Journal of Hazardous Materials 303 (2016) 28–40

Contents lists available at ScienceDirect

Journal of Hazardous Materials

journal homepage: www.elsevier.com/locate/jhazmat

A Quantitative Structure Activity Relationship for acute oral toxicity of pesticides on rats: Validation, domain of application and prediction

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

[a] Laboratoire des Biomatériaux et Phénomènes de Transport (LBMPT), Université de Médéa, Quartier Ain D'heb, 26000 Medea, Algeria
[b] 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

HIGHLIGHTS

- QSAR model for acute oral toxicity of pesticides is developed and proposed.
- This model has been developed and validated on the basis of the OECD principles for QSAR acceptance in regulation.
- The new nonlinear model showed excellent robustness, fitness and predictive ability.
- The majority of pesticides were located within the AD, further indicated the reliability of the prediction.

ARTICLE INFO

Article history:
Received 22 July 2015
Received in revised form 7 September 2015
Accepted 9 September 2015

Keywords:
Acute toxicity
Pesticides
QSAR
Prediction
External validation

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.

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

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.

* Corresponding author. Fax: +213 25 58 12 53.
E-mail addresses: mhamdeche@yahoo.fr (M. Hamadache), benkortbi_oth@yahoo.fr (O. Benkortbi), s_hanini2002@yahoo.fr (S. Hanini), abdeltif.amrane@univ-rennes1.fr (A. Amrane), latifa_khaouane@yahoo.fr (L. Khaouane), simoussa_cherif@yahoo.fr (C. Si Moussa).

http://dx.doi.org/10.1016/j.jhazmat.2015.09.021
0304-3894/

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 human 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 (e.g., 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 et al. [29], Guo et al. [30] and Garcıa-Domenech 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 Organization 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 $\log[1/(\text{mmol/kg})]$.

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