Journal of Theoretical Biology 380 (2015) 16–23

Contents lists available at ScienceDirect

Journal of Theoretical Biology

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

Development of quantitative interspecies toxicity relationship modeling of chemicals to fish

M.H. Fatemi*, E. Mousa Shahroudi, Z. Amini

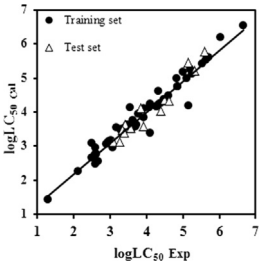
Laboratory of Chemometrics, Department of Chemistry, University of Mazandaran, Babolsar, Iran

HIGHLIGHTS

- Quantitative interspecies-toxicity relationship methodologies were used to improve the prediction power of interspecies toxicity model.
- By analyzing descriptors in the model, it was concluded that geometrical, topological and electronic interactions are important in the true estimation of the toxicity of chemicals.
- Developed QSTR model successfully was used to correlate fish toxicity of chemical to their molecular structural descriptors as well as their toxicity to *Daphnia magna*.

GRAPHICAL ABSTRACT

Quantitative interspecies-toxicity relationship methodologies were used to improve the prediction power of interspecies toxicity model.



ARTICLE INFO

Article history:
Received 10 January 2015
Received in revised form
5 May 2015
Accepted 9 May 2015
Available online 19 May 2015

Keywords:
Toxicity
Quantitative interspecies-toxicity relationship
Daphnia magna
Molecular descriptors

ABSTRACT

In this work, quantitative interspecies-toxicity relationship methodologies were used to improve the prediction power of interspecies toxicity model. The most relevant descriptors selected by stepwise multiple linear regressions and toxicity of chemical to *Daphnia magna* were used to predict the toxicities of chemicals to fish. Modeling methods that were used for developing linear and nonlinear models were multiple linear regression (MLR), random forest (RF), artificial neural network (ANN) and support vector machine (SVM). The obtained results indicate the superiority of SVM model over other models. Robustness and reliability of the constructed SVM model were evaluated by using the leave-one-out cross-validation method (Q^2=0.69, SPRESS=0.822) and Y-randomization test (R^2=0.268 for 30 trail). Furthermore, the chemical applicability domains of these models were determined via leverage approach. The developed SVM model was used for the prediction of toxicity of 46 compounds that their experimental toxicities to a fish were not being reported earlier from their toxicities to *D. magna* and relevant molecular descriptors.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Until recently medical substances (pharmaceuticals) have been exposed to the environment with very little attention (Fent et al., 2006; Jones et al., 2002; Sanderson and Thomsen, 2007). The major routes of entry of pharmaceuticals into the environment are from their use by individuals, either dispersed throughout the community or concentrated in medical centers of hospitals, and the disposal of unwanted or expired drugs by users. Therefore pharmaceuticals can be frequently found in rivers, streams, sewage effluents, surface water, and groundwater and even in drinking water, which is creating a big dilemma in water treatment for drinking purpose (Daughton and Jones-Lepp, 2001; Halling-Sørensen et al., 1998;

* Corresponding author. Tel.: +98 1135302395; fax: +98 1135302350.
E-mail address: mhfatemi@umz.ac.ir (M.H. Fatemi).

http://dx.doi.org/10.1016/j.jtbi.2015.05.017
0022-5193/© 2015 Elsevier Ltd. All rights reserved.

Heberer et al., 1997; Ternes, 1998). The first report regarding detection of pharmaceuticals in wastewater effluents and surface waters was published in the United States in the 1970s (Garrison et al., 1976; Hignite and Azarnoff, 1977; Tabak and Bunch, 1970). Then clofibric acid and diazepam were detected in treated drinking water in Milan, Italy (Zuccato et al., 2000). Moreover Heberer and colleagues have reported the presence of clofibric acid, propylphenazone, and diclofenac in the drinking water of Berlin in the concentration range of several hundreds of nanograms per liter (Heberer, 2002). Also Frick detected three widely used nonprescription drugs, caffeine, cotinine, and acetaminophenone, in samples of potable water collected near Atlanta, Georgia (Frick and Zaugg, 2003). Although the current concentrations of pharmaceutical chemicals in both surface waters and effluents are low, their possible adverse effects on aquatic life and ultimately on human health are not well understood (Cunningham et al., 2006). Determination of the toxic effects of the pharmaceuticals on aquatic life can be elicited by either ecotoxicity testing or by applicable models. The experimental studies on pharmaceuticals are in general very limited relative to the studies on industrial chemicals, therefore development of some theoretical models for estimation of the toxicological properties of these compounds is important. Today quantitative structure-toxicity relationships (QSTRs) studies have been used successfully for modeling in the field of toxicology and drug design (Kar and Roy, 2010; Länge and Dietrich, 2002). For example, Tugcu et al. (2012) developed a QSAR model to estimate the acute pharmaceutical toxicity on fish by using multiple descriptors (Tugcu et al., 2012). They concluded that the toxicity of studied chemicals mainly depends on their hydrophobicity and heteroatom-bonded carbon atom. Thomsen and Sanderson (2009) performed comparative analysis of pharmaceuticals versus industrial chemicals acute aquatic toxicity (Sanderson and Thomsen, 2009). The results of this analysis indicate that pharmaceuticals are generally not more hazardous than industrial chemicals relative to their acute aquatic toxicity. In 2001 Di Marzio et al. studied the application of WHIM molecular descriptors in QSAR modeling of the fish toxicity of some aromatic hydrocarbons (Di Marzio et al., 2001). Also Duchowicz et al. (2009) developed quantitative structure-toxicity relationship models for predicting the fish toxicity of a diverse set of 92 benzene derivatives. While there are many reports about the toxicity of chemicals against small biological species such as *Daphnia magna*, there is little report on bigger animals such as fishes due to ethical limitation in the experiments on bigger animals. Therefore it was very interesting to predict the toxicity of chemicals against animals from their toxicities against smaller biological species. Reviewing the literature indicates that very limited number of quantitative interspecies-toxicity relationship models have been reported on interspecies quantitative correlation of ecotoxicity of pharmaceuticals. The main aim of the present work was to develop a QSTR model for the prediction of toxicity of some organic chemicals on fish from their toxicities to *D. magna* and molecular structural descriptors.

2. Methodology

2.1. Data set

In this study, the data set was taken from Sanderson and Thomsen report (Sanderson and Thomsen, 2009). The toxicity of chemicals was reported as LC_{50} (mM) in logarithmic scale and was considered as a dependent variable. In this paper the values of LC_{50} for 77 chemicals on a fish and *D. magna* were reported that were selected as the main data set. Moreover the value of LC_{50} on *D. magna* for 59 chemicals was reported in reference (Kar and Roy, 2010) where their toxicities on fish were not experimentally determined. This set was considered as an external set, and the values of LC_{50} of these chemicals on fish were predicted form their toxicities on *D. magna* by our developed model. Two compounds were eliminated from the main data set due to the inconsistency between their names and CAS numbers. Before the division of compounds into training and test sets, principal component analysis (PCA) score plot was used to check the applicability range of data and detect the outlier compounds (Fig. 1). As can be seen in Fig. 1 six compounds were obviously located separately from other compounds and were eliminated from the data set. These compounds were Acroleine (4), Budesonide (12), Ethyl bromide (29), Famotidine (31), Finasteride (33) and Nitroglycerin (53). Therefore, the data set has 69 members. The names and specification of these compounds and those that were used as external test chemicals are shown in Table 1. The main data set was split into two independent subsets, trainings and test sets, by sample set partitioning based on joint *x*–*y* distances (SPXY) algorithm. This algorithm considers both distances in the dependent and independent variable spaces and has been expressed in detail by Galvão et al. (2005). The SPXY method reforms the Kennard–Stone algorithm by encircling both *x* and *y* differences in the computation of inter-sample distance. In this method, the selection of training set samples is performed by choosing two objects that are most distant from each other as starting points. Subsequently, at each step, the algorithm selects the samples that show the largest minimum distance with respect to any sample already selected. The algorithm adds the newly selected samples to the set of training samples until the number of samples predefined by the analyst is achieved. In this way 80% of chemicals (55) were selected as the training set and the rest (14) as the test set. The training set participated in the generation of the model, adjusting its parameters and independent set of samples in test set was used to evaluate the performance of the model.

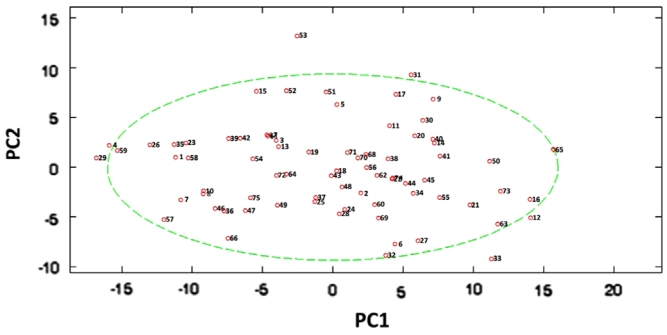


Fig. 1. PCA score plot for 75 compounds.

2.2. Descriptor calculation

In order to develop a QSAR model it is necessary to calculate the molecular structural features of interested chemicals. In this way the structures of compounds in data set were imported from the Pubchem website (Pubchem, 2013) and their validations were checked by their individual CAS numbers from the Lookchem website (Lookchem, 2013). Then, the structures of all compounds were sketched using Hyperchem (ver. 7.0) package and were geometrically optimized by employing the semi-empirical AM1 method by MOPAC (ver. 6.0) package. The molecular geometries corresponding to the lowest energy structure were selected for calculations of the molecular descriptors. The Hyperchem and MOPAC output files were used by Dragon (ver. 3.0) and Codessa (ver. 2.7.2) packages for obtaining the large number of descriptors (Katritzky et al., 1995). The prescreening of these descriptors were performed by eliminating: (i) those that are not available for all compounds; (ii) descriptors having a small variation in magnitude for all structures; and (iii) descriptors whose values are equal to zero for

Download English Version:

https://daneshyari.com/en/article/6369588

Download Persian Version:

https://daneshyari.com/article/6369588

Daneshyari.com