Chemosphere 198 (2018) 122–129

Contents lists available at ScienceDirect

Chemosphere

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

QSAR prediction of additive and non-additive mixture toxicities of antibiotics and pesticide

Li-Tang Qin [a, b, c], Yu-Han Chen [a], Xin Zhang [a], Ling-Yun Mo [b, c, **], Hong-Hu Zeng [a, b, c], Yan-Peng Liang [a, b, c, *]

[a] College of Environmental Science and Engineering, Guilin University of Technology, Guilin, 541004, China
[b] Guangxi Key Laboratory of Environmental Pollution Control Theory and Technology, Guilin University of Technology, Guilin, 541004, China
[c] Collaborative Innovation Center for Water Pollution Control and Water Safety in Karst Area, Guilin University of Technology, Guilin, 541004, China

HIGHLIGHTS

- QSAR model was developed based on binary and multi-component mixture.
- The proposed model presents high predictive ability for additive and non-additive toxicity.
- The predictive ability of QSAR model superior than concentration additive and independent action models.
- QSAR approach can fill the gaps in predicting non-additive toxicity of mixture.

ARTICLE INFO

Article history:
Received 5 October 2017
Received in revised form
1 January 2018
Accepted 27 January 2018
Available online 6 February 2018

Handling Editor: Frederic Leusch

ABSTRACT

Antibiotics and pesticides may exist as a mixture in real environment. The combined effect of mixture can either be additive or non-additive (synergism and antagonism). However, no effective predictive approach exists on predicting the synergistic and antagonistic toxicities of mixtures. In this study, we developed a quantitative structure-activity relationship (QSAR) model for the toxicities (half effect concentration, EC_{50}) of 45 binary and multi-component mixtures composed of two antibiotics and four pesticides. The acute toxicities of single compound and mixtures toward *Aliivibrio fischeri* were tested. A genetic algorithm was used to obtain the optimized model with three theoretical descriptors. Various internal and external validation techniques indicated that the coefficient of determination of 0.9366 and root mean square error of 0.1345 for the QSAR model predicted that 45 mixture toxicities presented additive, synergistic, and antagonistic effects. Compared with the traditional concentration additive and independent action models, the QSAR model exhibited an advantage in predicting mixture toxicity. Thus, the presented approach may be able to fill the gaps in predicting non-additive toxicities of binary and multi-component mixtures.

1. Introduction

The contamination of water environment with antibiotics and pesticides has drawn worldwide attention (Kemper, 2008; Gao et al., 2012; Topp et al., 2016). Contaminants are frequently exposed as a mixture derived from many different sources at varying doses. Considering the combination of mixtures is extremely large, testing the combined effects of mixtures by an experimental approach is unfeasible. Although two prominent reference models, namely, concentration addition (CA) (Loewe and Muischnek, 1926) and independent action (IA) (Bliss, 1939), are widely used in predicting the combined effect of chemical mixtures, the effective method for predicting interactive mixture toxicity is still rare. Thus, approximate methods for predicting mixture toxicity with synergism or antagonism should be developed (Iwasaki and Gauthier, 2016).

CA assumes that mixture components have the same or similar mode of action (MOA) (Loewe and Muischnek, 1926), whereas IA assumes they have different or dissimilar MOA (Bliss, 1939). A review regarding the use of IA and CA for predicting the joint effect of

* Corresponding author. College of Environmental Science and Engineering, Guilin University of Technology, Guilin, 541004, China.
** Corresponding author. Guangxi Key Laboratory of Environmental Pollution Control Theory and Technology, Guilin University of Technology, Guilin, 541004, China.
E-mail addresses: molingyun123@126.com (L.-Y. Mo), ypliang1980@163.com (Y.-P. Liang).

https://doi.org/10.1016/j.chemosphere.2018.01.142
0045-6535/

binary mixtures showed that approximately 20% of 158 mixtures were adequately predicted by using IA alone and 10% were adequately predicted by CA only (Cedergreen et al., 2008). Another research paper also showed that the toxicity of heavy metals and ionic liquids on photobacterium Q67 was underestimated by CA and IA models (Ge et al., 2014). These results indicated that CA and IA could not adequately describe all mixture toxicities. This condition is particularly true when synergism and antagonism are presented in a mixture. Certainly, CA and IA can only be used to predict the additive effect of mixtures. Although two-step prediction model (TSP) successfully predicted the toxicity of mixture contaminants (Altenburger et al., 2005; Ra et al., 2006; Tang and Escher, 2014), the data gaps of MOA from different species, varying doses, and different exposure time limited its application (Hertzberg and Macdonell, 2002; Syberg et al., 2008). By contrast, TSP was simplified to CA or IA model when it was used to predict binary mixtures. Integrated models that combined CA with IA were proposed for evaluating the toxicity of additive mixtures (Rider and LeBlanc, 2005; Qin et al., 2011). Linear CA (LCA) and Linear IA (LIA) for predicting additive and non-additive mixture toxicities were developed (Qin et al., 2015, 2017). Although LCA and LIA models are able to predict the combined toxicity of non-additive mixtures, the evidence supporting its predictive capabilities should still be tested.

Quantitative structure-activity relationship (QSAR) has been extensively used in predicting the activity of single chemicals. QSAR has also been used to predict the concentration effects of components in mixtures from the combined effects and defined concentration ratios of mixture components (Altenburger et al., 2003). Several QSAR models have been proposed for the combined effects of chemical mixtures from molecular descriptors that were calculated as composite properties based on the concentration ratios of mixture components (Ajmani et al., 2006; Wang et al., 2006; Lu et al., 2007; Toropova et al., 2012; Kim et al., 2013; Yao et al., 2013; Gaudin et al., 2015; Chang et al., 2016; Sobati et al., 2016; Soltanpour et al., 2016). However, the application of QSAR model on joint effects of mixtures is limited because it is mainly used in binary mixtures and additive toxicities of mixtures. Few models developed by Kim et al. (2013) and Sobati et al. (2016) were applied to multi-component mixtures. In addition, the QSAR model for predicting the toxicities of synergistic and antagonistic mixtures is not considered. Therefore, whether a QSAR model can be used to predict the non-additive effect of mixtures that can occur between chemicals and multi-component mixtures having three or more components should be tested.

In this study, we developed a predictive QSAR model for predicting the additive and non-additive toxicities of binary and multi-component mixtures. We tested the acute toxicities of antibiotics, such as chloramphenicol (CHL) and tetracycline hydrochloric (TET), and pesticides, such as dichlorvos (DIC), trichlorfon (TRC), metribuzine (MET), and linuron (LIN), and their binary, ternary, and quaternary mixtures to *Aliivibrio fischeri* at exposure time of 15 min. The basis for selecting these chemicals is because of the co-exposure of CHL and TET in the environment (Watkinson et al., 2009; Wu et al., 2016) and the wide use of MET (Christin et al., 2004) and LIN (Webster et al., 2015) in agriculture and their existence in surface water.

2. Materials and methods

2.1. Experimental design and toxicity test

Six testing chemicals, namely, TET, CHL, DIC, TRC, MET, and LIN with purity higher than 98% were purchased from Dr. Ehrenstorfer GmbH (Germany). The detailed information of chemicals is given in Table S1 (Supplementary data). The direct equipartition ray design (EquRay) (Dou et al., 2011) was used to design a binary mixture system (consisted of the same components) to obtain five mixture rays (named R1, R2, R3, R4, and R5) with different equivalent concentration ratios (p_i) (Liu et al., 2016b). Equivalent-effect concentration ratio (EECR) (Liu et al., 2016a) was used to design the ternary and quaternary mixture rays. Mixture system is defined as a pool of a number of mixtures having various concentration compositions, and mixture ray is defined as a series of mixtures having the same concentration ratio of components (Liu et al., 2016a). Five EECR mixture rays for each mixture system were designed at the fixed-concentration ratios of 50% concentration effect (EC_{50}, mol/L), 30% concentration effect (EC_{30}, mol/L), 20% concentration effect (EC_{20}, mol/L), 10% concentration effect (EC_{10}, mol/L), and 5% concentration effect (EC_5, mol/L), respectively. A total of 45 mixture rays (25 binary, 15 ternary, and five quaternary mixture rays) composed of antibiotics and pesticides was designed. The concentration ratios (p_i) of each chemical for 45 mixture rays are given in Table S2 (Supplementary data).

The freeze-dried luminescent *Aliivibrio fischeri*, which was purchased from Beijing Hamamatsu Photon Techniques INC. (Beijing, China), was used as an organism. The details of bacteria culture and toxicity test are given in Supplementary data. The luminescence inhibitions of individual chemicals and mixtures toward *Aliivibrio fischeri* were tested based on the method of microplate toxicity analysis (Zhang et al., 2008). The relative light unit of *Aliivibrio fischeri* was determined using a Synergy 2 (BioTek) apparatus after 15 min exposure at 22 ± 1 °C.

2.2. Concentration-effect fitting and toxicity data

The experimental concentration–effect curves of single compound and mixtures are described with log-logistic function (Eq. (1)) (Ritz, 2010), where parameter c (the response at infinite dose) and d (the mean response of the untreated control) in the original four-parameter log-logistic function are set to 1 and 0, respectively. The coefficient of determination (R^2) and parameters b and e for log-logistic function are given in Table 1. The EC_{50} value, which can be calculated from the inverse log-logistic function, is used to represent the mixture toxicity. The negative logarithmic EC_{50} ($pEC_{50,Obs}$) values for 45 mixtures are given in Table 1. The pEC_{50} values cover a range of 1.86–4.26. The 95% confidence intervals (CI) of pEC_{50} is also calculated and listed in Table 1.

$$y = 1 - \frac{1}{1 + (x/e)^b}, \tag{1}$$

where b is proportional to the slope around half effect concentration (EC_{50}), and e is EC_{50}. All the units of concentration effect are expressed in mol/L.

2.3. Toxicity interaction analysis

CA (Loewe and Muischnek, 1926) and IA (Bliss, 1939) were used as the reference addition models for analyzing the toxicity interaction of mixtures. The mathematical formulation of CA model, which is used to predict EC_{50} values of mixtures, is given in Eq. (2) (Altenburger et al., 2004):

$$EC_{50,CA} = \left(\sum_{i}^{n} \frac{p_i}{EC_{50,i}} \right)^{-1}, \tag{2}$$

where $EC_{50,CA}$ is the concentration of mixture provoking 50% effect predicted by CA, $EC_{50,i}$ is the concentration of component i provoking the same effect (50%) when applied individually (Table S1 in

Table 1
The observed and calculated pEC_{50} of 45 mixtures to *Aliivibrio fischeri* and the selected descriptors[a].

No.	Mixture rays	*b*	*e*	R^2	$pEC_{50,Obs}$	$pEC_{50,lower}$	$pEC_{50,upper}$	$pEC_{50,QSAR}$	$pEC_{50,CA}$	$pEC_{50,IA}$	Descriptors		
											RDF035m	*HATSs*	*H-047*
1[a]	TET-TRC (R1)	0.9004	1.16E-03	0.9921	2.94	2.81	3.06	2.55	3.04	3.13	15.676	54.828	6.369
2	TET-TRC (R2)	0.9719	2.61E-03	0.9939	2.58	2.49	2.68	2.41	2.73	2.86	16.690	55.411	6.156
3	TET-TRC (R3)	0.9303	4.26E-03	0.9959	2.37	2.30	2.44	2.37	2.54	2.67	17.059	55.609	6.079
4[a]	TET-TRC (R4)	1.2098	9.08E-03	0.9973	2.04	2.00	2.08	2.35	2.41	2.54	17.229	55.698	6.043
5	TET-TRC (R5)	1.2059	1.28E-02	0.9971	1.86	1.84	1.93	2.33	2.27	2.36	17.367	55.769	6.015
6	TET-DIC (R1)	1.1724	3.48E-04	0.9983	3.47	3.41	3.51	3.43	3.40	3.46	1.264	51.399	6.520
7	TET-DIC (R2)	1.1461	4.46E-04	0.9980	3.35	3.30	3.41	3.33	3.25	3.31	1.254	52.176	6.231
8	TET-DIC (R3)	1.0988	5.02E-04	0.9985	3.30	3.25	3.35	3.29	3.18	3.23	1.275	52.447	6.121
9	TET-DIC (R4)	1.0790	5.45E-04	0.9984	3.26	3.21	3.31	3.27	3.14	3.17	1.293	52.598	6.058
10	TET-DIC (R5)	1.0436	5.78E-04	0.9986	3.24	3.19	3.28	3.26	3.11	3.13	1.304	52.682	6.022
11	TET-LIN (R1)	0.9392	6.84E-05	0.9936	4.16	4.05	4.28	4.05	4.22	4.31	16.398	25.791	9.829
12	TET-LIN (R2)	0.9392	5.51E-05	0.9915	4.26	4.13	4.38	4.25	4.19	4.31	10.522	29.726	9.660
13	TET-LIN (R3)	1.0315	5.67E-05	0.9978	4.25	4.19	4.30	4.36	4.16	4.30	6.650	32.684	9.493
14	TET-LIN (R4)	1.0734	5.71E-05	0.9973	4.24	4.18	4.30	4.39	4.14	4.26	4.490	35.103	9.327
15[a]	TET-LIN (R5)	0.9836	5.94E-05	0.9989	4.23	4.19	4.26	4.34	4.11	4.20	3.870	37.130	9.163
16	CHL-DIC (R1)	0.9120	7.10E-04	0.9885	3.15	3.01	3.29	3.33	2.99	3.08	3.739	63.522	7.234
17	CHL-DIC (R2)	0.9460	6.15E-04	0.9954	3.21	3.12	3.30	3.36	3.03	3.11	1.832	59.642	6.790
18[a]	CHL-DIC (R3)	0.9683	5.58E-04	0.9972	3.25	3.19	3.32	3.34	3.06	3.12	1.291	57.061	6.495
19	CHL-DIC (R4)	1.0088	5.26E-04	0.9987	3.28	3.23	3.33	3.31	3.08	3.12	1.179	55.223	6.285
20	CHL-DIC (R5)	1.0315	5.23E-04	0.9990	3.28	3.24	3.32	3.28	3.09	3.11	1.219	53.811	6.124
21	CHL-LIN (R1)	0.7972	9.69E-04	0.9720	3.01	2.79	3.23	3.07	2.96	3.10	11.157	69.919	8.018
22	CHL-LIN (R2)	0.8684	7.70E-04	0.9858	3.11	2.97	3.25	3.11	3.07	3.23	10.669	69.485	8.043
23	CHL-LIN (R3)	0.8292	6.09E-04	0.9819	3.22	3.05	3.37	3.17	3.19	3.36	9.925	68.783	8.082
24[a]	CHL-LIN (R4)	0.9325	4.24E-04	0.9933	3.37	3.28	3.46	3.28	3.36	3.50	8.685	67.482	8.151
25[a]	CHL-LIN (R5)	0.9563	2.46E-04	0.9981	3.61	3.56	3.66	3.52	3.59	3.70	6.291	64.233	8.308
26	TET-CHL-MET (EE50)	0.7596	4.55E-04	0.9782	3.34	3.12	3.56	3.23	3.07	3.13	4.546	55.533	6.645
27[a]	TET-CHL-MET (EE30)	0.8141	4.13E-04	0.9911	3.35	3.24	3.52	3.18	3.08	3.13	4.363	50.912	6.223
28	TET-CHL-MET (EE20)	1.0362	8.66E-04	0.9869	3.06	2.93	3.18	3.11	3.08	3.12	4.578	46.642	5.831
29	TET-CHL-MET (EE10)	1.0349	7.51E-04	0.9910	3.12	3.02	3.23	3.01	3.09	3.11	5.276	39.974	5.283
30[a]	TET-CHL-MET (EE5)	1.0779	8.38E-04	0.9924	3.08	2.99	3.17	3.36	3.55	3.56	4.074	46.747	6.535
31	TET-CHL-LIN (EE50)	1.6319	2.39E-04	0.9272	3.62	3.26	3.99	3.30	3.40	3.60	8.850	67.404	8.224
32[a]	TET-CHL-LIN (EE30)	0.8416	4.78E-04	0.9845	3.32	3.16	3.47	3.37	3.91	3.99	8.196	66.504	8.295
33[a]	TET-CHL-LIN (EE20)	1.0653	3.14E-04	0.9549	3.50	3.24	3.74	3.44	3.56	3.75	7.620	65.563	8.370
34[a]	TET-CHL-LIN (EE10)	0.8315	6.20E-04	0.9757	3.21	3.00	3.41	3.58	3.69	3.86	6.604	63.343	8.535
35	TET-CHL-LIN (EE5)	1.0700	1.69E-04	0.9777	3.77	3.60	3.93	3.75	3.83	3.99	5.680	59.931	8.749
36[a]	TET-CHL-TRC (EE50)	1.0972	1.12E-03	0.9937	2.95	2.87	3.03	2.67	2.61	2.81	12.800	58.134	6.428
37	TET-CHL-TRC (EE30)	1.0885	2.68E-03	0.9959	2.57	2.51	2.64	2.67	2.66	2.85	12.925	57.948	6.428
38	TET-CHL-TRC (EE20)	1.1421	2.55E-03	0.9940	2.59	2.51	2.67	2.66	2.72	2.90	13.021	57.709	6.430
39	TET-CHL-TRC (EE10)	0.8161	1.67E-03	0.9780	2.78	2.58	2.97	2.68	2.84	3.00	13.089	57.351	6.466
40	TET-CHL-TRC (EE5)	0.9049	1.30E-03	0.9890	2.89	2.75	3.02	2.71	2.98	3.12	12.967	56.868	6.537
41[a]	TET-CHL-MET-LIN (EE50)	0.9899	5.06E-04	0.9958	3.30	3.22	3.37	3.29	3.20	3.34	4.221	55.075	6.765
42	TET-CHL-MET-LIN (EE30)	1.0688	5.28E-04	0.9967	3.28	3.22	3.34	3.25	3.20	3.33	4.065	50.580	6.366
43	TET-CHL-MET-LIN (EE20)	1.1153	6.39E-04	0.9949	3.19	3.12	3.26	3.18	3.19	3.30	4.303	46.426	5.995
44	TET-CHL-MET-LIN (EE10)	1.2975	6.24E-04	0.9943	3.20	3.13	3.27	3.08	3.17	3.27	5.003	39.979	5.457
45[a]	TET-CHL-MET-LIN (EE5)	1.0849	4.77E-04	0.9914	3.32	3.21	3.43	2.99	3.16	3.24	5.732	34.112	5.012

[a] Refers to the test set; Mixture rays: the concentration ratios (p_i) of each chemical for the 45 mixture rays were given in Table S2 (Supplemental data); TET: Tetracycline hydrochloric; CHL: Chloramphenicol; MET: Metribuzine; TRC: Trichlorfon; DIC: dichlorvos; LIN: Linuron; *RDF035m*: Radial Distribution Function-035/weighted by mass; *HATSs*: leverage-weighted total index/weighted by I-state; *H-047*: H attached to $C^1(sp3)/C^0(sp2)$; *b* is proportional to the slope around half effect concentration (EC_{50}, mol/L), and *e* is EC_{50} (mol/L) in the log-logistic equation.; $pEC_{50,upper}$ and $pEC_{50,lowe}$ are the 95% upper and lower observed confidence limits of $pEC_{50,obs}$, respectively; $pEC_{50,QSAR}$ is the pEC_{50} value predicted by the QSAR model, $pEC_{50,CA}$ is the pEC_{50} value predicted by the CA model, $pEC_{50,IA}$ is the pEC_{50} value predicted by the IA model.

Supplementary data), and p_i is the concentration ratio of the *i*th component in the mixture (Table S2 in Supplementary data). The negative logarithm of $EC_{50,CA}$ ($pEC_{50,CA}$) values calculated from the CA model is listed in Table 1.

The mathematical equation of IA model can be defined as (Bliss, 1939; Altenburger et al., 2004):

$$E(c_{mix}) = 1 - \prod_{i=1}^{n}(1 - E(c_i)), \quad (3)$$

where $E(c_{mix})$ is the total effect of mixture, c_{mix} is the concentration of mixture, $E(c_i)$ is the effect of *i*th component, and c_i is the concentration of *i*th component.

The toxicity interactions of mixtures were evaluated by the results of CA ($pEC_{50,CA}$) and IA ($pEC_{50,IA}$) predictions and 95% CI of $EC_{50,Obs}$. The additive effect is presented in the mixture if $pEC_{50,CA}$ or $pEC_{50,IA}$ value ranges between the lower ($pEC_{50,lower}$) and upper ($pEC_{50,upper}$) values of 95% CI. $pEC_{50,CA}$ or $pEC_{50,IA}$ less than $pEC_{50,lower}$ indicates the synergistic effect, whereas $pEC_{50,CA}$ or $pEC_{50,IA}$ larger than $pEC_{50,upper}$ indicates the antagonistic effect. Synergistic and antagonistic effects are considered as non-additive effect.

2.4. Calculation and screening of mixture descriptors

The mixture descriptors can be calculated from the molecular descriptors of individual chemicals. All the molecular structures of six individual chemicals were preoptimized using the MM2 molecular mechanics method. Dragon 7.0 software was used to calculate 5270 molecular descriptors for each chemical. These descriptors included constitutional descriptors, topological descriptors, connectivity indices, information indices, 2D autocorrelations, and atom-centered fragments. The original 5270

molecular descriptors for six chemicals were refined by the following principles (Qin et al., 2013a, 2013b): 1) descriptors with standard deviation less than 0.0001 were excluded, 2) descriptors with at least one missing value were deleted, 3) descriptors with (abs) pair correlation larger than or equal to 0.8 were excluded, and 4) descriptors with Pearson correlation coefficient ($|r|$) between descriptors and $pEC_{50,Obs}$ lower than 0.3 were deleted. The remaining 2009 descriptors were obtained to calculate the molecular descriptors of mixtures based on Eqs. (4)–(11), which were successfully used to characterize the molecular structure information on mixture (Altenburger et al., 2003; Gaudin et al., 2015; Sobati et al., 2016):

$$D_{mix,i} = \sum_{i=1}^{n} p_i x_i, \tag{4}$$

$$D_{mix,i} = \left(\sum_{i=1}^{n} p_i x_i \right)^2, \tag{5}$$

$$D_{mix,i} = \sqrt{\sum_{i=1}^{n} (p_i x_i)^2}, \tag{6}$$

$$D_{mix,i} = \sum_{i=1}^{n} \sqrt{p_i} x_i, \tag{7}$$

$$D_{mix,i} = \sum_{i=1}^{n} \sqrt{|x_i|}, \tag{8}$$

$$D_{mix,i} = \sum_{i=1}^{n} p_i (x_i)^2, \tag{9}$$

$$D_{mix,i} = \sum_{i=1}^{n} (p_i)^2 x_i, \tag{10}$$

$$D_{mix,i} = \sqrt[3]{\sum_{i=1}^{n} (p_i)^3 x_i}, \tag{11}$$

where $D_{mix,i}$ is the descriptor of one mixture, i is the ith component in a mixture, and x_i is the descriptor of the ith component.

A pool of 2009×8 descriptors resulted from eight equations (Eqs. (4)–(11)) was refined by the same principles (Qin et al., 2013a, 2013b), which was used for the abovementioned single chemicals. Consequently, the number of molecular descriptors for mixtures was reduced to 111, and these descriptors were used in the next step of feature selection.

2.5. Feature selection and development of QSAR model

The entire data set was divided into training and test sets, where 70% of the data set were randomly selected for the training set and the remaining 30% for the test set. The splitting of dataset was based on the following rules: 1) the additive, synergistic, and antagonistic effects of mixtures were presented in the training and test sets; 2) the binary, ternary, and quaternary mixtures should be included in the training and test sets.

The genetic algorithm (GA), which was performed in QSARINS software (Gramatica et al., 2013, 2014), was employed to select the best variables from the remaining descriptors. A list of multiple linear regression (MLR) model with 5 maximum variables was obtained. However, the GA-MLR model may not be the real model because the variables in a model may present sign change problem in the complete, training, and test sets (Kiralj and Ferreira, 2010). The sign change problem was tested based on the following conditions (Kiralj and Ferreira, 2010):

1) Pearson correlation coefficient for the complete (r_c), training (r_t), and test (r_e) sets equal or greater than 0.3: $|r_c|$ and $|r_t| \geq 0.3$;
2) the normalized regression coefficient of the descriptor for the complete (β_c) and training (β_t) sets equal or greater than 0.001: $|\beta_c|$ and $|\beta_t| \geq 0.001$;
3) absence of sign change problem: $\text{sign}(r_c) = \text{sign}(r_t) = \text{sign}(r_e)$; $\text{sign}(r_c) = \text{sign}(\beta_c) = \text{sign}(\beta_t)$.

2.6. Model validation

The statistical parameters for modeling and internal and external validations were adopted for evaluating the good of fitness, stability, and predictive capability of the QSAR model (Qin et al., 2013a). The quality parameters for modeling included R^2, adjusted coefficient of determination (R^2_{adj}), root mean square error ($RMSE$) in fitting, and F-value (F).

Internal validations were performed by leave-one-out (LOO) and leave-many-out (LMO) cross-validations (Q^2_{LOO} and Q^2_{LMO}), bootstrapping (R^2_{bstr} and Q^2_{bstr}) (Kiralj and Ferreira, 2009), and y-randomization test (R^2_{Yscr} and Q^2_{Yscr}) (Tropsha et al., 2003). In LMO cross-validation, M = 2–9 was used, and averaged Q^2_{LMO} was obtained. In bootstrapping, the complete data set was randomly split several times (100 runs) into training and test sets, and the averaged R^2_{bstr} and Q^2_{bstr} values were obtained (Qin et al., 2013a; Mo et al., 2015). In the y-randomization test, the dependent-variable vector was randomly scrambled for 1000 iterations, and new QSAR models were developed using the original independent-variable matrix (Tropsha et al., 2003). This approach was based on the absolute value of Pearson correlation coefficient (r) between the original vector $\boldsymbol{y}$ and randomized vectors $\boldsymbol{y}$. Two regression lines (r–R^2_{yrand} and r–Q^2_{yrand}) were drawn for the randomized models. The intercepts of equations obtained in two regression lines should be lower than 0.3 for R^2_{yrand} and 0.05 for Q^2_{yrand} (Eriksson et al., 2003).

External validation was evaluated by a test set. The parameters Q^2_{F1} (Shi et al., 2001), Q^2_{F2} (Schüürmann et al., 2008), Q^2_{F3} (Consonni et al., 2010), CCC (Chirico and Gramatica, 2011), $\overline{r^2_m}$, and Δr^2_m (Ojha et al., 2011) were used as the measures for the predictive power of a QSAR model. The proposed parameters by Golbraikh and Tropsha were also applied for the external validation criteria (Golbraikh and Tropsha, 2002): slope of regression line over external data (k and k'), coefficients of determination between predicted and observed activities (R^2_0), and coefficients of determination between observed and predicted activities (R'^2_0).

The validation criteria thresholds for the statistical parameters as expressed as follows:

1) $R^2 > 0.7$, Q^2_{LOO} and $Q^2_{LMO} > 0.6$, Q^2_{F1}, Q^2_{F2}, and $Q^2_{F3} > 0.7$, the difference between R^2 and Q^2_{LOO} is smaller than 0.1 (Chirico and Gramatica, 2011),
2) $\overline{r^2_m} > 0.65$,
3) $CCC > 0.85$, (Chirico and Gramatica, 2011), and
4) criteria recommended by Golbraikh and Tropsha (2002): $(R^2 - R^2_0)/R^2 < 0.1$ or $(R^2 - R'^2_0)/R^2 < 0.1$, $0.85 \leq k \leq 1.15$ or $0.85 \leq k' \leq 1.15$.

2.7. Applicability domain

The application domain of a QSAR model was defined by the leverage approach from the hat matrix (h_i in (Eq. (12)), which was calculated from the molecular descriptors of the mixture (Tropsha et al., 2003; Qin et al., 2013a) and by the identification of mixtures with LOO cross-validated standardized residuals greater than 2.5 standard deviation units. An outlier in the QSAR model is defined as h_i value larger than the warning leverage (h^*) and LOO standardized residuals greater than 2.5, which is graphically depicted in the Williams plot. Warning leverage h^* is fixed at $(3k)/n$, where k is the number variables in the model, and n is the number of mixtures used in the modeling.

$$h_i = x_i^T \left(X^T X\right)^{-1} x_i \quad (\mathrm{i} = 1,\ldots, \mathrm{n}), \tag{12}$$

where x_i is the descriptor row vector of the query compound; X is the ink matrix of k model descriptor values for n training set compounds. The superscript "T" refers to the transpose of the matrix/vector.

3. Results and discussion

3.1. Development of GA-MLR model

On the basis of the results of CA prediction, 18 mixture rays out of the entire mixture data set presented additive effect, nine mixture rays presented antagonistic effects, and 18 mixture rays presented synergistic effects. Therefore, 18 additive and 27 non-additive mixtures were classified from 45 mixture rays (25 binary and 20 multi-component mixture rays). By contrast, IA predictions showed that 27 mixtures presented non-additive effect.

The entire data set was randomly split into training (70% of the entire data set) and test sets (30% of the entire data set). The mixture rays used in the test set are marked in Table 1. GA was used to select the optimum descriptors from the set of all available mixture descriptors. First, the MLR models with one to five variables that obtained high R^2 and Q^2_{LOO} values were developed based on the training set (Table S3, Supplementary data). The real GA-MLR model, where the variables were without sign change problem, were obtained. The obtained 4-variable and 5-variable models presented sign change problem, and the 4-variable model did not obviously improve the precision of the 3-variable model. The MLR model with three descriptors of radial distribution function-035 (*RDF035m*), leverage-weighted total index (*HATSs*), and H attached to C1(sp3)/C0(sp2) (*H-047*) was considered as the final model. The values of *RDF035m* for mixtures are calculated from Eq. (9), and *HATSs* and *H-047* are calculated from Eq. (11). The three descriptors had low autocorrelation values, where the highest autocorrelation is 0.1131 between *RDF035* and *H-047*.

The QSAR model that used three mixture descriptors is given in Eq. (13). All the statistical parameters for the QSAR models based on the entire data, training, and test sets are shown in Table 2.

$$\begin{aligned} pEC_{50,mix} &= (2.2780\pm0.2173) \\ &-(0.0546\pm0.0050) \times (RDF035m)_{mix} \\ &-(0.0158\pm0.0024) \times (HATSs)_{mix} \\ &-(0.3124\pm0.0214) \times (H-047)_{mix} \end{aligned} \tag{13}$$

As shown in Table 2, the high coefficient of determination ($R^2 = 0.9366$) based on the training set indicated that the developed model presented a high predictive capability. The proposed models showed good fitting capability for the entire data and the training sets.

Table 2
Statistical parameters of the QSAR models.

Parameters[a]	Whole data set	Training set	Test set
n	45	31	14
m	3	3	–
R^2	0.8938	0.9366	–
R^2_{adj}	0.8860	0.9320	–
$RMSE$	0.1663	0.1345	0.2274
F	115.02	132.89	–
Q^2_{LOO}	0.8681	0.9087	–
$RMSE_{cv}$	0.1853	0.1614	–
Q^2_{LMO}	0.8604	0.9094	–
R^2_{bstr}	–	0.8981	–
Q^2_{bstr}	–	0.8600	–
R^2_{ext}	–	–	0.7774
Q^2_{F1}	–	–	0.7483
Q^2_{F2}	–	–	0.7477
Q^2_{F3}	–	–	0.8064
CCC	–	–	0.8802
$\overline{r^2_m}$	–	–	0.6852
Δr^2_m	–	–	0.0833
k	–	–	1.001
k'	–	–	0.9941
R^2_0	–	–	0.7732
R'^2_0	–	–	0.7478

[a] n is the number of sample in the data set, m is the number of variables, R^2 is coefficient of determination, R^2_{adj} is adjusted R^2, $RMSE$ is root mean square error, F is F-value, Q^2_{LOO} is explained variance in prediction leave-one-out, $RMSE_{cv}$ is root mean square error in cross-validation prediction, Q^2_{LMO} is explained variance in prediction leave-many-out, R^2_{bstr} and Q^2_{bstr} are R^2 and Q^2 in bootstrapping test, respectively, R^2_{ext} is external determination coefficient, Q^2_{F1}, Q^2_{F2}, and Q^2_{F3} are variance explained in the test set, CCC is concordance correlation coefficient, $\overline{r^2_m}$ and Δr^2_m are average and delta $\overline{r^2_m}$ values of Roy criteria (Ojha et al., 2011), k and k' are slope of the regression line over external data, and R^2_0 and R'^2_0 are R^2 values in Golbraikh & Tropsha criteria (Golbraikh and Tropsha, 2002).

3.2. Explanation of descriptors in the model

Three descriptors, namely, *RDF035m*, *HATSs*, and *H-047* selected by GA, were used in the developed model. *RDF035m* weighted by mass is a radial distribution function descriptor, which is based on radial distribution function. RDF descriptor indicates the probability distribution of finding an atom in a spherical volume of radius R. *HATSs* is a GETAWAY (Geometry, Topology and Atom-Weights AssemblY) descriptor, in which its chemical structure descriptor is derived from the molecular influence matrix. *HATSs* belongs to H indices and is indicates leverage-weighted total autocorrelation index/weighted by intrinsic state. *H-047* belongs to a block of atom-centered fragments, which are defined as specific atom types in a molecule. The types of one atom in the molecule are described by its adjacent atoms. *H-047* indicates that H^a is attached to C^1(sp3)/C^0(sp2). The superscript 'a' represents the formal oxidation number.

In the final model (Eq. (13)), the Pearson correlation coefficient between pEC_{50} and *H-047* is 0.7190, and the normalized regression coefficient of *H-047* is 0.7473, which is larger than *RDF035m* (−0.5413) and *HATSs* (−0.3259). The result indicates that *H-047* is the main factor influencing the mixture toxicity of two antibiotics and four pesticides. Considering the positive coefficient of *H-047*, the greater *H-047* values, the larger mixture toxicity toward *Aliivibrio fischeri*. The negative normalized regression coefficient of *RDF035m* and *HATSs* descriptors revealed that the lower values of *RDF035m* and *HATSs*, the larger mixture toxicity toward *Aliivibrio fischeri*.

3.3. Internal validation

LOO and LMO cross-validations and bootstrapping, which are

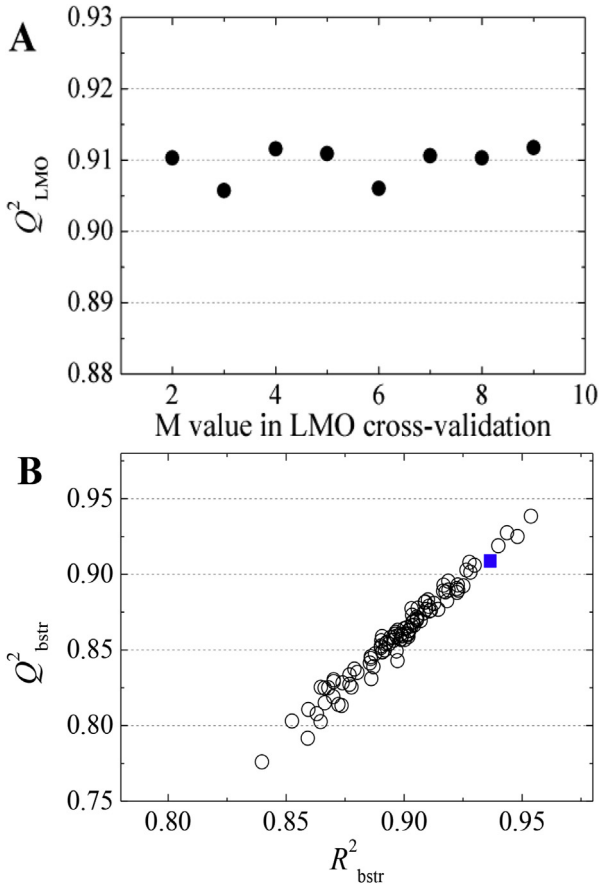

Fig. 1. Results of LMO cross-validation (**A**) and bootstrapping (**B**) for the training set. "■" refers to the model based on training set. In the LMO cross-validation, M = 2–9 (30% of the training set), Q^2_{LMO} is the average value from 100 iterations for each M.

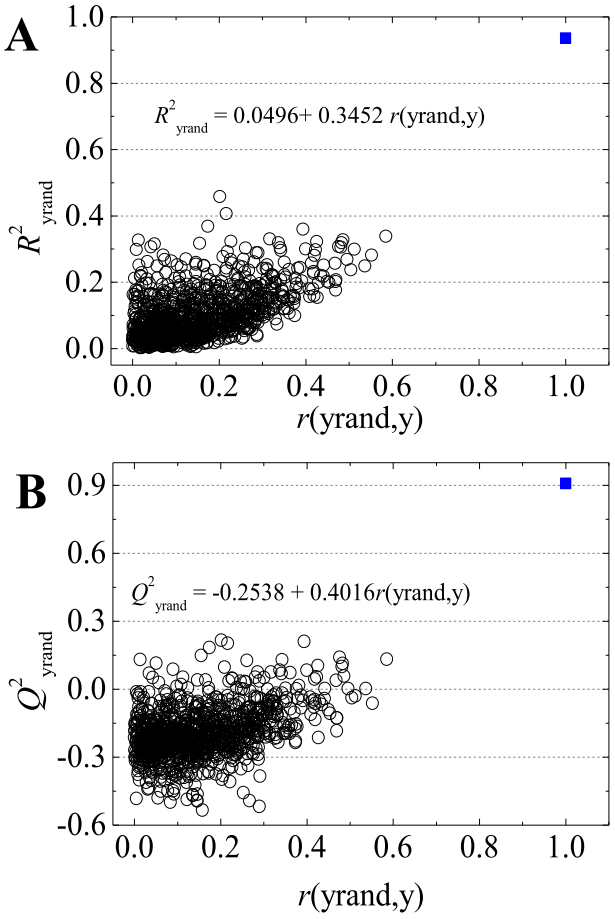

Fig. 2. Scatter plot of statistical parameters R^2_{yrand}, Q^2_{yrand}, and r(yrand, y) for y-randomization test (1000 iterations). "■" refers to the model based on the original training set. r(yrand, y) is the Pearson coefficient between randomize y and original y. The intercepts for regression lines of r–R^2_{yrand} (**A**) and r–Q^2_{yrand} (**B**) are 0.0496 (<0.3) and −0.2538 (<0.05), respectively.

indicators of internal robustness, were applied to test the model. The results in Table 2 and Fig. 1 indicated that the presented model was robust. In addition, y-randomization runs were repeated for 1000 times. Two regression lines, r-R^2_{yrand} and r-Q^2_{yrand}, are presented in Fig. 2. The intercepts of y-randomization test had low values of 0.0496 for r-R^2_{yrand} regression line and 0.2538 for r-Q^2_{yrand} regression line, which were lower than the threshold values of 0.3 and 0.05 (Eriksson et al., 2003), respectively. The results of y-randomization test revealed that the final model presented no chance correlation.

3.4. External validation

The test set was used to validate the external predictive capability of the model based on the training set. The predictive capability of the model was evaluated by external validation parameters Q^2_{F1} (Shi et al., 2001), Q^2_{F2} (Schüürmann et al., 2008), Q^2_{F3} (Consonni et al., 2010), CCC (Chirico and Gramatica, 2011), $\overline{r^2_m}$, Δr^2_m (Ojha et al., 2011), and the criteria recommended by Golbraikh and Tropsha (2002), as shown in Table 2. The reliabilities of variables used in the model are tested, and the result is shown in Table 3. The results satisfied the validation criteria thresholds (Chirico and Gramatica, 2011): Q^2 (Q^2_{F1}, Q^2_{F2}, and Q^2_{F3}) > 0.7, $\overline{r^2_m}$ > 0.65, and CCC > 0.85 along with four passed constraints (Golbraikh and Tropsha, 2002), which revealed that the developed model presented a significant predictive capability.

All internal and external validations clearly prove that the developed model is valid and can be utilized to predict the toxicities of mixtures. The predicted values of pEC_{50} in comparison with the experimental data are listed in Fig. 3A, which indicates that the additive and non-additive mixture toxicities are well predicted by the QSAR model.

Table 3
The sign change, training correlations, external correlation and contributions to the model[a].

Descriptors	r_c	r_t	r_e	β_c	β_t	F_1	F_2	F_3	F_4
RDF035m	−0.47	−0.45	−0.61	−0.53	−0.57	0.46	0.54	0.49	0.52
HATSs	−0.45	−0.48	−0.34	−0.35	−0.37	0.46	0.39	0.40	0.41
H-047	0.68	0.66	0.81	0.72	0.75	0.67	0.74	0.70	0.71

[a]: r_c, r_t, and r_e are the Pearson coefficients for the complete (r_c), training (r_t), and external validation (r_e), respectively; β_c are β_t are regression coefficients of descriptors for the complete and training sets, respectively. F-functions were defined as $F_1 = \text{sign}(r_c r_t)\sqrt{|r_c r_t|}$, $F_2 = \text{sign}(r_c r_e)\sqrt{|r_c r_e|}$, $F_3 = \text{sign}(r_c \beta_c)\sqrt{|r_c \beta_c|}$, $F_4 = \text{sign}(r_c \beta_t)\sqrt{|r_c \beta_t|}$.

3.5. Application domain

The application domain was defined by the leverage approach to verify the predictive reliability. To visualize the application domain of QSAR model, the Williams plot (Fig. 3B) was used, which was the plot of leverage (Hat diagonal) values versus standardized LOO cross-validated residuals (SR). Warning leverage h^* of the model

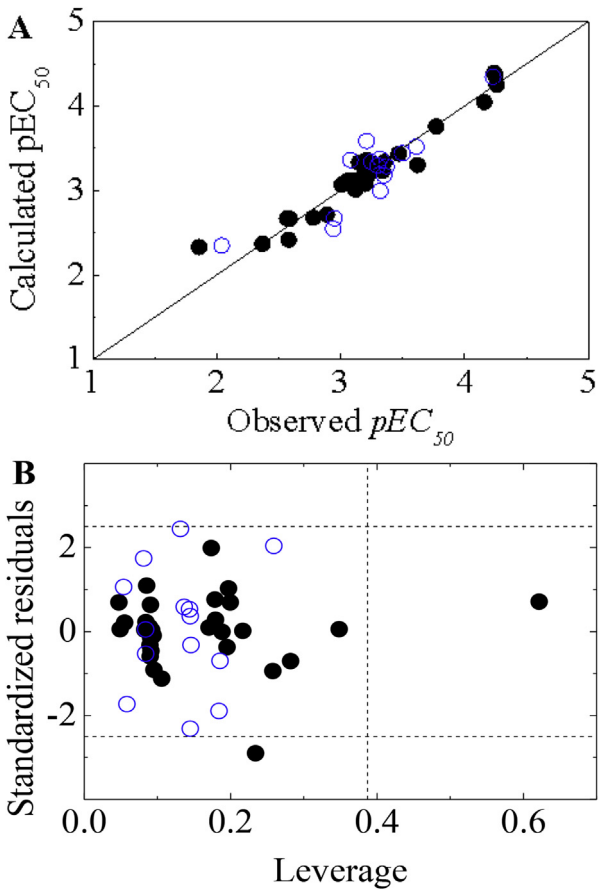

Fig. 3. Graph of observed versus calculated pEC_{50} (**A**) and Williams plot (**B**) for the training set ("•") and test set ("o").

based on training set was fixed at 0.3871. If the mixture has a high value of leverage (>0.3871) and SR (>2.5 or < −2.5), this mixture is considered as an outlier. All mixtures are located in the application domain based on Fig. 3B.

The QSAR model for the mixtures was built by using the molecular descriptors (*RDF035m*, *HATSs*, and *H-047*) of six single chemicals. The descriptors of any mixture combination were calculated. The developed QSAR model (Eq. (13)) was able to predict the toxicity of an unknown mixture with the combination of six compounds. The QSAR model was unable to predict any mixture combinations. This model only validated the unknown mixture located in the domain of $h < 0.3871$ and $2.5 < SR < 2.5$ (Fig. 3B), where h and SR values of an unknown mixture can be calculated based on Eq. (12) and the developed QSAR model (Eq. (13)), respectively.

3.6. Comparison of QSAR and CA/IA models

Although CA and IA models are widely used in the prediction of combined effect of chemical mixtures, they are only suitable for the additive effect of mixtures. Currently, few effective methods can be used to predict the interactive toxicity of mixtures. A total of 27 mixture rays out of 45 mixture rays presented non-additive effects, where the CA and IA predictions obviously deviated from the experimental pEC_{50} values. As shown in Table 1, the developed QSAR model predicted the toxicities of binary and multi-component mixtures.

A total of 12 pEC_{50} values out of 45 mixture toxicities predicted by the QSAR model are located out of the range of the lower and upper 95% CI, and 27 pEC_{50} values resulted from CA model are located out of the 95% CI. The 27 mixture toxicities cannot be predicted by CA and IA models accurately, and 15 mixture toxicities were accurately predicted by the QSAR model. *RMSE* can be used to evaluate the deviation between observation and calculation. The low *RMSE* indicates the close distance between observation and calculation. The *RMSE* value of the QSAR model based on the entire data set equals to 0.1663, which is lower than the *RMSE* value of CA (0.2045) and IA (0.2378). Therefore, the predictive capability of QSAR model is higher than the CA and IA models, which indicates that the QSAR model may be valuable in predicting the non-additive effects of mixtures.

4. Conclusions

In this study, several structural features of *RDF035m*, *HATSs*, and *H-047* related to the mixture toxicity of antibiotic and pesticides were provided to develop the QSAR model. Simple and accurate GA-MLR model was effectively presented. The training and test sets were applied to build and validate the model. Internal and external validations were employed to evaluate the predictive capability of the QSAR model, which exhibited high predictive capability for mixture toxicity with additive and non-additive effects. Considering that the QSAR model for mixtures relies on the molecular structures and concentration ratios of components, the estimated pEC_{50} values by the presented approach filled the gaps in the experimental mixture toxicity of two antibiotics and six pesticides mixed at any concentration ratios. The developed QSAR model was based on binary and multi-components mixtures, whereas the previous QSAR model was mainly developed for binary mixtures. In addition, the presented model provided more accurate predictive results for antagonistic and synergistic toxicities of mixtures compared with the CA and IA models. Therefore, the QSAR model can be used to predict the additive and non-additive toxicities of binary and multi-component mixtures.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (21407032), Provincial Natural Science Foundation of Guangxi (2017GXNSFAA198346), Science Research and Technology Development Project of Guilin (2016012505), and Special Funding for Guangxi 'BaGui Scholar' Construction Projects.

Appendix B. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.chemosphere.2018.01.142.

References

Ajmani, S., Rogers, S.C., Barley, M.H., Livingstone, D.J., 2006. Application of QSPR to mixtures. J. Chem. Inf. Model. 46, 2043–2055.
Altenburger, R., Nendza, M., Schüürmann, G., 2003. Mixture toxicity and its modeling by quantitative structure-activity relationships. Environ. Toxicol. Chem. 22, 1900–1915.
Altenburger, R., Schmitt, H., Schüürmann, G., 2005. Algal toxicity of nitrobenzenes: combined effect analysis as a pharmacological probe for similar modes of interaction. Environ. Toxicol. Chem. 24, 324–333.
Altenburger, R., Walter, H., Grote, M., 2004. What contributes to the combined effect of a complex mixture? Environ. Sci. Technol. 38, 6353.
Bliss, C.I., 1939. The toxicity of poisons applied jointly. Ann. Appl. Biol. 26, 585–615.
Cedergreen, N., Christensen, A.M., Kamper, A., Kudsk, P., Mathiassen, S.K., Streibig, J.C., Srensen, H., 2008. A review of independent action compared to concentration addition as reference models for mixtures of compounds with different molecular target sites. Environ. Toxicol. Chem. 27, 1621–1632.
Chang, C.M., Ou, Y.H., Liu, T.C., Lu, S.Y., Wang, M.K., 2016. A quantitative structure-activity relationship approach for assessing toxicity of mixture of organic

compounds. Sar Qsar Environ. Res. 27, 441–453.
Chirico, N., Gramatica, P., 2011. Real external predictivity of QSAR models: how to evaluate it? comparison of different validation criteria and proposal of using the concordance correlation coefficient. J. Chem. Inf. Model. 51, 2320–2335.
Christin, M.S., Menard, L., Gendron, A.D., Ruby, S., Cyr, D., Marcogliese, D.J., Rollins-Smith, L., Fournier, M., 2004. Effects of agricultural pesticides on the immune system of Xenopus laevis and Rana pipiens. Aquat. Toxicol. 67, 33–43.
Consonni, V., Ballabio, D., Todeschini, R., 2010. Evaluation of model predictive ability by external validation techniques. J. Chemom. 24, 194–201.
Dou, R.N., Liu, S.S., Mo, L.Y., Liu, H.L., Deng, F.C., 2011. A novel direct equipartition ray design (EquRay) procedure for toxicity interaction between ionic liquid and dichlorvos. Environ. Sci. Pollut. Res. 18, 734–742.
Eriksson, L., Jaworska, J., Worth, A.P., Cronin, M.T.D., McDowell, R.M., 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.
Gao, L., Shi, Y., Li, W., Liu, J., Cai, Y., 2012. Occurrence, distribution and bio-accumulation of antibiotics in the Haihe River in China. J. Environ. Monit. 14, 1248.
Gaudin, T., Rotureau, P., Fayet, G., 2015. Mixture descriptors toward the development of quantitative structure–property relationship models for the flash points of organic mixtures. Ind. Eng. Chem. Res. 54, 6596–6604.
Ge, H.L., Liu, S.S., Su, B.X., Qin, L.T., 2014. Predicting synergistic toxicity of heavy metals and ionic liquids on photobacterium Q67. J. Hazard Mater. 268, 77–83.
Golbraikh, A., Tropsha, A., 2002. Beware of q^2! J. Mol. Graph. Model. 20, 269–276.
Gramatica, P., Cassani, S., Chirico, N., 2014. QSARINS-chem: insubria datasets and new QSAR/QSPR models for environmental pollutants in QSARINS. J. Comput. Chem. 35, 1036–1044.
Gramatica, P., Chirico, N., Papa, E., Cassani, S., Kovarich, S., 2013. QSARINS: a new software for the development, analysis, and validation of QSAR MLR models. J. Comput. Chem. 34, 2121–2132.
Hertzberg, R.C., Macdonell, M.M., 2002. Synergy and other ineffective mixture risk definitions. Sci. Total Environ. 288, 31–42.
Iwasaki, Y., Gauthier, P., 2016. Concentration addition and response addition to analyze mixture toxicity: is it worth testing? Environ. Toxicol. Chem. 35, 526–527.
Kemper, N., 2008. Veterinary antibiotics in the aquatic and terrestrial environment. Ecol. Indicat. 8, 1–13.
Kim, J., Kim, S., Schaumann, G.E., 2013. Development of QSAR-based two-stage prediction model for estimating mixture toxicity. Sar Qsar Environ. Res. 24, 841–861.
Kiralj, R., Ferreira, M.M.C., 2009. Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. J. Braz. Chem. Soc. 20, 770–787.
Kiralj, R., Ferreira, M.M.C., 2010. Is your QSAR/QSPR descriptor real or trash? J. Chemom. 24, 681–693.
Liu, S.-S., Xiao, Q.-F., Zhang, J., Yu, M., 2016a. Uniform design ray in the assessment of combined toxicities of multi-component mixtures. Sci. Bull. 61, 52–58.
Liu, S.S., Li, K., Li, T., Qu, R., 2016b. Comments on "The synergistic toxicity of the multi chemical mixtures: implications for risk assessment in the terrestrial environment". Environ. Int. 94, 396–398.
Loewe, S., Muischnek, H., 1926. Effect of combinations: mathematical basis of problem. Naunyn-Schmiedebergs Arch. Exp. Pathol. Pharmakol. 114, 313–326.
Lu, G., Wang, C., Tang, Z., Guo, X., 2007. Joint toxicity of aromatic compounds to algae and QSAR study. Ecotoxicology 16, 485–490.
Mo, L.Y., Qin, R.Q., Qin, L.T., Zeng, H.H., Liang, Y.P., 2015. Study on the quantitative relationship between the structure and toxicity of organophosphorus pesticide. Chin. J. Struct. Chem. 34, 1473–1478.
Ojha, P.K., Mitra, I., Das, R.N., Roy, K., 2011. Further exploring r^{2m} metrics for validation of QSPR models. Chemometr. Intell. Lab. Syst. 107, 194–205.
Qin, L.-T., Liu, S.-S., Zhang, J., Xiao, Q.-F., 2011. A novel model integrated concentration addition with independent action for the prediction of toxicity of multi-component mixture. Toxicology 280, 164–172.
Qin, L.-T., Wu, J., Mo, L.-Y., Zeng, H.-H., Liang, Y.-P., 2015. Linear regression model for predicting interactive mixture toxicity of pesticide and ionic liquid. Environ. Sci. Pollut. Res. 22, 12759–12768.
Qin, L.-T., Zhang, X., Mo, L.-Y., Liang, Y.-P., Zeng, H.-H., 2017. Further exploring linear concentration addition and independent action for predicting non-interactive mixture toxicity. Chin. J. Struct. Chem. 36, 886–896.
Qin, L.T., Liu, S.S., Chen, F., Wu, Q.S., 2013a. Development of validated quantitative structure–retention relationship models for retention indices of plant essential oils. J. Sep. Sci. 36, 1553–1560.
Qin, L.T., Liu, S.S., Chen, F., Xiao, Q.F., Wu, Q.S., 2013b. Chemometric model for predicting retention indices of constituents of essential oils. Chemosphere 90, 300–305.
Ra, J.S., Lee, B.C., Chang, N.I., Kim, S.D., 2006. Estimating the combined toxicity by two-step prediction model on the complicated chemical mixtures from wastewater treatment plant effluents. Environ. Toxicol. Chem. 25, 2107.
Rider, C.V., LeBlanc, G.A., 2005. An integrated addition and interaction model for assessing toxicity of chemical mixtures. Toxicol. Sci. 87, 520–528.
Ritz, C., 2010. Toward a unified approach to dose-response modeling in ecotoxicology. Environ. Toxicol. Chem. 29, 220–229.
Schüürmann, G., Ebert, R.U., Chen, J.W., Wang, B., Kuhne, R., 2008. External validation and prediction employing the predictive squared correlation coefficient - test set activity mean vs training set activity mean. J. Chem. Inf. Model. 48, 2140–2145.
Shi, L.M., Fang, H., Tong, W., Wu, J., Perkins, R., Blair, R.M., Branham, W.S., Dial, S.L., Moland, C.L., Sheehan, D.M., 2001. QSAR models using a large diverse set of estrogens. J. Chem. Inf. Comput. Sci. 41, 186–195.
Sobati, M.A., Abooali, D., Maghbooli, B., Najafi, H., 2016. A new structure-based model for estimation of true critical volume of multi-component mixtures. Chemometr. Intell. Lab. Syst. 155, 109–119.
Soltanpour, S., Shahbazy, M., Omidikia, N., Kompany-Zareh, M., Baharifard, M.T., 2016. A comprehensive QSPR model for dielectric constants of binary solvent mixtures. Sar Qsar Environ. Res. 27, 165–181.
Syberg, K., Elleby, A., Pedersen, H., Cedergreen, N., Forbes, V.E., 2008. Mixture toxicity of three toxicants with similar and dissimilar modes of action to Daphnia magna. Ecotoxicol. Environ. Saf. 69, 428–436.
Tang, J.Y., Escher, B.I., 2014. Realistic environmental mixtures of micropollutants in surface, drinking, and recycled water: herbicides dominate the mixture toxicity toward algae. Environ. Toxicol. Chem. 33, 1427.
Topp, E., Renaud, J., Sumarah, M., Sabourin, L., 2016. Reduced persistence of the macrolide antibiotics erythromycin, clarithromycin and azithromycin in agricultural soil following several years of exposure in the field. Sci. Total Environ. 562, 136.
Toropova, A.P., Toropov, A.A., Benfenati, E., Gini, G., Leszczynska, D., Leszczynski, J., 2012. CORAL: models of toxicity of binary mixtures. Chemometr. Intell. Lab. Syst. 119, 39–43.
Tropsha, A., Gramatica, P., Gombar, V.K., 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.
Wang, B., Yu, G., Zhang, Z., Hu, H., Wang, L., 2006. Quantitative structure-activity relationship and prediction of mixture toxicity of alkanols. Sci. Bull. 51, 2717–2723.
Watkinson, A.J., Murby, E.J., Kolpin, D.W., Costanzo, S.D., 2009. The occurrence of antibiotics in an urban watershed: from wastewater to drinking water. Sci. Total Environ. 407, 2711–2723.
Webster, T.M.U., Perry, M.H., Santos, E.M., 2015. The herbicide linuron inhibits cholesterol biosynthesis and induces cellular stress responses in Brown trout. Environ. Sci. Technol. 49, 3110–3118.
Wu, M.H., Que, C.J., Xu, G., Sun, Y.F., Ma, J., Xu, H., Sun, R., Tang, L., 2016. Occurrence, fate and interrelation of selected antibiotics in sewage treatment plants and their receiving surface water. Ecotoxicol. Environ. Saf. 132, 132–139.
Yao, Z.F., Lin, Z.F., Wang, T., Tian, D.Y., Zou, X.M., Gao, Y., Yin, D.Q., 2013. Using molecular docking-based binding energy to predict toxicity of binary mixture with different binding sites. Chemosphere 92, 1169–1176.
Zhang, Y.H., Liu, S.S., Song, X.Q., Ge, H.L., 2008. Prediction for the mixture toxicity of six organophosphorus pesticides to the luminescent bacterium Q67. Ecotoxicol. Environ. Saf. 71, 880–888.