweighted). The analysis of the model applicability domain in the Williams plot reveals the presence of only one structural high leverage chemical in the training set (2,2′-methylenebis(3,4,6-trichlorophenol)) and only one response outlier in the test set (3-chloro-2-chloromethyl-1-propene).

As in the previous cases, we also developed a logP-free model with the following performance:

$$\begin{aligned}\text{Log}\,(1/\text{LC}_{50})_{96h} = {} & 2.2(\pm0.2) + 0.08(\pm0.01)\text{Ss} - \\ & 0.54(\pm0.07)\text{nHAcc} + \\ & 0.21(\pm0.03)\text{Tm} - 21.02(\pm4.6)\text{GGI8} - \\ & 1.06(\pm0.2)\text{HATS6u} + 0.86(\pm0.2)\ \text{nROR} \end{aligned} \quad (5)$$

$$n_{\text{training}} = 62;\ n_{\text{test}} = 19;\ R^2 = 0.82;$$
$$Q^2_{\text{LOO}} = 0.78;\ Q^2_{\text{BOOT}} = 0.77$$

$$Q^2_{\text{EXT}} = 0.77;\ \text{SDEC} = 0.468;\ \text{SDEP} = 0.417$$

$$\text{RMSE}_{\text{training}} = 0.47;\ \text{RMSE}_{\text{test}} = 0.54;$$
$$K_{\text{X}} = 37.81;\ K_{\text{XY}} = 41.03$$

This model presents better internal and external predictive power than the logP-based one (model 4) and is based on six theoretical descriptors: some dimensional (Tm (total size index/weighted by atomic masses), HATS6u (leverage weighted autocorrelation of lag 6), GGI8 (topological charge index of order 8)), one related to the electronic distribution (Ss (sum of Kier-Hall electrotopological states)), and one accounting for the number of acceptor atoms for H bonds (nHAcc), the least important nROR is the count of aliphatic ethers, useful for better modeling of these peculiar chemicals.

The analysis of the AD reveals two high leverage chemicals only in the training set: again 2,2′-methylenebis-(3,4,6-trichlorophenol) and *p-tert*-butylphenyl-*N*-methyl-carbamate. No outliers for the response are present, neither in the training set nor the validation set.

Specific Acting Compounds. The specific-acting class encompasses a range of action modes, including central nervous system seizure agents, respiratory blockers, acetylcholinesterase inhibitors, and so on. The underlying mechanisms of these specific modes of action are very diverse and could involve a very large number of receptor sites. Consequently, the actual mechanisms underlying these modes of action are usually poorly understood. Usually QSAR modeling of receptor−ligand complexes focuses mainly on docking procedures, in which the structure of the receptor must be known to assess all the steric and electronic relationships between the receptor and the toxicant, as also reported by Schultz et al.[17,20]

As the studied data set has few chemicals for any particular mechanism of action (11 phosphorylation uncouplers, 15 AchE inhibitors, 3 respiratory blockers, and 7 CNS seizure agents), there is no reason to develop a particular QSAR for each group of specific-acting compounds, as no significant structural information could be obtained for model development from such a small data set, and, in addition, any validation would fail. Thus, to develop a relatively simple general QSAR that accounts for heterogenic mechanisms of action, we tried to model the toxicity of a unique data set made up by all the different specific-acting compounds together. We used a training set of 29 chemicals and a test set of 7 molecules for the external validation procedure, selected by taking into account not only the structure but also, and mainly, their specific mechanism of action. The selected test set chemicals were pentachlorophenol (phosphorylation uncoupler), pentachloropiridine (phosphorylation uncoupler), chloroacetonitril (respiration inhibitor), resmethrin (central nervous system seizure agent), aldicarb (acetylcolinesterase inhibitor), malathion (acetylcolinesterase inhibitor), and carbaryl (acetylcolinesterase inhibitor).

For this class of toxicity we developed the following, externally validated, logP-based model:

$$\begin{aligned}\text{Log}\,(1/\text{LC}_{50})_{96h} = {} & 3.38(\pm0.26) + 0.56(\pm0.07)\text{ClogP} + \\ & 1.09(\pm0.29)\text{N} - 074 + 0.006(\pm0.002)\text{MPC09} \end{aligned} \quad (6)$$

$$n_{\text{training}} = 29;\ n_{\text{test}} = 7;\ R^2 = 0.78;$$
$$Q^2_{\text{LOO}} = 0.73;\ Q^2_{\text{BOOT}} = 0.63$$

$$Q^2_{\text{EXT}} = 0.91;\ \text{SDEC} = 0.616;\ \text{SDEP} = 0.671$$

$$\text{RMSE}_{\text{training}} = 0.67;\ \text{RMSE}_{\text{test}} = 0.33;$$
$$K_{\text{X}} = 22.22;\ K_{\text{XY}} = 32.8$$

Model performance appears satisfactory, considering mainly the good quality of the internal validation parameter Q^2_{LOO}, nevertheless the lowering of the Q^2_{BOOT} value could indicate a relative instability of the model, even for a very high Q^2_{EXT} (0.91). These discrepancies, particularly in this case, raise the need to carefully verify the model AD, the model having been developed on a very small and heterogeneous training set. It is evident that if a chemical with strong peculiarities is in the test set, but falls outside the chemical domain of the training set, then the training set lacks the information needed for prediction, and the predicted value for this chemical might be wrong. In fact, an analysis of the model AD in the Williams plot reveals the presence

of two high leverage compounds in the training set: rotenone and malononitrile, while no structurally influential or response outlier chemicals are present in the test set.

ClogP is the most important variable, but some other structural features are needed to improve the modeling of such a heterogeneous set of chemicals: N-074 (molecular fragment representing R═N− and R#N) related to the presence of oximes and nitriles in the data set and MPC09 (molecular path count of order 9).

The best logP-free model is

$$\text{Log}\,(1/\text{LC}_{50})_{96\text{h}} = 3.74(\pm 0.4) + 0.3(\pm 0.04)\text{S1K} - 0.91(\pm 0.1)\text{MAXDN} - 0.66(\pm 0.1)\text{O}- 058 + 9.3(\pm 3.21)\ \text{R2p}+ \quad (7)$$

$$n_{\text{training}} = 29;\ n_{\text{test}} = 7;\ R^2 = 0.82;$$
$$Q^2_{\text{LOO}} = 0.72;\ Q^2_{\text{BOOT}} = 0.67$$

$$Q^2_{\text{EXT}} = 0.75;\ \text{SDEC} = 0.553;\ \text{SDEP} = 0.690$$

$$\text{RMSE}_{\text{training}} = 0.69;\ \text{RMSE}_{\text{test}} = 0.56;$$
$$K_{\text{X}} = 29.97;\ K_{\text{XY}} = 38.35$$

Model performance appears satisfactory, considering the relatively high quality of the internal validation parameters (Q^2_{LOO} and Q^2_{BOOT}). The quality of Q^2_{EXT} (0.75) suggests that the selected training set contains all the necessary and useful information to assess the toxicity end point for the external test set data, in any case, and also here, the AD of this model must be carefully verified. An analysis of the model domain reveals the presence of a chemical that, at the same time, has high structural leverage and is a response outlier in the training set (azinphos methyl), while all the chemicals in the test set practically fall within the structural domain of the training set (only chloroacetonitrile very slightly exceeds the cut off value of the hat).

The higher Q^2_{EXT} value and the smaller RMSE value calculated for the test set of the logP-based model (6) highlight its better performance in predictions, at least for the selected test compounds, than the logP-free one (7). However, it must be pointed out that both models have a relative internal stability, as demonstrated by the quite low Q^2_{BOOT} values, and this instability is higher in the logP based model.

Direct Toxicity Prediction: General DTP Model. Finally, to propose a predictive tool with a wide applicability domain, applicable independently of the "a priori" knowledge of the MOA of chemicals, a general-direct toxicity prediction (DTP) model for $\log(1/\text{LC}_{50})$ in *Pimephales* has also been developed. As a preliminary step the whole data set of 468 compounds was considered, without the exclusion of chemicals identified as outliers in the MOA models as they could have been the result of errors in the "a priori" assignment of the MOA. Also chemicals previously identified as high leverage for their structure were not excluded at this first stage, since the general model was developed on a wider structural basis than those for each different MOA, thus all these chemicals could fall within the structural domain of this new model.

Also in this case external statistical validation was developed by a previous splitting of the original data set by applying Kohonen maps-artificial neural networks for the selection of representative and homogeneously distributed training and test sets. The statistical external validation was performed by checking the models on about 50% of the original data set.

After some attempts to model all the 468 chemicals together, 19 structurally heterogeneous outliers (9 molecules from the MOA 1 class, 3 from the MOA 3 class, and 7 from the MOA 4 class), strongly affecting the predictive performance of the developed general-DTP model, were excluded, thus the final studied data set consisted of 449 chemicals.

Some of those chemicals were already highlighted as problematic (outliers or high leverage) in the previous, already commented on, models (4,4′-isopropyldiene-bis-(2,6-dichlorophenol), triethylene glycol, 3-chloro-2-chloromethyl-1-propene, azinphos methyl, malononitrile). As these compounds fell outside the domain of both the specific and the general models, we related this behavior to questionable experimental data.

The final externally validated general-DTP logP-based model was then developed on a training set of 249 compounds (regression line in Figure 4):

$$\text{Log}\,(1/\text{LC}_{50})_{96\text{h}} = 2.9(\pm 0.25) + 0.56(\pm 0.04)\text{AlogP} + 0.34(\pm 0.04)\text{DP03} + 20.81(\pm 3.13)\text{H8m} - 0.79(\pm 0.11)\text{GATS1v} - 1.59(\pm 0.35)\text{R1v} \quad (8)$$

$$n_{\text{training}} = 249;\ n_{\text{test}} = 200;\ R^2 = 0.81;$$
$$Q^2_{\text{LOO}} = 0.80;\ Q^2_{\text{BOOT}} = 0.80$$

$$Q^2_{\text{EXT}} = 0.72;\ \text{SDEC} = 0.569;\ \text{SDEP} = 0.584$$

$$\text{RMSE}_{\text{training}} = 0.34;\ \text{RMSE}_{\text{test}} = 0.62;$$
$$K_{\text{X}} = 42.94;\ K_{\text{XY}} = 49.98$$

This model represents a very satisfactory result as it allows the contemporaneous modeling of a wide group of chemicals that are highly heterogeneous in their structure and that act by very different mechanisms.

Nicotine sulfate and fenvalerate are structurally influential in the training set but well predicted (they are "good leverage" chemicals); 3-(3,4-dichlorophenoxy)benzaldehyde (263) and chlorpyriphos are structural high leverage in the test set and the first one is also an outlier for the response. Other outliers can be identified: chloroacetonitrile (443), isovaleraldehyde (390), tripropargylamine (239), and rotenone (438) in the training set and ethanal (352) in the test set.

The model presents very good internal and external predictive power, and the best descriptor, AlogP, is combined here with four theoretical descriptors, which are mainly related to the dimensional features of the chemicals: DP03 (Randic molecular profile n°03), H8m (H autocorrelation of lag8/weighted on atomic masses), GATS1v (Geary autocorrelation of lag 1 weighted by atomic van der Waals volumes), and R1v (R autocorrelation of lag 1 weighted by atomic van der Waals volumes).

It is interesting to note that most of the chemicals falling outside the chemical domain of the model belong to the classes of reactive chemicals and specific acting compounds, whose structural heterogeneity and complexity in their mode

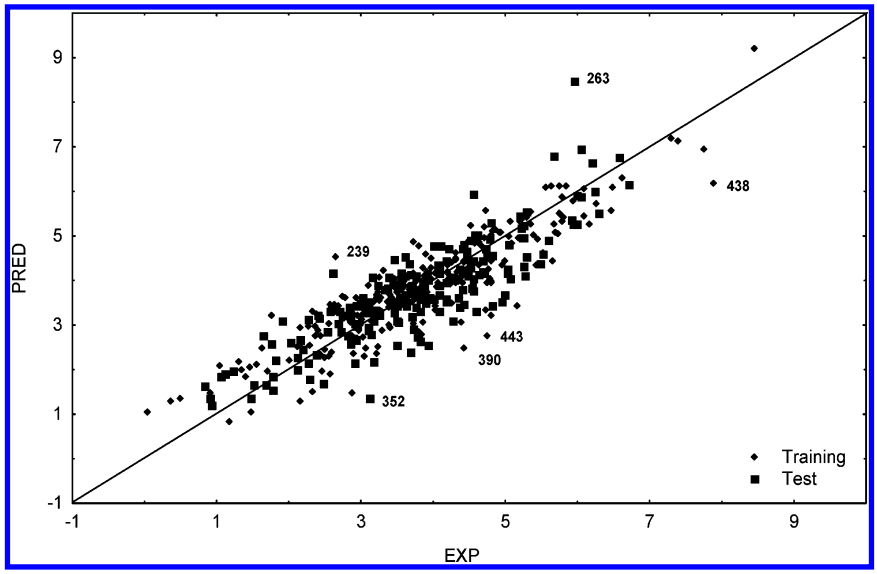

Figure 4. Regression line of the final externally validated general-DTP logP-based model developed on a training set of 249 compounds (eq 8).

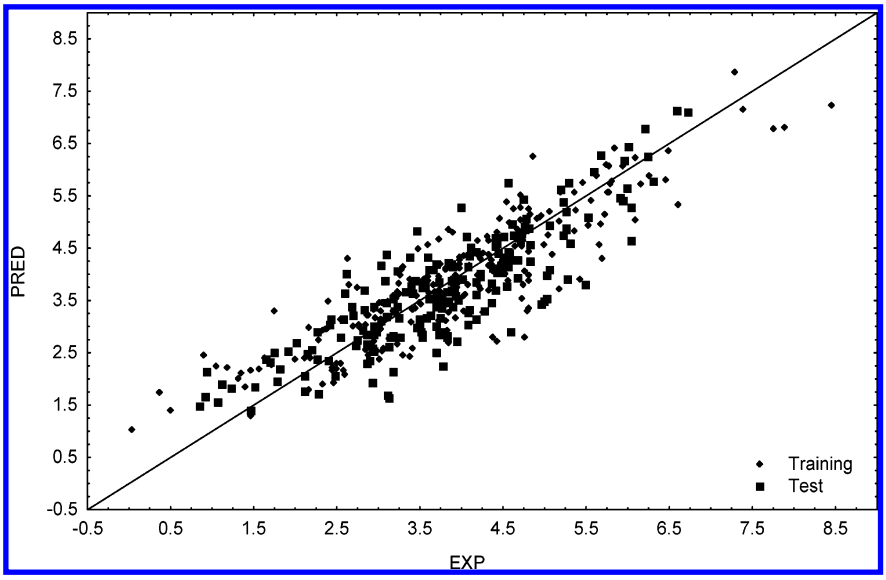

Figure 5. Regression line of the final externally validated general-DTP logP-free model developed on a training set of 249 compounds (eq 9).

of toxic action can affect model performance or force the selection of theoretical descriptors more able to describe their peculiar structure and/or toxicological behavior.

The best log-P free General-DTP model is (regression line in Figure 5)

$$\text{Log}\,(1/\text{LC}_{50})_{96h} = -2.54(\pm 0.4) + 0.91(\pm 0.06)\text{WA} + 6.2(\pm 0.6)\text{Mv} + 0.08(\pm 0.01)\text{H-046} + 0.22(0.03)\text{nCb} - 0.19(\pm 0.04)\text{MAXDP} - 0.33(\pm 0.06)\text{nN} \quad (9)$$

$$n_{\text{training}} = 249;\ n_{\text{test}} = 200;\ R^2 = 0.79;$$
$$Q^2_{\text{LOO}} = 0.78;\ Q^2_{\text{BOOT}} = 0.78$$

$$Q^2_{\text{EXT}} = 0.71;\ \text{SDEC} = 0.595;\ \text{SDEP} = 0.613$$

$$\text{RMSE}_{\text{training}} = 0.38;\ \text{RMSE}_{\text{test}} = 0.64;$$
$$K_{\text{X}} = 34.81;\ K_{\text{XY}} = 39.94$$

Good values for internal fitting and predictivity and a satisfying Q^2_{EXT} value confirm the reliability of the model in prediction, even when the mode of toxic action cannot be defined "a priori". This result appears even more appreciable considering the high number of chemicals in the test set (almost 50% of the complete data set) for the statistical external validation.

As expected the theoretical descriptors selected in this model are a combination of global structural features, able to represent the high structural heterogeneity of the training and test sets: WA (topological descriptor representing the mean Weiner index), Mv (mean atomic van der Waals volume), nCb- (number of C sp2 in substituted benzenes), H-046 (H attached to C−O sp3), MAXDP (maximal electrotopological positive variation), and nN (number of nitrogen atoms). The information related to dimensional features is condensed in WA and Mv that are related to the electronic distribution in MAXDP, while some counters (nN, nCb, and H-046) are mainly needed to model some peculiar chemicals in the data set.

On analyzing the model AD in the Williams plot, it is interesting to note that most of the chemicals falling outside

Table 2. Comparison of the RMSE Values of Specific Models for Each Considered MOA with Those Calculated for Different Classes of Toxicity by the General-DTP Model

MOA	model	RMSE training	RMSE test
M	LogP-based	0.29	0.34
O	LogP-free	0.55	0.64
A	DTP LogP-based	0.48	0.57
1	DTP LogP-free	0.56	0.52
M	LogP-based	0.31	0.44
O	LogP-free	0.40	0.37
A	DTP LogP-based	0.45	0.47
2	DTP LogP-free	0.42	0.43
M	LogP-based	0.54	0.56
O	LogP-free	0.47	0.54
A	DTP LogP-based	0.82	0.94
3	DTP LogP-free	0.89	0.92
M	LogP-based	0.67	0.33
O	LogP-free	0.69	0.56
A	DTP LogP-based	0.72	0.98
4	DTP LogP-free	0.86	0.93

the domain belong to specific acting compound class, nicotine sulfate again being the most influential chemical in the training set. The other structural high leverage chemicals in the training set are as follows: 2,2′-methylenebis(3,4,6-trichlorophenol), hexachloro-1,3-butadiene, pentachloropyridine, rotenone, and 2,6-di-*tert*-butyl-4-methylphenol. Only chloroacetonitrile is a slight outlier in the training set. The test set chemicals falling outside the structural domain of the model are as follows: tetrachloroethylene, hexachloroethane, 2,4,5-tribromoimidazole, 3-amino-5,6-dimethyl-1,2,4-triazine, caffeine, and pentabromophenol. Clearly, the predicted toxicity values for high leverage chemicals should be considered extrapolated, thus such values must be considered unreliable.

The models presented above have very similar RMSE values, but it is important to note the bigger difference between the RMSE values calculated for the training and test sets of models (8) and (9) than those obtained for the models developed on separate MOAs.

This result is not surprising considering the high heterogeneity in the structural features and modes of toxic action in the complete data set. Structural heterogeneity can affect model accuracy in prediction significantly more than it can in a model developed for a specific class of toxicity.

Table 2 allows a comparison of the RMSE values of specific models for each MOA with those calculated for the different classes of toxicity by the general-DTP model (predicted values for chemicals excluded from the previously developed DTP model, having been identified as outliers for the model AD, have not been considered in this calculation). It is possible to note slight differences in the RMSE values for the training and test sets, for either the specific or the general model considering MOA 1 and 2. On the contrary higher differences in the RMSE values in either the training or the test set are evident among the general model and the specific models for MOA 3 and 4. Thus, it is evident that the performance of the general model is comparable with that of the models specific for toxicity prediction of class 1 and 2 chemicals, while for MOA 3 and 4 the specific models result in more accurate predictions than the general one. Moreover, with the exception of models developed for narcotics (MOA1), there is not such a big difference in performance for logP based and logP free models, but it is quite clear that the accuracy of models developed on a specific MOA tends to be higher than the DTP models, when structural heterogeneity is associated with a strong heterogeneity of mechanisms involved in experimental response determination.

CONCLUSIONS

In this study we verified the effectiveness of Genetic Algorithms as a fast and efficient procedure to select (not by chance) significant variables to manage complex systems, even when large amounts of descriptors are used as input variables, such as different theoretical molecular descriptors and traditional "mechanistic interpretable" variables (logP and quantum chemical descriptors).

The applied GA-VSS procedure confirmed the relevance of the partitioning property to model the toxic power of various chemicals, and the result is consistent with mechanistic assumptions about logP as a good modeler of toxicity, mainly narcosis.

It was also demonstrated that the selection of the best logP from those considered here (AlogP, ClogP, MlogP) is not a simple choice, since it seems that it is necessary to use the appropriate one for each modeled MOA, while AlogP and ClogP have the best similar performance if applied to the complete data set. It is also important to note that in general, for each MOA class, the descriptors of the logP-free models have high correlation precisely with the specific logP selected by the GA-VSS in the corresponding logP-based models (values not commented on in the Results section).

However it is clear that the performances obtained using different logP also depend on the biological and structural features of the modeled data set.

In addition to logP some different theoretical molecular descriptors, calculated by the software DRAGON, can also be selected, improving the quality of the models based only on logP or substituting it, being able to furnish analogous information.

Many QSARs, according to the "a priori" defined MOA for the studied chemicals, were developed by combining different calculated logP and theoretical molecular descriptors or by just using theoretical molecular descriptors; they have been proposed here as predictive tools for acute toxicity, log-($1/LC_{50}$), in *Pimephales promelas*. Moreover, two general-DTP models for direct toxicity prediction, not based on defined modes of toxic action (MOA), have also been proposed as general predictive tools with a wide applicability domain independently of the "a priori" knowledge of the MOA of chemicals. A general analysis of the QSARs proposed here shows that high structural heterogeneity associated with strong heterogeneity of mechanisms involved in experimental response determination can affect the accuracy of a DTP model. Thus if comparable prediction can be obtained by applying MOA-based or DTP models for narcotics and polar narcotics, there is an increase in the accuracy of specific logP-free or logP-based models for the toxicity classes of reactive and specific acting compounds.

Finally the most important point: since the function of ecotoxicological QSARs is to predict toxicity accurately, the models must be validated for their predictivity on chemicals not used in model development. Thus, as the recently defined

OECD principles[8] require, all the obtained models were also statistically evaluated for their external predictive power and verified for their applicability domain.

Supporting Information Available: Information on data set, experimental and predicted data, and training and test set chemicals. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Zeeman, M.; Auer, C. M.; Clements, R. G.; Nabholz, J. V.; Boethling, R. S. U.S.EPA Regulatory perspectives on the use of QSAR for new and existing chemical evaluations. *SAR QSAR Environ. Res.* **1995**, *3*, 179−201.

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

(3) Bradbury, S. P.; Russom, C. L.; Ankley, G. T.; Schultz, T. W.; Walker, J. D. Overview of data and conceptual approaches for derivation of Quantitative Structure−Activity Relationships for ecotoxicological effects of organic chemicals. *Environ. Toxicol. Chem.* **2003**, *22*, 8, 1789−1798.

(4) European Commission. White Paper on a strategy for a future Community Policy for Chemicals, 2001. http://europa.eu.int/comm/enterprise/reach/.

(5) Toussaint, M. W.; Shedd, T. R.; Van der Schalie, W. H.; Leather, G. R. A comparison of standard acute toxicity tests with rapid-screening toxicity tests. *Environ. Toxicol. Chem.* **1995**, *14*, 5, 907−915.

(6) Kubinyi, H. From Narcosis to Hyperspace: The History of QSAR. *Quant. Struct.-Act. Rel.* **2002**, *21*, 348−356.

(7) http://ecb.jrc.it/QSAR/.

(8) http://www.oecd.org/document/23/0,2340,en_2649_201185_33957015_1_1_1_1,00.html.

(9) Tropsha, A.; Gramatica, P.; Gombar, V. K. The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR Comb. Sci.* **2003**, *22*, 1−9.

(10) Eriksson, L.; Jaworska, J.; Worth, A.; Cronin, M.; McDowell, R. M.; Gramatica, P. Methods for reliability, uncertainty assessment, and applicability evaluations of regression based and classification QSARs. *Environ. Health Perspect.* **2003**, *111*, 10, 1361−1375.

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

(12) Escher, B. I.; Hermens, J. L. M. Modes of Action in Ecotoxicology: Their Role in Body Burdens, Species Sensitivity, QSARs, and Mixture Effects. *Environ. Sci. Technol.* **2002**, *36*, 20, 4201−4217.

(13) Bradbury, S. P. Predicting modes of toxic action from chemical structure: An overview. *SAR QSAR Environ. Res.* **1994**, *2*, 89−104.

(14) Verhaar, H. J. M.; Van Leeuwen, C. J.; Hermens, J. L. M. Classifying environmental pollutants. 1: Structure−activity relationships for prediction of aquatic toxicity. *Chemosphere* **1992**, *25*, 4, 471−491.

(15) Verhaar, H. J. M.; Solbé, J.; Speksnijder, J.; Van Leeuwen, C. J.; Hermens, J. L. M. Classifying environmental pollutants: Part 3. External validation of the classification system. *Chemosphere* **2000**, *40*, 875−883.

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

(17) Schultz, T. W.; Cronin, M. T. D.; Walker, J. D.; Aptula, A. O. Quantitative structure−activity relationships (QSARs) in toxicology: a historical perspective. *J. Mol. Struct.-Theochem.* **2003**, *622*, 1−22.

(18) Posthumus, R.; Slooff, W. *Implementation of QSARs in ecotoxicological risk assessments*; RIVM report 601516003; 2001.

(19) Dearden, J. C. Prediction of Environmental Toxicity and Fate Using Quantitative Structure−Activity Relationships (QSARs). *J. Braz. Chem. Soc.* **2002**, *13*, 6, 754−762.

(20) Schultz, T. W.; Cronin, M. T. D.; Netzeva, T. I. The present status of QSAR in toxicology. *J. Mol. Struct.-Theochem.* **2003**, *622*, 23−38.

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

(22) Pontolillo, J.; Eganhouse, R. P. The Search for Reliable Aqueous Solubility (S_w) and Octanol−Water Partition Coefficient (K_{ow}) Data for Hydrophobic Organic Compounds: DDT and DDE as a Case Study. U.S. Geological Survey, Water-Resources Investigations Report 01-4201; USGS: Reston, VA, 2001.

(23) Renner, R. The K_{ow} controversy. Doubts about the quality of basic physicochemical data for hydrophobic organic compounds could be undermining many environmental models and assessments. *Environ. Sci. Technol.* **2002**, *36*, 410A−413A.

(24) Benfenati, E.; Gini, G.; Piclin, N.; Roncaglioni, A.; Varì, M. R. Predicting logP of pesticides using different software. *Chemosphere* **2003**, *53*, 1155−1164.

(25) Mannhold, R.; Petrauskas, A. Substructure versus Whole-molecule Approaches for Calculating LogP. *QSAR Comb. Sci.* **2003**, *22*, 4, 466−475.

(26) http://clogP.pomona.edu/medchem/chem/clogP/index.html.

(27) Klopman, G.; Li, J. K.; Wang, S.; Dimayuga, M. Computer Automated log P calculations based on an extended group contribution approach. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 752−781.

(28) Kaiser, K. L. E. The use of neural networks in QSARs for acute aquatic toxicological endpoints. *J. Mol. Struct.-Theochem.* **2003**, *622*, 85−95.

(29) http://www.epa.gov/med/databases/fathead_minnow.htm.

(30) Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M. DRAGON Software for the Calculation of Molecular Descriptors − ver. 5.2 for Windows, Talete S.r.l., Milan, Italy, 2005.

(31) HYPERCHEM/CHEMPLUS ver. 7.03 for Windows, Autodesk Inc., Sausalito, CA, U.S.A. 2002.

(32) Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: 2000; p 667.

(33) Consonni, V.; Todeschini, R.; Pavan, M. Structure/Response Correlation and Similarity/Diversity Analysis by GETAWAY descriptors. Part 1. Theory of the Novel 3D Molecular Descriptors. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 682−692.

(34) Consonni, V.; Todeschini, R.; Pavan, M.; Gramatica, P. Structure/Response Correlation and Similarity/Diversity Analysis by GETAWAY descriptors. Part 2. Application of the novel 3D-molecular descriptors to QSAR/QSPR studies. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 693−705.

(35) Todeschini, R.; Ballabio, D.; Consonni, V.; Mauri, A.; Pavan, M. MOBY DIGS − MOdels BY Descriptors In Genetic Selection − ver. 1 beta for Windows, Talete S.r.l., Milan, Italy, 2004.

(36) Leardi, R.; Boggia, R.; Terrile, M. Genetic Algorithms as a Strategy for Feature Selection. *J. Chemom.* **1992**, *6*, 267−281.

(37) Todeschini, R.; Maiocchi, A.; Consonni, V. The K Correlation Index: Theory Development and its Application in Chemometrics. *Chemom. Int. Lab. Syst.* **1999**, *46*, 13−29.

(38) Zupan, J.; Noviè, M.; Ruisánchez, I. Kohonen and counterpropagation artificial neural networks in analytical chemistry. *Chemom. Intell. Lab. Syst.* **1997**, *38*, 1−23.

(39) Todeschini, R. KOALA − Kohonen Artificial LAyers − ver. 1.1 for Windows, Talete S.r.l., Milan, Italy, 2001.

(40) Atkinson, A. C. *Plots, Transformation and Regression*; Clarendon Press: Oxford, U.K., 1985; p 282.

(41) SCAN − Software for Chemometric Analysis, ver. 1.1 for Windows, Minitab U.S.A., 1995.

(42) Rose, K.; Hall, L. H. E-State modeling of fish toxicity independent of 3D structure information. *SAR QSAR Environ. Res.* **2003**, *14*, 113−129.

(43) Sixt, S.; Altschunh, J.; Brüggemann, R. Quantitative structure-toxicity relationship for 80 chlorinated compounds using quantum chemical descriptors. *Chemosphere* **1995**, *30*, 12, 2397−2414.

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

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

(46) Guha, R.; Serra, J. R.; Jurs, P. C. Generation of QSAR sets with a self-organizing map. *J. Mol. Graphics Modell.* **2004**, *23*, 1−14.

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

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

(49) Karabunarliev, S.; Mekenyan, O. G.; Karcher, W.; Russom, C. L.; Bradbury, S. P. Quantum-chemical Descriptors for Estimating the Acute Toxicity of Electrophiles to the Fathead minnow (*Pimephales promelas*): An Analysis Based on Molecular Mechanisms. *Quant. Struct.-Act. Rel.* **1996**, *15*, 302−310.

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

CI050212L