Chemosphere 190 (2018) 463–470

Contents lists available at ScienceDirect

Chemosphere

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

Developing predictive models for toxicity of organic chemicals to green algae based on mode of action

Serge Bakire, Xinya Yang, Guangcai Ma, Xiaoxuan Wei, Haiying Yu*, Jianrong Chen, Hongjun Lin

College of Geography and Environmental Sciences, Zhejiang Normal University, Yingbin Avenue 688, 321004, Jinhua, PR China

HIGHLIGHTS

- Verhaar Scheme based toxicity predictive models for algae were developed.
- Molecular descriptors that characterize electronic and structural properties were used.
- The models were validated to be mechanistically explainable and statistically robust.
- The model can predict log EC_{50} values for algal growth inhibition (72 h).

ARTICLE INFO

Article history:
Received 24 July 2017
Received in revised form
3 October 2017
Accepted 4 October 2017
Available online 5 October 2017

Handling Editor: I. Cousins

Keywords:
Algae
Growth inhibition
Mode of action
In silico study
QSAR

ABSTRACT

Organic chemicals in the aquatic ecosystem may inhibit algae growth and subsequently lead to the decline of primary productivity. Growth inhibition tests are required for ecotoxicological assessments for regulatory purposes. *In silico* study is playing an important role in replacing or reducing animal tests and decreasing experimental expense due to its efficiency. In this work, a series of theoretical models was developed for predicting algal growth inhibition (log EC_{50}) after 72 h exposure to diverse chemicals. In total 348 organic compounds were classified into five modes of toxic action using the Verhaar Scheme. Each model was established by using molecular descriptors that characterize electronic and structural properties. The external validation and leave-one-out cross validation proved the statistical robustness of the derived models. Thus they can be used to predict log EC_{50} values of chemicals that lack authorized algal growth inhibition values (72 h). This work systematically studied algal growth inhibition according to toxic modes and the developed model suite covers all five toxic modes. The outcome of this research will promote toxic mechanism analysis and be made applicable to structural diversity.

1. Introduction

Acute toxicity is basic ecotoxicological information for the safety assessment of organic chemicals. According to REACH (Registration, Evaluation, Authorization, and Restriction of Chemicals), growth inhibition tests on algae are required for chemicals with annual production volumes above 1 t/a for regulatory purposes. Algae are dominant primary producers in the aquatic food web. The hazards of chemicals to algae can cause the depletion of dissolved oxygen and of primary productivity of the aquatic ecosystem (Geis et al., 2000; Jin et al., 2014; Singh et al., 2014). Numerous synthetic chemicals are produced in the modern industry and are widely used in daily life (Levet et al., 2016; UNEP, 2012). However, it is impractical to measure the toxicity of all existing and new chemicals using the standardized animal test protocol due to its limitations of being expensive, laborious, and time consuming.

In recent years, *in vitro* and *in silico* methods have been advocated as alternatives to reduce or replace animal experiments (Leusch et al., 2010; Tunkel et al., 2005). Among them, quantitative structure-activity relationship (QSAR) methods show strong potential to reduce the use of animals in experiments and fill in the toxicity data gaps under REACH legislation (Papa et al., 2005; Thomas et al., 2015; Zvinavashe et al., 2008). Several QSAR models have been developed to predict the acute toxicity of selected categories of chemicals to algae, such as alcohols (Chen et al., 2012), phenols (Aruoja et al., 2011; Chen and Lin, 2006; Lee

* Corresponding author.
E-mail address: yhy@zjnu.cn (H. Yu).

https://doi.org/10.1016/j.chemosphere.2017.10.028
0045-6535/

et al., 2006), benzoic acids (Lee and Chen, 2009), benzenes (Zeng et al., 2011), anilines (Aruoja et al., 2011), and nitriles (Huang et al., 2007). Only a few general models cover a large structural diversity in chemistry (Lessigiarska et al., 2004; Levet et al., 2016). Thus, almost all existing QSAR models for acute toxicity to algae are based on chemical species. However, modes of action (MOAs) are another crucial issue in predictive toxicity, as proved by previous research (Cronin and Schultz, 1996; Lyakurwa et al., 2014; Martin et al., 2013; Schultz et al., 1998). Considering MOAs is fundamental to developing mechanistically-based and statistically-robust QSAR models for predicting toxicity.

A number of methods that assign chemicals to a specific mechanism have been developed. They can be generally classified into molecular descriptor-based methods (Altenburger et al., 2003; Martin et al., 2013; Schüürmann et al., 2003; Spycher et al., 2005) and structural alerts-based methods (Nendza and Muller, 2007; von der Ohe et al., 2005). The former methods utilize a variety of molecular descriptors and statistical analyses, and the latter identify MOAs based on the presence or absence of certain atoms and specific functional groups. The Verhaar Scheme (Verhaar et al., 1992), which is widely used, employs structural rules to classify compounds. It has been coded into a software called Toxtree that is freely available from the European Chemicals Bureau. The Verhaar Scheme categorizes chemicals into five classes: inert chemicals or non-polar narcosis, less inert chemicals or polar narcotics, reactive chemicals, specifically acting chemicals, and chemicals that cannot be classified using the Verhaar Scheme. It has been successfully applied to the MOA decision for toxicity prediction to aquatic species at different trophic levels (Enoch et al., 2008; Lyakurwa et al., 2014).

The aim, therefore, of this study was to develop Verhaar Scheme-based theoretical models for predicting the acute toxicity of organic chemicals to algae (*Pseudokirchneriella subcapitata*) according to the Organization for Economic Co-operation and Development (OECD) guidelines (OECD, 2007). A further objective of the work was to analyze structure or property factors that govern the acute toxicity of organic chemicals to algae.

2. Material and methods

2.1. Experimental toxicity data

The chemical concentration that induces 50% growth inhibition of freshwater algae (*Pseudokirchneriella subcapitata*) after 72 h exposure in terms of log EC_{50}^{-1} was chosen as the toxicity endpoint in this work. The experimental log EC_{50}^{-1} values were collected from the previous research (Fu et al., 2015). The dataset was refined by: (1) removing ions, salts and mixtures; and (2) removing the experimental data from the CHRIP database (Chemical Risk Information Platform, https://www.env.go.jp/chemi/sesaku/02e.pdf). Since most of these experimental measurements were not confirmed by experts, the potential inaccuracies may be propagated during model development. After that, altogether 348 experimental log EC_{50}^{-1} values were obtained.

They are divided into five MOAs by Toxtree software (http://ecb.jrc.it/qsar/qsar-tools/index.php?c=TOXTREE): (1) Class 1: baseline toxicity chemicals (chemical number $n = 95$); (2) Class 2: less inert chemicals ($n = 86$); (3) Class 3: reactive chemicals ($n = 28$); (4) Class 4: specifically acting chemicals ($n = 12$); (5) Class 5: chemicals that are not possible to classify using the Verhaar Scheme ($n = 127$). Furthermore, the previous research (Lyakurwa et al., 2014) pointed out that the presence of nitrogen atoms and/or carbonyl groups was a key factor that influenced the model performance of Class 3 and Class 5. So they were further divided into the following: group 1: chemicals with nitrogen atoms (the number of nitrogen $n\mathrm{N} > 0$); and group 2: chemicals without nitrogen atoms and carbonyl groups ($n\mathrm{N} = 0$, the number of carbonyl group $n(\mathrm{C{=}O}) = 0$). The proposed method could not yield acceptable results for the chemicals with carbonyl groups but no nitrogen atoms ($n\mathrm{N} = 0$, $n(\mathrm{C{=}O}) > 0$), so these were not included in this study.

2.2. Molecular optimization and parameter calculation

Hydrophobicity and chemical reactivity have been identified as important issues for acute toxicities of xenobiotic chemicals (Bohme et al., 2016). The hydrophobicity of chemicals is generally quantified by the logarithmic octanol-water partition coefficient (log K_{ow}), which can be calculated by EPI Suite 4.11 (US EPA, 2015). In this work, the chemical reactivity is measured by quantum chemical descriptors that extract electronic, energy, and charge information of molecules including: the average molecular polarity (α, atomic unit), the energy of the highest occupied molecular orbital (E_{HOMO}, eV), the energy of the lowest unoccupied molecular orbital (E_{LUMO}, eV), softness (μ, eV), hardness (η, eV), electrophilicity index (ω, eV), the most positive net atomic charge on an atom (q^+, atomic charge unit), the most negative net charge on an atom (q^-, atomic charge unit), the most positive net atomic charge on a carbon atom (q_C^+, atomic charge unit), the most negative net atomic charge on a carbon atom (q_C^-, atomic charge unit), the most positive and negative values of the molecular surface potential ($V_{s,\max}$ and $V_{s,\min}$, eV), the average of the positive and negative potentials on the molecular surface ($\overline{V}_S^+$ and $\overline{V}_S^-$, eV), the average deviation of surface potential (Π, eV), and the balance parameter of surface potential (τ). The molecular structures of all compounds were optimized in their neutral electronic ground state by B3LYP/6-31G (d, p) algorithm, which was carried out by Gaussian 09 program package (Frisch et al., 2009). Then frequency analysis was conducted on the same basis set to validate the energetic minima. α, E_{HOMO}, E_{LUMO}, μ, η, ω, q^+, q^-, q_C^+, and q_C^- were directly obtained from the output files of the Gaussian computation. In this calculation, Mulliken charge was chosen as the atomic charge model. $V_{s,\max}$, $V_{s,\min}$, $\overline{V}_S^+$, $\overline{V}_S^-$, Π, and τ were computed through GsGrid software package (Version 1.7) (Lu, 2010).

In addition, dragon descriptors that reflect 0–3 dimensional geometric and topological information as well as molecular properties were also employed to further characterize molecular properties. Their potential to predict acute toxicities has been proved (Papa et al., 2005; Tebby et al., 2011). They were calculated by Dragon software 6.0 (2012) based on the optimized molecular structures from the Gaussian calculation. After removing constants (relative standard deviation < 0.001) and high correlated descriptors, 1203 dragon descriptors were maintained.

2.3. Model development and characterization

Every class (or group) was randomly divided into two subsets. One set contained 70% of the chemicals, and it was used as a training set to develop models. The other set, which contained the remaining 30% of the chemicals, was considered to be the external test set. Both subsets should cover the toxicity and structure ranges of the whole class. Multilinear regression (MLR) with a step-wise algorithm was applied for variable selection and model development by SPSS 21.0. The model performance was quantified by the squared correlation coefficient (R^2), the root mean square error (*RMSE*), the systematic error (*BIAS*), the maximum positive error (*MPE*), and the maximum negative error (*MNE*).

2.4. Model validation and outlier diagnosis

The predictive capability of model was tested by external