



MOA-based linear and nonlinear QSAR models for predicting the toxicity of organic chemicals to *Vibrio fischeri*

Shengnan Zhang¹ · Ning Wang² · Limin Su¹ · Xiaoyan Xu¹ · Chao Li¹ · Weichao Qin¹ · Yuanhui Zhao¹

Received: 24 May 2019 / Accepted: 9 October 2019 / Published online: 8 January 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Risk assessment of pollutants to humans and ecosystems requires much toxicological data. However, experimental testing of compounds expends a large number of animals and is criticized for ethical reasons. The *in silico* method is playing an important role in filling the data gap. In this paper, the acute toxicity data of 1221 chemicals to *Vibrio fischeri* were collected. The global models obtained showed that there was a poor relationship between the toxicity data and the descriptors calculated based on linear and nonlinear regression analysis. This is due to the fact that the studied compounds contain not only non-reactive compounds but also reactive and specifically acting compounds with different modes of action (MOAs). MOAs are fundamental for the development of mechanistically based QSAR models and toxicity prediction. To investigate MOAs and develop MOA-based prediction models, the compounds were classified into baseline, less inert, reactive, and specifically acting compounds based on the modified Verhaar's classification scheme. Satisfactory models were established by multivariate linear regression (MLR) and support vector machine (SVM) analysis not only for baseline and less inert chemicals, but also for reactive and specifically acting compounds. Compared with linear models obtained by the MLR method, the nonlinear models obtained by the SVM method had better performance. The cross validation proved that all of the models were robust except for those for reactive chemicals with nN (number of nitrogen atoms) = 0 and $n(C=O)$ (number of carbonyl groups) > 0 ($Q^2_{ext} < 0.5$). The application domains and outliers are discussed for those MOA-based models. The models developed in this paper are significantly helpful not only because the application domains for baseline and less inert compounds have been expanded, but also the toxicity of reactive and specifically acting compounds can be successfully predicted. This work will promote understanding of toxic mechanisms and toxicity prediction for the chemicals with structural diversity, especially for reactive and specifically acting compounds.

Keywords QSAR · Mode of action · Verhaar scheme · *Vibrio fischeri* · *In silico* method · Multivariate linear regression · Support vector machine

Responsible editor: Cinta Porte

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11356-019-06681-y>) contains supplementary material, which is available to authorized users.

✉ Limin Su
sulm932@nenu.edu.cn

¹ State Environmental Protection Key Laboratory of Wetland Ecology and Vegetation Restoration, School of Environment, Northeast Normal University, Changchun, Jilin 130117, People's Republic of China

² College of Environmental Science and Engineering, Ocean University of China, Qingdao, Shandong 266100, People's Republic of China

Introduction

New chemicals enter the market at a rate of about 1000–2000 per year (Judson et al. 2009). A number of synthetic chemicals are produced and introduced to the environment through daily life (Levet et al. 2016). Some herbicides, such as atrazine and striazine, can induce developmental toxicity (Diana et al. 2000; Allran and Karasov 2001; Saka et al. 2017). 4-Nitrophenol, a commercial compound with significant functions in many industries, was reported as a potential carcinogen, teratogen, and mutagen (Mitchell and Waring 2000). These chemicals can directly or indirectly harm humans and aquatic organisms through bioaccumulation or amplification. Therefore, there is greatly increasing concern about their

toxicity; thus, it is important to assess the hazard and risk of chemical substances released into the environment (Zhang et al. 2013).

The toxicological information required in risk assessment is usually obtained from experimental determination. However, most traditional acute toxicity tests using animals or organisms are material-consuming, time-consuming, and technically difficult (Luis et al. 2007; de Melo et al. 2016; Wang et al. 2019). They also violate the reduction, replacement, and refinement (3R) principle of animal experiments (Hamadache et al. 2016). Moreover, there is a large gap between existing experimental data and actual need for accurate toxicological information of chemicals for risk assessment. It is urgent that we develop predictive methods as an alternative to animal experimentation. An *in silico* method, the quantitative structure-activity relationship (QSAR), is a technique with wide application prospects and is one of the best methods to fill this gap (Papa et al. 2013; Zhang et al. 2019). The QSAR method cannot only predict and evaluate the ecological toxicity and environmental behavior of unknown chemicals, but also explore the toxic mechanism of pollutants and support the environmental risk assessment of organic chemicals (Tsakovska et al. 2008).

Many studies have been carried out to develop QSAR models for predicting the toxicity values of various chemicals to aquatic organisms, for example, *Vibrio fischeri* (*V. fischeri*). QSAR models were developed to predict the toxicity of a set of 52 aromatic sulfone chemicals against *V. fischeri*, and the results showed that the toxicity was markedly related to water solubility (de Melo et al. 2016). QSAR techniques were adopted to predict the toxicity of alkylated aromatic hydrocarbons towards *V. fischeri* with the partial least square (PLS) analysis, and goodness-of-fit was identified through a high statistical value ($R^2 = 0.956$) (Wang et al. 2016b). The toxicity of 24 bromide-based ionic liquids (Br-ILs) against *V. fischeri* was used to establish a QSAR model with a relatively high correlation coefficient; results indicated that the toxicity was related to the energy of the lowest unoccupied molecular orbitals and the volume of Br-IL cations (Wang et al. 2015a). Most QSAR models were established based on the similar functional groups or usages/applications of chemicals, such as alcohols (Belanger et al. 2018), anilines (Tugcu and Sacan 2018), phenols (Abbasitabar and Zare-Shahabadi 2017), specifically acting antibiotics (Neale et al. 2017), pesticides (Martin et al. 2017), surfactants in personal care products (Di Nica et al. 2017), and antidepressants (Minguez et al. 2018). However, the application domains of those models were limited to narrow ranges of chemical species. Thus, they cannot be used to predict toxicity for a large number of chemicals with structural diversity. It is necessary, therefore, to develop general models which can cover different chemicals with structural diversity (Lessigiarska et al. 2004; Levet et al. 2016; Bakire et al. 2018).

The mode of action (MOA) is essential in the understanding of toxic mechanisms. It also plays a key role in the development of QSAR models. A variety of methods for the assignment of chemicals to specific MOAs are available (Könemann 1981; McKim et al. 1987; Verhaar et al. 1992; Russom et al. 1997). The Verhaar scheme (Verhaar et al. 1992) is one of the most widely used methods in which chemicals can be classified based on physicochemical properties and structure rules. Chemicals are categorized using the Verhaar scheme into five different MOAs which are baseline chemicals, less inert chemicals, reactive chemicals, specifically acting chemicals, and chemicals that are not possible to classify using the Verhaar scheme, respectively. Up to now, the QSAR models for baseline chemicals against different aquatic organisms were mostly generated. A general baseline toxicity QSAR model for the fish embryo's acute toxicity has been developed with $R^2 = 0.97$ using liposome-water partition coefficient ($\log K_{lipw}$) instead of the octanol/water partition coefficient ($\log K_{ow}$) (Klüver et al. 2016). A relationship has been found between fish toxicity and theoretical Volsurf molecular descriptors for 36 baseline chemicals with a robustness of $R^2 = 0.823$ (de Moraise Silva et al. 2018). Although a number of QSAR models for baseline or less inert (polar narcotics) compounds have been established, fewer models have been developed for reactive and specifically acting compounds. Development of QSAR models for these compounds is crucial because of their greatly toxic effect to humans and environmental organisms.

In this paper, toxicity data of acute luminescence inhibition to *V. fischeri* were compiled for 1221 chemicals with structural diversity and different MOAs. The purpose of the paper is to develop linear and nonlinear QSAR models to predict the toxicity of more organic chemicals with different MOAs to *V. fischeri*. At the same time, the property or structure factors that attribute to the acute toxicity of organic chemicals to *V. fischeri* were investigated based on the MOA-based models. According to international principles of QSAR model development (OECD guideline, 2007), the robustness and application domains of developed models are discussed in this paper. This information is very valuable for risk assessment of organic chemicals in aquatic environment, specifically for reactive and specifically acting compounds.

Material and methods

Toxicity data to *V. fischeri*

The acute toxicity data of organic chemicals to *V. fischeri* for 15 or 30 min expressed in the logarithmic form of 50% inhibition concentration of bioluminescence ($\log I/IBC_{50}$, in the unit of mol/L) was collected from literature (Kaiser and Palabrica 1991; Cronin and Schultz 1998; Zhao et al. 1998a;

Cronin et al. 2000; Dearden et al. 2000; Terasaki et al. 2009; Qin et al. 2010; Aruoja et al. 2011; Jones et al. 2011; Shi et al. 2012; Villa et al. 2012). If acute toxicity data for both endpoints was present, the toxicity value at the 15-min endpoint was preferred. A total of 1221 organic compounds and their toxicity data were collated by (1) removing ions, salts, and mixtures and (2) using arithmetic means as the final toxicity data for chemicals with more than one experimental value. The 1221 organic chemicals were divided into five MOAs using Toxtree software (<http://ecb.jrc.it/qsar/qsar-tools/index.php?c=TOXTREE>). The details of classification, together with CAS numbers, are reported in **Online Resource 2**.

Calculation of molecular parameters

Hydrophobicity quantified by the logarithm of the octanol/water partition coefficient (logKow) was obtained from the KOWWIN program in EPISuite (version 4.0) (<http://www.epa.gov/oppt/exposure/pubs/episuitedl.htm>). When possible, measured log Kow values from experiments were used before turning to calculated values. The dragon descriptors were calculated by Dragon software (Version 6.0, Talete SRL, Milano, Italy). The MM2 method (Schnur et al. 1991) involved in the ChemBio3D Ultra (Version 12.0) (<http://www.cambridgesoft.com/services/>) was used to optimize the molecular structures. The initial descriptors were reduced by excluding three types of descriptors, namely, highly correlated descriptors with (abs) pair correlation larger than or equal to 0.95, descriptors with constant (relative standard deviation < 0.0001), and descriptors with at least one missing value. Thus, a total of 1379 descriptors were maintained and used for further analysis.

Model development and evaluation

The toxicity data was divided into the training and verification sets that the toxicity values and structures of the studied compounds were well covered in the ratio of 4:1 based on the random data segmentation (RSS) (Lyakurwa et al. 2014b). The training set containing 80% of the chemicals was used to develop models, and the verification set was used for the external test. Multilinear regression (MLR) with a step-wise algorithm in SPSS 19.0 software (SPSS Company, Chicago, IL, USA) was used for linear QSAR model development. The support vector machine (SVM) running in MATLAB 2014 which has been extensively applied for nonlinear analysis was employed to build the nonlinear QSAR models.

The linear models obtained in the analysis abide by following principles. Firstly, the number of predictor variables should be reduced to avoid over-fitting. The ratio between the number of chemicals in the training set and the number of selected descriptors should be more than 5:1 (Tropsha et al.

2003). Secondly, the model should have a higher adjusted determination coefficient (R^2_{adj}) and a lower root mean squared error (RMSE). Thirdly, because the probability of highly correlated descriptors rises with the increasing number of available descriptors for model, the variable inflation factors (VIF) for the parameters should be less than 10 to avoid the issue of collinearity. Fourthly, the QUIK rules (Stewart 1989) should be satisfied, i.e., K_x (intercorrelation of selected descriptors) < K_{yx} (the correlation of the x block with y), where x is the selected molecular descriptor matrix and y is the response variable vector (Li et al. 2014; Wang et al. 2015b; Luo et al. 2017). Last but not the least, the R^2_{YS} and Q^2_{YS} of the Y-scrambling technique should be lower than the criteria of 0.3 and 0.05, respectively (Eriksson et al. 2003).

The number of the terminated generation of the nonlinear models was set to 100, and the best nonlinear model was selected with minimum variance.

The performance of linear models was evaluated by the statistical parameters. The determination coefficient (R^2) and root mean square error (RMSE) mainly reflect the goodness of fit of the models. The Q^2_{LOO} (leave-one-out cross-validation) and Q^2_{BOOT} (the bootstrap method, to 1/5 cross-validation, repeated 5000 times) were used to evaluate the robustness of the developed models. In addition, the slopes of the experimental value against predicted value or predicted value against experimental value without intercept (expressed as k and k' , respectively) for the validation set were used to evaluate the predictive ability of the models. If $0.85 \leq k \leq 1.15$ or $0.85 \leq k' \leq 1.15$, models were considered acceptable (Golbraikh and Tropsha 2002). Furthermore, the external determination coefficient (R^2_{ext}), the external explained variance (Q^2_{ext}), and the root mean square error of verification sets (RMSE_{ext}) were also adopted to characterize the predictability of the models.

The performance of nonlinear models was evaluated through the determination coefficient (R^2), the external determination coefficient (R^2_{ext}), and cross-validated of verification set (Q^2_{ext}).

Applicability domain characterization

Application domains (ADs) were characterized using the leverage distance method and the Euclidean distance method. For leverage distance method, Williams plots of the standardized residual (s) versus the leverage (h) were used to characterize ADs and determine whether the outliers or influential chemicals exist. The influential chemicals can be identified by h_i value larger than h^* ($3p/n$, where n and p are the number of chemicals and descriptors, respectively) and the outliers were diagnosed by the criterion of the standardized residual (s) being larger than 3 units (Bakire et al. 2018; Li et al. 2014). Similar to the leverage distance method, for the

Euclidean distance method, plots of the standardized residual(s) versus Euclidean distance were used to characterize ADs and determine whether outliers existed. The largest value of Euclidean distance in the training set was set as the warning value (d^*) (Li et al. 2014).

Results and discussion

QSAR models for all the chemicals

MLR analysis was carried out for all the collected chemicals. The best global linear model obtained by the stepwise analysis was

$$\begin{aligned} \log 1/IBC_{50} = & 3.096 + 0.766 \text{ SpMax4_Bh(m)} + 0.166 \log \text{Kow} \\ & - 0.720 \text{ GATS1i} - 0.147 \text{ CATS2D.02_DA} \\ & - 0.468 \text{ GATS3e} - 0.265 \text{ nROH} + 1.969 \text{ nRNHR} \\ & + 1.946 \text{ B02[S-Cl]} + 0.587 \text{ S-106} - 0.098 \text{ F02[C-F]} \end{aligned} \quad (1)$$

$$\begin{aligned} n_{\text{tra}} = & 977, R^2 = 0.500, R^2_{\text{adj}} = 0.495, Q^2_{\text{LOO}} \\ = & 0.486, Q^2_{\text{BOOT}} = 0.797, \text{RMSE} = 0.817, K_{\text{yx}} \\ = & 0.228, K_{\text{x}} = 0.189, n_{\text{ext}} = 244, R^2_{\text{ext}} \\ = & 0.467, \text{RMSE}_{\text{ext}} = 0.754, Q^2_{\text{ext}} = 0.433, R^2_{\text{YS}} \\ = & 0.010, Q^2_{\text{YS}} = -0.013 \end{aligned}$$

As shown in model (1), ten descriptors were used in the linear equation (**Online Resource 1 Table S1**). This model could only account for 49.5% of the variance (R^2_{adj}), suggesting an unsatisfactory fitting ability of the model. Although the model and coefficient of variation (R^2) could be improved with the increasing number of descriptors, the value of R^2_{ext} was no higher than 0.6 even when more descriptors were introduced into the model. At the same time, the nonlinear model was developed for all the chemicals using the SVM method, and the model was also unsatisfactory with $C = 0.594$, $g = 2.226$, $R^2 = 0.581$, and $R^2_{\text{ext}} = 0.533$.

Considering the unsatisfactory results the global linear and nonlinear models had, the chemicals were classified into four classes with different MOAs (baseline, less inert, reactive chemicals, and specifically acting chemicals) according to the modified Verhaar scheme (Enoch et al. 2008b). It should be noted that the MOAs of some chemicals cannot be assigned, and they were not used in the model construction in the present study. A total of 606 chemicals, therefore, were used for QSAR model development based on classification and analysis of mode of action.

MOA-based linear QSAR models

Linear QSAR model for baseline chemicals

A total of 215 chemicals were assigned to baseline compounds. They were divided into the training and verification sets in the ratio of 4:1. The optimal linear model was constructed by the MLR method and the result is shown as follows:

$$\begin{aligned} \log 1/IBC_{50} = & 0.55 + 0.468 \log \text{Kow} + 0.293 \text{ SpDiam_AEA(ed)} \\ & + 0.237 \text{ N\%} - 0.741 \text{ O-057} - 2.14 \text{ B09[C-Cl]} \\ & + 0.455 \text{ Eig04_AEA(ed)} - 0.761 \text{ GGI3} \end{aligned} \quad (2)$$

$$\begin{aligned} n_{\text{tra}} = & 172, R^2 = 0.787, R^2_{\text{adj}} = 0.778, Q^2_{\text{LOO}} \\ = & 0.764, Q^2_{\text{BOOT}} = 0.783, \text{RMSE} = 0.579, K_{\text{yx}} \\ = & 0.379, K_{\text{x}} = 0.345, n_{\text{ext}} = 43, R^2_{\text{ext}} \\ = & 0.788, \text{RMSE}_{\text{ext}} = 0.578, Q^2_{\text{ext}} = 0.786, R^2_{\text{YS}} \\ = & 0.041, Q^2_{\text{YS}} = -0.072 \end{aligned}$$

Model (2) contains a total of seven descriptors and their detailed information is listed in **Online Resource 1 Table S2**. It can be found that, as a dominant descriptor, $\log \text{Kow}$ can account for the largest proportion ($t = 13.277$). This result is consistent with the observation published in the literature for nonpolar anesthetic compounds (He et al. 2014). The hydrophobicity expressed as $\log P$ or $\log \text{Kow}$ is regarded as one of the most common descriptors to predict the toxicity of organic chemicals to organisms and this property decided the process of a chemical passing through the cell membrane. Studies found that hydrophobicity had a good correlation with the acute toxicity of chemicals to many aquatic species, such as fish or embryo (*Poecilia reticulata* (Su et al. 2014), zebrafish embryo (Zhu et al. 2018)), tadpoles (*Rana chensinensis* and *Rana japonica*) (Wang et al. 2019), green algae (Bakire et al. 2018), *D. magna* (Zvinavashe et al. 2009), and *Tetrahymena pyriformis* (Enoch et al. 2008a). Model (2) indicates that the acute toxicity to *V. fischeri* is correlated not only with $\log \text{Kow}$, but also with other factors. In the present study, the correlation coefficient could increase from 0.543 to 0.787 when another six descriptors were employed in the model. The value of k is 0.986 and k' is 0.992 for the validation set (details in Online Resource 1 Fig. S4), indicating that the model predictive ability is acceptable. The Q^2_{LOO} of the equation (2) is 0.764 and the Q^2_{BOOT} is 0.783, indicating that the model yields a good statistical performance. Additionally, the R^2_{YS} and Q^2_{YS} of the Y-scrambling technique are 0.041 and -0.072 , respectively, suggesting the acceptable model. The plot of experimental and predicted values of $\log 1/IBC_{50}$ is Fig. 1a. The Pearson

correlation coefficient (expressed as R_p) of experimental and predicted values is 0.886, indicating that the model established for class 1 chemicals is suitable and robust.

It should be noted that the linear model is not for highly hydrophobic chemicals with logKow value over 5 (Lee et al. 2013). Due to the 15 PAHs in the data set with logKow > 5, it will have a greater impact on the development of baseline model. Therefore, those chemicals are not considered in the establishment of baseline model.

Linear QSAR model for less inert chemicals

The model constructed by the MLR method for less inert chemicals was as follows:

$$\begin{aligned} \log 1/IBC_{50} = & -4.626 + 1.951 \text{ EE_B(p)} + 0.804 \log \text{Kow} \\ & -0.612 \text{ X2v} - 3.98 \text{ Eig03_EA(dm)} \\ & -6.251 \text{ JG13} + 1.662 \text{ CATS2D_05_DP} \\ & + 0.102 \text{ GATS7e} \end{aligned} \quad (3)$$

$$\begin{aligned} n_{\text{tra}} = 133, R^2 = 0.737, R^2_{\text{adj}} = 0.723, Q^2_{\text{LOO}} \\ = 0.641, Q^2_{\text{BOOT}} = 0.790, \text{RMSE} = 0.454, K_{\text{yx}} \\ = 0.388, K_{\text{x}} = 0.348, n_{\text{ext}} = 33, R^2_{\text{ext}} \\ = 0.758, \text{RMSE}_{\text{ext}} = 0.431, Q^2_{\text{ext}} = 0.729, R^2_{\text{YS}} \\ = 0.053, Q^2_{\text{YS}} = -0.060 \end{aligned}$$

It was found that the model tested by Y-scrambling technique was acceptable with R^2_{YS} 0.053 and $Q^2_{\text{YS}} = -0.060$ with a total of seven descriptors. Detailed information is listed in **Online Resource 1** Table S3. This model could account for 72.3% of the variance (R^2_{adj}), indicating the preferable fitting ability of the model. The value of k is 0.978 and k' is 1.013 for the validation set (Online Resource 1 Fig. S5), indicating that the predictive ability of the regression for less inert compounds is acceptable. The plot of experimental and predicted values of log1/IBC₅₀ for less inert chemicals is shown in Fig. 1c and the outcome was acceptable ($R_p = 0.858$). The observed and predicted log1/IBC₅₀ values, together with the descriptors introduced to model (3), are reported in **Online Resource 2**. It was found that logKow still has a large influence on the toxicity of less inert chemicals. This situation was also observed in other aquatic organisms (Qin et al. 2010; Vighi et al. 2009; Fu et al. 2015). EE_B (p) ($t = 8$), the second most important descriptor in the model, is an estrada-like index (logarithmic form) from Burden matrix weighted by polarizability. Therefore, it is considered that the polarizability of less inert chemicals has a significant effect on the toxicity. A positive relationship indicates that the higher polarizability a chemical has, the more toxic the chemical is.

Linear QSAR models for reactive chemicals

For reactive compounds, poor relationship was usually found between toxicity values and descriptors (Zhu et al. 2018). The unified prediction model on reactive chemicals could not be obtained due to the fairly complex structures of the compounds in class 3. It has been identified that the toxicity would be influenced by some structural characteristics such as the number of nitrogen atoms (nN) or carbonyl groups (n(C=O)) (Lyakurwa et al. 2014a). Therefore, the chemicals in class 3 were further divided into three groups according to nN or n(C=O). The results were as follows:

Group 1: nN > 0

$$\begin{aligned} \log 1/IBC_{50} = & 2.509 + 21.815 \text{ X5Av} - 2.751 \text{ MATS3s} \\ & + 0.252 \text{ GATS8m} - 0.249 \text{ Eig11_EA(dm)} \\ & - 0.227 \text{ F06[C-O]} + 0.64 \text{ MLOGP} \end{aligned} \quad (4)$$

$$\begin{aligned} n_{\text{tra}} = 57, R^2 = 0.801, R^2_{\text{adj}} = 0.777, Q^2_{\text{LOO}} \\ = 0.729, Q^2_{\text{BOOT}} = 0.757, \text{RMSE} = 0.527, K_{\text{yx}} \\ = 0.299, K_{\text{x}} = 0.189, n_{\text{ext}} = 14, R^2_{\text{ext}} \\ = 0.738, \text{RMSE}_{\text{ext}} = 0.605, Q^2_{\text{ext}} = 0.715, R^2_{\text{YS}} \\ = 0.106, Q^2_{\text{YS}} = -0.199 \end{aligned}$$

The R^2_{YS} and Q^2_{YS} of the Y-scrambling technique are 0.106 and -0.199 lower than the criteria of 0.3 and 0.05, respectively, which confirms that the model is acceptable. Six descriptors were introduced to the model (4) and detailed information is listed in **Online Resource 1** Table S4. Although both MLOGP and logKow represent the octanol/water partition coefficient, the results from different calculation methods could be slightly different. MLOGP and X5Av had the fairly large t values, 7.692 and 4.501, respectively (Online Resource 1), demonstrating that the toxic values of chemicals in this group are mainly affected by the hydrophobicity and the average connectivity index of chemicals. The fitting ability and robustness of the model are acceptable ($R^2 = 0.801$, $\text{RMSE} = 0.527$, $Q^2_{\text{LOO}} = 0.729$, $Q^2_{\text{BOOT}} = 0.757$). The predictive ability is acceptable with $k = 0.969$ and $k' = 1.015$ (see Fig. S6 A and B in Online Resource 1). Moreover, the plot of experimental and predicted values of log1/IBC₅₀ for chemicals in group 1 is shown in Fig. S1A in **Online Resource 1** with $R_p = 0.877$, indicating that the model was statistically significant.

Group 2: nN = 0, n(C=O) = 0

$$\begin{aligned} \log 1/IBC_{50} = & 2.111 + 2.078 \text{ B03[C-C]} - 0.759 \text{ Hy} \\ & + 1.759 \text{ B06[C-Cl]} - 0.946 \text{ B06[O-F]} \end{aligned} \quad (5)$$

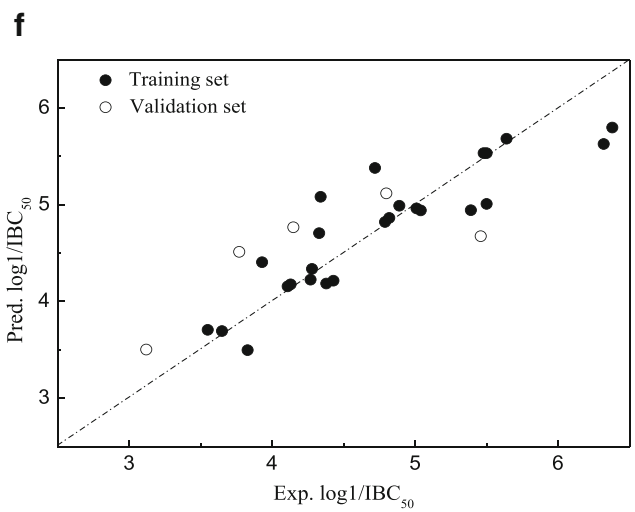
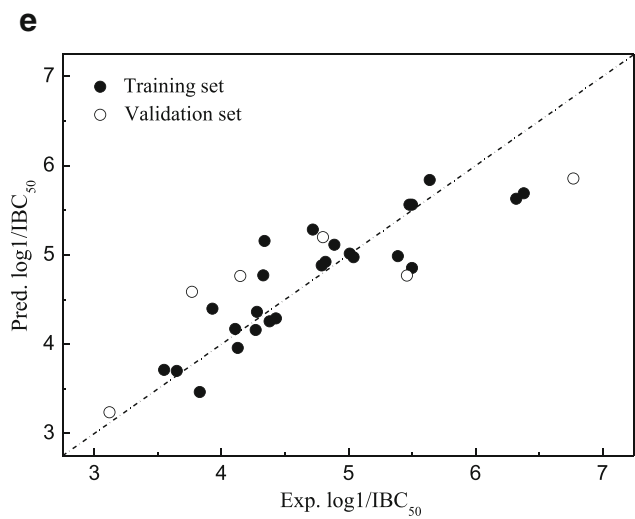
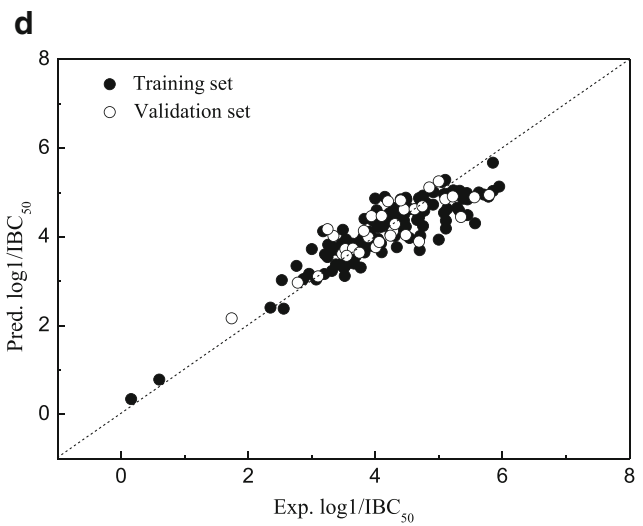
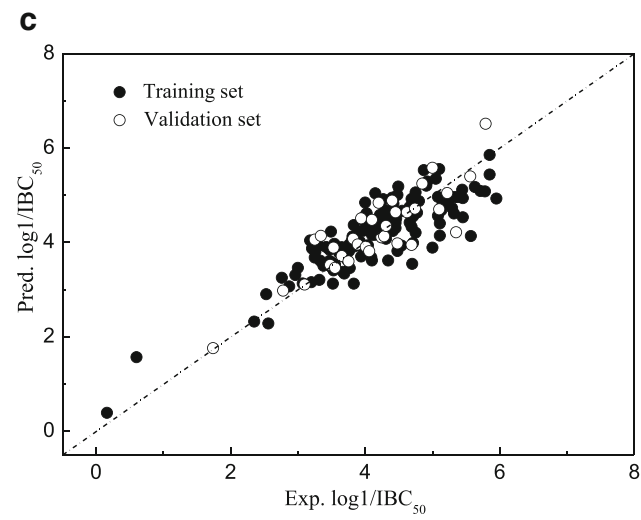
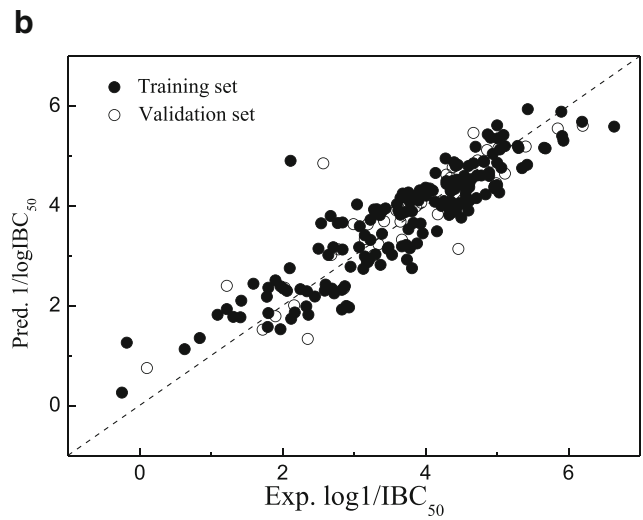
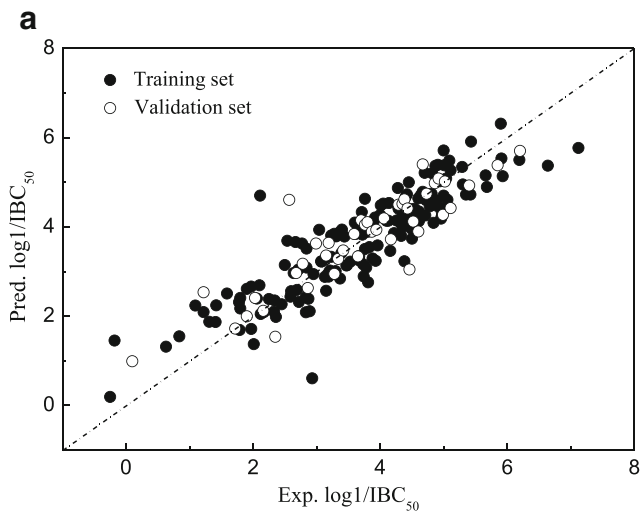


Fig. 1 Fitting plots of experimental values and predictive values for baseline, less inert, and specifically acting chemicals (**a** baseline chemicals by the MLR method; **b** baseline chemicals by the SVM method; **c** less inert chemicals by the MLR method; **d** less inert chemicals by the SVM method; **e** specifically acting chemicals by the MLR method; **f** specifically acting chemicals by the SVM method)

$$\begin{aligned}n_{\text{tra}} &= 29, R^2 = 0.864, R^2_{\text{adj}} = 0.841, Q^2_{\text{LOO}} \\ &= 0.749, Q^2_{\text{BOOT}} = 0.784, \text{RMSE} = 0.489, K_{\text{yx}} \\ &= 0.360, K_{\text{x}} = 0.191, n_{\text{ext}} = 7, R^2_{\text{ext}} = 0.833, \text{RMSE}_{\text{ext}} \\ &= 0.614, Q^2_{\text{ext}} = 0.720, R^2_{\text{YS}} = 0.143, Q^2_{\text{YS}} = -0.237\end{aligned}$$

The model is acceptable with the $R^2_{\text{YS}} = 0.143$ and $Q^2_{\text{YS}} = -0.237$. There are four descriptors in model (5) and detailed information is listed in **Online Resource 1** Table S5. B03[C-C] which is a 2D atom pair descriptor was positively correlated to $\log 1/\text{IBC}_{50}$ with the largest value of t ($t = 8.527$), manifesting that the toxicity of the compounds increased with the increasing number of the C–C structure fragment. This trend is opposite to that predicted by Hy which is a descriptor of hydrophilia. Bakire et al. (2018) found that, for reactive chemicals ($n_{\text{N}} = 0, n_{\text{C=O}} = 0$), only $\log K_{\text{ow}}$ observably related to the toxicity to green algae. Compared with the descriptor Hy in this model, $\log K_{\text{ow}}$ stands for hydrophobicity which is the opposite descriptor of Hy. It states clearly that hydrophobicity is positively related to the toxicity of reactive chemicals without nitrogen atoms and carbonyl groups ($n_{\text{N}} = 0, n_{\text{C=O}} = 0$). The highly hydrophobic chemicals would be more lipophilic and easier to pass through the membrane to cause toxicity. Relatively high negative correlation was found between $\log K_{\text{ow}}$ and Hy ($R_{\text{p}} = -0.778$). Hy instead of hydrophobicity introduced to the model (5) indicates that the more the solubility of a chemical is, the less the toxicity to *V. fischeri* is. A few 2D atom pairs descriptors (B06[C-Cl], B06 [O-F]) in the present study were also introduced into the model (5) and satisfactory performance was obtained ($R^2 = 0.864, R^2_{\text{ext}} = 0.833, Q^2_{\text{ext}} = 0.720$). The plot of experimental and predicted values of $\log 1/\text{IBC}_{50}$ is shown in Fig. S1C in **Online Resource 1** and significant correlation was found ($R_{\text{p}} = 0.919$). The values of k and k' (Online Resource 1 Fig. S6 C and D) for the validation set of reactive compounds in group 2 are 0.916 and 1.073, respectively. It is clear that model (5) has good predictive ability.

Group 3: $n_{\text{N}} = 0, n_{\text{C=O}} > 0$

$$\begin{aligned}\log 1/\text{IBC}_{50} &= -1.714 + 0.758 \text{ nRCO} - 0.482 \text{ nArOR} \\ &+ 11.071 \text{ SpPosA}_A - 1.799 \text{ Eta}_F_A \\ &+ 0.727 \text{ F02}[\text{O-Cl}] + 0.346 \text{ Eig02}_E\text{A}(\text{dm}) \\ &+ 0.327 \text{ Eig06}_E\text{A}(\text{dm})\end{aligned}\quad (6)$$

$$\begin{aligned}n_{\text{tra}} &= 69, R^2 = 0.734, R^2_{\text{adj}} = 0.704, Q^2_{\text{LOO}} \\ &= 0.659, Q^2_{\text{BOOT}} = 0.772, \text{RMSE} = 0.326, K_{\text{yx}} \\ &= 0.303, K_{\text{x}} = 0.246, n_{\text{ext}} = 18, R^2_{\text{ext}} \\ &= 0.447, \text{RMSE}_{\text{ext}} = 0.639, Q^2_{\text{ext}} = 0.441, R^2_{\text{YS}} \\ &= 0.096, Q^2_{\text{YS}} = -0.171\end{aligned}$$

Y-scrambling test indicates that the model is acceptable ($R^2_{\text{YS}} = 0.096$ and $Q^2_{\text{YS}} = -0.171$). A total of seven descriptors were used in model (6) and detailed information is listed in **Online Resource 1** Table S6. The number of carbonyl in the aliphatic chemicals represented by nRCO was positively correlated to $\log 1/\text{IBC}_{50}$, suggesting that the more carbonyl a chemical has, the more toxic the chemical is. Eig06_EA(dm) stands for an edge adjacency index weighted by the molecular dipole moment reflecting the polarity of a molecule (Luo et al. 2017). It indicated that the toxicity may be caused by the chemicals undergoing polar interactions with biomacromolecules. The plot of experimental and predicted values of $\log 1/\text{IBC}_{50}$ is shown in **Online Resource 1** Fig. S1E and R_{p} was equal to 0.812. The prediction of the model for training set was satisfactory; however, when it was used to predict the toxicity of the chemicals in validation set, an unsatisfactory outcome was obtained with $R^2_{\text{ext}} = 0.447$.

Linear QSAR model for specifically acting chemicals

The model by the MLR method for the class four chemicals was as follows:

$$\begin{aligned}\log 1/\text{IBC}_{50} &= -10.193 + 5.839 \text{ SpMAD}_A\text{EA}(\text{dm}) \\ &- 0.179 \text{ F03}[\text{C-N}] + 3.994 \text{ SpMAD}_B(\text{p})\end{aligned}\quad (7)$$

$$\begin{aligned}n_{\text{tra}} &= 25, R^2 = 0.766, R^2_{\text{adj}} = 0.733, Q^2_{\text{LOO}} \\ &= 0.665, Q^2_{\text{BOOT}} = 0.729, \text{RMSE} = 0.327, K_{\text{yx}} \\ &= 0.351, K_{\text{x}} = 0.172, n_{\text{ext}} = 6, R^2_{\text{ext}} = 0.749, \text{RMSE}_{\text{ext}} \\ &= 0.649, Q^2_{\text{ext}} = 0.703, R^2_{\text{YS}} = 0.127, Q^2_{\text{YS}} = -0.269\end{aligned}$$

Specifically acting chemicals are made up of various chemicals and the toxicity mostly ascribes to the interaction with the target receptors. A total of three descriptors were introduced into the model (7), namely, $\text{SpMAD}_A\text{EA}(\text{dm})$, $\text{F03}[\text{C-N}]$, and $\text{SpMAD}_B(\text{p})$. $\text{SpMAD}_A\text{EA}(\text{dm})$, which stands for spectral mean absolute deviation from augmented edge adjacency matrix weighted by dipole moment edge adjacency indices, has the largest influence on toxicity. $\text{F03}[\text{C-N}]$ is the frequency at which the C-N of the topological distance is 3, and $\text{SpMAD}_B(\text{p})$ represents the spectral mean absolute deviation from Burden matrix weighted by

polarizability. The values of R^2_{YS} and Q^2_{YS} by Y-scrambling test are 0.127 and -0.269 , respectively. The results indicate that model (7) is acceptable for the toxicity prediction of specifically acting chemicals to *V. fischeri*. The Q^2_{LOO} and Q^2_{BOOT} values ($Q^2_{LOO} = 0.665$, $Q^2_{BOOT} = 0.729$) indicate that the model has good robustness. The values of external prediction correlation coefficients ($R^2_{ext} = 0.749$ and $Q^2_{ext} = 0.703$) as well as k and k' (0.996 and 0.986) reflect the good external prediction ability of the model. The plot of experimental and predicted values of $\log 1/IBC_{50}$ is shown in Fig. 1e and R_p was 0.864.

Comparison with existing linear QSAR models

The existing linear QSAR models for *V. fischeri* have been collected and shown in Table 1. Inspection of the QSAR models shows that most of them were established based on chemical species (Cronin et al. 2000; Wang et al. 2016b; Wang et al. 2015a; de Melo et al. 2016). Although a few MOA-based models have been established, most of them are used to predict the toxicity baseline (nonpolar narcotics) or less inert (polar narcotics) compounds (Zhao et al. 1998b; Li et al. 2015; Wang et al. 2016a). Only one global model was

established. However, the AD of the model is limited because a small data set (102 chemicals) was used for building the model (Qin et al. 2010). To our knowledge, this is the first work to carry out an investigation on linear and nonlinear QSARs for the toxicity of *V. fischeri* by a large number of chemicals based on their MOAs. More importantly, the models developed include not only baseline and less inert compounds, but also reactive and specifically acting compounds. If a chemical can be classified into one of four MOAs (baseline, less inert, reactive chemicals, and specifically acting chemicals) based on its structural characters, the toxicity of the chemical to *V. fischeri* can be predicted using the MOA-based models in this study. It should be noted that, if the MOAs are unclear and cannot be identified using the Verhaar scheme, their toxicity cannot be well predicted by the models established in this study. In addition, reactive chemicals with $nN = 0$ and $n(C=O) > 0$ cannot be predicted either, because a satisfactory model has not been established for this group of chemicals. Compared with the established models presented in the references (Table 1), the ADs of the models in this study have been broadened to the chemicals with determined MOAs.

Table 1 QSAR models on the toxicity to *V. fischeri* from references

Chemicals	<i>n</i>	R^2	QSAR models ^a	Ref.
Baseline compound	77	0.87	$\log 1/IBC_{50} = 0.970 \log Kow + 0.98$	Li et al. 2015
Polar narcotics	15	0.84	$\log 1/EC_{50} = 2.50 + 0.645 \log P$	Qin et al. 2010
Non-polar narcotics	33	0.88	$\log 1/EC_{50} = 0.89 + 1.01 \log P$	
Non-polar and polar narcotics	48	0.86	$\log 1/EC_{50} = 0.76 + 0.790 \log P + 1.16 S$	
Non-polar, polar, and reactive chemicals	90	0.81	$\log 1/EC_{50} = 0.91 + 0.746 \log P + 1.12 S + 1.15 I_{NO_2}$	
Non-polar, polar, reactive, and ionizable chemicals	102	0.79	$\log 1/EC_{50} = 1.04 + 0.701 \log P + 1.11 S + 1.12 I_{NO_2} - 0.157 \log F_0$	
Non-polar narcosis	33	0.85	$\log 1/EC_{50} = 0.824 \log P + 1.29$	Zhao et al. 1998b
Polar narcosis (phenols and anilines)	12	0.85	$\log 1/EC_{50} = 0.617 \log P + 2.61$	
Haloalcohols	15	0.908	$pT_{15} = 1.15 \log P - 1.34$	Cronin et al. 2000
Halonitriles	11	0.954	$pT_{15} = 1.61 \log P - 0.501 (E_{LUMO}) - 1.19$	
HaloesterS	11	0.873	$pT_{15} = 0.473 \log P - 0.521 (E_{LUMO}) + 0.146$	
Diones	10	0.924	$pT_{15} = 1.18 \log P - 0.859 (E_{LUMO}) - 0.281$	
Alkylated aromatic hydrocarbons	17	0.956	$\log(1/EC_{50}, \text{free}) = 2.760 + 1.804 \times 10^{-3} V + 1.909 \times 10^{-3} \alpha + 8.580 \times 10^{-1} E_{HOMO} - 7.454 \times 10^{-1} \omega + 3.463 \times 10^{-3} E_{th} + 8.646 \times 10^{-3} C_v + 8.105 \times 10^{-3} S^o - 5.973 \times 10^{-4} E_T - 5.975 \times 10^{-4} H^+ + 5.974 \times 10^{-4} G^+$	Wang et al. 2016b
Bromide-based ionic liquids (Br-ILs)	24	0.954	$\log IC_{50} = 8.493 E_{LUMO} - 0.026 V + 7.846$	Wang et al. 2015a
Aromatic sulfone chemicals	35	0.804	$pEC_{50} = 4.603 - 0.005 P_VSA_v_3 - 0.352 SM02_AEA(dm) - 0.599 ATS4m - 0.009 CATS2D_07_LL$	de Melo et al. 2016
Baseline compound	97	0.89	$\log 1/IBC_{50} = 0.994 \log Kow + 0.863$	Wang et al. 2016a
Less inert compound	76	0.79	$\log 1/IBC_{50} = 0.708 \log Kow + 2.26$	

^a In different models, $\log 1/IBC_{50}$, pEC_{50} , $\log 1/EC_{50}$, pT_{15} , $\log(1/EC_{50}, \text{free})$, and $\log IC_{50}$ represented different toxicity endpoints or different expressions of the toxicity of chemicals to *V. fischeri*; $\log Kow$ (also expressed as $\log P$ in some references) is a descriptor related to hydrophobicity.

Nonlinear QSAR models

The nonlinear QSAR models were also investigated for chemicals based on classification and MOAs using the SVM method. The prediction ability and robustness as well as the calculating parameters are shown in Table 2.

The results in Table 2 show that greater performance was found from the nonlinear models developed with SVM when compared with models developed with the MLR method except for reactive chemicals without nitrogen atoms and carbonyl group.

High-influence chemicals and outliers diagnosis

Leverage distance and Euclidean distance methods were used to characterize application domains (ADs). For model established for baseline compounds, seven chemicals are defined as influential compounds with $h_i > h^*$ and $|s| < 3$ using the leverage distance method (see Williams plots in Fig. 2a). Those influential compounds are not outliers. Three compounds (acetic acid, octane in the training set, and 1-chlorooctane in the verification set) are identified as outliers with $|s| > 3$. The toxicity of 1-chlorooctane and octane is overestimated (Table 3) from the model. Outliers occurred for several reasons (Zhao et al. 2009; Hamadache et al. 2016; Wang et al. 2015b; Bakire et al. 2018). First, experimental errors or experimental uncertainty may be one reason for the deviation of 1-chlorooctane and octane. Apparently, the experimental toxicity sometimes does not really reflect “true” toxicity of the compounds because of experimental errors or experimental uncertainty (Zhao et al. 2009; Hamadache et al. 2016). The toxicity of octane and 1-chlorooctane (2.11 and 2.57) seems too low compared with that of nonane and heptane (5.93 and 4.96) although they are structurally similar compounds. Second, species sensitivity may be another reason for the outliers. Most straight-chain alkanes belong to the baseline mode for toxicity to fish, whereas more straight-chain alkanes including octane and 1-chlorooctane are identified as outliers for toxicity to *V. fischeri* (Wang et al. 2016a). This

means that *V. fischeri* is not sensitive to all alkanes which results in more outliers observed in *V. fischeri* toxicity than in fish toxicity. The predicted toxic value of acetic acid was underestimated and also identified as an outlier. The pka value of acetic acid is 4.9, whereas the pH of *V. fischeri* toxicity test is close to 7. The ionic form of acetic acid exists under the test condition. Higher toxicity of acetic acid was observed which may ascribe to the higher ionization. This makes it easier for the chemical to enter the cell tissue and interact with the organism, *V. fischeri*. This phenomenon has been observed by one other study (Zhang et al. 2010). In addition, the water phases of *V. fischeri* are the main storage sites, rather than the lipid tissue, for a highly hydrophilic compound. This would result in the underestimation of toxicity from log Kow (Wen et al. 2012).

The AD of nonlinear model for base line chemicals is shown in Fig. 2b; octane in the training set and 1-chlorooctane in the verification set were also considered as outliers with $|s| > 3$. The results were relatively matched by those of linear model.

For less inert chemicals, the ADs of linear and nonlinear models are shown in Fig. S2A and Fig. S2B (Online Resource 1), respectively. For linear model, four compounds (allylamine, 4-chloro-N-methylaniline, antioxidant 264, and p-aminodiphenylamine) in the training set and 4-n-nonylpheno in the verification set were defined as influential compounds with $h_i > h^*$ and $|s| < 3$. Only one compound (4-bromophenol) in the training set with $|s| > 3$ is regarded as the outlier of the model.

For reactive chemicals, it is found that three compounds (benzyl benzoate, metolachlor, and dithiocyanomethane in the training set) predicted by linear model are identified as influential chemicals with leverages exceeding the warning value ($h^* = 0.368$) in group 1 (Fig. S3A and Fig. S3B in Online Resource 1). Similarly, the leverage value of 1',4'-dichloro-p-xylylene in the training set exceeded the warning value in group 2 (Fig. S3C and Fig. S3D in Online Resource 1). However, the predicted result is not significantly affected. The standardized residuals obtained by the

Table 2 Parameters of nonlinear QSAR models

	R^2	Q^2_{cv}	R^2_{ext}	Q^2_{ext}	C	g
Baseline chemicals	0.814	0.774	0.792	0.782	0.616	0.957
Less inert chemicals	0.785	0.705	0.742	0.737	6.155	0.576
Group 1 of reactive chemicals	0.816	0.742	0.801	0.776	4.561	0.121
Group 2 of reactive chemicals	0.866	0.764	0.790	0.684	85.645	0.031
Group 3 of reactive chemicals	0.736	0.634	0.445	0.441	3.844	0.108
Specifically acting chemicals	0.799	0.732	0.767	0.675	35.841	0.068

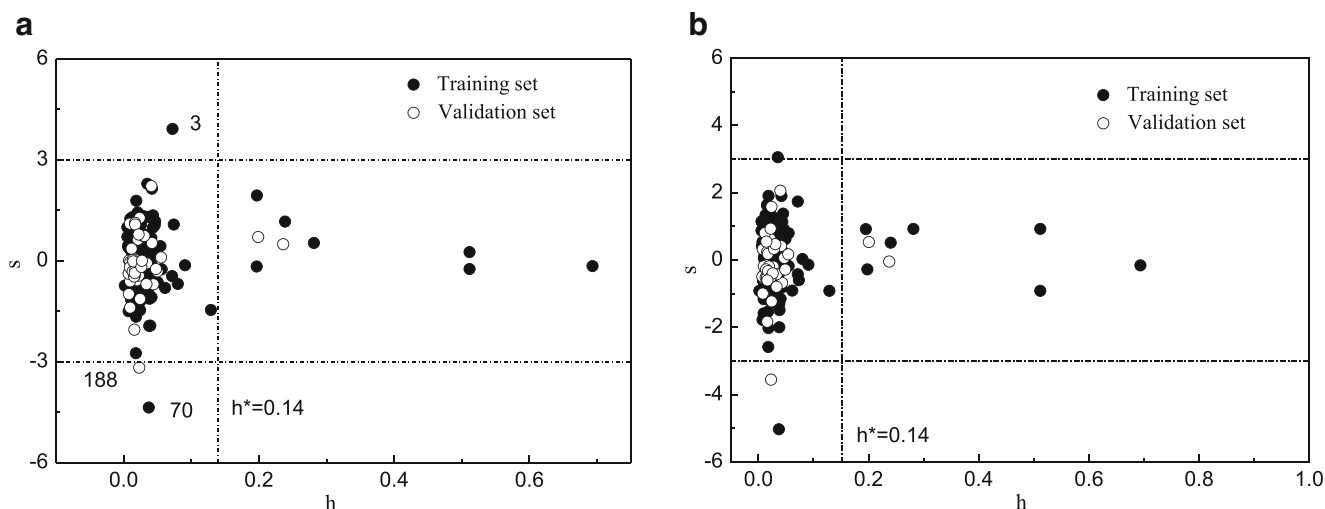


Fig. 2 Plots of ADs by the leverage distance method for baseline chemicals (**a** the MLR method; **b** the SVM method)

MLR method are similar to those obtained using the SVM method. There is no significant difference between ADs by the SVM method and by the MLR method. All chemicals are within the ADs by both methods. The Williams diagrams for linear model and nonlinear model of group 3 are shown in Fig. S3E and Fig. S3F in **Online Resource 1**. The result from the toxicity to organism of green algae is consistent with that from the present study and no acceptable models were established for group 3 of reactive chemicals ($nN = 0, n(C=O) > 0$) (Bakire et al. 2018).

For specifically acting chemicals, it has been verified that the 31 compounds covered by linear or nonlinear models are all within the AD of model (7) (Fig. S2C and Fig. S2D in **Online Resource 1**).

Based on the Euclidean distance method, plots of the standardized residual (s) versus Euclidean distance were used to characterize ADs and determine whether the outliers exist. All plots not only for linear models but also for nonlinear models are shown in Online Resource 1 Fig. S9 and Fig. S10. The outliers decided by Euclidean distance method are the same with those by leverage distance method. As the biggest value of Euclidean distance in the training set is set as the warning value (d^*), no influential chemicals exist for all the models.

The reason is probably due to the different warning values adopted by different methods.

Conclusions

This study demonstrates that the global linear and nonlinear models for all collected acute toxicity data of 1221 chemicals to *V. fischeri* were unsatisfactory for chemicals with structural diversity and different MOAs. Identification of MOA is crucial for the establishment of mechanistically based QSAR models. MOA-based linear and nonlinear models have been developed for baseline, less inert, reactive, and specifically acting compounds based on the modified Verhaar’s classification scheme. QSAR models based on MOAs were more predictable and robust not only for baseline and less inert chemicals, but also for reactive and specifically acting compounds. Compared with linear models obtained through the MLR method, the nonlinear models obtained by the SVM method had better performance. There was no significant difference between ADs determined by the SVM method and by the MLR method. The most extensive chemicals with toxic values to *V. fischeri* could be predicted when the MOA of a chemical was assigned. The descriptors selected in the models reveal that the acute toxicity of baseline compounds is dominated by the hydrophobicity. Also of note, chemical polarizability has an effect on the toxicity of acute exposure when dealing with less inert and reactive chemicals. The application domains of linear and nonlinear models and outliers have been discussed and explained. The models developed in this paper can be used to predict the toxicity not only for baseline and less inert compounds, but also for reactive and

Table 3 Outliers of the model for baseline chemicals

CAS	Compound	Experimental	Verhaar scheme	
			Class	Predicted
64-19-7	Acetic acid (acetate)	2.93	1	0.61
111-65-9	Octane	2.11	1	4.70
111-85-3	1-Chlorooctane	2.57	1	4.61

specifically acting compounds. This information is very valuable for the risk assessment of organic chemicals in an aquatic environment, specifically for reactive and specifically-acting compounds.

Funding information This work was supported by the National Natural Science Foundation of China (21777022 and 21107012) and the Fundamental Research Funds for the Central Universities (2412018ZD014)

References

- Abbasitabar F, Zare-Shahabadi V (2017) *In silico* prediction of toxicity of phenols to *Tetrahymena pyriformis* by using genetic algorithm and decision tree-based modeling approach. *Chemosphere* 172:249–259
- Allran JW, Karasov WH (2001) Effects of atrazine on embryos, larvae, and adults of anuran amphibians. *Environ Toxicol Chem* 20(4):769–775
- Aruoja V, Sihtmae M, Dubourguier HC, Kahru A (2011) Toxicity of 58 substituted anilines and phenols to algae *Pseudokirchneriella subcapitata* and bacteria *Vibrio fischeri*: Comparison with published data and QSARs. *Chemosphere* 84:1310–1320
- Bakire S, Yang XY, Ma GC, Wei XX, Yu HY, Chen JR, Lin HJ (2018) Developing predictive models for toxicity of organic chemicals to green algae based on mode of action. *Chemosphere* 190:463–470
- Belanger SE, Rawlings JM, Stackhouse R (2018) Advances in understanding the response of fish to linear alcohols in the environment. *Chemosphere* 206:539–548
- Cronin MTD, Schultz TW (1998) Structure-toxicity relationships for three mechanisms of action of toxicity to *Vibrio fischeri*. *Ecotoxicol Environ Saf* 39:65–69
- Cronin MTD, Bowers GS, Sinks GD, Schultz TW (2000) Structure-toxicity relationships for aliphatic compounds encompassing a variety of mechanisms of toxic action to *Vibrio fischeri*. *SAR QSAR Environ Res* 11:301–312
- de Melo EB, Martins JPA, Miranda EH, Ferreira MMC (2016) A best comprehension about the toxicity of phenylsulfonyl carboxylates in *Vibrio fischeri* using quantitative structure activity/property relationship methods. *J Hazard Mater* 304:233–241
- de Moraise Silva L, Alves MF, Scotti L, Lopes WS, Scotti MT (2018) Predictive ecotoxicity of MoA 1 of organic chemicals using in silico approaches. *Ecotoxicol Environ Saf* 153:151–159
- Dearden JC, Cronin MTD, Zhao YH, Raevsky OA (2000) QSAR studies of compounds acting by polar and non-polar narcosis: An examination of the role of polarisability and hydrogen bonding. *Quant Struct -Act Relat* 19(1):3–9
- Di Nica V, Gallet J, Villa S, Mezzanotte V (2017) Toxicity of Quaternary Ammonium Compounds (QACs) as single compounds and mixtures to aquatic non-target microorganisms: Experimental data and predictive models. *Ecotoxicol Environ Saf* 142:567–577
- Diana SG, Resetarits WJ, Schaeffer DJ, Beckmen KB, Beasley VR (2000) Effects of atrazine on amphibian growth and survival in artificial aquatic communities. *Environ Toxicol Chem* 19(12):2961–2967
- Enoch SJ, Cronin MTD, Schultz TW, Madden JC (2008a) An evaluation of global QSAR models for the prediction of the toxicity of phenols to *Tetrahymena pyriformis*. *Chemosphere* 71:1225–1232
- Enoch SJ, Hewitt M, Cronin MTD, Azam S, Madden JC (2008b) Classification of chemicals according to mechanism of aquatic toxicity: An evaluation of the implementation of the Verhaar scheme in Toxtree. *Chemosphere* 73:243–248
- Eriksson L, Jaworska J, Worth AP, Cronin MTD, McDowell RM, Gramatica P (2003) Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environ Health Perspect* 111:1361–1375
- Fu L, Li JJ, Wang Y, Wang XH, Wen Y, Qin WC, Su LM, Zhao YH (2015) Evaluation of toxicity data to green algae and relationship with hydrophobicity. *Chemosphere* 120:16–22
- Golbraikh A, Tropsha A (2002) Beware of q! *J Mol Graph Model* 20:269–276
- Hamadache M, Benkortbi O, Hanini S, Amrane A, Khaouane L, Moussa CS (2016) A Quantitative Structure Activity Relationship for acute oral toxicity of pesticides on rats: Validation, domain of application and prediction. *J Hazard Mater* 303:28–40
- He J, Fu L, Wang Y, Li JJ, Wang XH, Su LM, Sheng LX, Zhao YH (2014) Investigation on baseline toxicity to rats based on aliphatic compounds and comparison with toxicity to fish: Effect of exposure routes on toxicity. *Regul Toxicol Pharmacol* 70:98–106
- Jones D, Scarlett AG, West CE, Rowland SJ (2011) Toxicity of Individual Naphthenic Acids to *Vibrio fischeri*. *Environ Sci Technol* 45:9776–9782
- Judson R, Richard A, Dix DJ, Houck K, Martin M, Kavlock R, Dellarco V, Henry T, Holderman T, Sayre P, Tan S, Carpenter T, Smith E (2009) The Toxicity Data Landscape for Environmental Chemicals. *Environ Health Perspect* 117:685–695
- Kaiser KLE, Palabrica VS (1991) *Photobacterium phosphoreum* Toxicity Data Index. *Water PollResJCanda* 26(26):361–431
- Klüver N, Vogts C, Altenburger R, Escher BI, Scholz S (2016) Development of a general baseline toxicity QSAR model for the fish embryo acute toxicity test. *Chemosphere* 164:164–173
- Könemann H (1981) Quantitative structure-activity relationships in fish toxicity studies. Part 1: relationship for 50 industrial pollutants. *Toxicology* 19:209–221
- Lee SY, Kang HJ, Kwon JH (2013) Toxicity cutoff of aromatic hydrocarbons for luminescence inhibition of *Vibrio fischeri*. *Ecotoxicol Environ Saf* 94:116–122
- Lessigiarska CMTD, Worth AP, Dearden JC, Netzeva TI (2004) QSARs for toxicity to the bacterium *Sinorhizobium meliloti*. *SAR QSAR Environ Res* 15:169–190
- Levet A, Bordes C, Clement Y, Mignon P, Morell C, Chermette H, Marote P, Lanteri P (2016) Acute aquatic toxicity of organic solvents modeled by QSARs. *J Mol Model* 22:288
- Li C, Yang XH, Li XH, Chen JW, Qiao XL (2014) Development of a model for predicting hydroxyl radical reaction rate constants of organic chemicals at different temperatures. *Chemosphere* 95:613–618
- Li JJ, Wang XH, Wang Y, Wen Y, Qin WC, Su LM, Zhao YH (2015) Discrimination of excess toxicity from narcotic effect: influence of species sensitivity and bioconcentration on the classification of modes of action. *Chemosphere* 120:660–673
- Luis P, Ortiz I, Aldaco R, Irabien A (2007) A novel group contribution method in the development of a QSAR for predicting the toxicity (*Vibrio fischeri* EC₅₀) of ionic liquids. *Ecotoxicol Environ Saf* 67:423–429
- Luo X, Yang XH, Qiao XL, Wang Y, Chen JW, Wei XX, Peijnenburg WJGM (2017) Development of a QSAR model for predicting aqueous reaction rate constants of organic chemicals with hydroxyl radicals. *Environ Sci: Processes Impacts* 19: 350–356
- Lyakurwa F, Yang XH, Li XH, Qiao XL, Chen JW (2014a) Development and validation of theoretical linear solvation energy relationship models for toxicity prediction to fathead minnow (*pimephales promelas*). *Chemosphere* 96:188–194
- Lyakurwa F, Yang XH, Li XH, Qiao XL, Chen JW (2014b) Development of in silico models for predicting LSER molecular parameters and for acute toxicity prediction to fathead minnow (*Pimephales promelas*). *Chemosphere* 108:17–25

- Martin TM, Lilavois CR, Barron MG (2017) Prediction of pesticide acute toxicity using two-dimensional chemical descriptors and target species classification. SAR QSAR Environ Res 28:525–539
- McKim JM, Bradbury SP, Niemi GJ (1987) Fish acute toxicity syndromes and their use in the QSAR approach to hazard assessment. Environ Health Perspect 71:171–186
- Minguez L, Bureau R, Halm LMP (2018) Joint effects of nine antidepressants on *Raphidocelis subcapitata* and *Skeletonema marinoi*: A matter of amine functional groups. Aquat Toxicol 196:117–123
- Mitchell SC, Waring RH (2000) Ullmanns Encyclopedia of Industrial Chemistry. Weinheim, Germany
- Neale PA, Leusch FDL, Escher BI (2017) Applying mixture toxicity modelling to predict bacterial bioluminescence inhibition by non-specifically-acting pharmaceuticals and specifically-acting antibiotics. Chemosphere 173:387–394
- OECD (2007) Guidance document on the validation of (Quantitative) structure– activity relationships [(Q)SAR] models. Organisation for Economic Co-Operation and Development, Paris, France.
- Papa E, Kovarich S, Gramatica P (2013) QSAR prediction of the competitive interaction of emerging halogenated pollutants with human transthyretin. SAR QSAR Environ Res 24:599–615
- Qin WC, Su LM, Zhang XJ, Qin HW, Wen Y, Guo Z, Sun FT, Sheng LX, Zhao YH, Abraham MH (2010) Toxicity of organic pollutants to seven aquatic organisms: effect of polarity and ionization. SAR QSAR Environ Res 21:389–401
- Russom CL, Bradbury SP, Broderius SJ, Hammermeister DE, Drummond RA (1997) Predicting modes of toxicity action from chemical structure: acute toxicity in the fathead minnow (*Pimephales Promelas*). Environ Toxicol Chem 16:948–967
- Saka M, Tada N, Kamata Y (2017) Chronic toxicity of 1,3,5-triazine herbicides in the postembryonic development of the western clawed frog *Silurana tropicalis*. Ecotoxicol Environ Saf 147:373–381
- Schnur DM, Grieshaber MV, Bowen JP (1991) Development of an internal searching algorithm for parameterization of the MM2/MM3 force fields. Comput Chem 12:849
- Shi JQ, Cheng J, Wang FY, Flamm A, Wang ZY, Yang X, Gao SX (2012) Acute toxicity and n-octanol/water partition coefficients of substituted thiophenols: Determination and QSAR analysis. Ecotoxicol Environ Saf 78:134–141
- Stewart JJP (1989) Optimization of parameters for semiempirical methods I. Method Comp Chem 10:209–220
- Su LM, Liu X, Wang Y, Li JJ, Wang XH, Sheng LX, Zhao YH (2014) The discrimination of excess toxicity from baseline effect: Effect of bioconcentration. Sci Total Environ 484:137–145
- Terasaki M, Makino M, Tatarazako N (2009) Acute toxicity of parabens and their chlorinated by-products with *Daphnia magna* and *Vibrio fischeri* bioassays. J Appl Toxicol 29:242–247
- Tropsha A, Gramatica P, Gombar VK (2003) The importance of being earnest: validation is the absolute essential for successful application and interpretation of QSPR models. QSAR Comb Sci 22:69–77
- Tsakovska I, Lessigiarska I, Netzeva T, Worth AP (2008) A mini review of mammalian toxicity (Q)SAR models. QSAR Comb Sci 27:41–48
- Tugcu G, Sacan MT (2018) A multipronged QSAR approach to predict algal low-toxic-effect concentrations of substituted phenols and anilines. J Hazard Mater 344:893–901
- Verhaar HJM, van Leeuwen CJ, Hermens JLM (1992) Classifying environmental pollutants. Chemosphere 25:471–491
- Vighi M, Migliorati S, Monti GS (2009) Toxicity on the luminescent bacterium *Vibrio fischeri* (Beijerinck). I: QSAR equation for narcotics and polar narcotics. Ecotoxicol Environ Saf 72:154–161
- Villa S, Migliorati S, Monti GS, Vighi M (2012) Toxicity on the luminescent bacterium *Vibrio fischeri* (Beijerinck). II: Response to complex mixtures of heterogeneous chemicals at low levels of individual components. Ecotoxicol Environ Saf 86:93–100
- Wang C, Wei ZB, Wang LS, Sun P, Wang ZY (2015a) Assessment of bromide-based ionic liquid toxicity toward aquatic organisms and QSAR analysis. Ecotoxicol Environ Saf 115:112–118
- Wang Y, Chen JW, Yang XH, Lyakurwa F, Li XH, Qiao XL (2015b) In silico model for predicting soil organic carbons normalized sorption coefficient (Koc) of organic chemicals. Chemosphere 119:438–444
- Wang XH, Yu Y, Huang T, Qin WC, Su LM, Zhao YH (2016a) Comparison of toxicities to *Vibrio fischeri* and fish based on discrimination of excess toxicity from baseline level. PLoS One 11: e0150028
- Wang Y, Yang XH, Wang JY, Cong Y, Mu JL, Jin F (2016b) A DFT-based toxicity QSAR study of aromatic hydrocarbons to *Vibrio fischeri*: Consideration of aqueous freely dissolved concentration. J Hazard Mater 308:149–156
- Wang S, Yan LC, Zheng SS, Li TT, Fan LY, Huang T, Li C, Zhao YH (2019) Toxicity of some prevalent organic chemicals to tadpoles and comparison with toxicity to fish based on mode of toxic action. Ecotoxicol Environ Saf 167:138–145
- Wen Y, He J, Liu X, Li JJ, Zhao YH (2012) Linear and non-linear relationships between bioconcentration and hydrophobicity: Theoretical consideration. Environ Toxicol Pharmacol 34:200–208
- Zhang XJ, Qin HW, Su LM, Qin WC, Zou MY, Sheng LX, Zhao YH, Abraham MH (2010) Interspecies correlations of toxicity to eight aquatic organisms: Theoretical considerations. Sci Total Environ 408:4549–4555
- Zhang XJ, Qin WC, He J, Wen Y, Su LM, Sheng LX, Zhao YH (2013) Discrimination of excess toxicity from narcotic effect: Comparison of toxicity of class-based organic chemicals to *Daphnia magna* and *Tetrahymena pyriformis*. Chemosphere 93:397–407
- Zhang SN, Su LM, Zhang XJ, Li C, Qin WC, Zhang DM, Liang XX, Zhao YH (2019) Combined Toxicity of Nitro-Substituted Benzenes and Zinc to Photobacterium Phosphoreum: Evaluation and QSAR Analysis. Int J Environ Res Public Health 16:1041–1053
- Zhao YH, Cronin MTD, Dearden JC (1998a) Quantitative structure-activity relationships of chemicals acting by non-polar narcosis-theoretical considerations. Quant.Struct.-Act.Relat 17(2):131–138
- Zhao YH, Ji GD, Cronin MTD, Dearden JC (1998b) QSAR study of the toxicity of benzoic acids to *Vibrio fischeri*, *Daphnia magna* and carp. Sci Total Environ 216:205–215
- Zhao YH, Yuan X, Su LM, Qin WC, Abraham MH (2009) Classification of toxicity of phenols to *Tetrahymena pyriformis* and subsequent derivation of QSARs from hydrophobic, ionization and electronic parameters. Chemosphere 75:866–871
- Zhu D, Li TT, Zheng SS, Yan LC, Wang Y, Fan LY, Li C, Zhao YH (2018) Comparison of modes of action between fish and zebrafish embryo toxicity for baseline, less inert, reactive and specifically-acting compounds. Chemosphere 213:414–422
- Zvinavashe E, Du T, Griff T, van den Berg HHJ, Soffers AEMF, Vervoort J, Murk AJ, Rietjens IMCM (2009) Quantitative structure-activity relationship modeling of the toxicity of organothiophosphate pesticides to *Daphnia magna* and *Cyprinus carpio*. Chemosphere 75: 1531–1538

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.