Chemosphere 93 (2013) 1094–1103

Contents lists available at SciVerse ScienceDirect

Chemosphere

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

Quantitative structure–activity relationship to predict acute fish toxicity of organic solvents

A. Levet, C. Bordes *, Y. Clément, P. Mignon, H. Chermette, P. Marote, C. Cren-Olivé, P. Lantéri

Université de Lyon, F-69622 Villeurbanne, France
Université Claude Bernard Lyon 1, Institut des Sciences Analytiques, UMR CNRS 5280, F-69622 Villeurbanne, France

HIGHLIGHTS

- A reliable QSAR is proposed to predict fish LC50 of organic solvents.
- Solvents are described both by physicochemical and quantum theoretical descriptors.
- Log*P*, LUMO, dielectric constant and surface tension were selected in the QSAR.

ARTICLE INFO

Article history:
Received 7 February 2013
Received in revised form 30 May 2013
Accepted 2 June 2013
Available online 15 July 2013

Keywords:
Ecotoxicity
QSAR
Organic solvents
Fish LC50
DFT
ECOSAR

ABSTRACT

REACH regulation requires ecotoxicological data to characterize industrial chemicals. To limit *in vivo* testing, Quantitative Structure–Activity Relationships (QSARs) are advocated to predict toxicity of a molecule. In this context, the topic of this work was to develop a reliable QSAR explaining the experimental acute toxicity of organic solvents for fish trophic level. Toxicity was expressed as log(LC50), the concentration in mmol.L^{-1} producing the 50% death of fish. The 141 chemically heterogeneous solvents of the dataset were described by physico-chemical descriptors and quantum theoretical parameters calculated via Density Functional Theory. The best subsets of solvent descriptors for LC50 prediction were chosen both through the Kubinyi function associated with Enhanced Replacement Method and a stepwise forward multiple linear regressions. The 4-parameters selected in the model were the octanol–water partition coefficient, LUMO energy, dielectric constant and surface tension. The predictive power and robustness of the QSAR developed were assessed by internal and external validations. Several techniques for training sets selection were evaluated: a random selection, a LC50-based selection, a balanced selection in terms of toxic and non-toxic solvents, a solvent profile-based selection with a space filling technique and a D-optimality onions-based selection. A comparison with fish LC50 predicted by ECOSAR model validated for neutral organics confirmed the interest of the QSAR developed for the prediction of organic solvent aquatic toxicity regardless of the mechanism of toxic action involved.

1. Introduction

The European regulation REACH (Registration, Evaluation, Authorization and restriction of Chemicals) aims to control the industrial use of chemicals since 2007 (EC, 2006). All the substances used have to be registered to the European CHemicals Agency (ECHA, http://echa.europa.eu/). Chemicals registration especially requires toxicological and ecotoxicological informations such as acute aquatic toxicity on three trophic levels, i.e., algae, invertebrates and fish.

REACH regulation focusing on *in vivo* testing limitation, alternative methods are encouraged for chemicals risk assessment. In this context, Quantitative Structure–Activity Relationship (QSAR) methods are advocated to predict the biological activity of a molecule through a validated mathematical relationship including its physico-chemical properties (EC, 2006) (ECHA, http://echa.europa.eu/). QSARs are also recommended by the Globally Harmonized System of classification and labeling of chemicals (EC, 2008) and by the Technical Guidance document on risk assessment (EC, 2003).

QSAR development requires several steps: data collection, molecule descriptor selection, model implementation and validation. Once validated, the obtained relationships can be used to predict the activity of other molecules on the basis that similar compounds have similar activities or properties (Katritzky et al., 1995; Tropsha, 2010). Several QSARs have been developed to predict toxicological

* Corresponding author at: Institut des Sciences Analytiques, UMR 5280, 5 rue de la Doua, 69100 Villeurbanne, France. Tel.: +33 4 37 42 35 36; fax: +33 4 37 42 35 62.
E-mail address: bordes@univ-lyon1.fr (C. Bordes).

0045-6535/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.chemosphere.2013.06.002

(Tsakovska et al., 2008; Gu et al., 2010) and ecotoxicological (Huang et al., 2003; Luis et al., 2010) activities.

In the ecotoxicology field, predicted parameters generally concern NOEC (No observed effect concentration) (Chen et al., 2009), biodegradation (Toropov et al., 2012), EC50 (median effect concentration) of different trophic levels such as fish (Konemann, 1981; Mazzatorta et al., 2003a), invertebrates (Katritzky et al., 2009; Toropova et al., 2012a) or algae (Hsieh et al., 2006; Lu et al., 2007).

QSAR models for fish LC50 include one (Voutchkova et al., 2011), two (Lessigiarska et al., 2004) or more descriptors (Wang et al., 2010). They may take into account the mechanism of toxic action (non-polar or polar narcosis, electrophilic process) (Tan et al., 2010) or not (Netzeva et al., 2005). They are generally dedicated to specific chemical families: ester (Papa et al., 2005a), hydrocarbon (Di Marzio and Saenz, 2004), phenol (Rose and Hall, 2003), pharmaceuticals (Tugcu et al., 2012), etc.

Only few general models have been developed for a large set of chemicals using physico-chemical and theoretical parameters for modeling fish LC50 (Mazzatorta et al., 2003a,b; Lessigiarska et al., 2004; Wang et al., 2010; Voutchkova et al., 2011). However, to our knowledge, no model is specific to solvents.

Octanol–water partition coefficient (Log$P$) is the most commonly used descriptor (Mazzatorta et al., 2004; Voutchkova et al., 2011), with many one-parameter QSAR models based on Log$P$ (Parkerton and Konkel, 2000; Tremolada et al., 2004) as in ECOSAR program (http://www.epa.gov/oppt/newchems/tools/21ecosar.htm), which is extensively used and recognized to be highly validated for ecotoxicological profile assessment (Reuschenbach et al., 2008). In addition to Log$P$ parameter, ecotoxicological models may include various physico-chemical properties such as electrophilic (LUMO (Zhang et al., 2010) or HOMO energy (Voutchkova et al., 2011)), atomic charges (Verhaar et al., 1996), dipole moment (Verhaar et al., 1996), topological (Wang et al., 2010), constitutional (Gong et al., 2008), and/or physico-chemical (Di Marzio and Saenz, 2004) descriptors.

The purpose of this study is first to develop QSAR in order to predict the fish acute toxicity of a diverse and large set of organic solvents, regardless of the mechanism of toxic action involved. Organic solvents have been studied since they are essential products in many sectors of industry and everyday life such as detergents, agrochemicals, cosmetics, pharmaceuticals, paints, varnishes, and inks. Reliable QSAR development requires a training set representative of the initial dataset. Therefore this study aims to evaluate different strategies of training set design, namely random selection, LC50-based selection, space-filling or D-optimality approaches. Then, the prediction ability of the developed QSAR was compared to that of the ECOSAR program.

2. Materials and methods

2.1. Dataset

The dataset is described in Table 1 and contains 141 solvents being chemically heterogeneous and belonging to different chemical families such as halogenated hydrocarbons, alcohols, ethers, amines, acids, ketones, sulfur compounds, amides, aromatic hydrocarbons, nitrogen compounds, esters, hydrocarbons, carbonate esters, aldehydes, anhydride acids and glycols.

The experimental acute toxicity was expressed as log(LC50) for fish trophic level which is representative for ecotoxicological evaluation of industrial chemicals. Different species were considered: *Pimephales promelas, Brachydanio rerio* and *Cyprinus carpio*.

LC50 denotes the concentration in mmol.L^{-1} producing the 50% of fish death referred to a test period of 96 h (EC, 2003). The ecotoxicological literature data were collected in INERIS (http://www.ineris.fr/), ESIS (http://ecb.jrc.ec.europa.eu/esis/) and ECHA (http://echa.europa.eu/) databases as recommended by the INERIS Institute for good reliability of the data.

2.2. Solvent descriptors

2.2.1. Physico-chemical descriptors

Various experimental physico-chemical properties were selected as descriptors on the basis of literature dealing with solvent properties (Katritzky et al., 2004; Hansen, 2007), classification (Chastrette et al., 1985; Durand et al., 2011) and QSARs (Wang et al., 2010; Voutchkova et al., 2011). Classical solvent descriptors such as octanol/water partition coefficient (Log$P$), molecular weight (M_w), boiling point (b_p), density (d), molar volume (v_m), dipole moment (μ), dielectric constant (ε), refractive index (n_r), surface tension (γ) and vapor tension (P_{vap}) were found in several databases (Weast, 1975; Smallwood, 1996; Cheremisinoff, 2003) (https://reaxys.com).

Some Log$P$ values were also verified with counter current chromatography measurements. This chromatographic technique is characterized by the use of a mobile and a stationary phase which are both liquid (Ruiz-Angel et al., 2011).

Hansen and Hildebrand solubility parameters were also included in the database since these properties are widely used in solvent substitution procedures (Bordes et al., 2010). Hildebrand parameter (δ) reflects the amplitude of cohesion intermolecular forces in liquids (Hildebrand and Scott, 1950). Hansen suggested the splitting of Hildebrand solubility parameter into three parts derived from different types of cohesive forces (a disperse part, a polar part and a hydrogen-bonding part) according to (Hansen, 2007): $\delta = (\delta_d^2 + \delta_p^2 + \delta_h^2)^{1/2}$ where δ_d corresponds to the so-called London interaction resulting from the existence of induced dipoles as two molecules approach one another (disperse part), δ_p corresponds to Keesom forces occurring when two permanent dipoles are present (polar part) and δ_h represents hydrogen bonding forces.

2.2.2. Theoretical descriptors

QSARs require also structural and topological molecule descriptors. The 3D chemical structures of all solvents were initially built and optimized in gas phase using ADF software (http://www.scm.com). This software is based on density functional theory (DFT) which is the most efficient way to calculate accurate and reliable electronic properties of systems with respect to the computational cost. Descriptor calculations were performed with the PBE (Perdew, Burke and Ernzerhof) (Perdew et al., 1996) GGA exchange-correlation functional (Generalized Gradient Approximation) method and a TZP (triple zeta) basis set. For given molecules several conformers were computed and only the one exhibiting the closest molecular dipole with the experimental one was considered.

From the results of DFT calculations, 15 descriptors were selected: the energy of the highest occupied molecular orbital (HOMO), the energy of the lowest unoccupied molecular orbital (LUMO), the polarisability (α), the maximal (q_{max}) and minimal (q_{min}) atomic Mulliken charges, the maximal (V_{max}) and minimal (V_{min}) electrostatic potential values computed on an isodensity surface around the molecule, the surface area (Surf), the so-called molecular volume (Vol) and the molecule's ovality (Ov). DFT based descriptors are often used to describe the reactivity of the molecule (Chermette, 1999; Geerlings et al., 2003) in particular the hardness computed as the difference between HOMO and LUMO energies is related to the resistance of the density against a change in electron number, i.e. the larger the hardness the more stable the electronic distribution. Another reactivity parameter is calculated, namely