Ecotoxicology and Environmental Safety 153 (2018) 151–159

Contents lists available at ScienceDirect

Ecotoxicology and Environmental Safety

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

Predictive ecotoxicity of MoA 1 of organic chemicals using *in silico* approaches

Luana de Morais e Silva[a], Mateus Feitosa Alves[b], Luciana Scotti[c], Wilton Silva Lopes[a], Marcus Tullius Scotti[c,*]

[a] Post-Graduate Program in Science and Environmental Technology, Department of Sanitary and Environmental Engineering, State University of Paraíba, 58429500 Campina Grande, PB, Brazil
[b] Pharmacy Department, Federal University of Paraiba, 58051900 João Pessoa, PB, Brazil
[c] Post-Graduate Program in Natural and Synthetic Bioactive Products, Federal University of Paraíba, 58051-900 João Pessoa, PB, Brazil

ARTICLE INFO

Keywords:
QSAR
Fish acute toxicity
Molecular descriptors
Volsurf software
MoA 1

ABSTRACT

Persistent organic products are compounds used for various purposes, such as personal care products, surfactants, colorants, industrial additives, food, pesticides and pharmaceuticals. These substances are constantly introduced into the environment and many of these pollutants are difficult to degrade. Toxic compounds classified as MoA 1 (Mode of Action 1) are low toxicity compounds that comprise nonreactive chemicals. *In silico* methods such as Quantitative Structure–Activity Relationships (QSARs) have been used to develop important models for prediction in several areas of science, as well as aquatic toxicity studies. The aim of the present study was to build a QSAR model-based set of theoretical Volsurf molecular descriptors using the fish acute toxicity values of compounds defined as MoA 1 to identify the molecular properties related to this mechanism. The selected Partial Least Squares (PLS) results based on the values of cross-validation coefficients of determination (Q_{cv}^2) show the following values: Q_{cv}^2 = 0.793, coefficient of determination (R^2) = 0.823, explained variance in external prediction (Q_{ext}^2) = 0.87. From the selected descriptors, not only the hydrophobicity is related to the toxicity as already mentioned in previously published studies but other physicochemical properties combined contribute to the activity of these compounds. The symmetric distribution of the hydrophobic moieties in the structure of the compounds as well as the shape, as branched chains, are important features that are related to the toxicity. This information from the model can be useful in predicting so as to minimize the toxicity of organic compounds.

1. Introduction

Organic compounds represent a wide range of chemicals with different physicochemical properties, whether from synthetic or natural sources. These include persistent organic products, which are compounds used for various purposes, such as personal care products, surfactants, dyes, industrial additives, food, pesticides and pharmaceuticals (Isidori et al., 2016; Ribeiro et al., 2015; Jurado et al., 2012; Tamura et al., 2017). These substances are constantly introduced into the environment, originating mainly from industrial, domestic, hospital and surface effluent releases from agricultural and livestock areas (Esteban et al., 2014; Mousel et al., 2017; Serra-Roig et al., 2016). Its occurrence is widely detected in different aquatic matrices, surface water, groundwater and oceans (Busch et al., 2016; Gros et al., 2012, 2016). Many of these pollutants are difficult to degrade, having high chemical stability. Exposure of chemical contaminants to the aquatic environment poses serious threats to the preservation of environmental quality, as well as issues related to human health and is recognized by experts from several countries as a major global problem (Bourgin et al., 2013).

Adverse effects have already been identified and continue to be related to the presence of pollutants in water, such as inhibition of neurotransmitters (de Oliveira et al., 2016), mutagenicity (Lutterbeck et al., 2015), carcinogenicity, biomagnification in marine trophic nets (Xue et al., 2017), phytotoxicity (Richter et al., 2016), feminization of fish (Hicks et al., 2017) and development of bacteria resistant to antibiotics (Miranda et al., 2016).

Studies developed in toxicology have used physicochemical information aiming to classify chemicals by the common mode of action (MoA) to understand better the interaction by which a chemical can cause an adverse effect (Nendza et al., 2014). The MoA can be defined as a common set of physiological and behavioral signs produced in an exposed organism that characterize a type of adverse biological response (Borgert et al., 2004; Schlosser and Bogdanffy, 1999; Rand et al.,

* Corresponding author.
E-mail address: mtscotti@ccae.ufpb.br (M.T. Scotti).

https://doi.org/10.1016/j.ecoenv.2018.01.054
Received 5 September 2017; Received in revised form 29 December 2017; Accepted 29 January 2018

1995).

According to Verhaar (1992), the MoA can be categorized into four classes based on identification of chemical groups and/or structural fragment: MoA 1 (nonpolar narcosis)—it is also called baseline toxicity or minimum toxicity and comprises nonreactive chemicals; MoA 2 (polar narcosis) are classified as ionic and inert organic compounds, which present low toxicity and these chemicals are usually characterized as possessing hydrogen-bond donor acidity; MoA 3 (reactive chemicals)—applies to chemicals that react unselectively with certain chemical structures commonly found in biomolecules or chemicals that are metabolized and MoA 4 (specifically acting chemicals)—inclusion in this class must, and should, be based on specific knowledge on mode of toxic action of groups of chemicals. Compounds that cannot be classified as belonging to classes 1, 2 or 3 and that are not known to act by a specific mechanism can only be classified as "not possible to classify according to these rules" (Verhaar et al., 1992, 2000).

The use of *in silico* methods, such as *(Quantitative) Structure–Activity Relationships [(Q)SARs]*, has become increasingly helpful in understanding many aspects of chemical–biological interactions in drug and pesticide research, as well as in the areas of toxicology (Hansch and Verma, 2009). (Q)SAR is based on quantitative models, derived from application of mathematical and statistical tools, resulting in correlation between physicochemical properties and biological activities (e.g., toxicity) from a variety of chemical classes, numerically encoded by one or more molecular descriptors that are used to predict the properties of interest (Consonni et al., 2002; Dearden, 2016; Gasteiger, 2016; OECD, 2014; Todeschini and Consonni, 2009). This method provides an important and useful alternative means of testing; it can be employed to deal with large quantities of data, for assessing the potential hazards of chemicals, as well as to avoid expensive and time-consuming experiments and is an important alternative to reduce animal experimental testing (Chen et al., 2015; Gramatica et al., 2016). QSAR models have been introduced in chemical safety assessment and regulatory decision support, required by European Legislation on Registration, Evaluation, Authorization, and restriction of CHemicals (REACH) and by the United States Environmental Protection Agency (US EPA), for comprehensive protection of public health and the environment (Delgado et al., 2012; ECHA, 2014, US EPA, 2016).

Recent QSAR models have been proposed for (eco)toxicity prediction of chemical substances, including fish toxicity studies (Toropova et al., 2012; Burden et al., 2016; Cassani et al., 2013; Cassotti et al., 2016; Gramatica et al., 2016; Kluver et al., 2016; Levet et al., 2013; Nendza et al., 2017; Singh et al., 2014; Schüürmann et al., 2011; Tugcu et al., 2012). Singh et al. (2013) constructed a probability function-based neural networks model for predicting the toxicity of diverse chemical compounds. The models showed good predictive and generalization abilities for predicting toxicities. Another study demonstrated a classification scheme to discriminate between baseline and excess toxicants for replacing fish acute toxicity tests with QSAR predictions for baseline toxicants (Nendza et al., 2017). Burden et al. (2016) reported a predictive study to evaluate the fish acute toxicity of metabolites derived from plant protection product active substances. The work demonstrated a high correlation of QSAR-predicted versus experimentally derived fish acute toxicity values.

The aim of this study was to build a QSAR model-based set of theoretical molecular descriptors using acute fish toxicity values for compounds defined as MoA 1 to identify the molecular properties related to this mechanism and predict the fish toxicity of untested compounds. The QSAR model was validated according to the principles of validation for regulatory purposes and the acceptability of (Q)SARs, proposed by the Organization for Economic Cooperation and Development–OECD (OECD, 2007).

2. Methods

2.1. Data collection

For the development of QSAR models, the dataset comprised 61 organic compounds with experimental values on acute fish toxicity. The dataset was split prior to model development into training and test subsets. The experimental data for the training used in the present study were constituted by over 36 compounds (Tables S1 and S2), obtained from the database reported in the literature containing 86 different experimental toxicological values of compounds classified in MoA 1 (nonpolar narcosis), all information of duplicates being used (Thomas et al., 2015). The predictivity of the model was determined by an external validation based on the predictions performed for the independent test set with 25 individuals and 25 unique experimental values of fish toxicity with the aim of assessing the robustness of the developed models (Konemann, 1981; Konemann and Musch, 1981). The chemicals included in the prediction sets were not used in the model development step to select the modeling descriptors. The experimental values measured for EC_{50} (effective concentration) (mol/L) were converted to a molar basis and then the logarithmically transformed data [$-\log EC_{50} = pEC_{50}$] were used as response variables.

2.2. Molecular descriptors

To generate the molecular descriptors, the individual structures were defined as *Simplified Molecular Line Entry System (SMILES)* codes. The chemical structures were converted into a.mol file (MDL format) and used as input in the program for calculation. The Standardizer tool was used to convert the chemical structures into customized canonical representations, [JChem 16.1.11.0, 2016), ChemAxon (http://www.chemaxon.com)], ensuring comparability of all molecular representations; stripping salts, adding explicit hydrogens and aromatizing were the standardizations applied to every structure. This tool also was used to generate three-dimensional (3D) structures.

Using the VolSurf+ program v. 1.0.7 (http://www.moldiscovery.com), 3D structures of the compounds were used as input data and were subjected to molecular interaction fields (MIFs) to generate descriptors using the following probes: N1 (amide nitrogen), O (carbonyl oxygen), H_2O (water probe) and DRY (hydrophobic probe) (Cruciani et al., 2000). Additional non-MIF-derived descriptors were generated to create a total of 128 descriptors, including descriptors that quantify molecular size, shape, hydrophilic and hydrophobic regions, interaction energy moments, capacity factors, amphiphilic moments, hydrophobic–lipophilic balance and other descriptors. These descriptors have been selected for the present study because they are simple to use and make it easy to interpret and understand the mechanism of action and/or physical meaning. Volsurf descriptors have been previously used to build a QSSR (Quantitative Structure–Sorption Relationships) model to predict K_{oc} values for the chemical structures of the active ingredients found in pesticides (Soares et al., 2014).

2.3. Statistical analysis and model validation

QSAR models were developed using Partial Least Squares (PLS) included in the VolSurf+ software. PLS is a statistical procedure based on linear regression that allows extracting and rationalizing the multivariate information, to explain the maximum correlation between the descriptors matrix X and response matrix Y by calculating a new set of orthogonal variables, therefore uncorrelated, called latent variables (LVs). This method is suitable when the number of variables is greater than the number of samples and there is multicollinearity among the independent variables (Baroni et al., 1993; Geladi and Kowalski, 1986; Wold et al., 1984, 2001). The autoscaling preprocess, in which the mean is subtracted from the variable values and the resultant values are divided by the standard deviation, for all independent variables was