Chemosphere 96 (2014) 188–194

Contents lists available at ScienceDirect

Chemosphere

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

Development and validation of theoretical linear solvation energy relationship models for toxicity prediction to fathead minnow (*pimephales promelas*)

Felichesmi 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

- Verhaar classification scheme based toxicity predictive models were developed.
- The electron donor–acceptor descriptors were introduced into the TLSER models.
- The E-TLSER models have high goodness of fit, robustness and predictive capacity.
- The cavity term was the most significant descriptor in the models.

ARTICLE INFO

Article history:
Received 22 August 2013
Received in revised form 10 October 2013
Accepted 11 October 2013
Available online 9 November 2013

Keywords:
TLSER models
Fathead minnow
Electron donor–acceptor interaction
Mode of action
Applicability domain

ABSTRACT

The acute toxicity predictive models are vitally important for the toxicological information used in the ecological risk assessments. In this study, we used Verhaar classification scheme to group compounds into five modes of toxic action. The quantum chemical descriptors that characterize the electron donor–acceptor property of the compounds were introduced into the theoretical linear solvation energy relationship (TLSER) models. The predictive models have relatively larger data sets, which imply that they cover a wide applicability domain (AD). All models were developed following the Organization for Economic Co-operation and Development (OECD) QSAR models development and validation guidelines. The adjusted determination coefficient (R^2_{adj}) and external explained variance (Q^2_{EXT}) of the models were ranging from 0.707 to 0.903 and 0.660 to 0.858, respectively, indicating high goodness-of-fit, robustness and predictive capacity. The cavity term (McGowans volume) was the most significant descriptor in the models. Moreover, the electron donor–acceptor (E-TLSER) models are comparable to the TLSER models for the toxicity prediction to fathead minnow. Thus, the E-TLSER models developed can be used to predict acute toxicity of new compounds within the AD.

1. Introduction

A large number of synthetic chemicals with toxic effects to aquatic species are produced and used in our daily lives (UNEP, 2012; Stenzel et al., 2013). The biological endpoints used for the aquatic ecological risk assessment of chemicals are based on the effective concentrations of few species typically algae, crustaceans and fish (van Leeuwen and Vermeire, 2007). The median lethal concentration (LC_{50}) can be used in conjunction with an assessment factor to estimate the predicted no effect concentrations of a specific chemical in preliminary effects assessment (Jager et al., 2006). The LC_{50} data are usually obtained from the experimental tests using standardized test protocols. However, with limitations of cost and time, it is unrealistic to identify all the potentially harmful chemicals following the standardized animal test protocols. Therefore, fast and less expensive alternative methods including *in silico* technology, e.g. quantitative structure activity relationship (QSAR) have been proposed (Cronin et al., 2009).

In recent years, *in silico* technology including QSARs (Deng et al., 2012), read-across (Schüürmann et al., 2011) and chemical category formation (Dom et al., 2012) have been applied to predict LC_{50} for fish. The LC_{50} data for fish are usually diverse in terms of chemical structures and modes of action (MOA), consequently only few QSAR models with a large data set showed good predictability. Moreover, high quality acute toxicity predictive models are usually obtained from complex modeling methods such as counter propagation artificial neural network (Wang et al., 2010). However, these methods are not transparent and it is difficult to interpret the mechanisms. Thus, constructing QSAR models based on the MOA and transparent regression methods is a substitute to obtain excellent toxicity predictions.

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

0045-6535/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.chemosphere.2013.10.039

Among the toxicity classification methods by the MOA, Verhaar Scheme is best recognized and has been used extensively (Verhaar et al., 1992; Netzeva et al., 2007). This method is transparent and readily available for public use in contrast to other classification methods that could only be implemented with commercial softwares. QSAR models for predicting the toxicity of chemicals to the aquatic species are easily generated from compounds with unreactive MOA including inert (baseline) and less inert toxicity compared to QSAR models constructed from compounds whose toxicity mechanism involves electro(nucleo)philic covalent reactivity (Qin et al., 2010; Yong et al., 2012). Introducing molecular descriptors that can characterize the electron donor–acceptor property of a compound may improve the predictability of the models for the reactive compounds and compounds that are not possible to classify by Verhaar scheme.

The solvation models including the theoretical linear solvation energy relationships (TLSER) and linear solvation energy relationships (LSER) are considered as one type of the concise modeling methods (Kühne et al., 2013). The LSER parameters are usually obtained from the experimental measurements. These parameters have better predictability but most of the experimental measurements are limited to high demand of time and cost and are prone to human errors (Stenzel et al., 2013). However, TLSER parameters which are calculated by the quantum chemical methods have a comparable model quality to experimental LSER parameters (Wilson and Famini, 1991; Huddleston et al., 2004).

The TLSER approaches have been used to predict the partition coefficients of solutes and acute toxicity for the aquatic species (Liu et al., 2003; Huang et al., 2004). The hydrogen bonding terms of TLSER models are either described by the covalent basicity and the acidity of the solute (Wilson and Famini, 1991; Liu et al., 2003) or energies of the highest occupied molecular orbital (E_{HOMO}) and the lowest unoccupied molecular orbital (E_{LUMO}) (Chen and Wang, 1997). Nevertheless, most of the TLSER models are developed for predicting acute toxicity of the inert chemicals (baseline toxicity) and less inert chemicals (Ramos et al., 1998; Liu et al., 2003). Furthermore, the TLSER models are established based on a small number of compounds, as a result they have a narrow applicability domain (AD). Therefore, the objectives of this study were: (1) to introduce new quantum chemical descriptors that characterize the electron donor–acceptor property of a compound for the acute toxicity prediction to fathead minnow (*pimephales promelas*) and (2) to develop the Verhaar scheme based predictive models following the Organization for Economic Co-operation and Development (OECD) QSAR model development and validation guidelines (OECD, 2007).

2. Material and methods

2.1. Data set

The experimental data for the 96 h fish toxicity toward 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 for the aquatic species. The quality checks of 720 chemicals in the database were performed so as to eliminate metals and salts. Thereafter, 646 chemicals with acute toxicity – logLC_{50} (M) were classified into five MOA (Table S3).

2.2. Calculation of the molecular structure descriptors

All the molecular structures of the studied compounds were obtained from the EPI Suit™ Version 4.10 (US EPA, 2012). The octanol–water partition coefficient (logK_{OW}) values were estimated using the KOWIN (Version 1.68) module within the EPI Suite. The molecular structures were optimized with Gaussian 09 program package prior to the calculations of the molecular descriptors at the B3LYP/6–31 + g (d, p) level (Frisch et al., 2009). The integral equation formalism polarized continuum model (IEFPCM) was used to account for the solvent effects of water. The quantum chemical descriptors including E_{HOMO}, E_{LUMO}, the most positive net atomic charge of a hydrogen atom (q^+) and most negative net atomic charge of an atom (q^-) were calculated from the Gaussian output files. The covalent basicity descriptor (ε_b) was obtained from the difference between E_{HOMO} of a target compound and E_{LUMO} of water (Wilson and Famini, 1991). The covalent acidity descriptor (ε_a) was calculated from the difference of E_{LUMO} of a target compound and E_{HOMO} of water. The optimized structures of the Gaussian 09 program were used to calculate the McGowan volume using Dragon software version 6.0 (Talete srl, 2012).

Another five quantum chemical descriptors including the ionization potential (I), electron affinity (A), chemical potential (μ), chemical hardness (η) and electrophilicity index (ω), which describes the electron donor–acceptor interactions were selected (Thanikaivelan et al., 2000; Enoch, 2010). The names/meanings of the descriptors and descriptor values for each compound are listed in Tables S1 and S2 in Supporting information.

2.3. Development of the model

In this study, the Verhaar classification scheme was applied to classify compounds into five MOA, namely baseline toxicity (class 1, 120 compounds), less inert (class 2, 79 compounds), reactive (class 3, 128 compounds), chemicals that act by a specific mechanism (class 4, 5 compounds) and chemicals that are not possible to classify by the Verhaar scheme (class 5, 311 compounds) (Verhaar et al., 1992). The Verhaar classification was performed in the toxtree software (http://ecb.jrc.it/qsar/qsar-tools/index.php?c=TOXTREE). There were only five chemicals in class 4, thus no TLSER model was developed. For the acute toxicity modeling, the data set for each class was divided into a training set and validation set in the ratio of 4:1.

According to Famini et al. (1993) the (theoretical) linear solvation energy relationship (T/LSER) approach assumes that solute–solvent interactions are based on several additive properties of the molecules that can be separated as formal descriptors in the solvation process. The conventional LSER model is given by the following solvation equation:

$$\text{Property} = \text{cavity term} + \text{polarity term} + \text{hydrogen bonding} + \text{constant} \quad (1)$$

In the scheme of TLSER model (Wilson and Famini, 1991), McGowan volume (V) and dipolarizability/polarizability (π) describe the cavity and polarity terms, respectively. The hydrogen bonding term is described by ε_a and q^+ (hydrogen bond donor) as well as ε_b and q^- (hydrogen bond acceptor). In this study, quantum chemical descriptors that encode electron donor–acceptor property of the compounds were introduced into the TLSER models. With the electron donor–acceptor descriptors, we have the following model (termed E-TLSER):

$$\text{Property} = \text{cavity term} + \text{polarity term} + \text{hydrogen bonding} + \text{electron donor-acceptor term} + \text{constant} \quad (2)$$

where the electron donor–acceptor terms are presented by the chemical potential (μ, $(E_{LUMO} + E_{HOMO})/2$), chemical hardness (η, $(E_{LUMO} - E_{HOMO})/2$), electrophilicity index (ω, $\mu^2/2\eta$), electron affinity (A, $-E_{LUMO}$) and the ionization potential (I, $-E_{HOMO}$).