Chemosphere 108 (2014) 17–25

Contents lists available at ScienceDirect

Chemosphere

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

Development of *in silico* models for predicting LSER molecular parameters and for acute toxicity prediction to fathead minnow (*Pimephales promelas*)

Felichesmi Selestini Lyakurwa, Xianhai Yang, Xuehua Li *, Xianliang Qiao, Jingwen Chen

Key Laboratory of Industrial Ecology and Environmental Engineering (MOE), School of Environmental Science and Technology, Dalian University of Technology, Dalian 116024, China

HIGHLIGHTS

- Predictive models for LSER molecular parameters were developed.
- LSER acute toxicity models of Verhaar classes and subgroups were developed.
- Nitrogen atoms and carbonyl group are vital for modeling reactive MOA.
- LSER models have satisfactory goodness-of-fit, robustness and predictive ability.
- Degradation of ketones cause difficulties in modeling toxicity of reactive chemicals.

ARTICLE INFO

Article history:
Received 25 December 2013
Received in revised form 22 February 2014
Accepted 23 February 2014

Handling Editor: A. Gies

Keywords:
Mode of action
LSER parameters
QSAR
Acute toxicity
Applicability domain

ABSTRACT

Many chemicals with toxic effects to aquatic species are produced every year. To date, linear solvation energy relationship (LSER) models for toxicity prediction to aquatic species are limited to non-polar and polar narcotic compounds. In this study, the Verhaar scheme was used to classify chemicals into five modes of toxic actions. LSER models for predicting acute toxicity to fathead minnow were developed by identifying chemical functional groups that influence toxicity prediction of reactive chemicals. Moreover, the predictive models that can be used to estimate LSER molecular parameters have been developed by using quantum chemical and Dragon descriptors. All the predictive models were developed following the OECD guidelines for QSAR model development and validation, with a satisfactory goodness-of-fit, robustness and predictive ability. The McGowans volume was the most significant descriptor in the toxicity models. This study also inferred that, compounds with carbonyl group have different behaviors such that some can biodegrade in the organism while others do not biodegrade, which might be the reason for the difficulties in modeling the acute toxicity of reactive chemicals.

1. Introduction

To date, more than 140 000 synthetic chemicals that might have toxic effects to aquatic species are used in consumer products (EC, 2010; Rudén and Hansson, 2010). The ecological risk assessment (ERA) of chemicals is a necessary measure to control and prevent risks of synthetic chemicals to the aquatic species. The medium lethal dose (LC_{50}) of aquatic species (e.g., fish) is an indispensable biological endpoint for the ERA (Van Leeuwen and Vermeire, 2007). However, the available eco-toxicity data are very sparse. For example, the publicly available toxicity data of the biological endpoints including LC_{50} for 11 300 chemicals in the Canadian domestic substance list are <9% (Weisbrod et al., 2007). The LC_{50} data are usually obtained from experimental measurements following the standardized test protocols which in most cases are costive and time consuming. Hence, it is unrealistic to get the LC_{50} data of all the existing and new chemicals following the standardized animal test protocols. Nevertheless, the international regulations including Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) have approved the application of *in silico* technologies to substitute animal testing for the ERA (Cronin et al., 2009).

Among the *in silico* approaches, quantitative structure–activity relationship (QSAR) is one of the non-animal test method that provide toxicity data in-time and at low-cost (EC, 2010; Papa et al., 2013). QSAR models which are built from complex modeling methods including support vector machine have excellent toxicity

* Corresponding author. Tel./fax: +86 411 8470 7894.
E-mail address: lixuehua@dlut.edu.cn (X. Li).

http://dx.doi.org/10.1016/j.chemosphere.2014.02.076
0045-6535/

prediction (Wang et al., 2010). However, these methods are ambiguous and hard to interpret the mechanism. Likewise, mode of action (MOA) based QSAR models have satisfactory correlations with the biological endpoints (In et al., 2012). Hence, with the knowledge of the chemicals' toxicity mechanism and their inherent risk factors, it is confident to get accurate toxicity predictions.

To date, there are few number of QSAR models that are constructed based on a specific MOA for the molecular structure discrimination analysis (Yao et al., 2005; Wang et al., 2010) and mechanistically structural alerts (e.g., Verhaar scheme) (Nendza and Muller, 2007). The former uses complex machine learning methods and they are less applied for regulatory purposes. The latter employs a transparent and easy to interpret regression models and they are widely applied for regulatory purposes (Nendza and Muller, 2007). The Verhaar scheme was used to classify chemicals into five MOAs including baseline toxicity, less inert, reactive, chemicals that act by specific mechanism, and chemicals that are not possible to classify by Verhaar scheme (Verhaar et al., 1992). QSAR models for the aquatic species are easily built from the chemicals with unreactive MOAs including baseline and less inert toxicity, compared with the QSAR models that are constructed from the reactive MOAs (Hoover et al., 2005). In our previous study, we developed theoretical linear solvation energy relationship (TLSER) models for the chemicals with reactive MOAs (Lyakurwa et al., 2014). The introduction of electron donor – acceptor quantum chemical descriptors into the TLSER models revealed slight statistical improvements of the models for the reactive chemicals, and chemicals that are not possible to classify by Verhaar scheme. Likewise, there is no detailed analysis in the previous studies to unveil the reason for the difficulties in modeling toxicity of reactive chemicals. Therefore, we hypothesized that identifying chemical functional groups that influence toxicity prediction of the reactive chemicals can unveil the reason.

Linear solvation energy relationship (LSER) method has been useful in modeling partition coefficients of solutes and toxicity to aquatic species (Hoover et al., 2005; Abraham and Acree Jr, 2013). However, this method is limited to the LSER parameters that are obtained from the experimental measurements which are costive, laborious and time consuming (Paul et al., 2010). Thus, estimating LSER parameters from the molecular descriptors is important. Interestingly, LSER models which are developed from the predicted parameters have a comparable model quality to the LSER models that are constructed from the experimental parameters (Zissimos et al., 2002; Jover et al., 2004).

LSER models that have been constructed for the toxicity prediction to aquatic species are limited to the non-polar and polar narcotic compounds (Feng et al., 1996; Hoover et al., 2005). Considering non-polar narcosis and polar narcosis are same as the baseline toxicity and less inert MOAs, respectively, there are no previous LSER models developed for the reactive chemicals, and chemicals that are not possible to classify by Verhaar scheme. It was also acknowledged that, many LSER models for the toxicity prediction to aquatic species were not developed following the Organization for Economic Co-operation and Development (OECD) QSAR model development and validation guidelines (OECD, 2007). Therefore, the objectives of this study were: (1) To develop models for predicting LSER molecular parameters, (2) To identify key chemical functional groups that have a significant influence on the predicted toxicity of reactive chemicals, (3) To develop LSER models of Verhaar classes and subgroups of chemicals that are not possible to classify by Verhaar scheme, and reactive chemicals for predicting acute toxicity to fathead minnow (*Pimephales promelas*) following the OECD guidelines.

2. Material and methods

2.1. Development of models for predicting LSER parameters

2.1.1. Collection of LSER parameters

A total of 950 chemicals with experimental LSER parameters were collected from various literatures, which are presented in Table S1 of the supporting information. Among the 950 chemicals, 301 chemicals have both toxicity data for the fathead minnow and experimental LSER parameters.

2.1.2. Calculation of molecular descriptors

The molecular structures of the compounds were optimized at the B3LYP/6-31+g (d,p) level by employing Gaussian 09 program (Frisch et al., 2009). Based on the optimized structures, Dragon software (Ver. 6.0) was applied to generate molecular descriptors for predicting LSER parameters (Talete srl, 2012). Also, Jover et al. (2004) concluded that LSER parameters such as the excess molar refraction (*E*) and polarizability/dipolarizability (*S*) are related to the quantum chemical descriptors including the average valency of an atom and molecular electron repulsion (i.e., average valency of a carbon atom (*Cv*) and total molecular 1-Center E-E repulsion (*CEE1*)). Thus, we calculated similar descriptors with MOPAC-2012 (Stewart, 2012) using keywords: PM3 EF VECTORS PRECISE BOND GNORM = 0.01 PI POLAR ENPART EPS = 78.4 DEBUG.

2.1.3. Construction of the models for predicting LSER parameters

The data set for each LSER parameter was divided into training and validation sets in the ratio of 4:1 (Table S1). The stepwise multiple linear regression analysis (MLR) was applied to generate correlations between experimental parameters (dependent variable) and molecular descriptors (predictor variable). We evaluated the robustness and external predictive ability of the models by using cross and external validation. The meaning of the descriptors that were used to predict LSER parameters are presented in Table S2.

2.2. Development of models for predicting acute toxicity to fathead minnow

2.2.1. Collection of the toxicity data

The experimental data for 96 h fish toxicity towards fathead minnow in terms of LC_{50} values were collected from EPAFHM (http://www.epa.gov/ncct/dsstox/sdf_epafhm.html) and CERCE (www.cerc.usgs.gov/data.html) database. Excluding chemicals with missing CAS numbers and inorganic compounds, a total of 696 chemicals with toxicity ($-\log LC_{50}$) values in mol L^{-1} was grouped into five MOAs (Table S3).

2.2.2. Development of the acute toxicity predictive models

For toxicity modeling, the Verhaar scheme was applied to classify chemicals into five MOAs as described in our previous paper (Lyakurwa et al., 2014). The following steps were performed to identify the key chemical functional groups that have significant influence on the predicted toxicity of the reactive chemicals and chemicals that are not possible to classify by the Verhaar scheme. According to Papa et al. (2005), the number of nitrogen atoms (nN) has significant effect on the predicted acute toxicity to fathead minnow. Hoover et al. (2005) concluded that the removal of the esters and amides lead to improved acute toxicity models of the non-polar and polar narcotic compounds towards six fish species. Both esters and amides contain carbonyl group in their molecular structures. Despite of the fact that the carbonyl group comprises other chemical groups such as carboxylic acid, aldehydes, ketones and enone, yet there is no analysis performed to assess their influence on the predicted toxicity to aquatic species. Moreover, most of the