A Quantitative Structure Activity Relationship for acute oral toxicity of pesticides on rats: Validation, Domain of Application and Prediction

Mabrouk Hamadache[1*], Othmane Benkortbi[1], Salah Hanini[1], Abdeltif Amrane[2], Latifa Khaouane[1], Cherif Si Moussa[1]

[1] Laboratoire des Biomatériaux et Phénomènes de Transport (LBMPT), Université de Médéa, Quartier Ain D'heb, 26000, MEDEA, Algérie

[2] Ecole Nationale Supérieure de Chimie de Rennes, Université de Rennes 1, CNRS, UMR 6226, 11 allée de Beaulieu, CS 50837, 35708 Rennes Cedex 7, France

L. KHAOUANE : latifa_khaouane@yahoo.fr ; O. BENKORTBI: benkortbi_oth@yahoo.fr ;
C. SI MOUSSA : simoussa_cherif@yahoo.fr ; S. HANINI : s_hanini2002@yahoo.fr ;
A. AMRANE: abdeltif.amrane@univ-rennes1.fr

*Corresponding author: Mabrouk HAMADACHE;
mhamdeche@yahoo.fr
Tel: + 213 07 78 12 37 50; Fax: +213 25 58 12 53

ABSTRACT

Quantitative Structure Activity Relationship (QSAR) models are expected to play an important role in the risk assessment of chemicals on humans and the environment. In this study, we developed a validated QSAR model to predict acute oral toxicity of 329 pesticides to rats because a few QSAR models have been devoted to predict the Lethal Dose 50 (LD_{50}) of pesticides on rats. This QSAR model is based on 17 molecular descriptors, and is robust, externally predictive and characterized by a good applicability domain. The best results were obtained with a 17/9/1 Artificial Neural Network model trained with the Quasi Newton back propagation (BFGS) algorithm. The prediction accuracy for the external validation set was estimated by the Q^2_{ext} and the Root Mean Square error (RMS) which are equal to 0.948 and 0.201, respectively. 98.6% of external validation set is correctly predicted and the present model proved to be superior to models previously published. Accordingly, the model developed in this study provides excellent predictions and can be used to predict the acute oral toxicity of pesticides, particularly for those that have not been tested as well as new pesticides.

Keywords

Acute toxicity, Pesticides, QSAR, Prediction, External validation

Abbreviations : QSAR, quantitative structure-activity relationship; LD_{50}, lethal dose 50; ANN, artificial neural networks; BFGS, Quasi-Newton back propagation algorithm; RMS, root mean square error; REACH, registration, evaluation, authorization and restriction of chemicals; OECD, organization for economic cooperation and development; LOO, leave-one-out; CV, cross-validation; AD, applicability domain; VIF, variation inflation factors; MLP, multi-layer perceptron.

1. Introduction

Pesticides are widely used in agriculture for plant protection and for increasing production yields and quality of agricultural products but also in domestic applications. They are also used to slow the spread of insects, to maintain lawns, recreational areas and highways. Pesticides have also contributed to the control of many human diseases transmitted by insects. The most common pesticides are herbicides, insecticides and fungicides. However, despite these advantages, pesticides have a major drawback such as toxicity **[1]**. Due to the excessive use of these products, they are found as well as residue in the environment (water, soil, air) than in terrestrial and aquatic food chains **[2, 3]**. In addition, they also pose a threat to the environment, humans, animals and other organisms **[4, 5]**. Many studies made internationally highlight the environmental pollution by pesticides. The consequences of this pollution are the widespread presence of residues in air, water, soil and foodstuffs **[6-13]**.

Long-term exposure to pesticides can cause harm to human life and can disrupt the functioning of various organs in the body. This significant relationship between exposure to pesticides and some chronic diseases has been the subject of several scientific publications. Exposure to these persistent pesticides has been associated with health effects including cancer, headache, skin and eye irritation, immune system problems, stomach, kidney, Parkinson and Alzheimer's disease, reproductive difficulties, birth defects, diabetes, cataracts and anemia **[14-17]**.

As seen, humans and the environment are exposed to thousands of pesticides. This pollution caused by pesticides has become an important issue affecting the survival and development of humain being. It is evident that risk assessment for pesticides can provide a precaution against the corresponding pollution. One of the procedures currently used for human and environmental risk assessment is the determination of the acute toxicity of pesticides **[18]**. Unfortunately, experimental determination of the toxicity takes time, requires a high expense and poses an ethical problem (demands to reduce or abolish the use of animals). Also, there is a very large body of research going on in many countries with the aim of replacing in vivo tests by in silico prediction methods according to the European Directive on the Protection of Laboratory Animals **[19]** and the Registration, Evaluation, Authorization and restriction of Chemicals (REACH) regulation **[20]**. Despite being significantly cheaper than in vivo study, in vitro tests are still costly compared with in silico methods **[21]**. The use of in silico predictive methods, based on computer tools, offers a rapid, cost-effective and ethical alternative to testing toxicity of chemical substances in animals **[22]**. These methods include the Quantitative Structure–Activity Relationship (QSARs) models. To establish a QSAR model, three elements

are necessary. The first relates to the biological activity (eg toxicity) measured for a set of molecules. The second concerns the descriptors. Finally, the third must be a statistical learning method that is used to connect the first two elements.

The acute toxicity still remains the object of interest in QSAR model building. To date, a large number of QSARs models for predicting the acute toxicity of chemical substances have been developed **[23, 24]**. Unfortunately, few studies have been devoted to the acute toxicity of pesticides on rats. For example, Enslein et al. **[25, 26]** developed regression analysis models using two large data sets (425 and 1851 various chemicals, respectively). The R^2 value for the test set is 0.33, which means that these models are characterized by low power external prediction. A very marked improvement in R^2 coefficient was obtained following the QSAR models developed with 44, 54, 67, 30 and 62 pesticides by Zakaria et al. **[27]**, Eldred and Jurs **[28]**, Zahouily **[29]**, Guo et al. **[30]** and Garcia et al. [31] respectively. Recent studies devoted to pesticides **[32, 33]** have proposed QSAR models with values of 0.93 (27 herbicides) and 0.96 (62 herbicides) for the R^2 coefficient. The conclusion which can be draw from these studies is that most QSAR models developed are distinguished by two major shortcomings: lack of validation test on the one hand, and a limited field of application because these studies included a relatively small number of pesticides on the other hand.

Since the prediction of potential risks to human health is based on the assumption that test results seen in high-dose animal tests are predictive of effects that will occur in human populations exposed to much lower levels **[34]**, our main goal in this work is to establish a robust QSAR model to predict acute toxicity (log [1/LD_{50}]) of pesticides on rats. The database used consists of 329 pesticides. The QSAR model established by using artificial neural networks and molecular descriptors satisfies the guidelines required by the Organisation for Economic Cooperation and Development (OECD), namely: (1) a defined endpoint; (2) an unambiguous algorithm; (3) a defined domain of applicability; (4) appropriate measures of goodness of fit, robustness, predictability; (5) a mechanistic interpretation, if possible.

2. Materials and method

2.1 Data set

It is well known that high-quality experimental data are essential for the development of high quality QSAR models **[35]**. If they are unreliable, the model will be unreliable. The rat lethal dose 50 (LD_{50} - rat, male via oral exposure) values were retrieved from Pesticide Properties Database **[36]**. The LD_{50} correspond to the concentration (mg/kg) of pesticide that

lead to the death of 50% of rat. All values of oral acute toxicity were first converted into mmol/kg and then translated to log [1/ (mmol/kg)].

The initial database that included 907 pesticides was rigorously reviewed and "cleaned" by removing pesticides whose LD_{50} was not experimentally determined or whose LD_{50} was not determined in the same experimental conditions. A total of 329 pesticides with experimental data were selected to form the final database (**Table 1**). The basis of 329 pesticides was divided into 2 lots. The first with 258 pesticides was dedicated to develop the QSAR model. The second which included 71 pesticides that had not been used for the development of the QSAR model, was left for the external validation.

2.2 Molecular descriptors

One important step in obtaining a QSAR model is the numerical representation of the structural features of molecules, which were named molecular descriptors. Nowadays, there are more than 4000 of molecular descriptors which can be used to solve different problems in Chemistry, Biology and related sciences [1]. In the specific case of this study, for each molecule, 1664 molecular descriptors were calculated, which belong to many classes. All descriptors were obtained through the online program E-Dragon 1.0 (http: //www.vcclab. org/lab/edragon).

To avoid the phenomenon of overfitting, the number of descriptors must be reduced. Several methods to simplify a database are used. The method used to select the most significant descriptors was described previously **[32]**. In the first step, invariant descriptors, namely those with absent values (represented by the code ''999''), were manually removed. Next, any descriptor that had identical values for 75% of the samples and any descriptors with a relative standard deviation < 0.05 were removed. Finally, half of the descriptors showing an absolute value of the Pearson correlation coefficient > 0.95 were also removed. The number of descriptors obtained after the selection was 95. For relevant descriptors selection, stepwise regression was then used **[37]**. Twenty nine descriptors were selected.

2.3 Model development

In this work, all calculations were run on a Sony personal computer with a Core (TM) i3 and windows XP as operating system. The Artificial Neural Networks (ANN) which has extensive applicability in solving non-linear systems was employed to build the QSAR model between the molecular relevant descriptors and the toxicity of pesticides. A three-layer feed-

forward neural network utilizing back-propagation algorithm was employed. The typical back-propagation network consists of an input layer, an output layer and at least one hidden layer. Each layer contains neurons and each neuron is a simple micro-processing unit which receives and combines signals from many neurons.

The use of a neuronal regression goes through the choice of the input parameters but also by optimizing the architecture of the neural network itself. The optimization of both the distribution of the database, the number of hidden layers, the number of neurons per hidden layer, the transfer functions as well as algorithms was carried after extensive testing. The design of the neural model is to evaluate the components of the network according to the desired performance modeling. Model performance is evaluated in terms of root mean square error (RMS) **[38]** and was calculated with the following equation:

$$RMS = \sqrt{\frac{\sum_{i=1}^{n}(y_i^{exp} - y_i^{pred})^2}{n}} \quad (1)$$

where *n* is the number of compounds in the dataset, and y_i^{pred}, y_i^{exp} are the predicted and the experimental values, respectively.

2.4 Model validation

For the validation of the predictive power of a QSAR model, two basic principles (internal validation and external validation) are available. The quality is always judged by the statistical parameters, for instance, the squared R (R^2) and root mean square error (RMS). These parameters mainly reflect the goodness of fit of the models. However, recent studies **[38]** have indicated that the internal validation is considered to be necessary for model validation. In the present study, we took the leave-one-out (LOO) cross-validation (CV) for the internal validation to evaluate the internal predictive ability of the developed model, and its result was defined as Q^2_{LOO}, which could be calculated according to the following equation **[38]**:

$$Q^2_{LOO} = 1 - \frac{\sum_{i=1}^{training}(y_i^{exp} - y_i^{pred})^2}{\sum_{i=1}^{training}(y_i^{exp} - \bar{y})^2} \quad (2)$$

where y_i^{exp}, y_i^{pred} and $\bar{y}$ are the experimental, predicted, and average log (1/LD50) values of the samples for the training set, respectively. A value of $Q^2_{LOO} > 0.5$ is considered satisfactory, and a Q^2_{LOO} value > 0.9 is excellent **[39]**.

Furthermore, the external validation is a significant and necessary validation method used to determine both the generalizability and the true predictive ability of the QSAR models

for new chemicals, by splitting the available dataset into a training set and an external prediction set. As mentioned above, the whole dataset in this work has been randomly divided into a training set with 258 compounds for model development, and a prediction set with 71 compounds for model external validation. The external predictive ability of the developed models on the external prediction set was evaluated by Q^2_{ext}, which could be calculated as follows [38]:

$$Q^2_{ext} = 1 - \frac{\sum_{i=1}^{prediction}(y_i^{exp} - y_i^{pred})^2}{\sum_{i=1}^{prediction}(y_i^{exp} - \overline{y_{tr}})^2} \quad (3)$$

where y_i^{exp}, y_i^{pred} are the experimental and predicted log(1/LD_{50}) values of the samples for the prediction set, and $\overline{y_{tr}}$ is the mean experimental log(1/LD_{50}) values of the samples for the training set.

2.5 Applicability domain

Even the most comprehensive and validated models cannot predict reliably properties for all existing compounds. The QSAR model is not intended to be used outside its domain of applicability, that is to say, outside of the chemical space covered by the training set. Also, the applicability domain (AD) of models must be defined and the predictions of the molecules in this area can be considered admissible. The determination of AD is therefore of great importance [40].

The AD is a theoretical region in the space defined by the descriptors of the model and the modeled response, for which a given QSAR should make reliable predictions. This region is defined by the nature of the compounds in the training set, and can be characterized in various ways. In our work, the AD was verified by the leverage approach. The leverage h_i is defined as follows [41]:

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \quad (4)$$

Where $\boldsymbol{x_i}$ is the descriptor value of the ith object, and $\bar{x}$ is the average value of the descriptor in the training set, and *n* is the number of substances in the training set. The warning leverage h* is, generally, fixed at 3(p + 1)/n, where n is the total number of samples in the training set and p is the number of descriptors involved in the correlation.

The applicability domain (AD) of QSAR model is defined from the Williams plot. The plot of leverage values versus standardized residuals (Williams plot) was used to give a

graphical detection of both the response outliers (Y outliers) and the structurally influential compounds (X outliers). In this plot, the two horizontal lines indicate the limit of normal values for Y outliers (i.e. samples with standardized residuals greater than 3.0 standard deviation units, ±3.0s); the vertical straight lines indicate the limits of normal values for X outliers (i.e. samples with leverage values greater than the threshold value, h > h*). For a sample in the external test set whose leverage value is greater than h*, its prediction is considered unreliable, because the prediction is the result of a substantial extrapolation of the model. Conversely, when the leverage value of a compound is lower than the critical value, the probability of accordance between predicted and experimental values is as high as that for the compounds in the training set **[42]**.

3. Results and discussion

3.1 Selection of relevant descriptors

To select the most important descriptors and the optimal number, the influences of the number of descriptors on the correlation coefficients [R^2 and adjusted R^2 **(**R^2_{adj})] and the RMSE were investigated for 1–29 descriptors. R^2 and R^2_{adj} increased with increasing number of descriptors. However, the values of RMSE decreased with increasing number of descriptors. Models with 18–29 descriptors did not significantly improve the statistics of the model. For these reasons, the number of descriptors used to develop the model was 17. Let us note that n / k is greater than 5 **[43]** where n (258) and k (17) are respectively the number of compounds and the number of descriptors used in the QSAR model.

Multi-collinearity between the 17 descriptors was detected by calculating their variation inflation factors (VIF). If VIF falls into the range of 1–5, the related model is acceptable. All the descriptors have VIF values < 2.873, indicating that the obtained model has statistical significance, and the descriptors were found to be reasonably orthogonal. Order to study the correlation between the selected descriptors, the correlation matrix has been established using the XLSTAT software. The results show that these descriptors are not correlated owing to the fact that the greatest value of the correlation coefficient is 0.512. The list of descriptors used in the development of QSAR model is given in **Table 2**.

3.2 QSAR modeling

The main objective of this phase of the study is to find the optimal architecture of the neural network to predict the acute oral toxicity of pesticides on rats. A typical multilayer

perceptron (MLP) three-layered network with an input layer, a hidden layer and an output layer is adopted in this work. Increasing the number of the hidden layers decreases the learning accuracy. Theoretical works have shown that a single hidden layer is sufficient for the ANN to approximate to any complex nonlinear function and many experimental results seem to confirm that one hidden layer may be enough for most forecasting problems **[44]**. The use of a neuronal regression requires the selection of input parameters, but also the optimization of the neural network architecture. Before training the network, the database distribution, the activation functions (for hidden neurons and output neurons), the number of neurons in the hidden layer and the learning algorithms were optimized after many trials. The optimal model performance is evaluated in terms of root mean square error (RMS) **[45, 46]**. The results of this study and the ANN network optimal adopted are given in **Table 3**.

The selected parameters (Table 3) were used to develop nonlinear model. The seventeen relevant descriptors were used as inputs to the network. Before training the network, the number of nodes in the hidden layer was optimized, because it is an important parameter influencing the performances of the ANN. Thus, a 17-9-1 network architecture was obtained after trial and error procedure. The main performance parameters of MLP-ANN model are shown in **Table 4**. The predictive results from the MLP-ANN model for the entire dataset (329 compounds) are obtained and presented in Table 1. Figure 1 and 2 shows the regression line of the model equation, i.e. predicted *vs* experimental results for the training and validation set highlighted by different symbols.

Fig.1 and **Fig.2** indicates that there is a significant correlation between experimental values and predicted values of log (1/ LD_{50}). As can be seen from Table 4, the non-linear MLP-ANN model give good results with higher correlation coefficients (R^2 and R^2_{ext}), lower RMS, as well as better robustness (Q^2) in both training set and validation set, which indicated that the MLP-ANN not only performed well in model development, but also had excellent prediction. This fact suggested a non-linear correlation between the acute toxicity and the relevant descriptors. In addition, the residual of the predicted values of the toxicity data against the experimental values for the present model is shown in **Fig. 3**. As most of the calculated residuals are distributed on two sides of the zero line, a conclusion may be drawn that there is no systematic error in the development of the present model.

To see the importance of each descriptor for the prediction of LD_{50} oral toxicity of pesticides towards rats, the relative contributions **[47]** of the seventeen descriptors to the MLP-ANN model were determined and are plotted in **Fig.4**. The contribution of the descriptors

decreased in the order: HATS0m (12.81%) > E1u (7.98%) > MATS2p (7.74%) > HATSe (7.63%) > Mor15m (7.14% > RDF030e (6.48%) > H6m (6.27%) > Mor23u (6.12%) > Du (5.88%) > nS (5.58%) > PJI3 (5.10%) > N-072 (4.68%) > RDF020e (4.29%) > MATS1m (3.47%) ≈ nArX (3.45%) > Mor26u (2.93%) > H-046 (2.45%). The most significant descriptor in the model was therefore HATS0m. It should be noted that for the majority of the descriptors, the difference between two descriptors contribution was not significant, indicating that all selected descriptors were needed in the development of QSAR predictive model.

Generally, QSAR models are functions of a molecule's structure, electronic properties and hydrophobicity **[48]**. In the present model, HATS0m, E1u, Mor15m, H6m, Mor23u, Du, nS, PJI3, N-072, MATS1m, nArX, Mor26u and H-046 involve the structure while MATS2p, HATSe, RDF030e and RDF020e represent the electronic properties.

Descriptors used in our model have been used in previous QSAR models in the literature. Hamadache et al. **[32]** have used MATS2p, HATSe, HATS0m, nS, E1u and N-072 in their MLR and ANN models to predict rat oral acute toxicity of 62 herbicides. In a study by Habibi-Yangjeh and Danandeh-Jenagharad **[49]**, the MATS1m, H-046, Mor23u and PJI3 descriptors were used for global prediction of the toxicity of 250 phenols to *Tetrahymena pyriformis* in a linear and nonlinear model. In a QSAR model of acute toxicity LD_{50} for rats caused by aromatic compounds, Bakhtiyor et al. **[50]** found that the descriptor MATS2p significantly contributes to the toxicity of these compounds. In a study on the penetration of the blood–brain barrier, the human intestinal absorption and the hydrophobicity, Soto et al. **[51]** proposed linear and nonlinear QSAR/QSPR models that include the descriptor MATS2p. A QSA(P)R model with high internal and external statistical quality for predicting toxicity was developed by Borges **[52]** with MATS2p for a set of 28 alkyl (1-phenylsulphonyl)-cycloalkane-carboxilates. A QSAR model on rat oral LD_{50} data of 58 per- and polyfluorinated chemicals developed by Bhhatarai and Gramatica **[53]** employed E1u; the authors concluded that E1u is one of the most important descriptors.

Moreover, some authors **[48, 54-57]** found that among the descriptors that affect the toxicity of the compounds studied, a substantial number belong to the categories of WHIM descriptors, GETAWAY descriptors, 2D autocorrelations, and Atom-centered fragments. In our study, a large number of descriptors involved in the present model also belong to this category. It is obvious that the descriptors in this category have major significance in the toxicity of pesticides

3.3 Applicability domain

The applicability domain of the model was analysed using a Williams plot (**Fig.5**), where the vertical line is the critical leverage value (h*), and the horizontal lines are 3s the cut off value for Y space. As seen in **Fig.5**, one can observe that none of the pesticides compounds in the training set and validation set have a leverage higher than the warning h* value of 0.16. In the Williams plot, three pesticides can be considered as response outlier (in the Y-response space). In the training set, one pesticide (Pyrazophos: 225) was overestimated, while another pesticide was underestimated (Oxycarboxine: 201). However, in the region of underestimated pesticides, Pyrazophos (329) was from the validation set. These three response outlier (in the Y-response space) could be associated with errors in the experimental values.

It should be noted that 98.6% of the domain was covered by the model when it was applied to predict the acute oral toxicity of the 71 pesticides in the validation set. Thus, these results show that MLP-ANN model complies with the third principle of the OECD. Accordingly, the model developed in this study provides excellent predictions for 329 pesticides. It can be used to predict the acute oral toxicity of pesticides, particularly for those that have not been tested as well as new pesticides.

3.4 Comparison with different models

As indicated in the introduction, there are a limited number of QSAR models available in the literature for predicting the oral acute toxicity of pesticides to rats. The evaluation of their advantages and disadvantages is quite difficult, because each published study used different data sets and a different modeling approach (chemical descriptors, algorithms, etc.). However, it would be worthwhile to evaluate the performance of our model (present work) in light of the few QSAR models published in the literature over the last few years. Our main aim is to compare the predictive power of each model, which gives an estimation of the fitting of the model and its robustness.

It should be noted that the most of these QSAR models were obtained using small databases **[33]** and generally with structurally similar chemicals such as amide herbicides **[27, 58]**, benzimidazoles herbicides **[59]** or phenylurea herbicides **[60]**. Also, the number of statistical parameters used for validation of this QSAR models is limited, especially in old publications. Devillers **[61]** developed a QSAR model for acute oral toxicity in rodents (rats). He used artificial neural networks (ANN) to predict the LD_{50} values of organophosphate

pesticide. The 51 chemicals of the training set and the nine compounds of the external testing set were described by a set of descriptors. The acute toxicities (1/log LD_{50}) were converted to mmol/kg and a series of 8 descriptors has been used. The best results were obtained with an 8/4/1 ANN model. The root mean square error (RMS) values for the training set and the external testing set equaled 0.29 and 0.26, respectively. This study demonstrated the usefulness of descriptors such as lipophilicity and molar refractivity.

Structure-toxicity relationships were studied for a set of 47 insecticides with three-layer perceptron and use of a backpropagation algorithm **[29]**. A model with three descriptors showed good statistics in the artificial neural network model with a configuration of 3/5/1 (r = 0.966, RMS = 0.200 and Q^2 = 0.647). The statistics for the prediction on toxicity [log LD_{50}, oral, rat)] in the test set of 20 organophosphorus insecticides derivatives was r = 0.748, RMS = 0.576). The model descriptors indicate the importance of molar refraction toward toxicity of organophosphorus insecticides derivatives used in this study. Otherwise, different topological descriptors were used by Garcia-Domenech et al. **[31]** in the prediction of the oral acute toxicity (LD_{50}) of 62 organophosphorus pesticides on rats. The LD_{50} values were expressed in mmol/kg with a logarithmic transformation before use. A model with eight variables (r = 0.906, Q^2 = 0.701) was generated. Zhu et al. **[62]** have developed a number of QSAR models for acute oral toxicity in rats using large datasets (7385 compounds). Several sets of descriptors and different modeling methods were used. It should be noted an improvement of the prediction compared to other works. However, the complexity of the modeling approach, while being interesting and promising, renders these models little useful in practice.

The statistical parameters of the results obtained from the present study and studies published in the literature are shown in **Table 5**. It is possible to observe that all of those models could give high prediction ability (correlation coefficient R^2**,** Q^2). However, our model exceeds the previously published models in all statistical indices available for comparison. Indeed, it gives the higher correlation coefficient and the lower RSM error if compared to the other models. It can be seen that the database for this study (training set and validation set) was wider than that of previous models with the exception of the base used by Zhu et al. **[62]**. According to these results, the present model can be promisingly used for predicting the toxicity of new chemicals, thus contributing to the risk assessment, saving substantial amounts of money and time.

4. Conclusion

The aim of the present work was to develop a QSAR study and to predict the oral acute toxicity of pesticides to rats. This study involved 258 pesticides with an additional external set of 71 pesticides modelled for their oral acute toxicity on rat based on the artificial neural network (multi-layer perceptron: MLP-ANN) with descriptors calculated by Dragon software and selected by a stepwise MLR method. The seventeen selected descriptors showed that the electronic properties and the structure of the molecule play a main role in the toxicity activity of the pesticides. The built MLP-ANN model was assessed comprehensively (internal and external validations). It showed good values of R^2 = 0.963 and Q^2_{LOO} = 0.962 for the training set and high predictive R^2_{ext} and Q^2_{ext} values (0.950 and 0.948) for the validation set. All the validations indicate that the built QSAR model was robust and satisfactory. Based on the comparison with models previously published, the proposed QSAR model achieved good results and provided 98.6% predictions that belong to the applicability domain. In conclusion, the model developed in this study meets all of the OECD principles for QSAR validation and can be used to predict the acute oral toxicity of pesticides, particularly for those that have not been tested as well as new pesticides and thus help reduce the number of animals used for experimental purposes.

References

[1] A. Speck-Planche, V.V. Kleandrova, F. Luan, M.N.D.S. Cordeiro, Predicting multiple ecotoxicological profiles in agrochemical fungicides: A multi-species chemoinformatic approach, Ecotoxicol. Environ. Saf. 80 (2012) 308–313.

[2] M.L. Gómez-Pérez, R. Romero-González, J.L. Martínez Vidal, A. Garrido Frenich, Analysis of pesticide and veterinary drug residues in baby food by liquid chromatography coupled to Orbitrap high resolution mass spectrometry, Talanta, 131 (2015) 1–7.

[3] K. Müller, A. Tiktak, T.J. Dijkman, S. Green, B. Clothier, Advances in Pesticide Risk Reduction. Encyclopedia of Agriculture and Food Systems, (2014) 17-34.

[4] J. Regueiro, O. López-Fernández, R. Rial-Otero, B. Cancho-Grande, J. Simal-Gándara, A Review on the Fermentation of Foods and the Residues of Pesticides-Biotransformation of Pesticides and Effects on Fermentation and Food Quality, Crit. Rev. Food Sci. 55:6 (2015), 839-863.

[5] M. T. Wan, Ecological risk of pesticide residues in the British Columbia environment: 1973–2012, J. Environ. Sci. Heal. B 48:5 (2013) 344-363.

[6] Y. Moussaoui, L. Tuduri, Y. Kerchich, B.Y. Meklati, G. Eppe, Atmospheric concentrations of PCDD/Fs, dl-PCBs and some pesticides in northern Algeria using passive air sampling, Chemosphere 88 (2012) 270–277.

[7] C.B. Choung, R.V. Hyne, M.M. Stevens, G.C. Hose, The ecological effects of a herbicide-insecticide mixture on an experimental freshwater ecosystem, Environ. Pollut. 172 (2013) 264-274.

[8] E.T. Rodrigues, I. Lopes, M.Â. Pardal, Occurrence, fate and effects of azoxystrobin in aquatic ecosystems: A review, Environ. Int. 53 (2013) 18–28.

[9] E. Herrero-Hernandez, M.S. Andrades, A. Alvarez-Martin, E. Pose-Juan, M.S. Rodriguez-Cruz, M.J. Sanchez-Martin, Occurrence of pesticides and some of their degradation products in waters in a spanish wine region, J. Hydrol 486 (2013) 234–45.
[10] O. Oukali-Haouchine, E. Barriuso, Y. Mayata, K.M. Moussaoui, Factors affecting Métribuzine retention in Algerian soils and assessment of the risks of contamination, Environ. Monit. Assess. 185 (2013) 4107–4115.
[11] A. Moretto, Pesticide Residues: Organophosphates and Carbamates. Encyclopedia of Food Safety, 3 (2014) 19-22.
[12] A. Nougadère, V. Sirot, A. Kadar, A. Fastier, E. Truchot, C. Vergnet, F. Hommet, J. Baylé, P. Gros, J. C. Leblanc, Total diet study on pesticide residues in France: Levels in food as consumed and chronic dietary risk to consumers, Environ. Int. 45 (2012) 135–150.
[13] J. Stanley, K. Sah, S.K. Jain, J.C. Bhatt, S.N. Sushil, Evaluation of pesticide toxicity at their field recommended doses to honeybees, Apis cerana and A. mellifera through laboratory, semi-field and field studies, Chemosphere 119 (2015) 668–674.
[14] S. H. Shojaei, M. Abdollahi, Is there a link between human infertilities and exposure to pesticides, Int. J. Pharmacol. 8 (2012) 708–710.
[15] S. Mostafalou, M. Abdollahi, Pesticides and human chronic diseases: evidences, mechanisms, and perspectives, Toxicol. Appl. Pharmacol. 268 (2013) 157-77.
[16] E.J. Mremaa, F.M. Rubino, G. Brambilla, A. Morettoc, A.M. Tsatsakis, C. Colosio, Persistent organochlorinated pesticides and mechanisms of their toxicity, Toxicology 307 (2013) 74– 88.
[17] G. Van Maele-Fabry, P. Hoet, F. Vilain, D. Lison, Occupational exposure to pesticides and Parkinson's disease: a systematic review and meta-analysis of cohort studies, Environ. Int. 46 (2012) 30–43.
[18] A. A. Lagunin, A. V. Zakharov , D. A. Filimonov, V. V. Poroikov, A new approach to QSAR modelling of acute toxicity , SAR QSAR Environ. Res. 18 (2007) 285-298.
[19] A. Golbamaki, A. Cassano, A. Lombardo, Y. Moggio, M. Colafranceschi, E. Benfenati, Comparison of in silico models for prediction of Daphnia magna acute toxicity, SAR QSAR Environ. Res. 25 (2014) 673-694.
[20] M. Cassotti, V. Consonni, A. Mauri, D. Ballabio, Validation and extension of a similarity-based approach for prediction of acute aquatic toxicity towards Daphnia magna, SAR 2 QSAR Environ. Res. 25 (2014) 1013-1036.
[21] A. Sazonovas, P. Japertas, R. Didziapetris, Estimation of reliability of predictions and model applicability domain evaluation in the analysis of acute toxicity (LD50), SAR QSAR Environ. Res. 21(2010) 127-148.
[22] K.M. Sullivan, J.R. Manuppello, C.E. Willett, Building on a solid foundation: SAR and QSAR as a fundamental strategy to reduce animal testing, SAR QSAR Environ. Res. 25 (2014) 357-365.
[23] F. Cheng, W. Li, G. Liu, Y. Tang, In silico ADMET prediction: recent advances, current challenges and future trends, Curr. Top. Med. Chem. 13 (2013) 1273-89.
[24] F. Dulin, M. P. Halm-Lemeille, S. Lozano, A. Lepailleur, J. Sopkova-de Oliveira Santos, S. Rault, R. Bureau, Interpretation of honeybees contact toxicity associated to acetylcholinesterase inhibitors, Ecotox. Environ. Safe. 79 (2012) 13–21.
[25] K. Enslein, P. N. Craig, A toxicity estimation model, J. Environ. Pathol. Toxicol. 2 (1978) 115–121.
[26] K. Enslein, T. R. Lander, M. E. Tomb, P. N. Craig, A Predictive Model for Estimating Rat Oral LD_{50} Values, Princeton Scientific Publishers, Princeton (1983).

[27] D. Zakarya, E.M. Larfaoui, A. Boulaamail, T. Lakhlifi, Analysis of structure–toxicity relationships for a series of amide herbicides using statistical methods and neural network, SAR QSAR Environ. Res. 5 (1996) 269–279.
[28] D.V. Eldred, P.C. Jurs, Prediction of acute mammalian toxicity of organophosphorus pesticide compounds from molecular structure, SAR QSAR Environ. Res. 10 (1999) 75–99.
[29] M. Zahouily, A. Rhihil, H. Bazoui, S. Sebti, D. Zakarya, Structure–toxicity relationships study of a series of organophosphorus insecticides, J. Mol. Model. 8 (2002) 168–172.
[30] J. X. Guo, J. J. Wu, J. B. Wright, G. H. Lushington, Mechanistic insight into acetylcholinesterase inhibition and acute toxicity of organophosphorus compounds: A molecular modeling study, Chem. Res. Toxicol. 19 (2006) 209–216.
[31] R. Garcıa-Domenech, P. Alarcon-Elbal, G. Bolas, R. Bueno-Marı, F.A. Chorda-Olmos, S.A. Delcour, M.C. Mourino, A. Vidal, J. Galvez, Prediction of acute toxicity of organophosphorus pesticides using topological indices, SAR QSAR Environ. Res. 18 (2007) 745–755.
[32] M. Hamadache, L. Khaouane, O. Benkortbi, C. Si Moussa, S. Hanini, A. Amrane, Prediction of Acute Herbicide Toxicity in Rats from Quantitative Structure–Activity Relationship Modeling, Environ. Eng. Sci. 31(2014) 243-252.
[33] A. Can, I. Yildiz, G. Guvendik, The determination of toxicities of sulphonylurea and phenylurea herbicides with quantitative structure-toxicity relationship (QSTR) studies, Environ. Toxicol. Pharmacol. 35 (2013) 369-79.
[34] M.E. Andersen, M. Al-Zoughool, M. Croteau, M. Westphal, D. Krewski, The Future of Toxicity Testing, J. Toxicol. Env. Heal. B,13 (2010) 163-196.
[35] M.T.D. Cronin, T.W. Schultz, Pitfalls in QSAR, J. Mol. Struct. 622 (2003) 39−51.
[36] PPDB (Pesticide Properties DataBase), http:// sitem.**herts.ac.uk/aeru/**footprint/ (accessed 14/05/2014).
[37] L. Xu, W. J. Zhang, Comparison of different methods for variable selection, Anal. Chim. Acta 446 (2001) 477-483.
[38] R. Wang, J. Jiang, Y. Pan, H. Cao, Yi. Cui, Prediction of impact sensitivity of nitro energetic compounds by neural network based on electrotopological-state indices, J. Hazard. Mater. 166 (2009) 155–186.
[39] L. Eriksson, J. Jaworska, AP. Worth, MT. Cronin, R. M. McDowell, P. Gramatica, Methods for reliability and uncertainty assessment and for applicability evaluations of classification and regression-based QSARs, Environ. Health Perspect. 111 (2003) 1361–1375.
[40] OECD principles for the Validation, for Regulatory Purposes, of (Quantitative) Structure-Activity Relationship Models, (2009).
[41] E.M. De Haas, T. Eikelboom, T. Bouwman, Internal and external validation of the long-term QSARs for neutral organics to fish from ECOSAR, SAR QSAR Environ. Res. 22 (2011) 545–559.
[42] T.I. Netzeva, A.P. Worth, T. Aldenberg, R. Benigni, M.T.D. Cronin, P. Gramatica, J.S. Jaworska, S. Kahn, G. Klopman, C.A. Marchant, G. Myatt, N. Nikolova-Jeliazkova, G.Y. Patlewicz, R. Perkins, D.W. Roberts, T.W. Schultz, D.T. Stanton, J.J.M. Van De Sandt, W. Tong, G. Veith, C. Yang, Current status of methods for defining the applicability domain of (quantitative) structure–activity relationships, Altern. Lab. Anim. 33 (2005) 155–173.
[43] A. Tropsha, P. Gramatica, 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) 69–77.

[44] F. Othman, M. Naseri, Reservoir inflow forecasting using artificial neural network, Int. J. Phys. Sci. 6 (2011) 434-440.
[45] T.L. Lee, Back-propagation neural network for the prediction of the short-term storm surge in Taichung harbor, Taiwan. Eng. Appl. Artif. Intell. 21 (2008) 63-72.
[46] A. Sedki, D. Ouazar, E EI Mazoudi, Evolving neural network using real coded genetic algorithm for daily rainfall-runoff forecasting, Expert Syst. Appl. 36 (2009) 4523-4527.
[47] F. Zheng, E. Bayram, S.P. Sumithran, J.T. Ayers, C. G. Zhan, J.D. Schmitt, L.P. Dwoskin, P.A. Crooks, QSAR modeling of mono- and bis-quaternary ammonium salts that act as antagonists at neuronal nicotinic acetylcholine receptors mediating dopamine release, Bioorg. Med. Chem. 14 (2006) 3017–3037.
[48] G. Tugcu, M. Türker Saçan, M. Vracko, M. Novic, N. Minovski, QSTR modelling of the acute toxicity of pharmaceuticals to fish, SAR QSAR Environ. Res. 23 (2012) 297-310.
[49] A. Habibi-Yangjeh, M. Danandeh-Jenagharad, Application of a genetic algorithm and an artificial neural network for global prediction of the toxicity of phenols to Tetrahymena pyriformis, Monatsh Chem. 140 (2009) 1279-1288.
[50] R. Bakhtiyor, H. Kusic, D. Leszczynska, J. Leszczynski, N. Koprivanac, QSAR modeling of acute toxicity on mammals caused by aromatic compounds: the case study using oral LD50 for rats, J. Environ. Monit. 12 (2010) 1037-1044.
[51] A. J. Soto, R. L. Cecchini, G. E. Vazquez, I. Ponzoni, Multi-Objective Feature Selection in QSAR Using a Machine Learning Approach, QSAR Comb. Sci. 28 (2009) 1509-1523.
[52] Eduardo Borges de Melo, Modeling physical and toxicity endpoints of alkyl (1-phenylsulfonyl) cycloalkane carboxylates using the Ordered Predictors Selection (OPS) for variable selection and descriptors derived with SMILES, Chemometr. Intell. Lab. 118 (2012) 79-87.
[53] B. Bhhatarai, P. Gramatica, Oral LD50 toxicity modeling and prediction of per- and polyfluorinated chemicals on rat and mouse, Mol. Divers. 15 (2011) 467-476.
[54] J. Xu, L. Zhu, D. Fang, L. Wang, S. Xiao, Li. Liu, W. Xu, QSPR studies of impact sensitivity of nitro energetic compounds using three-dimensional descriptors, J. Mol. Graph. Model. 36 (2012) 10–19.
[55] Ning-Xin Tan, Ping Li, Han-Bing Rao, Ze-Rong Li, Xiang-Yuan Li, Prediction of the acute toxicity of chemical compounds to the fathead minnow by machine learning approaches, Chemom. Intell. Lab. Syst. 100 (2010) 66-73.
[56] P. R. Duchowicz, J. Marrugo, J. H. Erlinda, V. Ortiz, E. A. Castro, R. Vivas-Reyes, QSAR study for the fish toxicity of benzene derivatives, J. Argentine Chem. Soc. 97 (2009) 116-127.
[57] H. Du, J. Wang, Z. Hu, X. Yao, X. Zhang, Prediction of fungicidal activities of rice blast disease based on least-squares support vector machines and project pursuit regression, J. Agric. Food Chem. 56 (2008) 10785-10792.
[58] J. D. Gough, L. H. Hall, Modeling the toxicity of amide herbicides using the electrotopological state, Environ. Toxicol. Chem. 18 (1999) 1069–1075.
[59] G.W. Adamson, D. Bawden, D.T. Saggers, Quantitative structure–activity relationship studies of acute toxicity (LD50) in a large series of herbicidal benzimidazoles, Pestic. Sci. 15 (1984) 31–39.
[60] M. Nendza, B. Dittrich, A. Wenzel, W. Klein, Predictive QSAR models estimating ecotoxic hazard of plant protecting agents: target and non-target toxicity, Sci. Tot. Environ. 109/110 (1991) 527-535.
[61] J. Devillers, Prediction of mammalian toxicity of organophosphorus pesticides from QSTR modeling, SAR QSAR Environ. Res. 15 (2004) 501–510.

[62] H. Zhu, L. Ye, A. Richard, A. Golbraikh, F.A. Wright, I. Rusyn, A. Tropsha, A novel two-step hierarchical quantitative structure-activity relationship modeling work flow for predicting acute toxicity of chemicals in rodents, Environ. Health Persp. 117 (2009) 1257-1264.

Table 1.

Observed (experimental) log ($1/LD_{50}$). predicted log ($1/LD_{50}$) and leverage of pesticide compounds.

No.	Compound	Type	log [$1/LD_{50}$] (mmol/kg)$^{-1}$		Leverage (h_i)
			Observed	Predicted	
		Training set			
1	1,2-Dichloropropane	Insecticide	−1.24	−1.26	0.010
2	2,4,5-Trichlorophenol	Herbicide	−0.62	−0.30	0.005
3	2,4-DB	Herbicide	−0.55	−0.53	0.005
4	2,4-Dimethylphenol	Fongicide	−0.30	−0.28	0.004
5	2-Amino butane	Fongicide	−0.68	−0.59	0.005
6	Acephate	Insecticide	−0.71	−0.73	0.006
7	Acetamiprid	Insecticide	0.02	−0.08	0.003
8	Acetochlor	Herbicide	−0.85	−0.84	0.007
9	4-CPA	Herbicide	−0.66	−0.63	0.005
10	Acrolein	Herbicide	0.29	0.32	0.003
11	Alachlor	Herbicide	−0.54	−0.71	0.005
12	Alanycarb	Insecticide	0.08	0.12	0.003
13	Aldicarb	Insecticide	2.31	2.41	0.024
14	Aldrin	Insecticide	0.97	0.94	0.006
15	Allyxycarb	Insecticide	0.49	0.27	0.004
16	Alpha-endosulfan	Insecticide	1.03	0.90	0.007
17	Amicarbazone	Herbicide	−0.62	−0.75	0.005
18	Amidithion	Insecticide	−0.34	−0.08	0.004
19	Aminocarb	Insecticide	0.84	0.82	0.006
20	Amiprofos-methyl	Herbicide	−0.01	−0.04	0.003
21	Amitraz	Insecticide	−0.44	−0.37	0.004
22	Ancymidol	Herbicide	−0.83	−0.77	0.006
23	Anilazine	Fongicide	−1.22	−1.41	0.010
24	Anilofos	Herbicide	−0.11	0.13	0.003
25	Asomate	Fongicide	0.11	0.21	0.003
26	Azaconazole	Fongicide	−0.01	0.02	0.003
27	Azametiphos	Insecticide	−0.56	−0.59	0.005
28	Azinphos-methyl	Insecticide	1.55	1.48	0.012
29	Benalaxil	Fongicide	−0.32	−0.65	0.004
30	Bendiocarb	Insecticide	0.82	0.83	0.005
31	Benfuracarb	Insecticide	0.30	0.00	0.003
32	Benquinox	Fongicide	0.38	0.36	0.003
33	Bentazone	Herbicide	−0.32	−0.27	0.004

No.	Compound	Type	log [1/LD_{50}] (mmol/kg)$^{-1}$		Leverage (h_i)
			Observed	Predicted	
34	Benzthiazuron	Herbicide	−0.79	−0.48	0.006
35	Binapacryl	Fongicide	0.75	0.45	0.005
36	Brodifacoum	Rodonticide	3.12	3.16	0.042
37	Bromacil	Herbicide	−0.70	−0.60	0.005
38	Bromocyclen	Insecticide	−1.50	−1.43	0.013
39	Bromophos	Insecticide	−0.64	−0.73	0.005
40	Bromophos-ethyl	Insecticide	0.88	0.91	0.006
41	Bromoxynil	Herbicide	0.53	0.86	0.004
42	Bromoxynil heptanoate	Herbicide	0.13	0.09	0.003
43	Bromoxynil octanoate	Herbicide	0.23	0.17	0.003
44	Bromuconazole	Fongicide	0.06	0.25	0.003
45	Bronopol	Fongicide	−0.10	−0.04	0.003
46	Bupirimate	Fongicide	−1.10	−0.96	0.009
47	Butachlor	Herbicide	−0.81	−0.97	0.006
48	Butamifos	Herbicide	−0.28	−0.07	0.003
49	Butylate	Herbicide	−1.21	−1.29	0.010
50	Butocarboxim	Insecticide	0.16	0.01	0.003
51	Butonate	Insecticide	−0.53	−0.48	0.004
52	Butoxycarboxim	Insecticide	−0.31	−0.10	0.004
53	Butralin	Herbicide	−0.55	−0.66	0.005
54	Cadusafos	Insecticide	0.95	1.05	0.006
55	Camphechlor	Insecticide	0.92	0.41	0.006
56	Carbanolate	Insecticide	0.85	0.82	0.006
57	Carbaryl	Insecticide	−0.48	−0.39	0.004
58	Carbetamide	Herbicide	−0.86	−1.01	0.007
59	Carbofuran	Insecticide	1.50	1.38	0.012
60	Carbophenothion	Insecticide	1.54	1.44	0.012
61	Carbosulfan	Insecticide	0.58	0.82	0.004
62	Carboxin	Fongicide	−1.04	−0.85	0.008
63	Chlordane	Insecticide	−0.05	0.06	0.003
64	Chlordecone	Insecticide	0.73	0.77	0.005
65	Chlorethoxyfos	Insecticide	2.27	2.28	0.023
66	Chlorfenac	Herbicide	−0.87	−0.72	0.007
67	Chlorfenethol	Insecticide	−0.27	−0.53	0.003
68	Chloridazon	Herbicide	−0.98	−1.01	0.008
69	Chlorobenzilate	Insecticide	−0.93	−0.95	0.007
70	Chloromethiuron	Insecticide	−2.04	−2.10	0.021
71	Chlorophacinone	Rodonticide	2.08	1.98	0.020

No.	Compound	Type	log [1/LD_{50}] (mmol/kg)$^{-1}$		Leverage (h_i)
			Observed	Predicted	
72	Chloropicrin	Insecticide	−0.18	−0.22	0.003
73	Chlorpyrifos	Insecticide	0.74	0.52	0.005
74	Chlorpyrifos-methyl	Insecticide	−0.94	−1.07	0.007
75	Chlorthiamid	Herbicide	−0.56	−0.54	0.005
76	Chlorthion	Insecticide	−0.47	−0.68	0.004
77	Clethodim	Herbicide	−0.50	−0.38	0.004
78	Clodinafop-propargyl	Herbicide	−0.60	−0.65	0.005
79	Cloethocarb	Insecticide	0.86	1.08	0.006
80	Clomazone	Herbicide	-0.76	-0.55	0.006
81	Coumachlor	Rodonticide	1.33	1.13	0.010
82	Crotoxyphos	Insecticide	0.68	0.85	0.005
83	Cyanazine	Herbicide	−0.08	−0.23	0.003
84	Cyanophos	Insecticide	−0.40	−0.30	0.004
85	Cycloxydim	Herbicide	−1.08	−1.15	0.008
86	Cyhexatin	Insecticide	0.16	0.41	0.003
87	Cymoxanil	Fongicide	−0.58	−0.79	0.005
88	Cypermethrin	Insecticide	0.16	0.11	0.003
89	Cyphenothrin	Insecticide	0.07	−0.05	0.003
90	Cyprofuram	Fongicide	0.21	0.39	0.003
91	Cyromazine	Insecticide	−1.31	−1.40	0.011
92	Dalapon	Herbicide	−1.81	−1.72	0.018
93	Dazomet	Insecticide	−0.41	−0.05	0.004
94	Deltamethrin	Insecticide	0.76	1.01	0.005
95	Demeton-*S*-methyl sulfone	Insecticide	0.91	1.08	0.006
96	Desmetryn	Herbicide	−0.81	−0.88	0.006
97	Diafenthiuron	Insecticide	−0.73	−0.65	0.006
98	Di-allate	Herbicide	−0.16	−0.44	0.003
99	Dibromochloropropane	Insecticide	0.14	−0.14	0.003
100	Dichlone	Fongicide	0.15	0.21	0.003
101	Dichlorprop	Herbicide	−0.55	−0.65	0.005
102	Dichlorvos	Insecticide	0.44	0.56	0.004
103	Dicofane	Insecticide	0.50	0.35	0.004
104	Dicofol	Insecticide	−0.19	0.02	0.003
105	Dicrotophos	Insecticide	1.14	1.25	0.008
106	Dienochlor	Insecticide	−0.82	−0.93	0.006
107	Diethatyl ethyl	Herbicide	−0.87	−0.74	0.007
108	Difenamide	Herbicide	−0.61	−0.68	0.005
109	Diflovidazin	Insecticide	−0.29	−0.20	0.004

No.	Compound	Type	log [1/LD_{50}] (mmol/kg)$^{-1}$		Leverage (h_i)
			Observed	Predicted	
110	Diflumetorim	Fongicide	−0.14	−0.33	0.003
111	Dimetachlor	Herbicide	−0.80	−0.68	0.006
112	Dimethenamid	Herbicide	−0.16	−0.22	0.003
113	Dimethenamid-*P*	Herbicide	−0.19	−0.31	0.003
114	Dimethomorph	Fongicide	−1.00	−0.99	0.008
115	Dimethylvinphos	Insecticide	0.53	0.71	0.004
116	Dimexano	Herbicide	−0.05	−0.09	0.003
117	Dinobuton	Fongicide	0.37	0.27	0.003
118	Dinoseb	Herbicide	0.98	1.11	0.007
119	Dinoterb	Insecticide	0.98	0.99	0.007
120	Dioxathion	Insecticide	1.30	1.09	0.009
121	Diphacinone	Rodonticide	2.17	2.18	0.021
122	Diquat	Herbicide	−0.06	−0.11	0.003
123	Dithianon	Fongicide	−0.01	−0.07	0.003
124	Diuron	Herbicide	−0.27	−0.47	0.003
125	Edifenphos	Fongicide	0.32	0.17	0.003
126	Endothal	Herbicide	0.56	0.35	0.004
127	EPN	Insecticide	1.36	1.26	0.010
128	EPTC	Herbicide	−0.68	−0.86	0.005
129	Ethanedial	Herbicide	−0.31	−0.24	0.004
130	Ethoate-methyle	Insecticide	−0.15	0.14	0.003
131	Ethoxysulfuron	Herbicide	−0.91	−0.94	0.007
132	Fenamidone	Fongicide	−0.81	−0.77	0.006
133	Fenchlorphos	Insecticide	−0.19	−0.17	0.003
134	Fenobucarb	Insecticide	−0.48	−0.22	0.004
135	Fenoprop	Herbicide	−0.38	−0.63	0.004
136	Fenpropathrin	Insecticide	−0.40	−0.38	0.004
137	Fenpropidin	Fongicide	−0.73	−0.71	0.006
138	Fenpropimorph	Fongicide	−0.74	−0.53	0.006
139	Fensulfothion	Insecticide	2.15	2.18	0.021
140	Fentin acetate	Fongicide	0.47	0.62	0.004
141	Fenvalerate	Insecticide	−0.03	−0.12	0.003
142	Fipronil	Insecticide	0.68	0.85	0.005
143	Florasulam	Herbicide	−1.14	−1.17	0.009
144	Fluazifop-butyl	Herbicide	−0.90	−1.11	0.007
145	Fluchloralin	Herbicide	−0.64	−0.59	0.005
146	Flucythrinate	Insecticide	0.83	0.84	0.005
147	Flufenacet	Herbicide	−0.22	−0.21	0.003

No.	Compound	Type	log [1/LD_{50}] (mmol/kg)$^{-1}$		Leverage (h_i)
			Observed	Predicted	
148	Flumorph	Fongicide	−0.86	−1.01	0.007
149	Fluoroacetamide	Insecticide	0.77	0.66	0.005
150	Fluquinconazole	Fongicide	0.53	0.41	0.004
151	Flusilazole	Fongicide	−0.33	−0.27	0.004
152	Fluvalinate	Insecticide	0.28	0.18	0.003
153	Fomesafen	Herbicide	−0.45	−0.38	0.004
154	Fonofos	Insecticide	1.56	1.55	0.012
155	Formetanate	Insecticide	1.17	1.21	0.008
156	Formothion	Insecticide	−0.15	−0.25	0.003
157	Fospirate	Insecticide	−0.45	−0.34	0.004
158	Fosthiazate	Insecticide	0.70	0.90	0.005
159	Furathiocarb	Insecticide	0.86	0.65	0.006
160	Furfural	Fongicide	0.17	−0.06	0.003
161	Gamma-cyhalothrine	Insecticide	0.91	0.76	0.006
162	Halfenprox	Insecticide	0.56	0.61	0.004
163	Halosulfuron-methyl	Herbicide	−1.25	−1.11	0.010
164	Heptenophos	Insecticide	0.42	0.39	0.004
165	Hexaconazole	Fongicide	−0.84	−0.88	0.006
166	Hexazinone	Herbicide	-0.83	-0.81	0.006
167	Hymexazol	Fongicide	−1.21	−1.43	0.010
168	Icaridin	Insecticide	−0.99	−1.07	0.008
169	Imiprothrin	Insecticide	−0.45	−0.29	0.004
170	Ioxynil	Herbicide	0.46	0.62	0.004
171	Iprobenfos	Fongicide	−0.37	−0.59	0.004
172	Isocarbophos	Insecticide	0.76	0.58	0.005
173	Isoprocarb	Insecticide	−0.32	−0.41	0.004
174	Isoprothiolane	Fongicide	−0.61	−1.11	0.005
175	Isoproturon	Herbicide	−0.95	−0.73	0.007
176	Isoxathion	Insecticide	0.45	0.56	0.004
177	Kelevan	Insecticide	0.42	0.40	0.004
178	Lambda-cyhalothrin	Insecticide	0.91	1.03	0.006
179	Lindane	Insecticide	0.25	0.14	0.003
180	Linuron	Herbicide	−0.66	−0.79	0.005
181	Malathion	Insecticide	−0.73	−0.44	0.006
182	MCPA-thioethyl	Herbicide	−0.26	−0.33	0.003
183	MCPB	Herbicide	−1.27	−1.31	0.010
184	Mecarbam	Insecticide	0.96	0.96	0.006
185	Mepiquat	Herbicide	−1.12	−1.23	0.009

No.	Compound	Type	log [1/LD_{50}] (mmol/kg)$^{-1}$		Leverage (h_i)
			Observed	Predicted	
186	Metalaxyl	Fongicide	−0.36	−0.29	0.004
187	Metamitron	Herbicide	−0.77	−0.59	0.006
188	Methomyl	Insecticide	0.73	0.98	0.005
189	Metominostrobin	Fongicide	−0.40	−0.24	0.004
190	Metsulfovax	Fongicide	−1.23	−1.43	0.010
191	Mevinphos	Insecticide	1.81	1.88	0.016
192	Monocrotophos	Insecticide	1.20	1.26	0.008
193	Morphothion	Insecticide	0.18	0.33	0.003
194	Naled	Insecticide	0.66	0.78	0.004
195	Naptalam	Herbicide	−0.78	−0.87	0.006
196	Nithiazine	Insecticide	−0.15	0.07	0.003
197	Nitrapyrin	Bactéricide	−0.49	−0.19	0.004
198	Nitrofen	Herbicide	−0.97	−1.13	0.007
199	Octhilinone	Fongicide	−0.41	−0.35	0.004
200	Ofurace	Fongicide	−0.97	−1.06	0.007
201	Oxycarboxin	Fongicide	−0.79	−0.09	0.006
202	Oxydemeton-methyl	Insecticide	0.71	0.61	0.005
203	Paraquat	Herbicide	0.23	0.32	0.003
204	Parathion	Insecticide	2.16	2.28	0.021
205	Parathion methyl	Insecticide	1.94	1.84	0.018
206	Pebulate	Herbicide	−0.74	−0.88	0.006
207	Pethoxamid	Herbicide	−0.52	−0.22	0.004
208	Phenkapton	Insecticide	0.93	0.45	0.006
209	Phenthoate	Insecticide	0.11	0.08	0.003
210	Phosalone	Insecticide	0.49	0.53	0.004
211	Picloram	Herbicide	−1.22	−1.18	0.010
212	Piperophos	Herbicide	0.04	−0.08	0.003
213	Pirimicarb	Insecticide	0.22	0.46	0.003
214	Plifenate	Insecticide	−1.47	−1.41	0.013
215	Prallethrin	Insecticide	−0.18	−0.06	0.003
216	Pretilachlor	Herbicide	−1.29	−1.27	0.011
217	Prometon	Herbicide	−0.83	−0.77	0.006
218	Propanil	Herbicide	−0.64	v0.68	0.005
219	Propargite	Insecticide	−0.88	−0.81	0.007
220	Propiconazole	Fongicide	−0.45	−0.25	0.004
221	Propoxur	Insecticide	0.62	0.36	0.004
222	Prosulfuron	Herbicide	−0.11	−0.26	0.003
223	Prothiofos	Insecticide	−0.43	−0.66	0.004

No.	Compound	Type	log [1/LD_{50}] (mmol/kg)$^{-1}$		Leverage (h_i)
			Observed	Predicted	
224	Pymetrozine	Insecticide	−1.43	−1.46	0.012
225	Pyrazophos	Fongicide	0.39	−0.18	0.003
226	Pyrazoxyfen	Herbicide	−0.61	−0.77	0.005
227	Pyridaben	Insecticide	0.36	0.37	0.003
228	Pyridafenthion	Insecticide	-0.35	-0.34	0.004
229	Pyrifenox	Fongicide	−0.99	−1.07	0.008
230	Pyrimethanil	Fongicide	−1.32	−1.34	0.011
231	Pyroquilone	Fongicide	−0.27	−0.52	0.003
232	Quinalphos	Insecticide	0.62	0.41	0.004
233	Quinclorac	Herbicide	−1.04	−1.01	0.008
234	Sethoxydim	Herbicide	−1.06	−1.08	0.008
235	Simetryn	Herbicide	−0.38	−0.58	0.004
236	Sulfotep	Insecticide	1.74	1.71	0.015
237	Sulfoxaflor	Insecticide	−0.49	−0.27	0.004
238	Sulprofos	Insecticide	0.24	0.32	0.003
239	Tebuconazole	Fongicide	−0.68	−0.83	0.005
240	Tecloftalam	Fongicide	−0.95	−1.02	0.007
241	Tecnazene	Fongicide	−0.52	−0.47	0.004
242	Tefluthrin	Insecticide	1.31	1.27	0.009
243	Thiocarboxime	Insecticide	1.16	0.90	0.008
244	Thiodicarb	Insecticide	0.64	0.77	0.004
245	Thiofanox	Insecticide	1.46	1.38	0.011
246	Thiometon	Insecticide	0.79	0.73	0.005
247	Tolfenpyrad	Insecticide	−0.05	−0.12	0.003
248	Tralkoxydim	Herbicide	−0.15	−0.11	0.003
249	Tri-allate	Herbicide	−0.44	−0.34	0.004
250	Tribufos	Herbicide	0.04	−0.15	0.003
251	Trichlorfon	Insecticide	0.20	0.13	0.003
252	Trichloronate	Insecticide	1.03	0.89	0.007
253	Tricyclazole	Fongicide	0.01	−0.14	0.003
254	Tridiphane	Herbicide	−0.88	−0.91	0.007
255	Trietazine	Herbicide	−0.08	−0.18	0.003
256	Triflumizole	Fongicide	−0.47	−0.68	0.004
257	Trimethacarb	Insecticide	−0.25	−0.23	0.003
258	Vamidothion	Insecticide	0.65	1.02	0.004
		Validation set			
259	2,4-D	Herbicide	−0.33	−0.39	0.004

No.	Compound	Type	log [1/LD_{50}] (mmol/kg)$^{-1}$		Leverage (h_i)
			Observed	Predicted	
260	Aldoxycarb	Insecticide	0.92	0.87	0.006
261	Allethrin	Insecticide	v0.35	−0.14	0.004
262	Alpha-cypermethrin	Insecticide	0.86	0.86	0.006
263	Azinphos-ethyl	Insecticide	1.46	1.42	0.011
264	Barban	Herbicide	−0.31	−0.31	0.004
265	Bensulide	Herbicide	0.17	0.21	0.003
266	Bensultap	Insecticide	−0.41	−0.16	0.004
267	Beta-cypermethrin	Insecticide	0.65	0.52	0.004
268	Chlorbromuron	Herbicide	−0.86	−0.85	0.007
269	Chlorbufam	Herbicide	−1.03	−0.88	0.008
270	Chlorpropham	Herbicide	−1.29	−1.58	0.011
271	Closantel	Insecticide	0.40	0.12	0.003
272	Crimidine	Rodonticide	2.14	2.18	0.021
273	Demeton-*S*-methyl	Insecticide	0.76	1.05	0.005
274	Dichlorprop-*P*	Herbicide	−0.38	−0.76	0.004
275	Dimethoate	Insecticide	−0.03	−0.03	0.003
276	Dinocap	Fongicide	−0.52	−0.54	0.004
277	Dioxabenzophos	Insecticide	0.24	0.26	0.003
278	Ditalimfos	Fongicide	−1.22	−1.15	0.010
279	DNOC	Herbicide	0.90	0.87	0.006
280	Endosulfan	Insecticide	1.03	0.71	0.007
281	Etaconazole	Fongicide	−0.61	−0.52	0.005
282	Ethiofencarb	Insecticide	0.05	−0.05	0.003
283	Ethiprole	Insecticide	−1.25	−1.25	0.010
284	Fenarimol	Fongicide	−0.88	−1.11	0.007
285	Fenazaquin	Acaricide	0.36	0.38	0.003
286	Fenitrothion	Insecticide	−0.08	−0.08	0.003
287	Flonicamid	Insecticide	−0.59	−0.47	0.005
288	Fluazifop-*P*-butyl	Herbicide	−0.81	−0.51	0.006
289	Fluoroglycofen	Herbicide	−0.55	−0.43	0.005
290	Furalaxyl	Fongicide	−0.50	−0.75	0.004
291	Furmecyclox	Fongicide	−1.18	−1.36	0.009
292	Glufosinate	Herbicide	−0.95	−0.91	0.007
293	Glutaraldehyde	Fongicide	−0.13	0.05	0.003
294	Halofenozide	Insecticide	−0.94	−1.08	0.007
295	Imazalil	Fongicide	0.12	0.35	0.003
296	Indoxacarb	Insecticide	0.29	0.19	0.003
297	Isofenphos-methyl	Insecticide	1.19	1.37	0.008

No.	Compound	Type	log [1/LD_{50}] (mmol/kg)$^{-1}$		Leverage (h_i)
			Observed	Predicted	
298	Leptophos	Insecticide	0.98	0.76	0.007
299	MCPA	Herbicide	−0.68	−0.79	0.005
300	Mecoprop	Herbicide	−0.73	−1.09	0.006
301	Metazachlor	Herbicide	−1.10	−0.95	0.009
302	Metconazole	Fongicide	−0.27	−0.06	0.003
303	Methazole	Herbicide	−0.47	−0.61	0.004
304	Methidathion	Insecticide	1.08	1.03	0.007
305	Metolachlor	Herbicide	−0.63	−0.63	0.005
306	Metribuzin	Herbicide	0.83	0.52	0.005
307	Molinate	Herbicide	−0.41	−0.18	0.004
308	Monolinuron	Herbicide	−0.99	−0.63	0.008
309	Nitenpyram	Insecticide	−0.76	−0.85	0.006
310	Oxadixyl	Fongicide	−0.82	−1.12	0.006
311	Oxamyl	Insecticide	1.94	2.04	0.018
312	Pendimethalin	Herbicide	−1.05	−1.27	0.008
313	Phosmet	Insecticide	0.45	0.04	0.004
314	Profenofos	Insecticide	0.02	−0.01	0.003
315	Promecarb	Insecticide	0.77	0.58	0.005
316	Propazine	Herbicide	−1.22	−1.49	0.010
317	Prosulfocarb	Herbicide	-0.86	-0.80	0.007
318	Prothoate	Insecticide	1.55	1.14	0.012
319	Tebutam	Herbicide	−1.43	−1.51	0.012
320	Tebuthiuron	Herbicide	−0.16	−0.07	0.003
321	Tepraloxydim	Herbicide	−1.36	−1.48	0.011
322	Terbufos	Insecticide	2.25	2.14	0.023
323	Tetraconazole	Fongicide	−0.46	−0.79	0.004
324	Thiacloprid	Insecticide	−0.18	−0.24	0.003
325	Thiobencarb	Herbicide	−0.44	−0.66	0.004
326	Tralomethrin	Insecticide	0.57	0.65	0.004
327	Triazamate	Insecticide	0.71	0.49	0.005
328	Tridemorph	Fongicide	−0.19	−0.31	0.003
329	Vernolate	Herbicide	−0.72	−0.12	0.006

Table 2.

List of descriptors used in the development of QSAR model.

Category	Descriptor	Description
2D Autocorrelations indices	MATS2p	Moran autocorrelation of lag 2 weighted by polarizability
	MATS1m	Moran autocorrelation of lag 1 weighted by mass
Atom-centred fragments	N-072	RCO-N</>*N* − *X* = *X*
	H-046	H attached to C0(sp3) no *X* attached to next C
Geometrical descriptors	PJI3	3D Petitjean shape index
Getaway descriptors	H6m	H autocorrelation of lag 6/weighted by mass
	HATSe	Leverage-weighted total index/weighted by Sanderson electronegativity
	HATS0m	Leverage-weighted autocorrelation of lag 0/weighted by mass
RDF descriptor	RDF020e	Radial distribution function—020/weighted by Sanderson electronegativity
	RDF030e	Radial distribution function—030/weighted by Sanderson electronegativity
3D-Morse descriptor	Mor15m	Signal 15/weighted by mass
	Mor23u	Signal 23/unweighted
	Mor26u	Signal 26/unweighted
Whim descriptors	Du	*D* total accessibility index/unweighted
	E1u	1st component accessibility directional WHIM index/unweighted
Functional group counts	nArX	Number of *X* on aromatic ring
Constitutional indices	nS	Number of sulfur atoms

Table 3.

Selected parameters of the optimal multi-layer perceptron.

Parameters studied	MSE (minimum value)	Selected parameters
The database distribution		
Training (80%) and validation (20%)	0.0311	
Training (79%) and validation (21%)	0.0317	
Training (78.5%) and validation (21.5%)	**0.0295**	Training (78.5%) and validation (21.5%)
Training (78%) and validation (22%)	0.0345	
Training (77%) and validation (23%)	0.0382	
Activation functions (hidden neurons/output neurons)		
Sigmoid–sigmoid	0.0291	
Sigmoid–linear	0.0293	
Sigmoid–tangent hyperbolic	0.1054	
Tangent hyperbolic–sigmoid	0.1719	
Tangent hyperbolic–linear	**0.0290**	Tangent hyperbolic–linear
Tangent hyperbolic–tangent hyperbolic	0.0293	
Linear–sigmoid	0.1563	
Linear–tangent hyperbolic	0.0306	
Linear–linear	0.0299	
Number of neurons in the hidden layer		
1–16	0.0290	9 Neurons
Learning algorithms		
Quasi–Newton back propagation (BFGS)	**0.0290**	
Levenberg–Marquardt (LM)	0.0293	
Scaled conjugate gradient (SCG)	0.0395	Quasi–Newton back propagation (BFGS)
Conjugate gradient descent (CGP)	0.0346	

Table 4.

Performance of MLP-ANN model for pesticides.

Training set (*n* = 258)	R^2	0.963
	Q^2_{LOO}	0.962
	RMS	0.164
Validation set (*n* = 71)	R^2_{ext}	0.95
	Q^2_{ext}	0.948
	RMS	0.201