Ecotoxicology and Environmental Safety 166 (2018) 92–101

Contents lists available at ScienceDirect

Ecotoxicology and Environmental Safety

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

Chemometric modeling of aquatic toxicity of contaminants of emerging concern (CECs) in *Dugesia japonica* and its interspecies correlation with daphnia and fish: QSTR and QSTTR approaches

Kazi Amirul Hossain[a], Kunal Roy[b,*,1]

[a] Department of Pharmacoinformatics, National Institute of Pharmaceutical Educational and Research (NIPER), Kolkata 700032, India
[b] Drug Theoretics and Cheminformatics Laboratory, Department of Pharmaceutical Technology, Jadavpur University, Kolkata 700032, India

ARTICLE INFO

Keywords:
QSTR
QSTTR
PLS
Planarian
CEC
Consensus

ABSTRACT

The contaminants of emerging concern (CEC) are universally detected in surface water and soil. They can affect the wild life, and their subsequent translocation through the food chain can affect human health, which is an issue of serious concern. Very few amounts of ecotoxicological data are available on the environmental behavior and ecotoxicity of CEC, thus modeling approaches are essential to bridge the existing gap in experimental data. In this present study, we have developed quantitative structure-toxicity relationship (QSTR) models using a data set of 75 compounds for the prediction of aquatic ecotoxicity of CECs on fresh water planarian (*Dugesia japonica*) by partial least squares (PLS) regression algorithm using simple molecular descriptors selected by genetic algorithm approach. We also explore the correlations between toxicity against *D. japonica* and those against daphnia (*D. magna*) and fish (*P. promelas*), and these were improved on addition of a few molecular descriptors (B08[C-O] and B09[N-O] in case of daphnia and C-006 and H-052 in case of fish) which allowed us to develop predictive interspecies quantitative structure toxicity-toxicity relationship (QSTTR) models, allowing to extrapolate data from one endpoint to another endpoint. The QSTR (Q^2_{LOO} ranging from 0.630 to 0.720 and R^2_{pred} ranging from 0.723 to 0.798) and QSTTR (Q^2_{LOO} = 0.60 and 0.67, R^2_{pred} = 0.88 and 0.84) models have desirable statistical qualities and acceptable internal and external validation measures, meeting rigorous criteria of different validation metrics and showing acceptability for regulatory purposes as proposed by Organization for Economic Cooperation and Development (OECD). Consensus predictions were also performed based on multiple models generated in this study by using the "Intelligent Consensus Predictor" (ICP) tool to enhance the prediction quality for external set compounds.

1. Introduction

Chemical Industries process raw materials and convert them into desired products, which we use in our day-to-day life. Chemicals play a major role in all facets of our life as these are used in pharmaceutical products, agriculture, food, cosmetics, etc. However, a useful chemical can also show hazardous effects on the ecosystem at a certain concentration, and this is a matter of great concern. Contaminants of emerging concern (CECs) refer to any chemical found in water or in the environment that is only present at relatively low to moderate levels. CECs include pharmaceuticals and personal care products (PPCPs), UV filters, hormones and endocrine disrupting chemicals (EDCs), pesticides and surfactants etc. Many CECs have been shown to be endocrine disruptors. The emerging contaminants may also demonstrate low acute toxicity but cause significant reproductive effects at very low levels of exposure (Maruya et al., 2013). The most challenging thing about them is their low detectable level. These chemicals have features that require additional consideration when applying existing ambient water quality criteria for the protection of aquatic life, using EPA's 1985 Guidelines for Deriving Numerical National Water Quality Criteria for the Protection of Aquatic Life and Their Uses (https://www.epa.gov/wqc/contaminants-emerging-concern-including-pharmaceuticals-and-personal-care-products). CECs are continuously entering water sources throughout the world because of their widespread use. Conventional wastewater and recycled water treatment are only partially effective in their removal or for their degradation, so they are discharged into the environment with treated wastewater effluent, recycled water, and wastewater plant sludge (Raghav et al., 2013). Effects of CECs on

* Corresponding author.
E-mail address: kunal.roy@jadavpuruniversity.in (K. Roy).
[1] <http://sites.google.com/site/kunalroyindia/>.

https://doi.org/10.1016/j.ecoenv.2018.09.068
Received 13 August 2018; Received in revised form 11 September 2018; Accepted 15 September 2018

human and ecosystem health are largely unknown, and relatively little is known about the ways they travel through the environment or how they may be transformed or degraded in the course of their travels. Some studies have shown that even very low exposure to certain CECs can have impacts on biological systems (Raghav et al., 2013).

More than 200 pharmaceuticals alone have been reported in river water globally, with concentrations up to a maximum of 6.5 mg/l for the antibiotic ciprofloxacin (Hughes et al., 2013). Imperfect removal results in pharmaceuticals being reported in receiving surface waters in the ng to μg/L range. The analgesic tramadol has been observed in river water at the highest concentration up to a maximum of 7731 ng/L (Kasprzyk-Hordern et al., 2008a). The sunscreen agent 4-benzophenone has been observed at mean final effluent concentrations ranging from 3597 to 5790 ng/L (Kasprzyk-Hordern et al., 2008b, 2009). In a whole lake experiment, Kidd et al. (2007) found that exposure of fathead minnows (*Pimephales promelas*) to 5–6 ng/L of 17α-ethinylestradiol resulted in feminized male fish with arrested gonadal development, which within two years led to recruitment failure in the population. This observation indicates the potential toxic effects of CECs at very low concentration.

Approximately 1,004,837 animals are used in toxicological and other safety assessment which is 8.75% of a total number of animals used for the scientific purposes (Directive, 1980). While experimental tests are expensive, time-consuming and actively fought against by Animal Rightists, it is always important to find an alternative way to study these chemicals for environmental monitoring and risk assessment. Quantitative Structure-Toxicity Relationship (QSTR) models allow us to estimate the toxicity end points with fair degree of precision and reliability. QSTR techniques have been proved to be a valuable approach in predictive toxicology research of chemical compounds in respect to their prospective harmful effects on the living system (Farahani et al., 2018). The limitation of ecotoxicological testing lies in generating toxicity prediction values for unknown, untested compounds, which can only be tackled through QSTR and interspecies quantitative structure-toxicity-toxicity relationships (QSTTR) modeling. In 2006, the European Commission finalized the Registration, Evaluation, Authorization and Restriction of Chemicals (REACH) legislation, which came into force from 1st June 2007. This legislation requires toxicological hazard and risk assessment of all new and existing chemicals, which is very expensive due to high degree of experimental and administrative work. Thus, the need of alternative in silico methods like QSAR (Quantitative Structure-Activity Relationship) is explicitly encouraged and even required in the REACH regulation to reduce or replace animal testing (Gozalbes and de Julian-Ortiz, 2018). The Organization for Economic Cooperation Development (OECD) (Assay, 2004) has provided several guidelines for the development of robust (QSAR)/QSTR models.

With this background, we have developed QSTR models of **75** CECs against *Dugesia japonica,* which is a species of fresh water triclad widely distributed in East Asia and considered the most common freshwater planarian in Japan (E.O.L). This planarian has a remarkable regeneration ability and it is very sensitive to CECs (Orii et al., 2002). The major advantage of using *D. japonica* model is that it is a surrogate of animal model and it can be maintained in laboratory easily and inexpensively. Development of QSTR model requires a chemical dataset with experimental quantitative toxicity values; in this study the toxicity data of 75 chemicals against *D. japonica* was obtained from reliable sources (Li, 2008, 2012a, 2012b, 2013a, 2013b; Hagstrom et al., 2015). Toxicity assessment of these chemicals and other mixture of chemicals seem to be ideal for the regulatory purposes. As the magnitude of the numbers of both untested and newly introduced chemicals is very large, toxicity testing alone cannot be enough for the production of new data and the reduction in gaps in existing toxicity data. Hence, gathering data using in silico techniques on ecotoxicity potential of chemicals against representative organisms is essential to fulfill the environmental toxicity data gap as much as possible. However, the QSTR models should generally be of a diverse chemical domain for a toxicity end point against a single species. The models should be developed by utilizing reliable software packages for calculation of descriptors and statistical analysis. Onlu and Sacan recently developed a QSTR model for the toxicity of **55** CECs to *D.japonica* with a mixture of 2D and 3D descriptors and using experimentally derived octanol-water partition coefficient ($\log K_{o/w}$) as the most important predictor variable for their model (Onlu and Sacan, 2018). However, it is difficult to use their model for precise predictions for compounds without having any experimentally derived $\log K_{o/w}$ values. Moreover, 3D descriptors require conformational analysis and geometry optimization, which may lead to ambiguities in the values of 3D descriptors. In the present work, we have developed QSTR models with an extended list of CEC compounds using only 2D descriptors (excluding $\log K_{o/w}$ and all 3D descriptors).

The lacking resources for providing ecotoxicological data against *D. japonica* for CECs paved way for the development of interspecies models on general species, thus filling data gap. To address data gaps in species sensitivity, the Interspecies Correlation Estimations (ICE) application was developed by the U.S. Environmental Protection Agency (US EPA) (https://www3.epa.gov/webice/). Besides QSTR, Quantitative Structure Toxicity-Toxicity Relationship (QSTTR) is becoming an important tool for the toxicity prediction of chemical compounds through interspecies relationship, which extrapolates data for one toxicity endpoint to those for another toxicity endpoint. This approach has the potential to fill the gap where toxicity data is scarce. In interspecies QSTTR model, toxicity endpoint for a particular species acts as a predictor variable along with other descriptors. The toxicity endpoint, which acts as a predictor variable, can emphasize the biological effect of a particular compound to some extent as it is derived by standard experimental bioassay, while other descriptors as used in standard QSAR/QSTR models are obtained purely from chemical structure. QSTTR modeling can promote reduction in uses of higher level organisms for toxicity testing and it gives an understanding of mechanism of action of toxic chemicals (Das et al., 2015; Kar et al., 2016). Although some reports can found on interspecies correlation of various compounds to various organisms, there is only one report found on the interspecies correlation of CECs between *D. japonica* and *D. magna,* which describes a Quantitative Toxicity-Toxicity Relationship (QTTR) model using **26** CECs (Onlu and Sacan, 2018). In the present study, we developed interspecies QSTTR correlation models for **47** CECs between *D. japonica* and *D. magna* (daphnia) and for **19** compounds between *D.japonica* and *P. promelas* (fish). The aim of the present study has been to predict the toxicity of CEC compounds to *D. japonica* using available experimental toxicity values of CEC compounds to *D. magna* and *P. promelas* and/or molecular structure information.

2. Materials and methods

2.1. The dataset

The toxicity data of 75 CEC compounds to *D. japonica*, 47 compounds to *D. magna* and 19 compounds against *P. promelas* (the common compounds in latter two cases with the list of 75 CECs) have been taken from the literature (Li, 2008, 2012a, 2012b, 2013a, 2013b; Sanderson and Thomsen, 2009). The considered CECs include ionic and non-ionic surfactants, UV filters, hormones and endocrine disrupting agents (EDCs), preservatives, pharmaceuticals, and organophosphates. The modeled toxicity data was carefully cross-checked with the source articles to remove any discrepancy that could arise due to false data implementation in modeling. All the structures were drawn manually in Marvin Sketch (Csizmadia, 1999) and cross-checked from chemical databases like Chemical Book (http://www.chemicalbook.com). The toxicity data for all compounds for a particular endpoint were reported to have been measured using the same experimental protocol. The reported concentration causing 50% mortality (LC_{50}) at **48** h were uniformly converted from different units into micro molar (μM) unit and