Predicting the aquatic toxicity mode of action using logistic regression and linear discriminant analysis

Article *in* SAR and QSAR in environmental research · September 2016

DOI: 10.1080/1062936X.2016.1229691

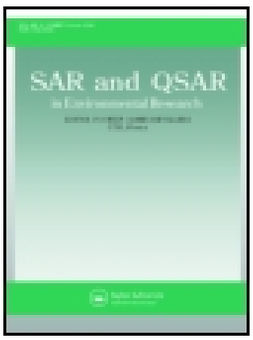
CITATIONS
10

READS
205

6 authors, including:

Yueying Ren
Lanzhou Jiaotong University
17 PUBLICATIONS **242** CITATIONS

SEE PROFILE



SAR and QSAR in Environmental Research

ISSN: 1062-936X (Print) 1029-046X (Online) Journal homepage: http://www.tandfonline.com/loi/gsar20

Predicting the aquatic toxicity mode of action using logistic regression and linear discriminant analysis

Y. Y. Ren, L. C. Zhou, L. Yang, P. Y. Liu, B. W. Zhao & H. X. Liu

To cite this article: Y. Y. Ren, L. C. Zhou, L. Yang, P. Y. Liu, B. W. Zhao & H. X. Liu (2016): Predicting the aquatic toxicity mode of action using logistic regression and linear discriminant analysis, SAR and QSAR in Environmental Research, DOI: 10.1080/1062936X.2016.1229691

To link to this article: http://dx.doi.org/10.1080/1062936X.2016.1229691

SAR AND QSAR IN ENVIRONMENTAL RESEARCH, 2016
http://dx.doi.org/10.1080/1062936X.2016.1229691

Taylor & Francis
Taylor & Francis Group

Predicting the aquatic toxicity mode of action using logistic regression and linear discriminant analysis

Y. Y. Ren[a], L. C. Zhou[b], L. Yang[a], P. Y. Liu[a], B. W. Zhao[a] and H. X. Liu[c]

[a]School of Environmental and Municipal Engineering, Lanzhou Jiaotong University, Lanzhou, P.R. China; [b]College of Chemistry and Chemical Engineering, Lanzhou University, Lanzhou, P.R. China; [c]School of Pharmacy, Lanzhou University, Lanzhou, P.R. China

ABSTRACT

The paper highlights the use of the logistic regression (LR) method in the construction of acceptable statistically significant, robust and predictive models for the classification of chemicals according to their aquatic toxic modes of action. Essentials accounting for a reliable model were all considered carefully. The model predictors were selected by stepwise forward discriminant analysis (LDA) from a combined pool of experimental data and chemical structure-based descriptors calculated by the CODESSA and DRAGON software packages. Model predictive ability was validated both internally and externally. The applicability domain was checked by the leverage approach to verify prediction reliability. The obtained models are simple and easy to interpret. In general, LR performs much better than LDA and seems to be more attractive for the prediction of the more toxic compounds, i.e. compounds that exhibit excess toxicity versus non-polar narcotic compounds and more reactive compounds versus less reactive compounds. In addition, model fit and regression diagnostics was done through the influence plot which reflects the hat-values, studentized residuals, and Cook's distance statistics of each sample. Overdispersion was also checked for the LR model. The relationships between the descriptors and the aquatic toxic behaviour of compounds are also discussed.

ARTICLE HISTORY
Received 2 June 2016
Accepted 22 August 2016

KEYWORDS
Logistic regression; linear discriminant analysis; modes of aquatic toxic action; CODESSA; DRAGON; influence plot

Introduction

The environment is regularly exposed to an increasing and diverse variety of industrial chemicals, and many of these exposures may adversely affect human health or the environment. Therefore it is very important to evaluate the potential environmental impact of a particular chemical prior to its release into an ecosystem [1]. In order to assess hazard and risk of chemical substances to marine and freshwater organisms living in the water column, information derived from studies on the physical chemical characteristics of compounds and from laboratory-based toxicity tests on aquatic toxicity is required [2]. Study of the relationships between a compound's structure and its aquatic toxicity has become a subject of great importance in environmental sciences and toxicology, and significant progress has been

CONTACT Y. Y. Ren renry04.my265@aliyun.com
The supplementary material for this paper is available online at http://dx.doi.org/10.1080/1062936X.2016.1229691.

made in classifying chemical compounds according to their mode/mechanism of toxicity and in screening them for assessment of their environmental risk. Nevertheless, as Seward et al. pointed out, the classification of chemicals according to mode/mechanism of action is not an easy task [3], since it is based not solely on the chemical itself but on an understanding of the interaction between the chemical and the living organism.

Mode of action (MOA) is always a difficult concept; it is likely to be species and compound dependent, probably also concentration dependent, and can be defined at different tropic levels. Aquatic toxicity MOAs have been classified in different ways [4–6]. Four modes of action were described in Verhaar et al. [6] as class I, non–polar narcosis (inert) chemicals which exhibit baseline or minimum toxicity; class II, polar narcosis (less inert) chemicals, which exhibit effects similar to those of non-polar narcotics, but their toxicities are slightly higher than baseline toxicity; class III, reactive chemicals; and class IV, specifically acting chemicals. This classification system could not afford the prediction for new chemicals with measured toxicity data. Based on joint toxic action studies, the establishment of toxicodynamic profiles, and observed 96-h LC50 values (50% lethality concentration) for the fathead minnow (*Pimephales promelas*) assay, Russom et al. proposed new classifications of modes of action as narcosis (non-polar, polar and ester), uncoupling of oxidative phosphorylation, respiratory inhibition, electrophilic/nucleophilic reactivity mechanisms, acetylcholinesterase inhibition and central nervous system seizure/stimulant mechanisms [4]. The need to distinguish between non-polar and polar narcosis has long been argued. Some authors suggest that there may be no difference between polar and non-polar narcosis mechanisms, based on a good correlation between the measured membrane potential effects (pEC^{W}_{50}) values and the pLC_{50} values which cover both polar and non-polar narcotics [7, 8]. However, later studies on the same data in the above two works [7, 8] proved that there is a real mechanistic difference between non-polar and polar narcotics, which is manifested by (1) significantly different quantitative structure–activity relationships (QSARs), even when based on log K_{ow}, for non-polar and polar narcotics treated separately; (2) differences in fish acute toxicity syndrome (FATS); and (3) non-additivity between general narcotics and polar narcotics in mixture toxicity studies [9, 10].

Besides direct experimental classification like concentration addition in joint toxic action models [4] or the surrogate testing in simple organisms or in glass, theoretical methods could provide/derive structural rules or information affecting the aquatic toxic MOA and therefore are very helpful in the assessment of hazards of substances. QSARs utilizing chemical structure-based descriptors have the benefit of avoiding needless expenditure, reducing the number of animals sacrificed during testing, and saving time on synthesizing and marketing drug candidates that may be found to cause severe effects later. In addition, they can be used in predictive toxicology because the descriptors can be calculated relatively easily with commercially available computer software. In the field of aquatic toxicology, QSARs have developed as scientifically credible tools for predicting the acute toxicity of chemicals when few or no empirical data are available [11], and QSAR studies have been carried out to predict the aquatic toxicity MOA of chemicals in the past decades. It is expected that once the MOA of a new chemical entity is predicted, the QSAR for the particular MOA can be used to estimate the toxic potential of the chemical more effectively [12]. The ability to correctly identify the MOA of a compound is critical, since a poor choice of MOA could lead to significant overestimation or underestimation of toxicity. The so developed MOA-based QSAR can be also applied to extrapolate toxic effects across species and exposure regimes when

limited experimental data are available [4, 6, 13]. An exhaustive review by Netzeva et al. [14] gave not only the approaches utilized in the literature for estimating the aquatic toxicity of chemical substances, but also directions and recommendations for further research in this field. Typically, such studies involved the modelling of aquatic toxicity for particular chemical classes, for example phenols, to *Tetrahymena pyriformis* [13, 15–20]. The authors utilized different statistical methods, such as linear discriminant analysis (LDA) [16], binary logistic regression (LR) [16, 17], radial basis function neural networks (RBFNNs) and support vector machines (SVM) [18], decision tree models [19], and Adaboost [20], and achieved 86–89% overall correct classification for the validation data sets. Also, other researchers have performed QSAR modelling of the toxic MOA for more diverse data sets than particular group-based chemicals like phenols [6, 12, 21–30]. Basak et al. [12] used 60 topological indices and neural networks and discriminant analysis to classify 283 chemicals according to the fathead minnow aquatic toxicity MOA. Rates of correct classification ranged from 65% to 95% for these chemicals. The results demonstrated that electrophiles and proelectrophiles were not satisfactorily separated from narcotics. Michielan et al. [28] extended the data set in Basak et al. [12] to 617 chemicals and derived a robust classification model by combining SVM analysis with 18 toxicokinetic-like descriptors and toxicodynamic-like descriptors. Very satisfactory results were obtained for the training set, with rates of correct classification ranging from 88.6% (electrophile or proelectrophile phosphorylation) to 100% (e.g. uncoupler of oxidative phosphorylation and central nervous system seizure mechanisms) for these chemicals. However, the results for the test set were not so good, with rates of correct classification ranging from 21.56% (electrophile or proelectrophile phosphorylation) to 66.67% (baseline narcosis). By using discriminant analysis and LR, Ren and Schultz [22] developed QSAR models to distinguish between narcotic and reactive compounds for 88 samples based on four calculated hydrophobicity and electrophilicity descriptors. Logistic regression and discriminant analysis gave small total classification error rates of 10.23% and 11.22%, respectively. Shortly after, this group performed a similar study [24] to predict aquatic toxic MOA of 337 compounds (actually there are 329 samples) using the hydrophobicity index (the logarithm of octanol–water partition coefficients, log K_{ow}) and the experimental toxicity data obtained from the *Pimephales promelas* and the *Tetrahymena pyriformis* assays as discriminating variables. In separating non-polar narcotic compounds from other compounds, discriminant analysis and LR both resulted in a total classification error rate of approximately 23%. However, in separating less reactive compounds from more reactive compounds, discriminant analysis and LR resulted in higher total error rates of 37% and 30%, respectively. Only 40.62% (39/96) and 54.17% (52/96) more reactive compounds were correctly classified by discriminant analysis and LR, respectively. Ren also utilized solute descriptors and discriminant analysis to classify and predict the non-polar, polar and ester narcosis toxicity mechanisms for 194 organic compounds [23]. Martin et al. [29] used LDA and random forest (RF) to develop models for assigning aquatic toxicity MOA: acetylcholinesterase (AChE) inhibition, nicotinic acetylcholine receptor (nAChR) agonism, narcosis, neurotoxicity, reactivity and uncoupling oxidative phosphorylation. The majority of the MOAs were predicted accurately (the LDA and RF models produced overall prediction accuracies ranging from 84.5 to 87.7% for the validation set) with the exception of reactive compounds and nAChR agonists, especially the unsatisfied differentiation between polar and non-polar narcosis. Ensemble learning methods-based classification and regression QSAR models such as decision treeboost and decision tree forest were also developed using the algae (*Pseudokirchneriella subcapitata*)

experimental toxicity data, and were applied to predict algae, daphnids, fish and bacteria toxicity of wide groups of chemicals [30]. The developed QSAR models exhibited good predictive and generalization abilities in different test species of varied trophic levels. Analysis of the two works by Ren et al. [22, 24] indicates that the predictor variables in the latter work are not good enough (at least for the discrimination between the less reactive compounds and the more reactive compounds). Also, the successful use of calculated descriptors in their former work demonstrates that it is very necessary to introduce other kinds of variables, for example other physicochemical parameters, toxicity data of other assays, or chemical structure-based descriptors to encode as much information contributing to the aquatic toxic action as possible.

In the present study, a new investigation was carried out for the same data studied by Ren et al. [24] with the main goal of building reliable classification QSAR models to predict the mode of toxic action of chemicals, so that the QSAR of the chemical class with correct MOA can be used in the estimation of their potential toxicity. Thus, essentials accounting for a reliable model should be considered carefully according to the Organization for Economic Co-operation and Development (OECD) principles [31]. In order to interpret the structural characteristics of compounds from all aspects, CODESSA software [32] and DRAGON software [33] were used to calculate different kinds of molecular descriptors. For the purpose of comparison with the results in Ren et al. [24], the logarithm of octanol–water partition coefficients (log K_{ow}) and two well-defined and quantifiable toxicity endpoints – the 96-h LC50 values for the fathead minnow (*Pimephales promelas*) assay and IGC50 values (50% inhibitory growth concentration) for the protozoans (*Tetrahymena pyriformis*) assay – were treated as independent variables. The best predictor variables were selected from a combined pool of three independent variables and the chemical structure-based descriptors. Logistic regression and discriminant analysis were chosen to do statistical analysis because (1) they are regularly used in common practice to perform supervised classification and prediction; and (2) both methods provide directly the posterior probability that a compound falls into a certain group and predictions are made based on this probability. The availability of such probabilities also provides a means of overcoming one of the shortcomings of models that employ cut-off values to classify chemicals, namely, the treatment of borderline predictions [34]. This is particularly useful in avoiding making false classifications. In addition, an important problem in QSAR studies is model validation, and the models in the abovementioned three works [22–24] were not validated externally. As Gramatica stated, 'evidence is presented that only models that have been validated externally, after their internal validation, can be considered reliable and applicable for both external prediction and regulatory purposes' [35], therefore the generalization ability of these models for the compounds that were not used in the model development was uncertain and thus limited their applications, for example the regulatory assessment of chemicals. To assess how the different models will accurately classify the type of chemicals for a given input, in the present study model predictive ability was validated both internally and externally, as is requested by the OECD principles for validation [31] and recommended and/or used by many researchers [29, 30, 35–44]. The applicability domain (AD) of the models was checked using the leverage approach to verify prediction reliability [35]. Optimal models were selected on the basis of predictive/classifier accuracy in the validation data which has been used as the main, and often only, evaluation criterion for the predictive performance of classification learning algorithms. All models were validated in two ways: an internal leave-one-out cross-validation

(LOO–CV) procedure combined with a test set. The results are discussed in the light of the main factors that influence the property under investigation and its modelling. Finally, a separate subset of the data, which was collected from other literature [4, 28] and therefore was excluded during the model training process, was used as external validation data to check the generalization ability of optimum model. Therefore, respective QSAR models were constructed first to separate non-polar narcotic compounds that exhibit baseline toxicity from other compounds that exhibit excess toxicity, and then to separate less reactive compounds from more reactive compounds. We hope that the obtained classification models have comparative predictive ability for aquatic toxicity MOA of the studied chemicals and, furthermore, that they can give some insight into what structural features are related to the respective aquatic toxicity MOA.

Materials and methods

Experimental data and generation of molecular descriptors

Compounds involved in this study consist of 482 observations. In order to compare the predicted results with those in Ren et al. [24], the log K_{ow} values, the experimental toxicity data from the *Pimephales promelas* and the *Tetrahymena pyriformis* assays as well as the MOAs of these compounds were taken from published papers [4, 24, 28] (salts, chemicals without the CAS number, undefined structure, or without an 'a priori' defined MOA were excluded from this study). Since log K_{ow} values were taken from a range of literature sources, there might by some uncertainty among these data, and this needs to be checked. For this purpose, we also calculated clog*P* values for these compounds using ChemDraw software [45]. Examination of these data revealed that except for a few compounds, log K_{ow} and clog*P* are almost the same for each compound. The largest difference between log K_{ow} and clog*P* is for *p*-Chlorophenyl-*o*-nitrophenyl ether (log K_{ow} = 4.74, clog*P* = 8.368). Since this compound is in the external data set and would not be involved in the model development, its reported log K_{ow} value is retained. Therefore, the reported log K_{ow} values were used in this study. The compounds were 'a priori' classified as non-polar narcotics (MOA 1), polar narcotics (MOA 2), less reactive chemicals (including MOA 2, polar narcosis; MOA 3, ester narcosis and MOA 4, amine narcosis), or more reactive compounds (including MOA 5, weak acid respiratory uncoupling; MOA 6 electrophilicity; MOA 7, pro-electrophilicity; and MOA 8, nucleophilicity) according to Ren et al. [24], based on the 96-h LC50 values (mmol) for the fathead minnow (*Pimephales promelas*) assay. Of them, eight replicate compounds in Ren et al. [24] were discarded and the remaining 329 observations were used in the present study. Besides these 329 ones, a further 153 compounds taken from the literature [4, 28] were used as external test data to check the generalization ability of optimum models. The structures, log K_{ow} values, toxicity data and the MOAs of these compounds are shown in Table S1 in Supplementary Materials (available via the Supplementary Content tab on the article's online page).

Bioreactivity can be thought of as the electronic or steric interaction between a chemical compound and a biological system. To search for the structural features affecting the aquatic toxic MOA it is necessary to calculate the molecular structural features of the studied chemicals. Due to the diversity of the molecules, a wide set of different descriptors was calculated. CODESSA, developed by the Katritzky group [32], can calculate more than 500 different

constitutional, topological, geometrical, electrostatic, quantum-chemical and thermodynamical descriptors. Since CODESSA cannot calculate some new generated 3D descriptors, DRAGON software, developed by Todeschini et al. in the Milano Chemometrics group [33], was used for this purpose. Twenty-nine blocks with a total of more than 4800 descriptors can be calculated by using DRAGON 6.0 [33]. Therefore, the molecular descriptors were calculated using the software CODESSA 2.0 [32] and DRAGON 6.0 [33] on the minimal energy conformations optimized with AM1 method in Hyperchem 6.0 [46]. A total of 447 descriptors were generated using the CODESSA 2.0 package. In DRAGON, 1663 descriptors were calculated, belonging to different logical molecular descriptors blocks, for example constitutional indices, information indices, matrix-based descriptors (2D and 3D), autocorrelations (2D and 3D), P_VSA-like descriptors, atom-centred fragments, charge descriptors, ETA indices, edge adjacency indices, geometrical descriptors, atom pairs (2D and 3D), 3D-MoRSE descriptors, WHIM descriptors, GETAWAY descriptors and molecular properties, etc. These descriptors encode various aspects of structural information, for example information on the shape, branching, symmetry, distribution of charges, quantum-chemical properties, thermodynamic properties and the inter-atom connections of the molecule, etc. Many of the calculated descriptors carry redundant or highly correlated information and their existence would result in a chance correlation during the construction of the model. Therefore, descriptors with constant (or near constant) values or missing values, or that are highly correlated with other descriptor(s) were removed in a pre-reduction step, and 1125 descriptors were left and used in subsequent variable selection.

The whole data set was split into two subsets. The first subset was used as training data to obtain appropriate models, and the second subset was retained as test data. Subset selection is of crucial importance in the development and validation of reliable QSARs. The rational division of the training set and the test set should satisfy, on one side, the diversity of the training set, which is a necessary condition for the construction of a QSAR model applicable to further compounds of interest in the same chemical domain and, on the other side, the closeness of the representative points of both the training set and the test set in the descriptor space to ensure a proper validation of the model [47]. Principal component analysis (PCA) is a very useful method proved by our previous works to assist the data splitting [48, 49]. Therefore, PCA was performed in the present study, using descriptors calculated by CODESSA and DRAGON, to detect the homogeneities in the data set (i.e. to identify possible outliers and clusters), and to show spatial location of compounds to analyse the relationship between the compounds corresponding to training set and the test set as well. Throughout this paper, 67% of the original data was chosen as our training data and the remaining 33% as the test data.

Linear discriminant analysis

LDA is a pattern-recognition method providing a classification model based on the combination of variables that best predicts the category or group to which a given compounds belongs. The basic theory of LDA is to classify the dependent variable by dividing an *n*-dimensional descriptor space into two regions that are separated by a hyperplane defined by a linear discriminant function. In this study, the variables used to compute the linear discriminant function are chosen in stepwise manner using the Wilks' Lambda criterion: at each step, the variable that minimizes the overall Wilks' Lambda is entered. The *F* value, which

can be calculated using Wilks' Lambda, allows the assessment of relative importance among the candidate variables. This criterion was used for entering and removing variables: a variable is entered into the model if its *F* value is greater than the Entry value (F_{min}); thereafter *F* values of the rest variables in the model are recalculated and those with *F* values less than the Removal value (F_{max}) are removed. This procedure is continued until *F* values of the rest of the variables are all less than the defined F_{min}. The quality of the discriminant equation was evaluated using Wilks' U-statistical parameter (Wilks' Lambda), a multivariate analysis of variance that tests the equality of group means for the variables in the discriminant equation. The Wilks' Lambda for the overall discrimination can take values in the range of 0 (perfect discrimination) to 1 (no discrimination). Statistical analyses were performed using the SPSS statistical software.

Logistic regression

LR is within the framework of generalized linear models (GLMs) and is frequently applied to situations in which the response variable (*Y*) is dichotomous (for example, 1/0, yes/no, passed/failed, lived/died etc.). Unlike the standard linear model, there is no requirement that *Y* is normally distributed. Instead, it can be assumed that *Y* follows a distribution that is a member of the exponential family (such as the binomial, Poisson, gamma, or inverse-Gaussian families of distributions, etc.) [50], and that you can fit a linear model of the form

$$\ln\left(\frac{\pi}{1-\pi}\right) = \alpha + \sum_{j=1}^{p} \beta_j X_j \tag{1}$$

where $\pi = \mu_Y$ is the conditional mean of *Y* (i.e. the probability that $Y = 1$ given a set of *X* values), $\pi/(1-\pi)$ is the odds that $Y = 1$, and $\ln(\pi/(1-\pi))$ is therefore the natural logarithm of the odds that $Y = 1$, or called *logit* of π. The *logit* is symmetric around 0, and unbounded both above and below, making it a good candidate for the response variable in a linear-like model [51, 52].

LR aims at establishing the relationship between the *logits* and the predictor variables. Predictor variables are pre-specified functions of the explanatory variables and therefore may include quantitative explanatory variables, transformation of quantitative explanatory variables, polynomial or regression-spline regressors, interactions, and so on [52]. The obtained equation is linear in its parameters ($\alpha, \beta_1, \ldots \beta_p$), which give the change in *logits* in the response variable for a unit change in the predictor variable and help to see how each predictor variable influences the probability of the response, holding all other predictor variables to their average values. *Logits* are difficult to interpret; however, you can exponentiate them to put the results on an odds scale. So, increasing X_i by one unit changes the *logit* by βj and multiplies the odds of $Y = 1$ by $e^{\beta j}$, holding the other *X*s fixed. After specifying the link function and the probability distribution, the parameters and consequently the probabilities are derived through iterative maximum likelihood estimation procedure. In this case, $\ln(\pi/(1-\pi))$ is the link function and the probability distribution is binomial [53, 54].

The assessment of model adequacy is as important for GLMs as it is for standard linear models (OLS). However, there is less agreement in the statistical community regarding appropriate assessment procedures. In general, most of the standard diagnostics for linear models extend relatively straightforwardly to GLMs, taking advantage of the computation of

maximum likelihood or quasi-likelihood estimates for GLMs by iterated weighted least squares (WLS) [52, 55]. Available diagnostics include a screening for unusual observations – namely outliers, high-leverage observations and influential observations – as well as the multicolinearity of the predictor variables. Since there is no general consensus on cut-off values for identifying problematic observations, statistic values have to be judged relative to each other. The most commonly used approach in this situation is to create index plots for each statistic and search for unusually large values. An omnibus plot (namely influence plot) combines studentized residuals, hat-values and Cook's distances for the logistic model, and is used for this purpose [56]. This plot can also be used to visualize the AD based on both the leverages and residuals. In this graph, the horizontal axis is the leverage, the vertical axis is the studentized residual, and the plotted symbol is proportional to the Cook's distance for the observations. Outliers are observations that are not predicted well by the model. Roughly speaking, an observation with a standardized residual greater than 3 or less than –3 is worth attention. A Bonferroni outlier test can be utilized as assistance for this purpose, which tests the single largest absolute studentized residual for significance as an outlier [56]. High-leverage observations have an unusual combination of predictor values and are identified through the hat-values. For a given data set, the average hat-value (h^*) is p/n, where p is the number of parameters in the model (including the intercept) and n is the sample size. A rough rule of thumb is that hat-values that are larger than 2 or 3 times the average hat-value should be examined. Influential observations are observations that have a disproportionate impact on the determination of the model parameters. Generally, Cook's distance values greater than $4/(n - k - 1)$, where n is the sample size and k is the number of predictor variables, indicate influential observations. However, Kabacoff found that a cut-off of 1 is more generally useful than $4/(n - k - 1)$ when searching for influential observations [53]. High-leverage observations may or may not be influential observations. That will depend on whether they are also outliers.

Another issue in assessing the adequacy of GLMs is overdispersion. Overdispersion occurs when the observed variance of the response variable is larger than what would be expected from a binomial distribution. It can lead to distorted test standard errors and inaccurate tests of significance. Therefore overdispersion is also checked for the obtained model, using the ratio of the residual deviance with the residual degrees of freedom in the obtained model as criterion [53, 57]. Due to the great computational requirement, the number of variables is limited in performing LR. Therefore, variable selection was done through LDA. The LR method was written in R-language utilized in *R* software [58] based on *car* package [56]. R is free, open-source software offering all manners of data analytic techniques.

Results and discussion

Separating non-polar narcotics and other compounds

Results of LDA

The whole data set was divided into a training set of 220 compounds to build the models and a test set of the remained 109 compounds to evaluate its prediction ability. In addition, it was verified that each set contained exactly the same ratio of polar/ non-polar compounds (83/137 = 0.606 for training set, 41/68 = 0.603 for test set). PCA gives 122 significant principal components (PCs) with eigenvalues >1. Of them, the first two PCs explain 49.20% of the

variation in the data (36.48% and 12.72%, respectively). The distribution of compounds over the first two PCs is shown in Figure S1 in Supplementary Materials (available online). As can be seen in this figure, the compounds in each subset seem to be relatively well balanced over the space of the PCs, and the training set is not very different from the test set and can represent the test set well. It can confirm the representative ability of the compounds in each subset during data set splitting. For the purpose of modelling, a value of 1 was assigned to non-polar narcotic compounds and a value of 0 was assigned to other compounds. The LOO–CV procedure was used to evaluate the predictive ability of the model.

In LDA, the prior probabilities were computed from group size (0.377 for polar compounds and 0.623 for non-polar compounds). One- to eight-parameter models were obtained by a stepwise discriminant analysis (see Table S2 in Supplementary Materials, available online); the last model containing eight parameters has the smallest Wilks' Lambda and it was chosen as the best linear discriminant function (shown in Table S3 in Supplementary Materials, available online). Statistical parameters of this model were as follows: $n = 220$, $F = 23.857$, Wilks'-$\lambda = 0.528$. Eigenvalue = 0.894, Canonical Corr = 0.687, Chi-square =136.696, $p < 0.0001$. Cross-validation results indicated that this model was able to correctly classify 83.18% of the 220 compounds in the training set; 65 of the 83 (78.31%) non-polar narcotic compounds and 118 of the 137 (86.13%) compounds exhibiting excess toxicity were correctly classified, respectively. For the test set, 31 of the 41 (75.61%) non-polar narcotic compounds and 62 of the 68 (91.18%) compounds exhibiting excess toxicity were correctly classified, respectively. With a total of 16 misclassifications in 109 classifications, the total predicts accuracy of the LDA for the test set is 85.32%. In summary, 83.89% of the whole data set compounds were correctly classified; 96 of 124 (77.42%) non-polar narcotic compounds and 180 of 205 (87.80%) compounds exhibiting excess toxicity were correctly classified, respectively.

Results of LR

The LR method was then performed in *R* software [56, 58] using the above eight variables. The *p*-values for the regression coefficients (not shown) indicate that three descriptors, namely nOHt, C038 and CATS2D-06-PL, may not make a significant contribution to the equation (we cannot reject the hypothesis that the parameters are 0). Therefore these descriptors were discarded and a new LR was performed with the remaining five predictors. According to the analysis of variable effects, each regression coefficient in the reduced model is statistically significant at 99% confidence level. Because the two above LR models are nested, a chi-square test was applied to compare them. The resulting non-significant chi-square value (*p*-value = 0.542) suggests that the reduced LR model with five predictors fits almost as well as the eight-predictor LR model, reinforcing our belief that nOHt, C038 and CATS2D-06-PL do not add significantly to the prediction above and beyond the other variables in the equation. Therefore, the five-predictor LR model was chosen to do further research (shown in Table 1).

Using the regression parameters of involved predictors, *logit* can be calculated for each chemical. *Logit* represents the log ratio of the probability that a compound exhibits baseline toxicity to the probability that it exhibits excess toxicity. The influence of each predictor on this ratio is reflected by the respective coefficient of predictors in the model. The positive signs of the coefficients indicate that this ratio increases as toxicity increases, whereas the negative signs of the coefficients indicate that this ratio decreases as toxicity increases. This model has Cox–Snell r^2 of 0.419, Nagelkerke r^2 of 0.573 and Akaike Information Criterion

Table 1. Logistic regression model in separating nonpolar narcotics and other compounds.

Predictor variables	*Meaning*	*B*	*S.E.*	*z-value*	*df*	*Sig.*	*Exp(B)*
log K_{ow}	Logarithm of the octanol–water partition coefficient	1.116	0.326	3.422	1	0.001	3.053
log ($1/LC_{50}$)	Logarithm of the 50% lethality concentration for the *Pimephales promelas* assay (mmol).	−1.826	0.402	−4.545	1	0.000	0.161
IC1	The first order neighbourhood Information Content	−1.088	0.429	−2.540	1	0.011	0.337
MATS1s	2D autocorrelation at lag 1	4.694	1.348	3.482	1	0.000	109.289
SAacc	All acceptor atom surface area	−0.035	0.010	−3.452	1	0.001	0.966
Constant		2.830	1.205	2.349	1	0.019	16.945

The output provides the deviances, regression parameters, standard errors, and tests that these parameters are 0. Note that each of the predictor variables is significant at the $p < 0.05$ level.

(AIC) of 183.43. A new LDA model based on the above five predictor variables is shown in Table S4 in the Supplementary Materials (available online).

The correlation matrix (see Table S5 in the online Supplementary Materials) shows that strong (negative) correlations exist between log ($1/LC_{50}$) and log K_{ow}. This does not mean that selection of the two variables into the model is irrational because discriminate analysis and LR are not based on correlations. The table also shows that the correlation coefficients between several variables (e.g. MATS1s and IC1, MATS1s and SAacc, log K_{ow} and IC1, etc.) are not statistically significant, implying that these quantities are orthogonal to each other. A summary of the predictions of MOAs using this LR model as well as the results by the new LDA are presented in Table 2. As seen in Table 2, 60 of the 83 (72.29%) non-polar narcotic compounds and 130 of the 137 (94.89%) compounds exhibiting excess toxicity were correctly classified, respectively. With a total of 30 misclassifications in 220 classifications, the total training accuracy of the LR model for the training set was 86.4%. Compared with the results by new LDA model, prediction for compounds exhibiting excess toxicity is much better (i.e. 94.89% by LR model versus 82.48% by the new LDA model).

The predictor variables involved in the above models include both experimental toxicity data (i.e. log K_{ow}, log ($1/LC_{50}$)) and chemical structure-based descriptors (IC1, SAacc and MATS1s). log K_{ow} represents the potential of a compound to bioaccumulate as residues in aquatic organisms from the surrounding medium, and is frequently used to quantify the hydrophobicity of chemical compounds [59]. The positive signs of log K_{ow} in the model indicate that the probability of a compound exhibiting baseline toxicity increase as its hydrophobicity increases. This is in good accordance with the mechanism of narcotic compounds, which are not reactive, do not interact with specific receptors in an organism [6] and are inert. Therefore narcotic toxicity is considered to be dependent entirely on the hydrophobicity of the compounds. The log ($1/LC_{50}$) is the logarithmic values of the $1/LC_{50}$ (50% lethal concentration) for the *Pimephales* assays. Log K_{ow} and log ($1/LC_{50}$) are also included in the models by Ren et al. [24] to group chemical compounds, their signs of the coefficients being the same as those in our *logit* model. The three chemical structure-based descriptors were calculated using DRAGON software. The first-order neighbourhood Information Content (IC1) is a neighbourhood symmetry index calculated for a hydrogen-included molecular graph and based on neighbour degrees and edge multiplicity [60]. This index represents a measure of structural complexity per vertex. The all acceptor atom surface area (SAacc), an approximation to the sum of van der Waals surface areas of pure hydrogen bond acceptors

Table 2. Summary of the results of linear discriminant analysis and logistic regression in separating non-polar narcotic compounds from other compounds.

	Data set	*Training set*			*Test set*			*Complete set*			*External set*		
Method	*Class*	*1*	*2*	*Total*	*1*	*2*	*Total*	*1*	*2*	*Total*	*1*	*2*	*Total*
LDA[*1]	1	64	24		28	6		92	30		74	0	
	2	19	113		13	62		32	175		41	11	
	Accuracy (%)	77.11	82.48	80.45	68.29	91.18	82.57	74.19	85.36	81.16	64.35	100	67.46
LR[*1]	1	60	7		27	2		87	9		75	0	
	2	23	130		14	66		37	196		40	11	
	Accuracy (%)	72.29	94.89	86.36	65.85	97.06	85.32	70.16	95.61	86.02	65.22	100	68.25
Non-DA[*2]	1							140	6				
	2							3	38				
	Accuracy (%)							97.90	86.36	95.18			
LDA[*3]	1							90	43				
	2							36	168				
	Accuracy (%)							71.43	79.62	76.6			
LR[*3]	1							78 (**77**) [*4]	25 (**24**) [*4]				
	2							48(**47**) [*4]	186(**181**) [*4]				
	Accuracy (%)							61.9(**62.1**) [*4]	88.15(**88.3**) [*4]	78.3 (**78.4**) [*4]			

Class 1: non-polar narcotics.
Class 2: other compounds.
[*1]Methods used in the present study.
[*2]Non-linear discriminant analysis used in Ren et al. [23]. Prediction for ester narcosis toxicity mechanism was not included.
[*3]Methods used in Ren et al. [24].
[*4]Results of LR using three variables in Ren et al. [24] by our calculation after removing eight duplicated compounds (bold items in brackets).

(not counting acidic atoms and atoms that are both hydrogen bond donors and acceptors such as -OH), is calculated by adding atomic surface areas on the basis of the formula implemented by P_VSA-like descriptors [61]. MATS1s is a 2D autocorrelation at lag 1, calculated by applying Moran coefficient to the molecular graph weighted by atomic properties as intrinsic state. It is related to the dimension and shape of the molecules [62]. These variables account for the electronic and steric interaction between a chemical compound and a biological system and, combined together with the two experimental toxicity data results, give a good result for the classification of non-polar narcotic compounds and other compounds.

The effect of predictor variables on *logits* can be seen from the odds ratios (namely the Exp(B), last column in Table 1). It means that the odds ratio of a compound exhibits baseline toxicity increase with MATS1s and log K_{ow}, and decrease with IC1, SAacc and log ($1/LC_{50}$), respectively. A unit increase in a compound's MATS1s or log K_{ow} makes it 109.29 or 3.051 times, respectively, more likely that this compound exhibits baseline toxicity, holding the other variables fixed. A unit increase in IC1 or SAacc or log ($1/LC_{50}$) makes it 0.337 or 0.965 or 0.161 times, respectively, less likely that it exhibits baseline toxicity or, in other words, 2.967 (1/0.337) or 1.036 (1/0.965) or 6.211 (1/0.161) times more likely that it exhibits excess toxicity, holding the rest of the variables fixed. Because the predictor variables cannot equal 0, the intercept is not meaningful in this case.

In addition, the impact of each predictor variable on the outcome can be explored easily in another term, namely the probability. For example, as IC1 increases from 0.72 (minimum) to 3.54 (maximum), the probability of a certain compound exhibiting baseline toxicity decreases from 0.6751 to 0.0879 (holding MATS1s, SAacc, log K_{ow} and log ($1/LC_{50}$) fixed). Conversely, the probability of a certain compound exhibiting baseline toxicity would increase from 0.0049 when log K_{ow} is –1.83 (minimum) to 0.9941 when log K_{ow} is 7.54 (maximum), holding the other variables fixed. This is consistent with the negative or positive coefficients of IC1 or log K_{ow} in the equation.

Figure 1 is the influence plot of the LR model. Here you can see that acrylamide has largest studentized residual (2.974), carbon tetrachloride has high leverage, and carbon tetrachloride and chloroacetonitrile have a large influence on the parameters estimates of the model. These compounds are particularly unusual when it comes to their predictor values. Acrylamide has a much lower MATS1s and log K_{ow} than other compounds, while having higher values for IC1 and SAacc. As mentioned above, the odds ratio estimate of MATS1s is 109.29. This indicates that if a unit increase in a compound's MATS1s multiplies the odds of compound exhibiting baseline toxicity by 109.29, a 10-unit increase would increase the odds by a factor of 109.29^{10}, or 2.43×10^{20}, holding the other predictor variables fixed. In this sense, it is not surprising that acrylamide has the largest residual. Nevertheless, acrylamide cannot be recognized as a statistically significant outlier (indeed, the outlier test shows no studentized residuals with Bonferonni $p < 0.05$). Carbon tetrachloride has smallest IC1, MATS1s and SAacc, while having a much higher log K_{ow} than most compounds. Chloroacetonitrile has much higher MATS1s and log ($1/LC_{50}$), while having lower log K_{ow} and SAacc than most compounds. These compounds are atypical compared with the other observations; however, none of the observations would appear to be an influential observation (given a criterion of Cook's distance $D = 1$) [53]. Therefore, none of these compounds were deleted from the model.

In addition, the ratio of the residual deviance with the residual degrees of freedom in this model is 0.8011 (171.43/214), which is smaller than 1, indicating that no overdispersion occurs.

When the test set consisting of 109 compounds was applied to the above model to evaluate the generalization performance using selected descriptors of its own space, satisfactory results were obtained (presented in Table 2). As seen in Table 2, 27 of the 41 (65.85%) non-polar narcotic compounds and 66 of the 68 (97.06%) compounds exhibiting excess toxicity were correctly classified, respectively. With a total of 16 misclassifications, the prediction accuracy for the test set is 85.32%. For the whole data set, 87 of the 124 (70.16%) non-polar narcotic compounds and 196 of the 205 (95.61%) compounds exhibiting excess toxicity were correctly classified, respectively. With a total of 46 misclassifications in 329 classifications, the total prediction accuracy is 86.02%. Compared with the new LDA model mentioned previously, the LR model shows much better prediction for compounds exhibiting excess toxicity (94.89% versus 82.48% for the training set, 97.06% versus 91.18% for the test set and 95.61% versus 85.36% for the whole set). In addition, the calculated leverage values for all of the compounds in the test set were below the critical value of 0.082 (3*6/220) and the residuals of these compounds were all less than 3, with malononitrile bearing the largest residual of 2.856. This demonstrates that all of the test set compounds fall within the AD of the training set and, consequently, also indicates that the predicted response is the outcome of interpolation of the model and therefore may be reliable.

In order to consider the differences between groups in the training set, the same predictors were used to perform a new LR model for the whole data set. Predictive accuracy obtained with the cross-validation procedure was used for the evaluation of prediction capabilities of this model. Eighty-three of the 124 (66.94%) non-polar narcotic compounds and 194 of the 205 (94.63%) compounds exhibiting excess toxicity were correctly classified. With a total of 52 misclassifications in 329 classifications, the total prediction rate of the LR is 84.19%. The above results were very similar to those obtained based on one training set/test set.

Detailed comparison of the predicted results by our LR model and that by the LR model in Ren et al. [24] are not practical, because (1) predicted results for each of the compounds in Ren et al. [24] are not accessible; (2) the compounds used to develop models in Ren et al. [24] and here in our study are not identical to each other; and (3) the statistical results presented in Ren et al. [24] must have some deviation from the real ones (though this may be very small), because of the several duplicated compounds in their work [24]. For the purpose of comparison, after removing the eight duplicated chemicals from the data set we re-ran the procedures described by Ren et al. [18] and developed a new LR model using the same three variables as in Ren et al. [24]. The results are included in Table 2. It shows that our LR model (Table 1) outperforms the so developed LR model (using three variables in Ren et al. [24]) in every aspect.

Another aspect that should be considered is the misclassified samples. Table 3 showed a full list of 47 commonly misclassified samples in two models, new LDA and LR, together with their a priori toxicity mechanism designations in the literature [4,5,24]. Among these compounds, 15 compounds have low confidence levels of C, and for two compounds (anthranilamide and 1-Benzoylacetone) no confidence was associated with assessments for which there were insufficient data. This indicates that the mechanisms assigned a priori to these compounds are questionable and may not be accurate. For some compounds, mechanism

Table 3. Misclassifications in LDA and LR in the separation of non-polar narcosis and other compounds that exhibiting excess toxicity.

	MOAs assigned a priori			
Compound	*Russom et al. [4]*	*Ren et al. [22–24]*	*MOA predicted* [*2]	*Data set*
Ethyl carbamate	non-polar narcosis (B) [*1]	non-polar narcosis	0	test
Benzamide	non-polar narcosis (C) [*1]	non-polar narcosis	0	train
α,α-Dimethylbenzene propanol	non-polar narcosis (B) [*1]	non-polar narcosis	0	test
2-Undecanone	non-polar narcosis (B) [*1]	non-polar narcosis	0	test
5-Chloro-2-pyridinol		non-polar narcosis	0	train
1-Hexen-3-ol	electrophilicity/proelectrophilicity (B)[*1]	non-polar narcosis	0	train
4-Bromophenyl-3-pyridylketone	non-polar narcosis (B) [*1]	non-polar narcosis	0	test
Acrylamide		non-polar narcosis	0	train
3-Nitrotoluene	non-polar narcosis (C) [*1]	non-polar narcosis	0	test
Dibutylfumarate	ester narcosis (C) [*1]	non-polar narcosis	0	train
2-Methyl-3-butyn-2-ol	non-polar narcosis (B) [*1]	non-polar narcosis	0	train
2-Phenyl-3-butyn-2-ol	non-polar narcosis (B) [*1]	non-polar narcosis	0	test
4-Pentyn-2-ol	electrophilicity/proelectrophilicity (B) [*1]	non-polar narcosis	0	train
3-Bromothiophene	electrophilicity/proelectrophilicity (B) [*1]	non-polar narcosis	0	test
Anthranilamide	non-polar narcosis (D) [*1]	non-polar narcosis	0	train
2-Acetyl-1-methylpyrrole	electrophilicity/proelectrophilicity (C) [*1]	non-polar narcosis	0	test
2',3',4'-Trichloroacetophenone	non-polar narcosis (C) [*1]	non-polar narcosis	0	train
2',4'-Dichloroacetophenone	non-polar narcosis (A) [*1]	non-polar narcosis	0	train
4-Nitrophenylphenylether	non-polar narcosis (C) [*1]	non-polar narcosis	0	train
Triphenylphosphate		non-polar narcosis	0	test
Carbontetrachloride		non-polar narcosis	0	train
Chloroacetonitrile	respiratory inhibition (C) [*1]	non-polar narcosis	0	train
Malononitrile	respiratory inhibition (C) [*1]	non-polar narcosis	0	test
1-Benzoylacetone	electrophilicity/proelectrophilicity (D) [*1]	non-polar narcosis	0	test
Acetaldoxime		non-polar narcosis	0	train
2-Methylimidazole	electrophilicity/proelectrophilicity (B) [*1]	non-polar narcosis	0	train
2,4,5-Tribromoimidazole	electrophilicity/proelectrophilicity (C) [*1]	non-polar narcosis	0	test
2-Methyl-2,4-pentanediol	non-polar narcosis (B) [*1]	non-polar narcosis	0	train
Methyl-4-chlorobenzoate	ester narcosis (A) [*1]	non-polar narcosis	0	train
3-Methylindole	electrophilicity/proelectrophilicity (B) [*1]	non-polar narcosis	0	train
Flavone	non-polar narcosis (B) [*1]	non-polar narcosis	0	train
3-(3-Pyridyl)-1-propanol	polar narcosis (C) [*1]	non-polar narcosis	0	test
N-Methylaniline	non-polar narcosis (C) [*1]	polar narcosis	1	train
2-Cyanopyridine	polar narcosis (A) [*1]	polar narcosis	1	train
Pyridine	polar narcosis (A) [*1]	polar narcosis	1	train
Hexylaldehyde	electrophilicity/proelectrophilicity (B) [*1]	ester narcosis	1	train
Methyl acetate	ester narcosis (A) [*1]	ester narcosis	1	train
Ethyl hexanoate	ester narcosis (C) [*1]	ester narcosis	1	test
Ethyl acetate	polar narcosis (A) [*1]	ester narcosis	1	train
(*tert*)Butylacetate	non-polar narcosis (C) [*1]	ester narcosis	1	train
Amylamine		amine narcosis	1	test
Hexylamine		amine narcosis	1	train
4-Phenoxybenzaldehyde	non-polar narcosis (C) [*1]	electrophilicity	1	train
2-Methylvaleraldehyde	electrophilicity/proelectrophilicity (B) [*1]	electrophilicity	1	test
2-Tolualdehyde	non-polar narcosis (C) [*1]	electrophilicity	1	train
Hexanoicacid		nucleophilicity	1	test
N,N-Diethylethanolamine	non-polar narcosis (B) [*1]	nucleophilicity	1	train

[*1]Level of confidence associated with an observed mode of action. A, high; B, moderate; C, low; D, No confidence was associated with assessments for which there were insufficient data; [*2]MOA predicted in this study by LDA model and LR model. 1, nonpolar narcosis; 0, other compounds.

designations in different literature sources are contradictory; for example, 1-hexen-3-ol, 4-pentyn-2-ol, 3-bromothiophene, 2-acetyl-1-methylpyrrole, 2-methylimidazole and 3-methylindole were assigned the electrophilicity/pro-electrophilicity mechanism with moderate confidence level of B in Russom et al. [4], while in the study of Ren et al. [24] these compounds were assigned the non-polar narcosis mechanism (MOA 1) according to Schultz et al. [5]. On the contrary, 2-tolualdehyde and N,N-diethylethanolamine were assigned the non-polar narcosis mechanism in Russom et al. [4], while in the study of Ren et al. [24] they were assigned the electrophilic reactivity mechanism (MOA 6) and nucleophilicity mechanism (MOA 8), respectively. Even methyl-4-chlorobenzoate and ethyl acetate, which were assigned the ester narcosis mechanism and polar narcosis mechanism with high confidence levels of A (the highest level) in Russom et al. [4], were assigned the non-polar narcosis mechanism (MOA 1) and ester narcosis (MOA 3) mechanism, respectively, in the study of Ren et al. [24]. In addition, due to the continuum between chemicals that act non-reactively and those that act reactively [59, 63, 64], overlapping of the MOAs should be anticipated. Because of disagreement, overlapping and low-level confidence in the a priori mechanism designations, it is hard to judge whether the toxicity mechanism predictions for such kinds of compounds are correct or not. In the remaining compounds, 2-methyl-2,4-pentanediol, ethyl carbamate, 2-cyanopyridine, pyridine and 2-methylvaleraldehyde can be classified correctly by LR model based on the three variables used in Ren et al. [24] (our calculation, results not shown). This implies that the misclassifications of these compounds here are due to the selected predictors that are not able to characterize the structural features of these compounds. Comparatively, the three variables (log K_{ow}, log (1/IGC50), and log (1/LC50)) used in Ren et al. [24] are better than the five variables (IC1, MATS1s, SAacc, log K_{ow} and log 1/LC50) in our LDA and LR model for predicting these compounds. Nevertheless, for most compounds investigated in this study, the five descriptors in our LDA and LR model are the better combination, and characterize the structural features of the studied compounds related to the aquatic toxic action more effectively and thus provide better prediction for their MOAs. Besides, it has been revealed that the false positive predictions for excess toxicity compounds are distributed mainly in three MOAs, that is, ester narcosis (MOA 3, with five compounds out of 23), amine narcosis (MOA 4, with two compounds out of 13) and nucleophiles (MOA 8, with two compounds out of five). All of the weak acid respiratory uncoupling compounds (MOA 5) and pro-electrophilic compounds (MOA 7) were predicted correctly by the LDA and LR model. In this sense, it may also imply that the five descriptors in our LDA and LR model better characterize the structural information related to the respective aquatic toxic action for weak acid respiratory uncoupling compounds (MOA 5) and pro-electrophilic compounds (MOA 7) than that for the ester narcosis (MOA 3) and nucleophiles (MOA 8). In addition, predictions using discriminant analysis and LR are based on the posterior probability that a compound falls into a certain group. If, for example, the probability of being a non-polar narcosis compound and the probability of being a polar narcosis compound were very close, it is hard to make a decision without additional information because the degree of confidence would be low no matter what classification was made.

Based on the above observations, it is concluded that the LR model is statistically significant, robust and has acceptable predictivity. A separate subset of 126 compounds collected from the literature [4] was then used as external validation data to check the generalization ability of this model. Seventy-five of the 115 (65.22%) non-polar narcotic compounds with baseline toxicity and all 11 compounds exhibiting excess toxicity were correctly classified

(shown in Table 2). Detailed predicted results for MOA of these compounds are included in Table S1 in Supplementary Materials (available online). It is interesting to note that among the nine compounds with log K_{ow} larger than 5, only three, 3-(3,4-dichlorophenoxy)benzaldehyde, (log K_{ow} = 5.49) 3-(4-*tert*-butylphenoxy)benzaldehyde (log K_{ow} = 5.93) and 4,4'-isopropylidene-bis-(2,6-dichlorophenol) (log K_{ow} = 6.44) in the external validation set were misclassified, whereas others in the training set and test set were all correctly classified. In addition, *p*-chlorophenyl-*o*-nitrophenyl ether was correctly predicted when its log K_{ow} value (4.79) was replaced by the clog*P* value (8.368), holding other variables fixed. It may reasonably be concluded that the log K_{ow} value for *p*-chlorophenyl-*o*-nitrophenyl ether is underestimated. On the other hand, this also demonstrates that log K_{ow} has a relationship with toxic effect. The partitioning of organic substances into the lipid membrane is a thermodynamically spontaneous process. Because sorption is a complex of both fast and slow processes, the chemical activity of a contaminant – like its accessibility – may slowly decrease over time [65]. Therefore, compounds with log K_{ow} values larger than 5 would show deviations from linearity in the standardized tests.

Separating less reactive and more reactive compounds

Results of LDA

In separating less reactive and more reactive compounds, the same procedure was applied. For each case, the dependent variable takes the value of 1 for less reactive compounds, and 0 otherwise. The 205 compounds were divided into a training set of 139 (67.80%) compounds and a test set of the remaining 66 (32.20%). In addition, it was verified that each set contained exactly the same ratio of less reactive/ more reactive compounds (77/62 = 1.24 for the training set and 35/31 = 1.29 for the test set). PCA gives 88 significant PCs (eigenvalues >1), with the first two PCs explaining 49.12% of the variation in the data (31.65% and 17.47%, respectively). Distribution of all 205 compounds over the first two PCs is shown in Figure S2 in the online Supplementary Materials. The figure shows that the compounds in each subset seem to be relatively well balanced over the space of the PCs, and the training set can represent the test set well. The prior probabilities were 0.556 for less reactive and 0.444 for more reactive compounds, respectively, which were computed from group size. A stepwise discriminant analysis gave a four-variable model (see Table S6 in Supplementary Materials, available online) with following statistical parameters: $n = 139$, $F = 28.104$, Wilks'-$\lambda = 0.544$, Eigenvalue = 0.839, Canonical Corr = 0.675, Chi-square = 82.24, $p < 0.0001$. It produced classification rates of 84.17%, 86.36% and 84.88% for the training set, the test set and the whole set, respectively. Less reactive compounds classification rates were 89.61%, 85.71% and 88.39%, while more reactive compounds classification rates were 77.42%, 87.10% and 80.65% in the training set, the test set and the whole set, respectively.

Results of LR

The correlation matrix (see Table S7 in the online Supplementary Materials) shows that the correlation coefficients between selected variables are not statistically significant except for RPCG and P_VSA_MR3, implying that these quantities are orthogonal to each other. These descriptors were used to develop a LR model thereafter, shown in Table 4. Each regression coefficient in the obtained model is statistically significant ($p < 0.05$). This model has a Cox–Snell r^2 of 0.445, Nagelkerke r^2 of 0.596 and AIC of 119.15. *Logit*, which represents the log

Table 4. Logistic regression model in separating less reactive and more reactive compounds.

Predictors	*Meaning*	*B*	*S.E.*	*z-value*	*df*	*Sig.*	*Exp(B)*
P_VSA_MR3	P_VSA-like descriptors for atomic molar refractivity 3	−0.033	0.010	−3.181	1	0.001	0.967
nArOH	number of aromatic hydroxyls in a molecule	1.490	0.517	2.881	1	0.004	4.437
NdsCH	the count of atom-type -CH=	−2.499	0.599	−4.173	1	0.000	0.082
RPCG	The relative positive charge	−19.709	4.460	−4.419	1	0.000	0.000
Constant		4.751	0.967	4.914	1	0.000	115.699

ratio of the probability that a compound being a less reactive compound to the probability that the compound is a more reactive compound, can be easily calculated using this model.

The predictor variables were all calculated using DRAGON software. The relative positive charge (RPCG) is the partial charge of the most positive atom divided by the total positive charge (Qpos) of the compound [66]. This descriptor was developed and used to account for the effects of polar intermolecular interactions. It also encodes indirect information on the size of the molecule via the sum of the partial positive charges, and thus shows the importance of the electrostatic interactions between the molecules. In the present study, RPCG was also computed by the CODESSA package. Comparison of the two sets values of RPCG (i.e. values calculated respectively by CODESSA and DRAGON software) reveals that for most compounds the RPCG values are identical to each other. Only a few compounds show a very small deviation of about 0.01 units. The two sets of RPCG values show a high correlation coefficient of 0.997. Here, RPCG calculated by DRAGON software was selected into the model, indicating to some extent that it is more suitable for describing these compounds' structural features such as polar intermolecular interactions. NdsCH means the count of atom-type -CH=. Atom-type counts are simple counts of the E-state atom-types in a molecule [67]. P_VSA-like descriptors are defined as the amount of van der Waals surface area (VSA) having a property P in a certain range [68]. Here, P_VSA_MR3 is P_VSA-like descriptors for atomic molar refractivity which are calculated on the basis of Ghose–Crippen group contribution models. nArOH counts the number of aromatic hydroxyls in a molecule. These descriptors encode for the presence of functional groups and bonds, electrostatic and steric interactions between a chemical compound and a biological system, which are relevant to improve the separation of less reactive and more reactive compounds. The odds ratio estimates of RPCG, NdsCH, P_VSA_MR3 and nArOH were 2.756×10^{-9}, 0.082, 0.967 and 4.437, respectively. It means that the odds of compound exhibiting less reactive toxicity increase with nArOH, and decrease with RPCG, NdsCH and P_VSA_MR3, respectively. A unit increase in a compound's nArOH makes it 4.437 times more likely that this compound exhibits less reactive toxicity (holding the other four variables fixed). Or in other words, the probability that a certain compound exhibits less reactive toxicity would increase from 0.4242 when nArOH is 0 (no aromatic hydroxyl existing in a molecule) to 0.9355 when nArOH is 2 (a molecule containing two aromatic hydroxyls), holding RPCG, NdsCH and P_VSA_MR3 fixed.

No overdispersion occurs as the ratio of the residual deviance with the residual degrees of freedom in this model is smaller than 1 (i.e. 109.15/134 = 0.815). The influence plot (Figure 2) shows that catechol has largest studentized residual, 4-(*tert*)Butyl-2,6-dinitrophenol has highest leverage, triethanolamine has the second highest leverage, and catechol, 2,4-Dinitrophenol and tetrachlorocatechol have a large influence on the parameters estimates of the model. Analysis of the predictor values of training set compounds indicates

that catechol has highest value of nArOH, triethanolamine has highest value of RPCG, and 2,4-Dinitrophenol and 4-(*tert*)Butyl-2,6-dinitrophenol have much higher P_VSA_MR3 and nArOH than most compounds. The above observation can give some insight (to some extent) into why these compounds are atypical compared with the others; however, none of them would appear to be an influential compound or a statistically significant outlier. Indeed, the outlier test shows no studentized residuals with Bonferonni $p < 0.05$, with catechol having largest studentized residual of –2.92 ($p = 0.486$). Consequently, all compounds were retained in the model.

The model-produced classification rates are 81.29% and 75.76% for the training set and the test set, respectively (shown in Table 5). Less reactive compounds classification rates were 77.92% in the training set and 60.00% in the test set, while more reactive compounds classification rates were 85.48% and 93.55% in the training set and the test set, respectively. As for all 205 samples, 72.32% less reactive compounds and 88.17% more reactive compounds were correctly classified, respectively. With a total of 42 misclassifications, the total prediction accuracy was 79.51%. Similar to the results for non-polar narcotic/other compounds separation in section 3.1, the LR model (Table 4) shows much better prediction for more reactive compounds (85.48% versus 77.42% for the training set, 93.55% versus 87.10% for the test set and 88.17% versus 80.65% for the whole set, respectively), as compared with the LDA model. For the purpose of comparison, after removing the six duplicated compounds, new LDA and LR models were developed respectively using log K_{ow} and log (1/LC50) in Ren et al. [24]. The results are included in Table 5. Also, compared with the poor prediction results by the LDA model (52.7%) and LR model (54.8%) in Ren et al. [24], our LR model (Table 4) proves to be a more attractive alternative for the prediction of more reactive compounds. In addition, the calculated leverage values for all of the compounds in the test set were below the critical value of 0.108 (3*5/139) and the residuals of these compounds were all less than 3, with 4-Phenylazophenol showing largest residual of –2.775. Therefore, all test set compounds fall within the AD of the training set and, consequently, this indicates that the predicted response is the outcome of interpolation of the model and therefore may be reliable.

The common misclassified 24 samples by the LDA and LR models are listed in Table 6. Of the 13 false negative predictions for the less reactive compounds (MOAs 2, 3, and 4), six are polar narcosis compounds (MOA 2) and the other seven are ester narcosis compounds (MOA 3). All of the amine narcosis compounds (MOA 4) were predicted correctly by the LDA and LR model. In the case of the 11 false positive predictions for more reactive compounds (MOAs 5, 6, 7 and 8), two are weak acid respiratory uncoupling compounds (MOA 5), five are electrophilic compounds (MOA 6) and the other two are pro-electrophilic compounds (MOA 7). All of the nucleophiles (MOA 8) were correctly predicted by our LR model, whereas in the LDA model two compounds were misclassified, indicating that LR is superior to LDA for the prediction of these compounds. Among these misclassified samples, some compounds have low confidence levels of C (e.g. 3,5-Dibromosalicylaldehyde, methyl-4-cyanobenzoate and 2-(Ethylamino)ethanol, etc.) or with no confidence associated with assessments (dimethyl nitroterephthalate), or their toxicity mechanisms (e.g. salicylaldoxime and diethyl chloromalonate, etc.) were not presented in Russom et al. [4]. For some compounds, contradictory a priori mechanisms were assigned in different literature sources, for example 2,4-Dinitrophenol was assigned the oxidative phosphorylation uncoupling with the highest level of A in Russom et al. [4], while in the study of Ren et al. [24] it was assigned ester narcosis (MOA 3) [5, 24]. Catechol was assigned polar narcosis in Russom et al. [4] and pro-electrophilicity (MOA 7)

Table 5. Summary of the results of linear discriminant analysis and logistic regression in separating less reactive and more reactive compounds.

	Data set	*Training set*			*Test set*			*Complete set*			*External set*		
Method	*Class*	*1*	*2*	*Total*	*1*	*2*	*Total*	*1*	*2*	*Total*	*1*	*2*	*Total*
LDA[*1]	1	69	14		30	4		99	18		7	8	
	2	8	48		5	27		13	75		4	19	
	Accuracy (%)	89.61	77.42	84.17	85.71	87.1	86.36	88.39	80.65	84.88	63.64	70.37	68.42
LR[*1]	1	60	9		21	2		81	11		7	6	
	2	17	53		14	29		31	82		4	21	
	Accuracy (%)	77.92	85.48	81.29	60	93.55	75.76	72.32	88.17	79.51	63.64	77.78	73.68
LDA[*2]	1							94 (**94**) [*3]	57 (**49**) [*3]				
	2							21 (**18**) [*3]	39 (**44**) [*3]				
	Accuracy (%)							81.74 (**83.9**) [*3]	40.63 (**52.7**) [*3]	62.56 (**69.8**) [*3]			
LR[*2]	1							96 (**93**) [*3]	44 (**42**) [*3]				
	2							19 (**19**)	52 (**51**) [*3]				
	Accuracy (%)							83.48 (**83.0**) [*3]	54.17 (**54.8**) [*3]	70.14 (**70.2**) [*3]			

Class 1: Less reactive compounds.
Class 2: More reactive compounds.
[*1]Methods used in the present study.
[*2]Methods used in Ren et al. [24].
[*3]Results of LDA and LR model using two variables in Ren et al. [24] by our calculation (bold items in bracket), after removing 6 duplicated compounds.

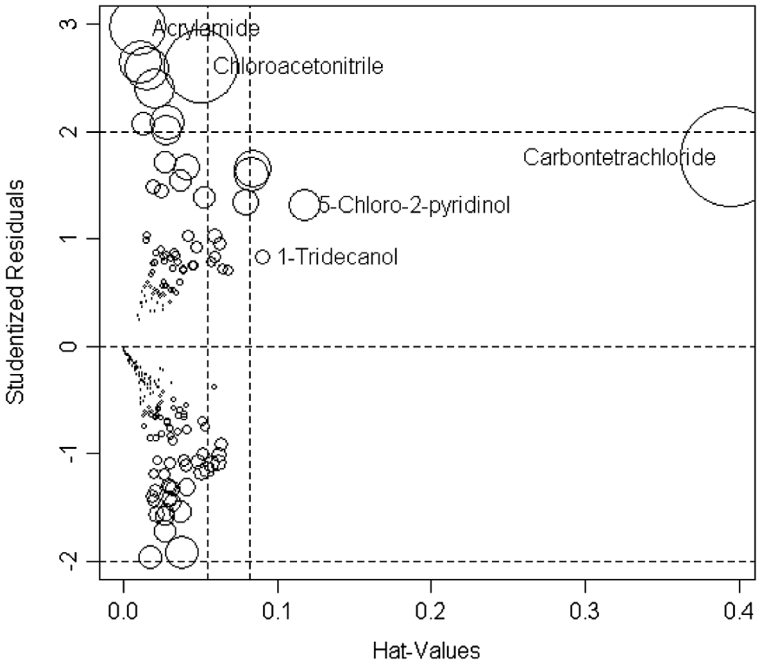


Figure 1. Influence plot of LR model for the separation of non-polar narcotics and other compounds.

in Ren et al. [24]. The disagreement, overlapping and low-level confidence in the a priori mechanism designations make it hard to judge whether the toxicity mechanism predictions for these compounds are correct or not. Taking phenolic compounds as an example, due to the effect of ionization, the toxicity mechanism of these compounds is very complicated, and can be categorized into different groups based on their p*K*a values (*K*a, the first dissociation constant) as suggested by Schultz et al. [5]. Catechol has high p*K*a values (where *K*a is the first dissociation constant) of 9.62 and it should act as a polar narcotic (as assigned by Russom et al. [4]) according to Schultz et al. [5]. In this sense, both LDA and LR models give correct classification for this compound (classified as a less reactive compound). However, catechol was classified as a more reactive compound by the LR model based on the two variables (log K_{ow} and log ($1/LC_{50}$)) used in Ren et al. [24] (our calculation, results not shown). For 2,4,6-Tribromophenol, however, it is not the case. 2,4,6-Tribromophenol has small p*K*a value of 6.31, which indicates that it has potential to act as reactive compound [5], and it was assigned the weak acid respiratory uncoupling mechanism in Ren et al. [24], whereas in Russom et al. [4] it was assigned the non-polar narcosis mechanism with the highest level of A. This compound was classified as less reactive in our LDA and LR models.

Also, the same predictors were used to perform a new LR model for the whole data set in order to consider the difference between groups in the training set. Respectively, 85 of the 112 (75.90%) less reactive compounds and 80 of the 93 (86.02%) more reactive compounds were correctly classified. With a total of 40 misclassifications in 205 classifications, the total prediction accuracy of the LR model is 80.50%. These results were very similar to those obtained based on one training set/test set.

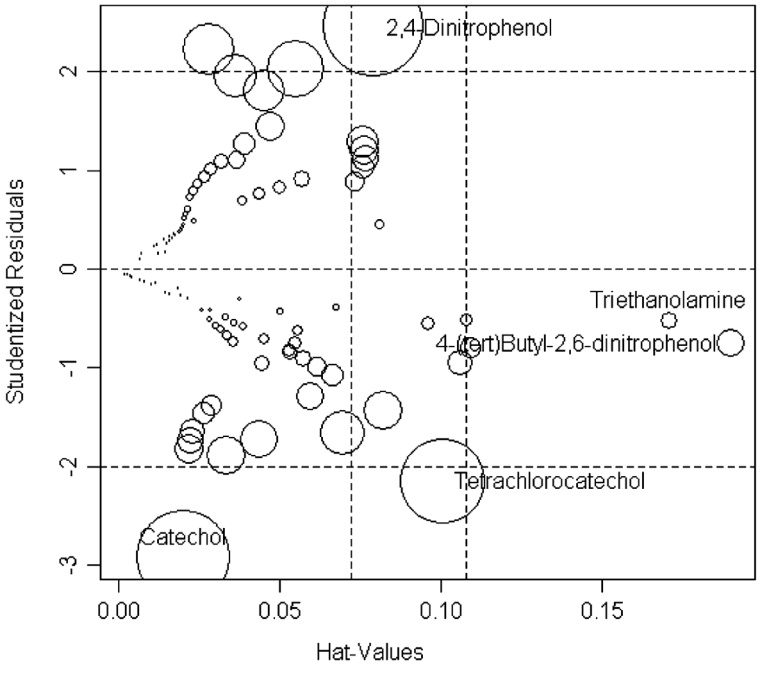


Figure 2. Influence plot of LR model for the separation of less reactive compounds and more reactive compounds.

The above observations demonstrate that the LR model we built is statistically significant, robust and has acceptable predictivity. An external validation data of 38 compounds collected from the literature [4, 28] was then used to check the generalization ability of this model. Satisfactory results were obtained. Seven of the 11 (63.64%) less reactive chemicals and 21 of the 27 more reactive compounds (77.78%) were correctly classified, shown in Table 5. The total classification rate is 73.68%.

The predicted accuracy for non-polar narcotic/other compounds and less reactive/more reactive compounds by the LDA and LR models is listed in Table 2 and Table 5, respectively. As Ren et al. [24] pointed out, there is not a general rule that LR always yields lower error rates than discriminant analysis or vice versa. Results in our study demonstrated this very well. As shown in Table 2, the accuracy for compounds exhibiting excess toxicity is generally better than that for non-polar narcotics in both models, and the results for compounds exhibiting excess toxicity by LR were better than that by LDA, while the results for non-polar narcotics by LR were worse than that by LDA. Also, the accuracy of more reactive compounds is generally better than that of less reactive ones when LR was used, the results for more reactive compounds by LR were better than those by LDA, and the results for less reactive ones by LR were worse than those by LDA (shown in Table 5). However, when it comes to the prediction of more toxic compounds, LR performs better than LDA in both cases. In assessing a toxicity hazard, a false negative is generally considered less acceptable from a risk management standpoint than a false positive. The former may lead to exposure and adverse health effects in human populations, whereas the latter may lead to a less than optimal allocation of resources by either triggering a testing requirement or eliminating or

Table 6. Misclassifications in LDA and LR in the separation of less reactive compounds and more reactive compounds.

	MOAs assigned a priori			
Compound	*Russom et al. [4]* [*1]	*Ren et al. [22–24]*	*MOA predicted* [*2]	*Data set*
3,5-Dibromosalicylaldehyde	electrophilicity/proelectrophilicity (C)[*1]	polar narcosis	0	test
Salicylaldoxime		polar narcosis	0	train
4-Ethoxy-2-nitroaniline	nonpolar narcosis (B)[*1]	polar narcosis	0	train
Methyl-4-cyanobenzoate	nonpolar narcosis (C)[*1]	polar narcosis	0	train
5-Bromosalicylaldehyde		polar narcosis	0	test
5-Chlorosalicylaldehyde		polar narcosis	0	train
2,4-Dinitrophenol	oxidative phosphorylation uncoupling (A) [*1]	ester narcosis	0	train
Hexylaldehyde	electrophilicity/proelectrophilicity (B)[*1]	ester narcosis	0	test
Methyl acetate	ester narcosis (A)[*1]	ester narcosis	0	train
2-(Ethylamino)ethanol	nonpolar narcosis (C)[*1]	ester narcosis	0	test
Ethyl acetate	ester narcosis (C)[*1]	ester narcosis	0	test
Isovaleraldehyde	electrophilicity/proelectrophilicity (B)[*1]	ester narcosis	0	train
Dimethyl nitroterephthalate	ester narcosis (D)[*1]	ester narcosis	0	train
2,4,6-tribromophenol	nonpolar narcosis (A)[*1]	Respiratory uncoupling	1	test
Pentabromophenol	oxidative phosphorylation uncoupling (B)[*1]	Respiratory uncoupling	1	train
4-Phenylazophenol	oxidative phosphorylation uncoupling (A)[*1]	Respiratory uncoupling	1	test
Pentachloropyridine	oxidative phosphorylation uncoupling (A)[*1]	Respiratory uncoupling	1	train
2,4-Pentanedione	electrophilicity/proelectrophilicity (C)[*1]	electrophilicity	1	train
Tetrachlorocatechol	oxidative phosphorylation uncoupling (B)[*1]	electrophilicity	1	train
Benzyl methacrylate	ester narcosis (B)[*1]	electrophilicity	1	train
3,5-Dibromo-4-hydroxybenzonitrile		electrophilicity	1	train
Diethyl chloromalonate		electrophilicity	1	train
Catechol	polar narcosis (A)[*1]	proelectrophilicity	1	train
2-Decyn-1-ol		proelectrophilicity	1	train

[*1]Level of confidence associated with an observed mode of action. A, high; B, moderate; C, low; D, No confidence was associated with assessments for which there were insufficient data; [*2]1, less reactive compounds; 2, more reactive compounds.

regulating chemical exposure [69]. In this sense, both LR models developed in this study appear to be more attractive.

Conclusions

In this study, LR was used to develop models to discriminate chemicals according to their modes of aquatic toxic action, giving full consideration to the OECD principles in regulation of QSAR acceptability during the model construction and assessment process. Satisfactory results were obtained. The obtained models are simple in formation and easy to interpret. They could provide some insight into what structural features are related to the aquatic toxic MOA of the compounds. Our results show that LR performs much better than LDA and seems to be more attractive, especially for the prediction of compounds exhibiting excess toxicity (95.60% versus 85.40% by LR and LDA, respectively) and the more reactive compounds (88.17% versus 80.65% by LR and LDA, respectively). In addition, influence plot and outlier

tests as well as the overdispersion test together demonstrate that the LR models we built are statistically significant, robust and acceptable.

Comparison of the predicted results by our LR models and those by Ren et al. [24] was done after removing the duplicated compounds from the data set. What should be noted is the large difference in the prediction results for more toxic compounds. For compounds exhibiting excess toxicity, the classification rates were 95.60% in our LR model (Table 1) and 88.3% in Ren et al. [24]. In the prediction of more reactive compounds, the respective classification rates were 88.17% in our LR model (Table 4) and 54.8% in Ren et al. [24]. In this sense, it can be concluded to some extent that both LR models in our study appear to be more attractive. It also proves that toxicity data used in the model construction by Ren et al. [24] are not good enough (at least in the case of the discrimination between the less reactive compounds and the more reactive compounds), and it is necessary to include other variables such as chemical structure-based descriptors into the models. In the present study, none of the CODESSA descriptors was selected into the models, and the chemical structure-based descriptors used are those calculated by DRAGON software. In this sense, this indicates that the DRAGON-calculated descriptors are more suitable for describing these compounds' structural features related to their aquatic toxic action. These descriptors, combined with the toxicity data (Table 1) or themselves solely (Table 4), give satisfactory results in the classification of compounds according to their MOAs.

In the present study, QSAR was performed using descriptors (including experimental data) and statistical techniques to classify the MOAs of chemicals for fathead minnow (*Pimephales promelas*). The applicability of the obtained models to other species was not checked due to lack of data. However, the concept of using statistical classification methods to identify the MOAs presented in this study can be extended directly to other species, though different MOAs may be involved in the toxicity for different tropic levels. In addition, the LDA and LR models for the discrimination between the less reactive compounds and the more reactive compounds have toxicity-predicting capacity, and can be used in predictive toxicology since they are structural descriptor-based models. With the correct identification of the toxic MOA, high-quality QSARs for the particular mode of action can be developed to estimate the toxic potential of the chemical more effectively. In summation, the proposed method is very useful for the classification of aquatic toxic MOA, and can also be extended in other QSAR investigations.

Acknowledgments

The authors thank the R Development Core Team for affording the free R software. The authors also wish to thank two anonymous reviewers who provided helpful comments on the manuscript.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the Young Scholars Science Foundation of Lanzhou Jiaotong University under Grant [2011009]; and the National Natural Science Foundation of China under [41261077].

ORCiD

Y. Y. Ren http://orcid.org/0000-0002-5323-2724

References

[1] E. Papa, F. Battaini, and P. Gramatica, *Ranking of aquatic toxicity of esters modelled by QSAR*, Chemosphere 58 (2005), pp. 559–570.
[2] M.N. Moorea, M.H. Depledge, J.W. Readmana, and P. Leonard, *An integrated biomarker-based strategy for ecotoxicological evaluation of risk in environmental management*, Mutat. Res. 552 (2004), pp. 247–268.
[3] J.R. Seward, G.D. Sinks, and T.W. Schultz, *Reproducibility of toxicity across mode of toxic action in the* Tetrahymena *population growth impairment assay*, Aquat. Toxicol. 53 (2001), pp. 33–47.
[4] C.L. Russom, S.P. Bradbury, S.J. Broderius, D.E. Hammereister, and R.A. Drummond, *Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow* (Pimephales Promelas), Environ. Toxicol. Chem. 16 (1997), pp. 948–967.
[5] T.W. Schultz, G.D. Sinks, and A.P. Bearden, *QSAR in aquatic toxicology: A mechanism of action approach comparing toxic potency to* Pimephales promelas, Tetrahymena pyriformis, *and* Vibrio fischeri, in *Comparative QSAR*, J. Devillers, ed., Taylor & Francis, New York, NY, 1998, pp. 51–109.
[6] H.J.M. Verhaar, C.J.V. Leeuwen, and J.L.M. Hermens, *Classifying environmental pollutants part 1: Structural activity relationship for prediction of aquatic toxicity*, Chemosphere 25 (1992), pp. 471–491.
[7] W.H.J. Vaes, E.U. Ramos, H.J.M. Verhaar, and J.L.M. Hermens, *Acute toxicity of nonpolar versus polar narcosis: Is there a difference?*, Environ. Toxicol. Chem. 17 (1998), pp. 1380–1384.
[8] B.I. Escher, R.I.L. Eggen, U. Schreiber, Z. Schreiber, E. Vye, B. Wisner, and R.P. Schwarzenbach, *Baseline toxicity (narcotics) of organic chemicals determined by in-vitro membrane potential measurements in energy-transducing membranes*, Environ. Sci. Technol. 36 (2002), pp. 1971–1979.
[9] J.C. Dearden, M.T.D. Cronin, Y. Zhao, and O.A. Raevsky, *QSAR studies of compounds acting by polar and non-polar narcotics: An examination of the role of polarisability*, Quant. Struct. Act. Relat. 19 (2000), pp. 3–9.
[10] D.W. Roberts and J.F. Costello, *Mechanisms of action for general and polar narcosis: A difference in dimension*, QSAR Comb. Sci. 22 (2003), pp. 226–233.
[11] C.M. Auer, J.V. Nabholz, and K.P. Baetcke, *Mode of action and the assessment of chemical hazards in the presence of limited data: Use of structure activity relationships (SAR) under TSCA Section 5*, Environ. Health Perspect. 87 (1990), pp. 183–187.
[12] S.C. Basak, G.D. Grunwald, G.E. Host, G.J. Niemi, and S.P. Bradbury, *A comparative study of molecular similarity, statistical, and neural methods for predicting toxic modes of action*, Environ. Toxicol. Chem. 17 (1998), pp. 1056–1064.
[13] S.P. Bradbury, *Predicting modes of toxic action from chemical structure: An overview*, SAR QSAR Environ. Res. 2 (1994), pp. 89–104.
[14] T.I. Netzeva, M. Pavan, and A.P. Worth, *Review of (quantitative) structure - activity relationships for acute aquatic toxicity*, QSAR Comb. Sci. 27 (2008), pp. 77–90.
[15] T.W. Schultz, G.D. Sinks, and M.T.D. Cronin, *Identification of mechanisms of toxic action of phenols to* Tetrahymena pyriformis *from molecular descriptors*, in *Quantitative Structure-Activity Relationships in Environmental Sciences – VII*, F. Chen and G. Schüürmann, eds., SETAC Press, Pensacola, FL, 1997, pp. 329–342.
[16] A.O. Aptula, T.I. Netzeva, I.V. Valkova, M.T.D. Cronin, T.W. Schultz, R. Kühne, and G. Schüürmann, *Multivariate discrimination between modes of toxic action of phenols*, Quant. Struct.-Act. Relat. 21 (2002), pp. 12–22.
[17] G. Schüürmann, A.O. Aptula, R. Kühne, and R.-U. Ebert, *Stepwise discrimination between four modes of toxic action of phenols in the* Tetrahymena pyriformis *assay*, Chem. Res. Toxicol. 16 (2003), pp. 974–987.
[18] X.J. Yao, A. Panaye, J.P. Doucet, H.F. Chen, R.S. Zhang, B.T. Fan, M.C. Liu, and Z.D. Hu, *Comparative classification study of toxicity mechanisms using support vector machines and radial basis function neural networks*, Anal. Chim. Acta 535 (2005), pp. 259–273.

[19] U. Norinder, P. Lidén, and H. Boström, *Discrimination between modes of toxic action of phenols using rule based methods*, Mol. Divers. 10 (2006), pp. 207–212.
[20] B. Niu, Y. Jin, W. Lu, and G. Li, *Predicting toxic action mechanisms of phenols using AdaBoost Learner*, Chemom. Intell. Lab. Syst. 96 (2009), pp. 43–48.
[21] M. Nendza and M. MQlle, *Discriminating toxicant classes by mode of action: 2. Physico-chemical descriptors*, Quant. Struct.-Act. Relat. 19 (2000), pp. 581–598.
[22] S. Ren and T.W. Schultz, *Identifying the mechanism of aquatic toxicity of selected compounds by hydrophobicity and electrophilicity descriptors*, Toxicol. Lett. 129 (2002), pp. 151–160.
[23] S. Ren, *Predicting three narcosis mechanisms of aquatic toxicity*, Toxicol. Lett. 133 (2002), pp. 127–139.
[24] S. Ren, P.D. Fraymer, and T.W. Schultz, *An exploratory study of the use of multivariate techniques to determine mechanisms of toxic action*, Ecotoxicol. Environ. Saf. 55 (2003), pp. 86–97.
[25] P.C. Von der Ohe, R. KQhne, R.-U. Ebert, R. Altenburger, M. Liess, and G. Schüürmann, *Structural alerts – a new classification model to discriminate excess toxicity from narcotic effect levels of organic compounds in the acute daphnid assay*, Chem. Res. Toxicol. 18 (2005), pp. 536–555.
[26] M. Nendza and M. Müller, *Discriminating toxicant classes by mode of action: 3, Substructure indicator*, SAR QSAR Environ. Res. 18 (2007), pp. 155–168.
[27] P.G. Polishchuk, E.N. Muratov, A.G. Artemenko, O.G. Kolumbin, N.N. Muratov, and V.E. Kuz'min, *Application of random forest approach to QSAR prediction of aquatic toxicity*, J. Chem. Inf. Model. 49 (2009), pp. 2481–2488.
[28] L. Michielan, L. Pireddu, M. Floris, and S. Moro, *Support vector machine (SVM) as alternative tool to assign acute aquatic toxicity warning labels to chemicals*, Mol. Inf. 29 (2010), pp. 51–64.
[29] T.M. Martin, C.M. Grulke, D.M. Young, C.L. Russom, N.Y. Wang, C.R. Jackson, and M.G. Barron, *Prediction of aquatic toxicity mode of action using linear discriminant and random forest models*, J. Chem. Inf. Model. 53 (2013), pp. 2229–2239.
[30] K.P. Singh, S. Gupta, A. Kumar, and D. Mohan, *Multispecies QSAR modeling for predicting the aquatic toxicity of diverse organic chemicals for regulatory toxicology*, Chem. Res. Toxicol. 27 (2014), pp. 741–753.
[31] *OECD principles for the validation, for regulatory purposes, of (quantitative) structure-activity relationship models*. Available at http://www.oecd.org/env/ehs/risk-assessment/37849783.pdf.
[32] A.R. Katritzky, V.S. Lobanov, and M. Karelson, *CODESSA (COmprehensive DEscriptors for Structural and Statistical Analysis), Version 2.0*, University of Florida, Gainesville, FL, 1994.
[33] *Talete srl, Dragon (Software for Molecular Descriptor Calculation), Version 6.0*, 2013; software available at http://www.talete.mi.it/.
[34] A.P. Worth and M.T.D. Cronin, *The use of discriminant analysis, logistic regression and classification tree analysis in the development of classification models for human health effects*, Mol. Struct. (Theochem) 622 (2003), pp. 97–111.
[35] P. Gramatica, *Principles of QSAR models validation: Internal and external*, QSAR Comb. Sci. 26 (2007), pp. 694–701.
[36] A. Tropsha, P. Gramatica, and V.K. Gombar, *The importance of being earnest: Validation is the absolute essential for successful application and interpretation of QSPR models*, QSAR Comb. Sci. 22 (2003), pp. 69–76.
[37] A. Golbraikh and A. Tropsha, *Beware of q2!*, J. Mol. Graph. Model. 20 (2002), pp. 269–276.
[38] L. Eriksson, J. Jaworska, A.P. Worth, M.T.D. Cronin, R.M. McDowell, and P. Gramatica, *Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs*, Environ. Health Perspect. 111 (2003), pp. 1361–1375.
[39] P. Gramatica, P. Pilutti, and E. Papa, *Validated QSAR prediction of OH tropospheric degradation of VOCs: Splitting into training-test sets and consensus modeling*, J. Chem. Inf. Comput. Sci. 44 (2004), pp. 1794–1802.
[40] T. Oberg, *A QSAR for baseline toxicity: Validation, domain of application and prediction*, Chem. Res. Toxicol. 17 (2004), pp. 1630–1637.
[41] E. Papa, F. Villa, and P. Gramatica, *Statistically validated QSARs based on theoretical descriptors for modeling aquatic toxicity of organic chemicals in* Pimephales promelas *(fathead minnow)*, J. Chem. Inf. Model. 45 (2005), pp. 1256–1266.

[42] E. Papa and P. Gramatica, *Screening of persistent organic pollutants by QSPR classification models: A comparative study*, J. Mol. Graph. Model. 27 (2008), pp. 59–65.
[43] B. Bhhatarai and P. Gramatica, *Prediction of aqueous solubility, vapor pressure and critical micelle concentration for aquatic partitioning of perfluorinated chemicals*, Environ. Sci. Technol. 45 (2010), pp. 8120–8128.
[44] Y.Y. Ren, B.W. Zhao, Q. Chang, and X.J. Yao, *QSPR modeling of nonionic surfactant cloud points: An update*, J. Colloid. Interf. Sci. 358 (2011), pp. 202–207.
[45] *ChemDraw (Chemical Structure Drawing Standard)*, Cambridge Soft Corporation, MA, USA, 1985–2003.
[46] *HyperChem, Release 6.0 for Windows*; Hypercube, Inc. Canada, 2000.
[47] I. Valkova, M. Vračko, and S.C. Basak, *Modeling of structure–mutagenicity relationships: Counter propagation neural network approach using calculated structural descriptors*, Anal. Chim. Acta 509 (2004), pp. 179–186.
[48] Y.Y. Ren, J. Qin, H.X. Liu, X.J. Yao, and M.C. Liu, *QSPR study on the melting points of a diverse set of potential ionic liquids by projection pursuit regression*, QSAR Comb. Sci. 28 (2009), pp. 1237–1244.
[49] Y.Y. Ren, B.W. Zhao, and X.J. Yao, *Use of advanced statistical learning methods and principal component analysis in quantitative structure-genotoxicity relationship study of amines*, Collect. Czech. Chem. Commun. 74 (2009), pp. 217–241.
[50] J.A. Nelder and R.W.M. Wedderburn, *Generalized linear models*, J. R. Stat. Soc. A. Stat. 135 (1972), pp. 370–384.
[51] P. McCullagh and J.A. Nelder (eds.), *Generalized Linear Models,* 2nd ed., Chapman & Hall/CRC, 1989.
[52] J. Fox, *Applied Regression analysis and Generalized Linear Models*, 3rd ed., Sage Publications, Inc., 2008.
[53] R.I. Kabacoff, *R in Action: Data analysis and graphics with R*, Manning, Shelter Island, NY, 2011.
[54] A.J. Dobson and A.G. Barnett, *An Introduction to Generalized Linear Models*, 3rd ed., Chapman & Hall/CRC, Boca Raton, FL, 2008.
[55] J.J. Faraway, *Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models*, Chapman & Hall/CRC, Boca Raton, FL, 2006.
[56] J. Fox and S. Weisberg, *An R Companion to Applied Regression*, 2nd ed., Sage, Thousand Oaks, CA, 2011. Available at: http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.
[57] D. Collett, *Modelling Binary Data*, 2nd ed., Chapman & Hall/CRC, London, 2003.
[58] *R 3.2.0*; software available at http://cran.R-project.org.
[59] O.G. Mekenyan and G.D. Veith, *Relationships between descriptors for hydrophobicity and soft electrophilicity in predicting toxicity*, SAR QSAR Environ. Res. 1 (1993), pp. 335–344.
[60] V.R. Magnuson, D.K. Harriss and S.C. Basak, *Topological indices based on neighborhood symmetry: Chemical and biological applications*, in *Studies in Physical and Theoretical Chemistry*, R.B. King, ed., Elsevier, Amsterdam, The Netherlands, 1983, pp. 178–191.
[61] R. Todeschini and V. Consonni (eds.), *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, Germany, 2000.
[62] P.A.P. Moran, *Notes on continuous stochastic phenomena*, Biometrika 37 (1950), pp. 17–23.
[63] J. Ferguson, *The use of chemical potentials as indices of toxicity*, P. Roy. Soc. Lon. 848 (1939), pp. 387–404.
[64] J.W. McFarland, *On the parabolic relationship between drug potency and hydrophobicity*, J. Med. Chem. 13 (1970), pp. 1192–1196.
[65] F. Reichenberg and P. Mayer, *Two complementary sides of bioavailability: Accessibility and chemical activity of organic contaminants in sediments and soils*, Environ. Toxicol. Chem. 25 (2006), pp. 1239–1245.
[66] D.T. Stanton and P.C. Jurs, *Development and use of charged partial surface area structural descriptors in computer-assisted quantitative structure–property relationship studies*, Anal. Chem. 62 (1990), pp. 2323–2329.
[67] D. Butina, *Performance of Kier-Hall E-state descriptors in quantitative structure activity relationship (QSAR) studies of multifunctional molecules*, Molecules 9 (2004), pp. 1004–1009.
[68] P. Labute, *A widely applicable set of descriptors*, J. Mol. Graph. Model. 18 (2000), pp. 464–477.
[69] A.M. Richard, *Commercial toxicology prediction systems: A regulatory perspective*, Toxicol. Lett. 102–103 (1998), pp. 611–616.