Available online at www.sciencedirect.com

SciVerse ScienceDirect

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

Exploring QSTR analysis of the toxicity of phenols and thiophenols using machine learning methods

*M. Asadollahi-Baboli**

Department of Science, Babol University of Technology, P.O. Box 47148-71167, Babol, Mazandaran, Iran

ARTICLE INFO

Article history:
Received 27 June 2012
Received in revised form
1 September 2012
Accepted 8 September 2012
Available online 15 September 2012

Keywords:
Quantitative structure–toxicity relationships
Phenols
Thiophenols
Classification and regression trees
Support vector regression
Photobacterium phosphoreum

ABSTRACT

There is an increasing need for the rapid safety assessment of chemicals by both industries and regulatory agencies throughout the world. In silico techniques are practical alternatives in the environmental hazard assessment. In this background, quantitative structure–toxicity relationship (QSTR) analysis has been performed on toxicity of phenols and thiophenols to *Photobacterium phosphoreum*. The techniques of classification and regression trees (CART) and least squares support vector regressions (LS-SVR) were applied successfully as variable selection and mapping tools, respectively. Four descriptors selected by the CART technique have been used as inputs of the LS-SVR for prediction of toxicities. The best model explains 91.8% leave-one-out predicted variance and 93.0% external predicted variance. The predictive performance of the CART-LS-SVR model was significantly better than the previous reported models based on CoMFA/CoMSIA and stepwise MLR techniques, suggesting that the present methodology may be useful to predict of toxicity, safety and risk assessment of chemicals.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

The environment is regularly exposed to phenols and thiophenols derivatives including nitro, chloro, fluoro and etc substitutes through their use in industrial processes. Information on aquatic toxicity is required in order to assess the hazard and risk of chemical substances to soil, marine and freshwater organisms living in the water column (Netzeva et al., 2008). Phenols and thiophenols are nearly ubiquitous pollutants in all aquatic and terrestrial ecosystems. A number of different mechanisms of toxic action have been recognized for these compounds including non-reactive polar narcosis, through to respiratory inhibition or uncoupling of oxidative phosphorylation, disruption of biological macromolecules due to the formation of free radicals or electrophilic alkylation, as well as endocrine disruption (Netzeva et al., 2008). The recent European Union REACH (Registration, Evaluation and Authorisation of Chemicals) legislation requires toxicological hazard and risk assessments for all new and existing chemicals. In addition, the use of in silico methods is explicitly encouraged and even required in the REACH regulation (Worth et al., 2007). Therefore, there is an essential need to use computation-based quantitative structure–toxicity relationship (QSTR) modeling, as an in silico technique, for providing information about the physicochemical properties of chemicals and their environmental fate as well as their human health effects (Lessigiarska et al., 2006). The use of QSTR modeling for toxicological predictions would help determine the potential adverse effects of chemical entities in risk assessment, chemical screening, and priority setting (Lessigiarska et al., 2006).

* Tel.: +98 1113232071x1721; fax: +98 1113234201.
E-mail address: asadollahi@nit.ac.ir
1382-6689/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.etap.2012.09.003

Photobacterium phosphoreum is a kind of luminescent bacterium in seawater and its character of luminous intensity change with toxic substance inhibition of growth (i.e., cell density) made it as an index for chemical toxicity measuring and water quality monitoring. For the development of QSTR models, *P. phosphoreum* has been largely used recently (Yu et al., 2009). An obvious advantage of this sort of studies is to minimize testing toxicity data in the laboratory and environmental damages. The ultimate role of formulating the QSTR is to suggest mathematical models estimating the toxicities by relying on the assumption that these are determined solely by the molecular structures of the chemicals. The structure is therefore translated into the so-called molecular descriptors, describing some relevant feature of the compounds, with mathematical formulae obtained from Chemical Graph Theory, Information Theory, Quantum Mechanics, etc. (Asadollahi-Baboli, 2011). In the next step, machine learning methods should be employed to select the most important descriptors that characterize the property under consideration in the best possible manner.

The objectives of this study were: (1) the QSTR analysis has been performed on phenols and thiophenols toxicity to *P. phosphoreum* on the base of chemical descriptors derived from molecular structures. CART algorithm combined with LS-SVR was used for variable selection and model developing. (2) The results of CART-LS-SVR were compared to those of the previous works. Also, the QSTR model was validated by internal validation, external validation and calculation of the applicability domain. (3) Besides, we employ the best robust QSTR model to estimate the unknown aqueous toxicity of 18 structures. Finally, the results encourage considering this alternative way for the prediction of toxicity using QSTR/CART-LS-SVR model without direct determination of toxicity.

2. Materials and methods

A data set of 51 unique phenols and thiophenols toxicity in terms of pEC_{50} (mol L^{-1}) was collected from the literature (Yu et al., 2009, 2012). The term half maximal effective concentration (EC_{50}) refers to the concentration of a toxicant which induces a response halfway between the baseline and maximum after 15 min. It has been proved that there is a good agreement between environmental toxicity and toxic potency of *P. phosphoreum* (Lin et al., 2011). The list of compounds along with their toxicity values are shown in Table 1. The total dataset are randomly divided into calibration set and prediction set considering chemical diversity in a ratio of approximate 75:25 (38 and 13 compounds, respectively). The calibration set is used to construct QSTR model and the prediction set to validate the external prediction ability of the resulting QSTR model.

In order to calculate the molecular descriptors, structures of phenols and thiophenols were built using ChemDraw Ultra (version 10.0) and then the molecular structures were optimized using semiempirical quantum-chemical method AM1 Hamiltonian (Zerner et al., 1991) implemented in HyperChem software (version 7.0) to generate the energy-minimized conformations. These conformations were used to calculate

Table 1 – The experimental and predicted pEC_{50} values of substituted phenol and thiophenol using CART-LS-SVR model.

No.	Chemical	Exp. pEC_{50}	Pred. pEC_{50}	No.	Chemical	Exp. pEC_{50}	Pred. pEC_{50}
1	Phenol	2.72	2.59	27	2,4-Difluorothiophenol	5.15	5.54
2	Catechol	3.14	3.49	28	2-Chlorothiophenol[a]	4.90	5.21
3	p-Nitrophenol[a]	3.72	4.12	29	3-Chlorothiophenol	5.03	5.24
4	o-Aminophenol	3.34	3.24	30	2,3-Dichlorothiophenol	4.91	4.65
5	p-Chlorophenol	3.88	3.96	31	4-Chlorothiophenol[a]	4.99	5.06
6	m-Cresol[a]	3.31	3.54	32	2,4-Dichlorothiophenol	5.59	5.89
7	Hydroquinone	3.14	3.18	33	3,5-Dichlorothiophenol	5.11	5.29
8	o-Cresol	3.35	3.73	34	2,5-Dichlorothiophenol	5.16	5.10
9	o-Chlorophenol	3.43	3.57	35	2,6-Dichlorothiophenol	4.99	4.97
10	2,3-Dimethylphenol	3.60	4.01	36	3-Bromothiophenol[a]	4.57	4.39
11	4-tert-Butylcatechol	5.87	5.74	37	4-Bromothiophenol	5.60	5.79
12	2,4,6-Trinitrophenol[a]	2.51	2.30	38	2-Bromothiophenol	4.88	4.95
13	o-Nitrophenol	3.48	3.76	39	2-Amino-4-chlorothiophenol	5.37	5.69
14	m-Nitrophenol[a]	3.31	3.45	40	2,4-Dimethylthiophenol	4.77	4.63
15	Resorcinol	2.22	2.61	41	2,5-Dimethylthiophenol[a]	4.66	4.44
16	p-Aminophenol	3.27	3.23	42	o-Methylthiophenol	4.49	4.31
17	p-Cresol	3.71	3.70	43	p-Methylthiophenol[a]	5.89	6.07
18	2,4-Dichlorophenol	4.01	4.17	44	2,6-Dimethylthiophenol	4.64	4.38
19	4-Fluorophenol[a]	2.22	2.63	45	m-Methylthiophenol	4.60	4.50
20	2,4-difluorophenol	2.33	2.73	46	3,4-Dimethylthiophenol	4.95	4.90
21	Thiophenol	5.71	5.97	47	2-Aminothiophenol	4.72	5.03
22	2-Fluorothiophenol	4.78	5.23	48	4-Aminothiophenol[a]	4.66	5.35
23	4-Fluorothiophenol	4.97	5.43	49	3-Methoxythiophenol	4.26	4.02
24	3-Fluorothiophenol[a]	5.06	4.90	50	3,4-Dimethoxythiophenol[a]	4.72	4.67
25	2,3,5,6-Tetrafluorothiophenol	4.86	4.79	51	4-Tertiary butylthiophenol	5.34	5.09
26	Pentafluorothiophenol	4.34	4.68				

[a] Prediction set chemicals.

Download English Version:

https://daneshyari.com/en/article/5849037

Download Persian Version:

https://daneshyari.com/article/5849037

Daneshyari.com