Science of the Total Environment 704 (2020) 135302

Contents lists available at ScienceDirect

Science of the Total Environment

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

Predicting acute contact toxicity of organic binary mixtures in honey bees (*A. mellifera*) through innovative QSAR models

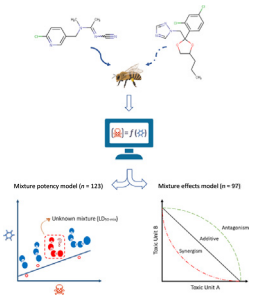
Edoardo Carnesecchi [a,b,*], Andrey A. Toropov [a], Alla P. Toropova [a], Nynke Kramer [b], Claus Svendsen [c], Jean Lou Dorne [d], Emilio Benfenati [a]

[a] Laboratory of Environmental Chemistry and Toxicology, Department of Environmental Health Sciences, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Via la Masa 19, 20156 Milan, Italy
[b] Institute for Risk Assessment Sciences (IRAS), Utrecht University, PO Box 80177, 3508 TD Utrecht, The Netherlands
[c] Centre for Ecology and Hydrology, Maclean Building, Benson Lane, Wallingford, Oxfordshire OX10 8BB, UK
[d] European Food Safety Authority (EFSA), Scientific Committee and Emerging Risks Unit, Via Carlo Magno 1A, 43126 Parma, Italy

HIGHLIGHTS

- Three QSAR models predict combined toxicity of binary mixtures in honey bees.
- Models were validated using robust internal and external parameters.
- Predictions of combined toxicity for untested binary mixtures are provided.
- The use of quasi-SMILES improves the statistical quality of the models.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:
Received 18 September 2019
Received in revised form 29 October 2019
Accepted 29 October 2019
Available online 19 November 2019

Editor: Damia Barcelo

Keywords:
Mixtures toxicity
Honey bees
Quantitative structure–activity relationship
CORAL software
Monte Carlo method

ABSTRACT

Pollinators such as honey bees are of considerable importance, because of the crucial pollination services they provide for food crops and wild plants. Since bees are exposed to a wide range of multiple chemicals "mixtures" both of anthropogenic (e.g. plant protection products) and natural origin (e.g. plant toxins), understanding their combined toxicity is critical. Although honey bees are employed worldwide as surrogate species for *Apis* and non-*Apis* bees in toxicity tests, it is practically unfeasible to perform *in vivo* tests for all mixtures of chemicals. Therefore, Quantitative Structure-Activity Relationships (QSAR) models can be developed using available data and can provide useful tools to predict such combined toxicity. Here, three different QSAR models within the CORAL software have been calibrated and validated for honey bees (*A. mellifera*) to predict the acute contact mixtures potency ($LD_{50\text{-mix}}$), in two regression based-models, and the nature of combined toxicity (synergism / non-synergism) in a classification-based model. Experimental data on binary mixtures (n = 123) ($LD_{50\text{-mix}}$) including dose response data (n = 97) and corresponding Toxic Unit values were retrieved from EFSA databases. The models were built using the principle of extraction of attributes from SMILES (or quasi-SMILES) while calculating so-called correlation weights for these attributes using Monte Carlo techniques. The two regression models were validated for their reliability and robustness (R^2 = 0.89, CCC = 0.92, Q^2 = 0.81; R^2 = 0.87, CCC = 0.89, Q^2 = 0.75). The classification model was validated using sensitivity (=0.86), specificity (=1), accuracy (=0.96), and Matthews correlation coefficient (MCC = 0.90) as qualitative statistical validation parameters. Results indicate that these QSAR models successfully predict acute contact toxicity of binary mixtures in honey bees and can support prioritisation of multiple chemicals of concerns. Data gaps and

* Corresponding author at: Laboratory of Environmental Chemistry and Toxicology, Department of Environmental Health Sciences – Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Via Mario Negri 2, 20156 Milan, Italy.
E-mail address: edoardo.carnesecchi@marionegri.it (E. Carnesecchi).

https://doi.org/10.1016/j.scitotenv.2019.135302

further development of QSAR models for honey bees are highlighted particularly for chronic and sub-lethal effects.

1. Introduction

Honey bees (*Apis mellifera*), solitary bees and bumble bees represent important environmental non-target species particularly because of their crucial pollination services for food crops and their contribution to the maintenance and reproduction of wild plant communities and biodiversity (Klein et al., 2007; Vanengelsdorp and Meixner, 2010; Lambert et al., 2012). Honey bees are employed worldwide as surrogate species for *Apis* and non-*Apis* bees to perform toxicity tests on single pesticides (EFSA, 2013; U. S. EPA, 2014). In addition, they also represent sentinel species together with their hive products as bioindicators (i.e. honey, propolis, pollen) to investigate environmental contamination by regulated products (e.g. Plant Protection Products (PPPs), veterinary drugs) or anthropogenic (polycyclic aromatic hydrocarbons, heavy metals) and natural contaminants (mycotoxins, plant alkaloids and flavonoids) (Lambert et al., 2012; Johnson et al., 2012, 2013; Bargańska et al., 2016; Tosi et al., 2018; Skorbiłowicz et al., 2018). As a matter of fact, bees are exposed to these as multiple substances "mixtures", either by foraging on contaminated areas or through contaminated food stored and consumed in the hive. Over the last decade, scientific advisory bodies and governmental agencies have developed methods and frameworks to assess such mixtures issues (U.S. EPA, 2003; Kemi, 2015; EFSA, 2009; EFSA PPR Panel, 2012, 2013; EFSA, 2014; Backhaus and Faust, 2012; Kienzler et al., 2016; Nys et al., 2018). In this context, the recent EFSA MIXTOX guidance document (More et al., 2019) has illustrated methods and case studies in bees providing opportunities to investigate their contribution to bee health compared with other stressors (e.g. varroa, viruses) and to develop holistic risk assessment methodologies (EFSA, 2013, 2016; EFSA, 2017; Rortais et al., 2017). In order to further understand combined toxicity in honey bees, a recent meta-analysis (Carnesecchi et al., 2019) of acute contact laboratory toxicity assays on PPPs and veterinary drugs highlighted synergisms and antagonisms in 72% and 11% of datasets, respectively. For most observed synergisms, cytochrome P450 (CYP) inhibition was the major mechanism resulting in a decrease in elimination and an increase in the toxicity of the binary mixture (Carnesecchi et al., 2019; Johnson et al., 2013; Wade et al., 2019). Although the authors identified numerous data gaps, such combined toxicity databases potentially allow developing predictive Quantitative Structure-Activity Relationships (QSAR) tools particularly because it is rather impossible to test all possible mixtures in bees for their acute or chronic effects. Such QSAR tools are only available for single chemicals but to date not for the prediction of combined toxicity (Venko et al., 2017; Singh et al., 2014; Hamadache et al., 2018; Como et al., 2017).

Hence, this manuscript describes the development and application of three innovative predictive QSAR models for honey bees within the CORAL software namely (i) two regression-based QSAR models predicting acute (contact) mixtures potency ($pLD_{50\text{-}mix}$) in a quantitative manner, and (ii) a classification-based model predicting the nature of combined toxicity for organic binary mixtures (i.e. synergism / non-synergism). Calibration and validation of the models are described using available experimental data, simplified molecular input-line entry system (SMILES) and attributes. Validation is assessed using independent datasets and associated statistics (i.e. correlation weights) (Toropova et al., 2012; Toropov et al., 2012a, 2019). Finally, conclusions highlight the potential application of such *in silico* tools for the hazard assessment and prioritisation of organic binary mixtures in ecological risk assessment.

2. Material and methods

2.1. Experimental data

Experimental data from laboratory studies on honey bees measuring the combined toxicity ($LD_{50\text{-}mix}$) and Toxic Units (TUs) following acute contact exposure to organic binary mixtures were retrieved from an EFSA database described in our recent meta-analysis (Carnesecchi et al., 2019). The database provides quantitative information (e.g. Toxic Unit, LD_{50} – 24 h) on 123 mixtures studies.

First, a simple regression model (Approach A) to predict mixture potency ($LD_{50\text{-}mix}$) was developed from the $LD_{50\text{-}mix}$ dataset (n = 123) while considering as input two chemical structures represented by simplified molecular input-line entry system (SMILES) (Table S1). As second step, a dataset including only dose response data (n = 97) on binary mixtures (Approach B) was created to develop (i) a regression-based model to predict the potency of the binary mixtures ($pLD_{50\text{-}mix}$) and (ii) a classification-based model to predict the nature of the combined toxicity (synergism/non-synergism), taking into account Toxic Units (TUs) for each chemical in the mixture (Table 1). Hence, in Approach B, quantitative data on TUs were used as additional features and were represented as quasi-SMILES (Toropov et al., 2018; Toropova et al., 2019a). The TU approach assumes that predictions for combined toxicity in the binary mixture follow the Concentration Addition (CA) model (Fig. 1) given the quantitative composition of each chemical within the binary mixture in relation to their relative potency (Jonker et al., 2005). A detailed account of the methodologies for TU calculation is provided elsewhere (Carnesecchi et al., 2019).

With regards to data pruning, no OECD guidelines are available for designing binary mixtures toxicity experiments in honey bees. Hence, it was not possible to follow any harmonised criteria in the data pruning steps, in contrast to what is available in the OECD guideline for testing single chemicals (OECD, 1998). Since the scientific literature most often reports *in vivo* $LD_{50\text{-}mix}$ as μg active substance/bee as a median lethal dose after 24 h, model results were expressed as $pLD_{50\text{-}mix}$ (i.e. negative decimal logarithm log $[1/LD_{50}]$), logarithm of the inverse of the lethal dose to kill 50% of honey bees in the tested sample (Toropova et al., 2012; Iwasa et al., 2004; Johnson et al., 2012, 2013).

2.2. Development of the models

2.2.1. Approach A – Regression model ($pLD_{50\text{-}mix}$)

The chemical structure of the two-component mixtures is represented using the disconnected simplified molecular input-line entry system (SMILES) (Table S1), where the "." (period or "dot") is used to represent disconnections (Hunter et al., 1987). The total number of data (n = 123) are randomly split into four sets: training set (≈25%; TRN), invisible training set (≈25%; iTRN), calibration set (≈25%; CLB), and validation set (≈25%; VLD) sets, each of which contains independent datasets and has a specific purpose:

Table 1
Overview of the QSAR models developed here. Input data (e.g. number and class of substances) and endpoints are reported for both models.

Input data (n)	Features	Statistical method	Classes and number (n) of substances	Endpoint/Model
Approach A				
123	• $SMILES_A$ $SMILES_B$	• Method Monte Carlo (MCC)	• Fungicide (13) • Insecticide/acaricide (12) • Synergist (3)	• Regression-based model ($pLD_{50\text{-}mix}$ 24 h)
Approach B				
97	• $SMILES_A$ $SMILES_B$ • Toxic Unit	• Method Monte Carlo (MCC) • Quasi-SMILES	• Insecticide/acaricide (11) • Fungicide (11) • Synergist (1)	• Regression- based model ($pLD_{50\text{-}mix}$ 24 h) • Classification-based model (synergism, non-synergism)

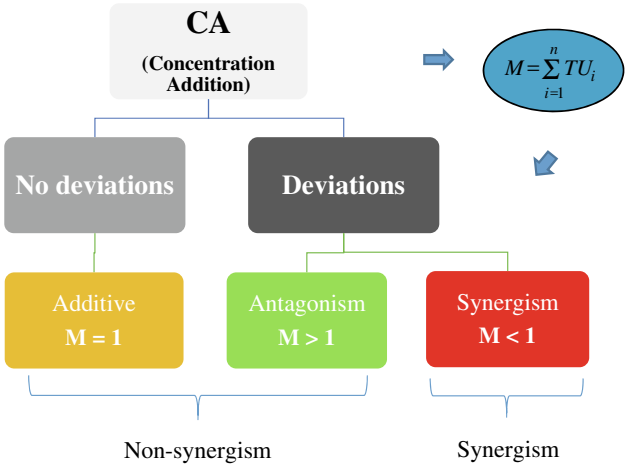


Fig. 1. Decision tree for predicting combined toxicity of binary mixtures (M) (i.e. dose addition, synergism and antagonism) according to Concentration addition (CA) model and Toxic Unit (TU) approach. Two classes (synergism and non-synergism) are identified as outputs of the classification model (Approach B).

1. The TRN set is the core dataset of the QSAR model. Compounds from this set are used to generate correlation weights giving maximal value of the target function using Monte Carlo optimisation;
2. The iTRN set inspects whether the model predictions are satisfactory using data for compounds that are independent from the TRN set;
3. The CLB set detects the start of the overtraining of the model.
4. The VLD set is used for the validation of the prediction model as a final step.

2.2.1.1. Optimal descriptor. The optimal descriptor used to develop the QSAR model for the combined toxicity of binary mixtures in honey bees (Approach A) is the following:

$$DCW(T^*,N^*) = \sum_{k=1} CW(S_k) + \sum_{k=1} CW(SS_k) + \sum_{k=1} CW(SSS_k) \quad (1)$$

The S_k is the "SMILES-atom" i.e. one or two symbols (e.g. 'C', 'N', 'O', etc.) and cannot be examined separately (e.g. 'Cl', 'Si', etc.). The SS_k is a combination of two SMILES-atoms. The CW(S_k), CW(SS_k), and CW(SSS_k) are so-called correlation weights of the above-mentioned attributes of SMILES. The numerical data on the CW (S_k), CW(SS_k), and CW(SSS_k) are calculated using the Monte Carlo method, i.e. the optimisation procedure which gives maximal value of target function (TF):

$$TF = r_{TRN} + r_{iTRN} - |r_{TRN} - r_{iTRN}| * 0.1 \quad (2)$$

where the r_{TRN} and r_{iTRN} are correlation coefficients between observed and predicted endpoints for the training and invisible training sets, respectively.

2.2.1.2. Statistical criteria. In order to evaluate a regression model on combined toxicity of binary mixtures ($pLD_{50\text{-}mix}$) in honey bees, the following statistical criteria are used: determination coefficient (R^2), cross-validated determination coefficient (Q^2) which measures prediction power, root mean squared error (RMSE), mean absolute error (MAE), Fischer F-ratio (F) and concordance correlation coefficient (CCC) (Roy et al., 2012; Chirico and Gramatica, 2012). The latter is defined as a complementary or alternative statistical criterion for external validation measures, particularly when other statistical criteria are in conflict. Results of the analyses are provided in Table 2.

2.2.2. Approach B – classification and regression models based on Toxic Unit

Approach B aims at developing two QSAR models on experimental dose response data (n = 97) while using the CORAL software. TUs for each chemical in the binary mixture are used as additional features to develop i) a regression-based model to predict the potency of the binary mixture ($pLD_{50\text{-}mix}$) and ii) a classification-based model to predict the nature of the combined toxicity (synergism/non-synergism). In both models, toxicity data for binary mixtures are represented by the so-called quasi–simplified molecular input-line entry system (quasi-SMILES) which is analogue of the traditional SMILES applied in QSPR/QSAR analyses but makes use of all available data (not only information about the molecular structure) (Toropov et al., 2018). The total number of data (n = 97) were randomly split into the training (≈25%), invisible training (≈25%), calibration (≈25%), and validation (≈25%) sets.

2.2.2.1. Optimal descriptor. Two kinds of optimal descriptors are calculated for i) the regression models and ii) the classification models, respectively:

$$DCW(T^*,N^*) = \sum_{k=1}^{NA} CW(S_k) + \sum_{k=1}^{NA-1} CW(SS_k) + \sum_{k=1}^{NA-2} CW(SSS_k) \quad (3)$$

$$DCW(T^*,N^*) = \sum_{k=1}^{NA} CW(S_k) + \sum_{k=1}^{NA-1} CW(SS_k) \quad (4)$$

All parameters are already described in Eq. (1). However, the number of attributes in SMILES (NA) is added for including TU values (TU_A, TU_B) as additional features. As a consequence, QSAR models are calculated with the Monte Carlo optimisation based on two kinds of target functions TF_1 (see Eq. (2)) and TF_2:

$$TF_2 = TF_1 + IIC_{CLB} * 0.1 \quad (5)$$

Where, the IIC_{CLB} is calculated with data on the calibration (CLB) set as follows:

$$IIC_{CLB} = r_{CLB} \frac{\min(-MAE_{CLB}, +MAE_{CLB})}{max(-MAE_{CLB}, +MAE_{CLB})} \quad (6)$$

$$-MAE_{CLB} = \frac{1}{-N}\sum_{k=1}^{-N} |\Delta_k|,\ \Delta_k < 0; -N \text{ is the number of } \Delta_k < 0 \quad (7)$$

$$+MAE_{CLB} = \frac{1}{+N}\sum_{k=1}^{+N} |\Delta_k|, \Delta_k \geqslant 0; +N \text{ is the number of } \Delta_k \geqslant 0 \quad (8)$$

$$\Delta_k = observed_k - calculated_k \quad (9)$$

The observed and calculated are corresponding values of the endpoint.

Having the numerical data on the $CW(S_k)$, $CW(SS_k)$ and $CW(SSS_k)$, the predictive model is calculated using the Least Squares method and data for compounds within the training set:

$$pEC_{50} = C_0 + C_1 * DCW(T^*, N^*) \quad (10)$$

2.2.2.2. Statistical criteria. Statistical criteria for the regression-based QSAR model (Approach B) have been applied as described in Section 2.2.1 (Approach A). In addition, TUs were used here as an additional feature to improve the performance of the regression and classification-based models developed according to Approach B. In addition, other statistical criteria were used namely: sensitivity, specificity, accuracy, and Mattews correlation coefficient (MCC) to in order to build up the classification model for two classes: synergism (1) and non-synergism (0) (Toropova and Toropov, 2017; Toropov et al., 2012b). Generally, the MCC coefficient is applied in machine learning to measure the quality of binary classifications and it can be used when the classes present very different sizes (Dao et al., 2011).

$$Sensitivity = \frac{TP}{TP + FN} \quad (11)$$

$$Specificity = \frac{TN}{TN + FP} \quad (12)$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (13)$$

$$MCC = \frac{TP \text{ x } TN - FP \text{ x } FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (14)$$

TP, TN, FP and FN represent the number of true positives, true negatives, false positives and false negatives, respectively, in a confusion matrix. MCC values range between −1 and +1, while the latter indicates a perfect prediction, a value of 0 indicate a prediction no better than a random one, and a value of −1 show total disagreement between predicted and observed values (Dao et al., 2011).

2.2.2.3. Model(s) validation and mechanistic interpretation (Approach A and B). Model validation for both approaches (A and B) used statistical parameters for modelling internal (i.e. calibration) and external validations in order to estimate the predictive capability and the goodness of fitness of the QSAR models (Qin et al., 2018). With regard to regression-based QSAR models, the statistical quality of the fitted equations was evaluated using the determination coefficient (R^2), concordance correlation coefficient (CCC), cross-validated determination coefficient (Q^2), root mean squared error (RMSE), mean absolute error (MAE) and Fischer F-ratio (F) (Roy et al., 2012). Results for the assessment of statistical quality of the regression models following approaches A and B are presented in Tables 2 and 3, respectively (Sections 3.1 and 3.2.1).

Similarly, the classification model (synergism/non-synergism) has been validated according to the following statistical parameters: sensitivity, specificity, accuracy, and MCC and results for the assessment of the statistical quality are presented in Table 4 (Section 3.2.2).

The "Mechanistic interpretation" is defined as the causality between a substance and its toxicity (or not-toxicity) and it is required when developing a QSAR model as described in the OECD guidelines (OECD, 2007), "*a mechanistic interpretation, if possible*" so that establishing a correlation and a causal relationship between the chemical structure of the compound and its toxicity (OECD principle 5) (Thoreau, 2016). The CORAL models provide the mechanistic interpretation in the form of promoters either increasing or decreasing potency of a chemical (Toropova and Toropov, 2017, 2018; Toropov et al., 2019). Here, the mechanistic interpretation is obtained by means of the results of several runs of the Monte Carlo optimisation. In particular, molecular features extracted from SMILES, providing stable positive correlation weights in several runs of the Monte Carlo optimisation, can be recognised as promoters of increase in the toxic potency of the mixture ($pLD_{50\text{-mix}}$). In contrast, molecular features presenting only negative correlation weights in several runs of the optimisation are promoters of a decrease in the toxic potency of the mixture.

2.2.2.4. Applicability domain (Approach A and B). The applicability domain is an important component of QSAR analyses (OECD, 2007). According to OECD principle 3, a QSAR model should have a well-defined applicability domain. Applicability domain is defined as the area or chemical space represented by the molecular properties or structural information of the chemicals used for the model development. A collection of conceptions of applicability domains for different QSAR approaches is available and include: (i) physico-chemical domain, (ii) structural domain, (iii) response domain and (iv) integrated methods (Gadaleta and Catto, 2016). However, for models developed in the CORAL software, the statistical defects of SMILES calculated according to the distribution of available data into the training (TRN), invisible training (iTRN), calibration (CLB), and validation (VLD) sets are the basis to define the applicability domain. Here for both approaches (A and B) the same methods (Toropova et al., 2018) were applied for the assessment of the applicability domain defined according to distribution of SMILES attributes in the training and calibration sets following two steps:

Step 1: the definition of statistical defect (d_k) for each SMILES attribute (A_k = S_k, SS_k, SSS_k) involved (non-blocked) to construct the model;

$$d_k = \frac{|P(A_k) - P'(A_k)|}{N(A_k) + N'(A_k)} \quad (15)$$

where $P(A_k)$ and $P'(A_k)$ are the probabilities of A_k in the training and calibration sets, respectively; $N(A_k)$ and $N'(A_k)$ are the frequencies of A_k in the training and calibration sets, respectively.

Step 2: The calculation for all substances of the statistical SMILES-defect (D_j):

$$D_j = \sum_{k=1}^{NA} d_k \quad (16)$$

where NA is the number of non-blocked SMILES attributes in the SMILES.

A substance falls in the domain of applicability if

$$Dj < 2 * \bar{D} \quad (17)$$

where $\bar{D}$ is average of the statistical SMILES-defect for the training set.

3. Results and discussion

3.1. Approach A – Regression model ($pLD_{50\text{-}mix}$)

3.1.1. Model validation and mechanistic interpretation

The regression-based QSAR model developed here aimed at predicting the combined toxicity ($pLD_{50\text{-}mix}$) of binary mixtures of organic compounds in honey bees (n = 123) while considering chemical structures of the two components represented by disconnected SMILES as features. The regression-based QSAR model has been built using three equations according to three random splits, which include the TRN, iTRN, CLB, and VLD sets. Results with the following equations for each random splits are:

$$pLD_{50-mix} = -3.6089(\pm 0.0708) + 0.1126(\pm 0.0020) * DCW(1,15) \quad (18)$$

$$pLD_{50-mix} = -3.2651(\pm 0.0668) + 0.1498(\pm 0.0038) * DCW(1,15) \quad (19)$$

$$pLD_{50-mix} = -3.9621(\pm 0.0855) + 0.1338(\pm 0.0030) * DCW(1,15) \quad (20)$$

Table 2 provides the results of the applications of statistical criteria and the characteristics of the regression-based models with the corresponding Eqs. (18)–(20). The predictability of the models has been assessed using (i) determination coefficient (R^2) (a model has desired predictability if $R^2 > 0.65$) (Roy et al., 2012); (ii) concordance correlation coefficient (CCC) which indicates good predictability of the model if CCC > 0.85 (Chirico and Gramatica, 2012); and (iii) cross-validated determination coefficient (Q^2) requiring a value larger than 0.70 to be interpreted as reliable models (Chirico and Gramatica, 2012). According to our statistics (Table 2), results can be considered satisfactory. The most reliable model "best split" is represented by Eq. (19) showing R^2 = 0.87, CCC = 0.89 and Q^2 = 0.75 (Table 2 and Fig. 2), respectively. Similarly, Eq. (20) provides good statistics for split 3 (R^2 = 0.83; CCC = 0.84; Q^2 = 0.72). To date, no QSAR models for predicting acute toxicity of organic binary mixtures in insects have been published in the scientific literature. Toropova et al. (2012) developed a QSAR model using CORAL software for toxicity of binary mixtures (expressed as pEC_{50} – decrease in light emission in *Photobacterium phosphoreum*) presenting $R^2 > 0.86$ across six different splits.

The mechanistic interpretation has been obtained through three runs of the Monte Carlo optimisation. Molecular features providing stable positive correlation weights have been recognised as promoters of increase in the toxic potency of the mixture ($pLD_{50\text{-}mix}$). In contrast, molecular features presenting only negative correlation weights in the three runs of the Monte Carlo optimisation are considered as promoters of decrease in the toxic potency of the mixture ($pLD_{50\text{-}mix}$). Hence, according to our results oxygen atoms connected with double bonds, presence of rings, as well as presence of atoms of nitrogen are promoters of $pLD_{50\text{-}mix}$ increase. Similarly, carbon atoms connected with double bonds, triple bonds, presence of fluorine as well as nitrogen involved in a ring are promoters of $pLD_{50\text{-}mix}$ decrease (Table S2). However, it is necessary to take into account the prevalence of corresponding features in the training set and validation set so that rare attributes are not considered as source of reliable heuristic hypotheses.

3.2. Approach B – classification and regression models based on Toxic Units

According to materials and methods, dose response data (n = 97) were used to develop two QSAR models taking into account mixtures ratios expressed as TUs. Results for each model to provide i) quantitative predictions of acute contact toxicity of binary mixtures ($pLD_{50\text{-}mix}$) and ii) classification of combined toxicity (qualitative) (synergism/non-synergism) are illustrated below in Sections 3.2.1 and 3.2.2, respectively.

3.2.1. Regression model ($pLD_{50\text{-}mix}$; toxic unit)

3.2.1.1. Model validation and mechanistic interpretation. The statistical quality of the regression-based QSAR model ($pLD_{50\text{-}mix}$) including TUs as additional features is shown in Table 3. The QSAR model is built on four statistical models according to four random splits, which include the TRN, iTRN, CLB, and VLD sets. Results with the following equations for each random splits are:

$$pLD_{50-mix} = -3.1822116(\pm 0.0231229) + 0.0791914(\pm 0.0006750) * DCW(1,15) \quad (21)$$

$$pLD_{50-mix} = -6.0814987(\pm 0.0642473) + 0.1193315(\pm 0.0013066) * DCW(1,15) \quad (22)$$

$$pLD_{50-mix} = -3.1281529(\pm 0.0260826) + 0.0720372(\pm 0.0007152) * DCW(1,15) \quad (23)$$

$$pLD_{50-mix} = -6.2876839(\pm 0.0806048) + 0.0469456(\pm 0.0006665) * DCW(1,15) \quad (24)$$

Results of statistical characteristics for the four models calculated with corresponding Eqs. (21)–(24) are presented in Table 3. Similarly to Approach A for the regression model (Section 3.1),

Table 2
Statistical quality of regression-based QSAR model for the prediction of acute contact toxicity of binary mixtures ($pLD_{50\text{-}mix}$) (Approach A). Statistical values highlighted in bold indicate the most reliable model "best split" across three different splits.

Split	Set	n*	R^2	CCC	Q^2	RMSE	MAE	F
1	TRN*	31	0.72	0.84	0.68	0.70	0.51	75
	iTRN	31	0.55	0.66	0.50	0.81	0.63	35
	CLB	31	0.74	0.82	0.71	0.47	0.40	82
	VLD	30	0.77			0.49	0.40	
2	TRN*	30	0.62	0.76	0.57	0.77	0.61	45
	iTRN	31	0.62	0.75	0.58	0.68	0.51	48
	CLB	31	0.79	**0.89**	**0.75**	0.51	0.39	111
	VLD	31	**0.87**			0.45	0.34	
3	TRN*	31	0.60	0.75	0.55	0.75	0.56	43
	iTRN	30	0.60	0.77	0.55	0.66	0.50	42
	CLB	31	0.75	0.84	0.72	0.57	0.48	88
	VLD	31	0.83			0.48	0.37	

* n = number of pairs of SMILES in a set; R^2 = determination coefficient; CCC = concordance correlation coefficient; Q^2 = cross-validated determination coefficient; RMSE = root mean squared error; MAE = mean absolute error; F = Fischer F-ratio.

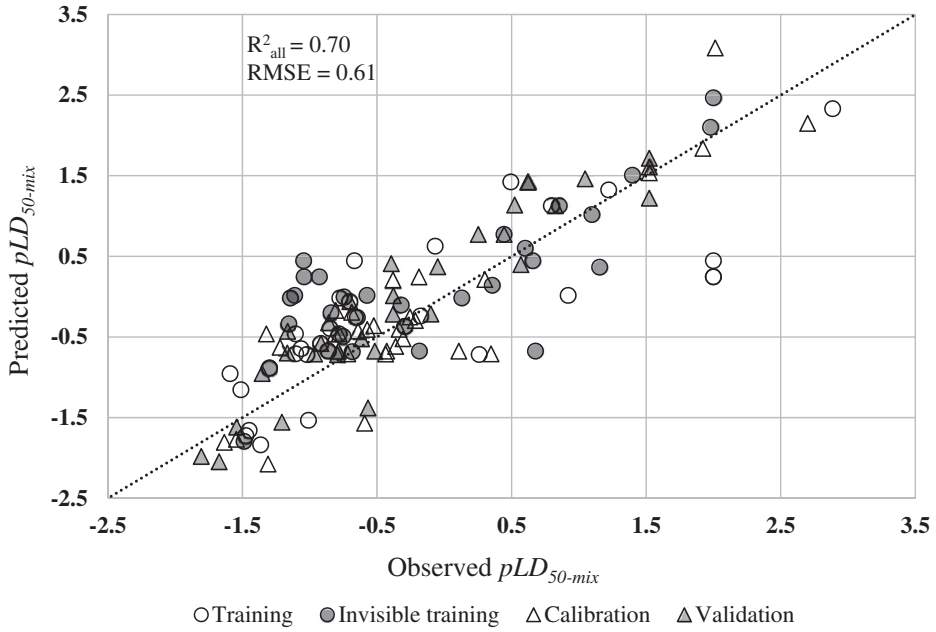


Fig. 2. Observed versus predicted log[1/$LD_{50\text{-}mix}$] of binary mixtures for the regression-based model (Approach A), split 2 "best split". R^2_{all} (determination coefficient) and RMSE (root mean squared error) are provided for all compounds (i.e. compounds from training, invisible training, calibration, and validation sets).

Table 3
Mechanistic interpretation and statistical robustness of the QSAR model for the prediction of acute contact toxicity of binary mixtures in honey bees based on Monte Carlo calculation with target functions 2 (TF_2). Statistical values highlighted in bold indicate the most reliable model "best split" across three different splits.

Split	Set	n	R^2	CCC	Q2	RMSE	F
1	TRN*	25	0.97	0.99	0.97	0.19	863
	iTRN	24	0.98	0.96	0.98	0.29	1140
	CLB	24	0.85	**0.92**	**0.81**	0.30	122
	VLD	24	**0.89**			0.53	
2	TRN	25	0.96	0.98	0.95	0.25	512
	iTRN	24	0.96	0.96	0.95	0.31	492
	CLB	24	0.79	0.88	0.73	0.35	81
	VLD	24	0.81			0.62	
3	TRN	24	0.96	0.98	0.95	0.21	522
	iTRN	24	0.95	0.95	0.94	0.33	462
	CLB	24	0.74	0.84	0.69	0.55	63
	VLD	25	0.75			0.61	
4	TRN	24	0.94	0.97	0.92	0.29	324
	iTRN	24	0.94	0.96	0.92	0.29	325
	CLB	24	0.86	**0.92**	**0.84**	0.45	133
	VLD	25	**0.78**			0.51	

* TRN, iTRN, CLB, and VLD are the training, invisible training, calibration, and validation sets, respectively; n is the number of mixtures in a set; R^2 is determination coefficient, CCC is concordance correlation coefficient; Q^2 is leave-one-out cross-validated correlation coefficient; RMSE is root means squared error; F is Fischer F-ratio.

the predictability of the models has been assessed according to: (i) determination of the R^2 coefficient so that the model has a desired predictability with R^2 > 0.65) (Roy et al., 2012); (ii) concordance correlation coefficient (CCC) for which good predictability is represented by CCC > 0.85) (Chirico and Gramatica, 2012); and (iii) cross-validated determination coefficient (Q^2) which is supposed to be larger than 0.70 for reliable models (Chirico and Gramatica, 2012). Hence, according to our statistics (Table 3 and Fig. 3), results can be considered satisfactory. In particular, the most reliable model is represented by Eq. (21) showing R^2 = 0.89, CCC = 0.92 and Q^2 = 0.81. Similarly, Eq. (24) provides good statistics (R^2 = 0.78; CCC = 0.92; Q^2 = 0.84), being R^2 slightly smaller than the one resulting from Eq. (21). Recently, Qin et al. (2018) developed a regression model for predicting mixture toxicities (additive and non-additive) of antibiotics and pesticide in *Aliivibrio fischeri* showing R^2_m = 0.68. However, results are not comparable due to the diversity of chemical classes and statistical approach used.

Mechanistic interpretation and statistical robustness of the CORAL model has been investigated using quasi-SMILES attributes and one run of Monte Carlo optimisation for the calculation of correlation weights for characterising either an overestimation or underestimation in predictions of the acute contact toxicity values for honey bees. According to our results (Tables S3 and S4), (i) positive correlation weights with SMILES attributes are interpreted as an increase in the acute contact toxicity (synergy) of the binary mixture; and (ii) negative correlation weights with SMILES attributes are associated with a decrease in the acute contact toxicity of the binary mixture (antagonism).

3.2.2. Classification model (synergism/non-synergism and toxic unit)
3.2.2.1. Model validation and mechanistic interpretation. Statistical robustness of the classification model on organic binary mixtures effect (synergism/non-synergism) has been tested and presented in Table 4. Table S5 contains the observed versus predicted syner-

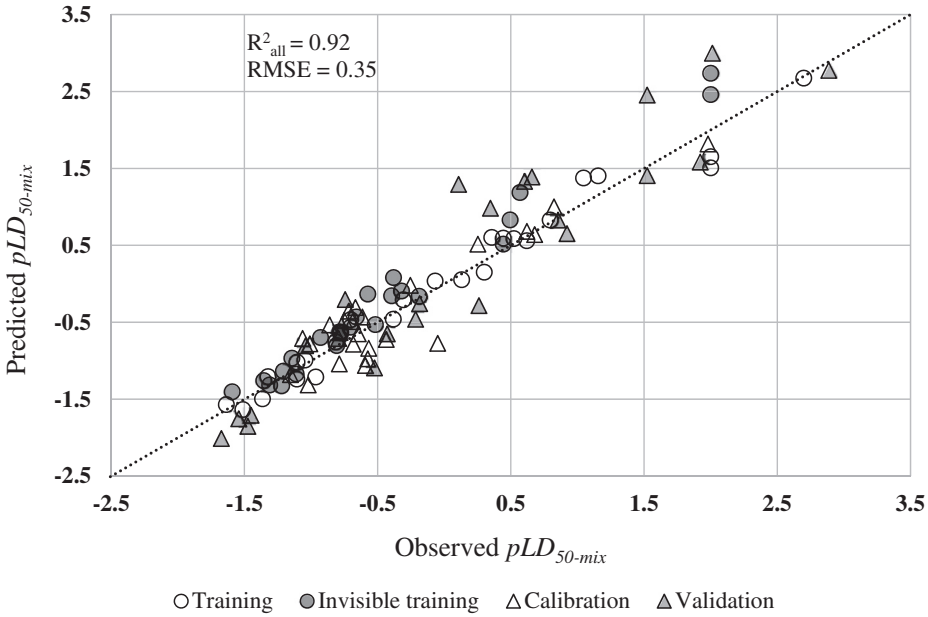


Fig. 3. Observed versus predicted log [1/$LD_{50\text{-mix}}$] of binary mixtures for the regression-based model (Approach B), split 1 "best split". R^2_{all} (determination coefficient) and RMSE (root mean squared error) are provided for all compounds (i.e. compounds from training, invisible training, calibration, and validation sets).

gism data. According to our results, the accuracy of the model in validation set is 0.96 and 0.87 in the calibration set. Similarly, specificity (=1.00), sensitivity (=0.86) and MCC (=0.90) for validation demonstrated robust results. To date, only classification models for single substances in honey bees are available in the literature such as Venko et al. (2017) and Como et al. (2017) presenting an accuracy = 0.84, sensitivity = 0.80, specificity = 0.86 and MCC = 0.67 (test set).

The logic behind the mechanistic interpretation of the classification model is based on three runs of Monte Carlo optimisation (see Section 3.2.1). Results on the classification model built up by means of the semi-correlation (Toropova et al., 2019b) are shown in Table 5. Similarly, the full list of correlation weights (CW) used for calculations of the classification model is provided in Table S6.

Table 4
Statistical quality of the classification-based model for binary mixtures effect (i.e. synergism/non-synergism).

Set	Statistical quality of model
Training	TP *= 5; TN = 19; FP = 0; FN = 0; N = 24 Sensitivity = 1.00 Specificity = 1.00 Accuracy = 1.00 MCC = 1.00
Invisible Training	TP = 5; TN = 18; FP = 0; FN = 1; N = 24 Sensitivity = 0.83 Specificity = 1.00 Accuracy = 0.96 MCC = 0.8885
Calibration	TP = 3; TN = 18; FP = 0; FN = 3; N = 24 Sensitivity = 0.50 Specificity = 1.00 Accuracy = 0.88 MCC = 0.66
Validation	TP = 6; TN = 18; FP = 0; FN = 1; N = 25 Sensitivity = 0.86 Specificity = 1.00 Accuracy = 0.96 **MCC = 0.90**

* TP, TN, FP and FN represent the number of true positives, true negatives, false positives and false negatives, respectively, in a confusion matrix.

3.3. Comparative assessment

3.3.1. Applicability of Approach A and Approach B

This manuscript presents three innovative QSAR models (two regression- and one classification-based) developed for the prediction of combined toxicity of binary mixtures in honey bees to date, notwithstanding that other models for predicting binary mixtures toxicity already exist in the literature (Kim et al., 2018; Qin et al., 2018; Muratov et al., 2012; Tian et al., 2013; Toropova et al., 2012;Wang et al., 2018a).

Two different approaches (A and B) have been applied and validated using the CORAL software for predicting both acute mixtures potency ($pLD_{50\text{-mix}}$) and the nature of combined toxicity (synergism/non-synergism). In particular, we demonstrated how depending on the availability of quantitative data on binary mixtures (e.g. dose-response, toxic unit, etc.), two different notation methods (traditional SMILES or quasi-SMILES) can be applied for developing robust regression- and classification-based QSAR models (Table 1).

Approach A (Section 3.1) allowed building up one regression-based QSAR model using as input the chemical structure of two-component mixtures codified as disconnected SMILES. Indeed, this is a simplistic representation of combined toxicity of binary mixture understanding, since the model does not take into account mixture ratio (e.g. toxic unit), thus ignoring the relative potency of each chemical contributing to the overall mixture toxicity (Bopp et al., 2015). Similar models have been already published in the literature for predicting binary mixtures toxicity in bacteria (Toropova et al., 2012; Wang et al., 2018a) and flammability of binary liquid mixtures (Toropova et al., 2019a). As a consequence, Approach A can be considered as valuable tool for the preliminary screening of chemical mixtures of concern in the case of lack of information on the potency of each component, by integrating available mechanistic and biological data on the specific target species (More et al., 2019; EFSA, 2013).

In contrast, Approach B considered additional quantitative information such as TUs for each chemical codified in quasi-SMILES (Toropov et al., 2018; Toropova et al., 2018). Hence, Approach B results of more interpretable than Approach A for pre-

Table 5
Mechanistic interpretation for categorical model "synergism / non-synergism" according to Approach B. Full list is provided in supplementary materials.

No.*	SA_k	CWs Run 1	CWs Run 2	CWs Run 3	N1	N2	N3	Dj
1	%.	1.67	1.07	1.58	24	24	24	0.000
2	=. . .1.	1.34	1.46	1.80	24	23	24	0.000
3	C. . .(.	0.40	0.25	0.29	24	24	24	0.000
4	C. . .2.	0.39	0.63	0.55	24	24	24	0.000
5	C. . . = ...	0.11	0.03	0.40	24	24	24	0.000
6	C. . .C.	0.24	0.37	0.43	24	24	24	0.000
7	^. . .%.	1.28	1.52	0.79	24	24	24	0.000
8	^.	0.89	0.85	1.59	24	24	24	0.000
9	^. . .C.	1.64	0.53	1.03	24	24	24	0.000
10	1. . .(.	0.52	0.58	0.93	23	22	22	0.001
11	5.	0.73	0.88	0.62	23	22	22	0.001
12	N. . .C.	0.58	0.37	0.39	21	15	18	0.003
13	3.	0.61	0.46	0.78	20	18	17	0.003
14	C. . .3.	0.46	0.06	0.59	20	18	17	0.003
15	1. . .%.	1.95	1.46	1.95	12	13	10	0.004
1	1.	−0.73	−0.30	−0.55	24	24	24	0.000
2	=.	−0.18	−0.08	−0.07	24	24	24	0.000
3	C.	−0.02	−0.30	−0.27	24	24	24	0.000
4	C. . .1.	−0.39	−0.42	−0.72	24	24	24	0.000
5	N. . .(.	−0.15	−0.29	−0.12	23	20	21	0.002
6	(. . .(.	−0.51	−0.31	−0.58	22	18	18	0.004
7	O. . . = ...	−0.54	−0.09	−1.05	22	22	20	0.002
8	3. . .(.	−0.12	−0.17	−0.46	20	18	17	0.003
9	C. . .#.	−0.39	−0.52	−0.47	16	10	14	0.003
10	N. . .#.	−0.41	−0.38	−0.42	16	10	14	0.003
11	F. . .(.	−0.06	−0.11	−0.32	15	9	12	0.005
12	Cl.0.1.	−0.54	−0.70	−0.50	10	7	8	0.005
13	4. . .%.	−0.42	−0.21	−0.23	5	5	2	0.018
14	Cl.0.2.	−0.39	−0.37	−0.37	5	3	4	0.005
15	2. . .%.	−0.28	−0.55	−0.41	4	3	7	0.011

* N1, N2, and N3 are number of SAk in the training, invisible training, and calibration sets, respectively. CW = Correleation Weights as in Eq. (1); Dj = SMILES-defect as in Eq. (16).

dicting (i) the potency of organic binary mixtures ($pLD_{50\text{-}mix}$) and (ii) the nature of combined toxicity (i.e. synergism / non-synergism). In addition, our results confirm that the statistical quality of regression-based model (Approach B) improved when including Toxic Unit as additional feature (Fig. 4). Regarding the classification-based model (Approach B), the scientific literature describes the effects of a given chemical mixture either as interactive (e.g. synergism, antagonism) or non-interactive (i.e. additive) when assuming concentration addition (CA) as default model (More et al., 2019). However, due to lack of experimental data, our classification model is based on two classes synergism (1) and non-synergism (0) (Toropova and Toropov, 2017). Nevertheless, this approach can result as highly conservative for predicting synergistic effects of binary mixture interactions. Indeed, a recent meta-analysis (Carnesecchi et al., 2019) confirmed that in honey bees, interactions were observed for 72% of cases as synergism and 28% as non-synergism (i.e. 17% additive, 11% antagonism). Similarly, such QSAR models can be applied to further refine current thresholds used in Model Deviation Ratio (MDR) and Estimated Mean Ratio (EMR) calculations (Belden et al., 2007; Carnesecchi et al., 2019; Cedergreen, 2014).

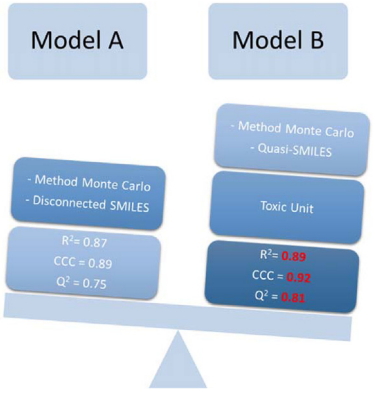


Fig. 4. Comparison of two QSAR regression models following Approaches A and B. Results of the statistical quality and statistical methods are presented. Results of the statistical quality are highlighted in red for the best model here developed. R^2- = determination coefficient; CCC = concordance correlation coefficient; Q^2 = cross-validated determination coefficient. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.3.2. SMILES attributes as drivers of binary mixture toxicity

Currently, most QSAR models investigating binary mixture toxicity provide qualitative or quantitative predictions of the toxicological endpoint without consideration of molecular feature(s) that may be responsible for an increase or a decrease in toxicity (Kim and Kim, 2015;Wang et al., 2018b). Here, this study demonstrates how the CORAL software can be applied to identify molecular features as drivers of binary mixture toxicity ($pLD_{50\text{-}mix}$) i.e. SMILES attributes (SAk) using Monte Carlo optimisation. According to our results (Tables S2–S4 and S6), the regression- and classification-based QSAR models (Approach A and B) showed that SMILES attributes with stable positive correlation weights (CWs) can be interpreted as promoters of an increase (synergism) in acute contact toxicity ($pLD_{50\text{-}mix}$) of the binary mixture in bees. In contrast, molecular features presenting negative CWs, from the Monte Carlo optimisation, can be interpreted as promoters of a decrease (antagonism) in the acute toxicity of the binary mixture. Generally speaking, it is recommended to avoid generating rules which are supported by few experimental data, and these can be identified by the software and qualified as "rare attributes". The attributes which are used by the QSAR models can be interpreted as associated to specific chemical features related to the acute contact toxicity of the binary mixture. From the statistical basis of the CORAL models, Fig. 5 illustrates examples of molecular features associated

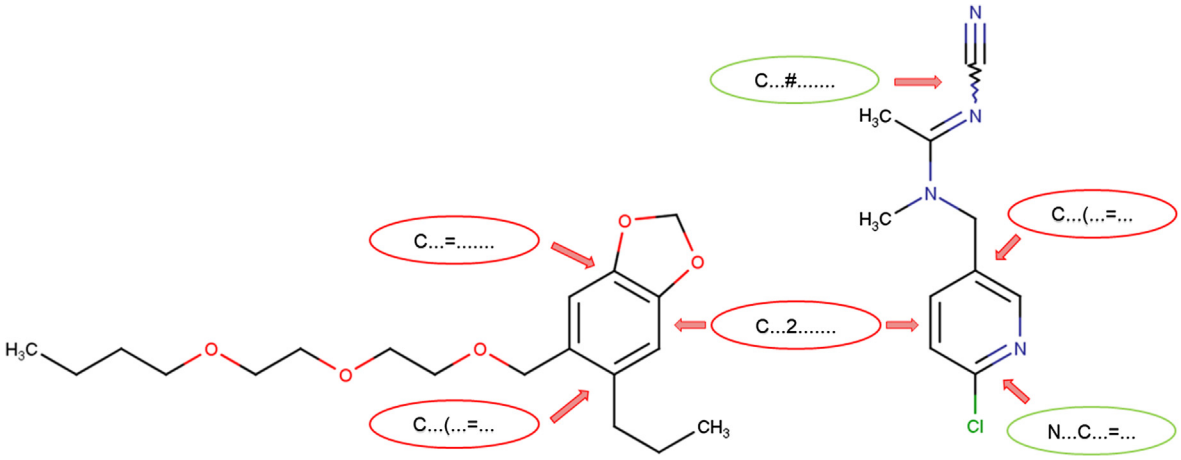


Fig. 5. Example of SMILES attributes associated with positive or negative correlation weights (CW) in binary mixture toxicity (piperonyl butoxide + acetamiprid). CW were demonstrated in two out of three the QSAR models following three runs of Monte Carlo optimisation. Red circles indicate SMILES attributes with positive correlation coefficient (i.e. increase combined toxicity). Green circles indicate SMILES attributes with negative correlation coefficient (i.e. decrease combined toxicity). [C. . .#. . ...] indicates the presence of carbon atom(s) connected with triple bond; [C... = ...] indicates the presence of carbon atom(s) connected with double bonds. [C...2.....] indicates the presence of two rings in the binary mixtures. [C...(... = ...] indicates the presence of carbon atom(s) with double bonds and branched chain. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

with either an increase or a decrease in the acute contact toxicity of the binary mixtures in bees. Overall, these innovative bee QSAR models developed within the CORAL software identify the most frequent molecular features that are associated with the binary mixture and its acute contact toxicity including the co-presence of functional groups associated with reactivity (e.g. double or triple bonds), the branching level of the molecule and steric components or the presence of certain atoms associated with polarity (e.g. nitrogen).

4. Conclusions and further perspectives

This manuscript has explored the development of innovative QSAR models for the prediction of acute contact toxicity of binary mixtures in bees through their calibration, validation and mechanistic interpretation:

- Two regression-based models predicting acute (contact) mixtures potency ($pLD_{50\text{-}mix}$) in a quantitative manner validated

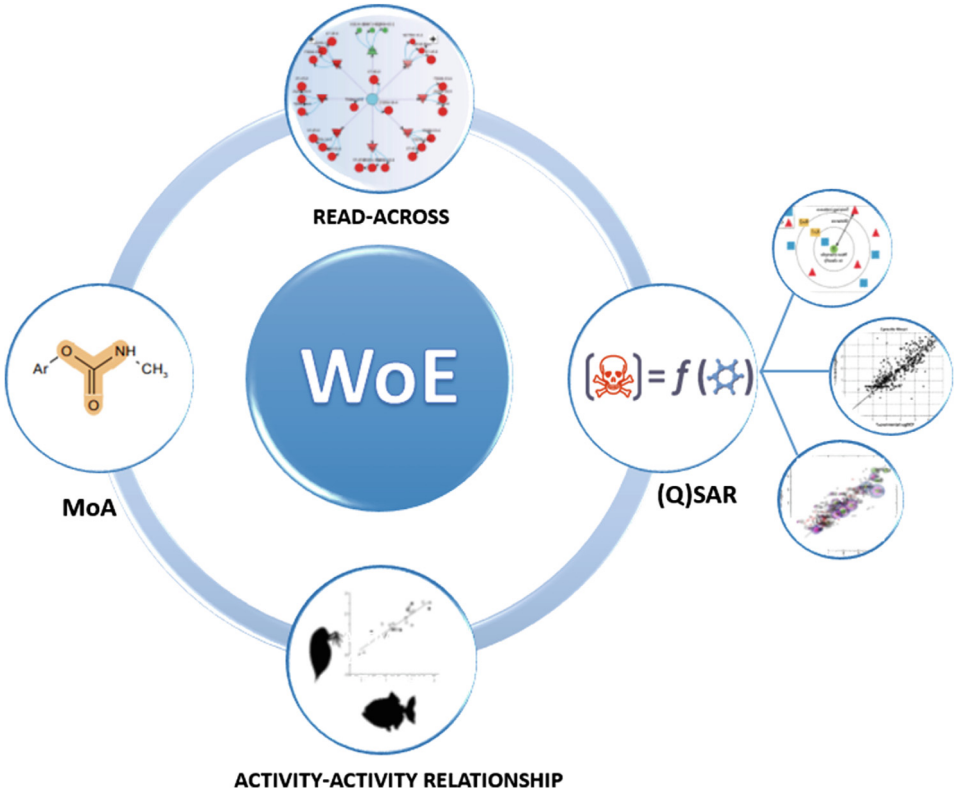


Fig. 6. Conceptual framework for integrating results from multiple *in silico* tools (Mode of Action, Read-across, (Q)SAR and Activity-Activity relationship) within a Weight of Evidence (WoE) approach (Hardy et al., 2017; Benfenati et al., 2019).

with statistical tests, which were demonstrated to have be reliable and robust (Approach A: R^2 = 0.87, CCC = 0.89, Q^2 = 0.75; Approach B: R^2 = 0.89, CCC = 0.92, Q^2 = 0.81).
- A classification-based model predicting the nature of the combined toxicity (synergism, non-synergism) validated using qualitative statistical validation parameters i.e. sensitivity (=0.86), specificity (=1.00), accuracy (=0.96), and Mattews correlation coefficient (MCC = 0.90). These models are currently being implemented within the VEGA-HUB platform (https://www.vegahub.eu/) as well as on CORAL software/databases (http://www.insilico.eu/coral/).

To date, such QSAR models were not previously available and this manuscript shows the potential use of *in silico* tools as part of New Approaches Methodologies (NAMs), particularly in the context of prioritisation and hazard assessment of potentially hazardous mixtures in honey bees (More et al., 2019). With regard to the chemical space of the models here developed, it is worth noting that the training sets are mostly constructed with available binary mixture data of limited toxicological and structural diversity (mostly conazole fungicides, pyrethroids and neonicotinoid insecticides) which reflects the limited applicability domain of these models. As a consequence, it is recommended to apply these QSAR models for predicting the combined toxicity ($pLD_{50\text{-}mix}$) of similar binary mixtures such as plant protection products (PPPs), veterinary drugs and their (potential) formulations in honey bees.

Finally, data gaps remain and still limit the development and broader applications of such QSAR models in honey bees:

- Acute contact toxicity data (e.g. mortality) are the major available datasets for a significant number of compounds and their mixtures but chronic toxicity oral data and toxicity data for sub-lethal effects in honey bees and wild bees are still lacking (Carnesecchi et al., 2019).
- Experimental toxicokinetic data (e.g. half-life, bioaccumulation) are also lacking particularly to develop ad-hoc QSAR models for further characterisation of the impact of persistence on combined toxicity of multiple chemicals in bees (including interactions). With such dataset, QSAR models can be integrated with Dynamic Energy Budget models to provide a refined understanding of combined toxicity at the honey bee population level (Spurgeon et al., 2017; Hesketh et al., 2016).
- Although QSAR models for the prediction of the Mode of Action (MoA) of chemicals have been developed for aquatic species (Kienzler et al., 2019), the integration of such (qualitative) information into QSAR models for predicting mixture toxicity results challenging. In this context, the ongoing EFSA's OpenFoodTox 2.0 "OptiTox" project (Fig. 6) aims to develop new predictive tools and integrate results from multiple *in silico* methods ((Q)SAR, MoA predictors, read-across and Activity-Activity relationship within a Weight of Evidence (WoE) strategy (Hardy et al., 2017; Benfenati et al., 2019).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The views expressed in this manuscript do not reflect the views of the European Food Safety Authority and are the authors only. This work was supported by the European Food Safety Authority (EFSA) [contract number: OC/EFSA/SCER/2018/01 and NP/EFSA/AFSCO/2016/02 (Edoardo Carnesecchi)]; and LIFE-VERMEER (LIFE16 ENV/IT/000167). Authors would like to thank Cosima Toma (European Chemicals Agency) for constructive criticism of the manuscript and Matteo Sironi (Istituto di Ricerche Farmacologiche Mario Negri IRCCS) for the graphic design support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2019.135302.

References

Backhaus, T., Faust, M., 2012. Predictive environmental risk assessment of chemical mixtures: a conceptual framework. Environ. Sci. Technol. 10.1021/es2034125.
Bargańska, Ż., Ślebioda, M., Namieśnik, J., 2016. Honey bees and their products: bioindicators of environmental contamination. Crit. Rev. Environ. Sci. Technol. 46, 235–248. https://doi.org/10.1080/10643389.2015.1078220.
Belden, J.B., Gilliom, R.J., Lydy, M.J., 2007. How well can we predict the toxicity of pesticide mixtures to aquatic life?. Integr. Environ. Assess. Manag. 3, 364–372. https://doi.org/10.1002/ieam.5630030307.
Benfenati, E., Chaudhry, Q., Gini, G., Dorne, J.L., 2019. Integrating in silico models and read-across methods for predicting toxicity of chemicals: a step-wise strategy. Environ. Int. 10.1016/j.envint.2019.105060.
Bopp, S., Berggren, E., Kienzler, A., van der Linden, S., Worth, A., 2015. Scientific methodologies for the assessment of combined effects of chemicals a survey and literature review. Publications Office of the European Union. doi: 10.2788/093511.
Carnesecchi, E., Svendsen, C., Lasagni, S., Grech, A., Quignot, N., Amzal, B., Toma, C., Tosi, S., Rortais, A., Cortinas-Abrahantes, J., Capri, E., Kramer, N., Benfenati, E., Spurgeon, D., Guillot, G., Dorne, J.L.C.M., 2019. Investigating combined toxicity of binary mixtures in bees: Meta-analysis of laboratory tests, modelling, mechanistic basis and implications for risk assessment. Environ. Int. 133. https://doi.org/10.1016/j.envint.2019.105256.
Cedergreen, N., 2014. Quantifying synergy: a systematic review of mixture toxicity studies within environmental toxicology. PLoS One 9,. https://doi.org/10.1371/journal.pone.0096580 e96580.
Chirico, N., Gramatica, P., 2012. Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. J. Chem. Inf. Model. 10.1021/ci300084j.
Como, F., Carnesecchi, E., Volani, S., Dorne, J.L., Richardson, J., Bassan, A., Pavan, M., Benfenati, E., 2017. Predicting acute contact toxicity of pesticides in honeybees (Apis mellifera) through a k-nearest neighbor model. Chemosphere. https://doi.org/10.1016/j.chemosphere.2016.09.092.
Dao, P., Wang, K., Collins, C., Ester, M., Lapuk, A., Sahinalp, S.C., 2011. Optimally discriminative subnetwork markers predict response to chemotherapy. Bioinformatics. https://doi.org/10.1093/bioinformatics/btr245.
EFSA (European Food Safety Authority), 2009. Risk assessment for birds and mammals. EFSA J. https://doi.org/10.2903/j.efsa.2009.1438.
EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2013. Guidance on tiered risk assessment for Plant Protection Products for aquatic organisms in edge-of-field surface waters. EFSA J. 11 (7), 3290, 268 pp. doi:10.2903/j.efsa.2013.3290.
EFSA PPR Panel (EFSA Panel on Plant Protection Products and their Residues), 2012. Scientific Opinion on the science behind the development of a risk assessment of Plant Protection Products on bees (Apis mellifera, Bombus spp. and solitary bees). EFSA J. 10 (5), 2668, 275 pp. doi:10.2903/j.efsa.2012.2668.
EFSA (European Food Safety Authority), 2013. EFSA Guidance Document on the risk assessment of plant protection products on bees (Apis mellifera, Bombus spp. and solitary bees). EFSA J. 11 (7), 3295. Doi: 10.2903/j.efsa.2013.3295 (268 pp.).
EFSA (European Food Safety Authority), 2014. Towards an integrated environmental risk assessment of multiple stressors on bees: review of research projects in Europe, knowledge gaps and recommendations. EFSA J. 12 (3), 3594, 102 pp. doi:10.2903/j.efsa.2014.3594.
EFSA Panel on Animal Health and Welfare (AHAW), 2016. EFSA Panel on Animal Health and Welfare (AHAW) (2016). Assessing the health status of managed honeybee colonies (HEALTHY-B): a toolbox to facilitate harmonised data collection. 10.2903/j.efsa.2016.4578.
EFSA (European Food Safety Authority), 2017. Specifications for field data collection contributing to honey bee model corroboration and verification. EFSA supporting publication 2017:EN-234.54pp. doi:10.2903/sp.efsa.2017.EN-1234.
Gadaleta, D., Catto, M., 2016. Applicability domain for QSAR models: where theory meets reality. Int. J. Quant. Struct. Relat. 1. https://doi.org/10.4018/IJQSPR.2016010102.
Hamadache, M., Benkortbi, O., Hanini, S., Amrane, A., 2018. QSAR modeling in ecotoxicological risk assessment: application to the prediction of acute contact toxicity of pesticides on bees (Apis mellifera L.). Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-017-0498-9.
Hardy, A., Benford, D., Halldorsson, T., Jeger, M.J., Knutsen, H.K., More, S., Naegeli, H., Noteborn, H., Ockleford, C., Ricci, A., Rychen, G., Schlatter, J.R., Silano, V., Solecki, R., Turck, D., Benfenati, E., Chaudhry, Q.M., Craig, P., Frampton, G., Greiner, M., Hart, A., Hogstrand, C., Lambre, C., Luttik, R., Makowski, D., Siani, A.,

Wahlstroem, Aguilera, J., Dorne, J., Fernandez Dumont, A., Hempen, M., Valtueña Martínez, S., Martino, L., Smeraldi, C., Terron, A., Georgiadis, N., Younes, M., 2018b. Guidance on the use of the weight of evidence approach in scientific assessments. EFSA J. https://doi.org/10.2903/j.efsa.2017.4971.

Hesketh, H., Lahive, E., Horton, A.A., Robinson, A.G., Svendsen, C., Rortais, A., Dorne, J.L., Baas, J., Spurgeon, D.J., Heard, M.S., 2016. Extending standard testing period in honeybees to predict lifespan impacts of pesticides and heavy metals using dynamic energy budget modelling. Sci. Rep. 10.1038/srep37655.

Hunter, R.S., Culver, F.D., Fitzgerald, A. 1987. SMILES User Manual. A Simplified Molecular Input Line Entry System. Includes extended SMILES for defining fragments. Review Draft, Internal Report, Montana State University, Institute for Biological and Chemical Process Control (IPA), Bozeman, MT.

Iwasa, T., Motoyama, N., Ambrose, J.T., Roe, R.M., 2004. Mechanism for the differential toxicity of neonicotinoid insecticides in the honey bee, Apis mellifera. Crop Prot. 23, 371–378. https://doi.org/10.1016/j.cropro.2003.08.018.

Johnson, R.M., Mao, W., Pollock, H.S., Niu, G., Schuler, M.A., Berenbaum, M.R., 2012. Ecologically appropriate xenobiotics induce cytochrome P450s in Apis Mellifera. PLoS ONE 7, 1–9. https://doi.org/10.1371/journal.pone.0031051.

Johnson, R.M., Dahlgren, L., Siegfried, B.D., Ellis, M.D., 2013. Acaricide, fungicide and drug interactions in honey bees (Apis mellifera). PLoS ONE 8. https://doi.org/10.1371/journal.pone.0054092.

Jonker, M.J., Svendsen, C., Bedaux, J.J.M., Bongers, M., Kammenga, J.E., 2005. Significance testing of synergistic/antagonistic, dose level-dependent, or dose ratio-dependent effects in mixture dose-response analysis. Environ. Toxicol. Chem. 10.1897/04-431R.1.

KEMI (The Swedish Chemicals Agency), 2015. An additional assessment factor (MAF) – A suitable approach for improving the regulatory risk assessment of chemical mixtures? Swedish Chemicals Agency. Print: Arkitektkopia, Stockholm 2015. ISSN 0284-1185. Article number: 361 159.

Kienzler, A., Bopp, S.K., van der Linden, S., Berggren, E., Worth, A., 2016. Regulatory assessment of chemical mixtures: requirements, current approaches and future perspectives. Regul. Toxicol. Pharmacol. https://doi.org/10.1016/j.yrtph.2016.05.020.

Kienzler, A., Connors, K.A., Bonnell, M., Barron, M.G., Beasley, A., Inglis, C.G., Norberg-King, T.J., Martin, T., Sanderson, H., Vallotton, N., Wilson, P., Embry, M. R., 2019. Mode of action (MOA) classifications in the EnviroTox database: development and implementation of a consensus MOA classification. Environ. Toxicol. Chem. 10.1002/etc.4531.

Kim, J., Kim, S., 2015. State of the art in the application of QSAR techniques for predicting mixture toxicity in environmental risk assessment. SAR QSAR Environ. Res. 26, 41–59. 10.1080/1062936X.2014.984627.

Kim, J., Fischer, M., Helms, V., 2018. Prediction of synergistic toxicity of binary mixtures to Vibrio fischeri based on biomolecular interaction networks. Chem. Res. Toxicol. https://doi.org/10.1021/acs.chemrestox.8b00164.

Klein, A.-M., Vaissière, B.E., Cane, J.H., Steffan-Dewenter, I., Cunningham, S.A., Kremen, C., Tscharntke, T., 2007. Importance of pollinators in changing landscapes for world crops. Proc. R. Soc. B Biol. Sci. 274, 303–313. https://doi.org/10.1098/rspb.2006.3721.

Lambert, O., Veyrand, B., Durand, S., Marchand, P., Le Bizec, B., Piroux, M., Puyo, S., Thorin, C., Delbac, F., Pouliquen, H., 2012. Polycyclic aromatic hydrocarbons: bees, honey and pollen as sentinels for environmental chemical contaminants. Chemosphere 86, 98–104. https://doi.org/10.1016/j.chemosphere.2011.09.025.

More, S.J., Hardy, A., Bampidis, V., Benford, D., Hougaard Bennekou, S., Bragard, C., Boesten, J., Halldorsson, T.I., Hernández-Jerez, A.F., Jeger, M.J., Knutsen, H.K., Koutsoumanis, K.P., Naegeli, H., Noteborn, H., Ockleford, C., Ricci, A., Rychen, G., Schlatter, J.R., Silano, V., Nielsen, S.S., Schrenk, D., Solecki, R., Turck, D., Younes, M., Benfenati, E., Castle, L., Cedergreen, N., Laskowski, R., Leblanc, J.C., Kortenkamp, A., Ragas, A., Posthuma, L., Svendsen, C., Testai, E., Dujardin, B., Kass, G.E.N., Manini, P., Zare Jeddi, M., Dorne, J.L.C., Hogstrand, C., 2019. Guidance on harmonised methodologies for human health, animal health and ecological risk assessment of combined exposure to multiple chemicals. EFSA J. https://doi.org/10.2903/j.efsa.2019.5634.

Muratov, E.N., Varlamova, E. V., Artemenko, A.G., Polishchuk, P.G., Kuz'Min, V.E., 2012. Existing and developing approaches for QSAR analysis of mixtures. Mol. Inform. 10.1002/minf.201100129.

Nys, C., Van Regenmortel, T., Janssen, C.R., Oorts, K., Smolders, E., De Schamphelaere, K.A.C., 2018. A framework for ecological risk assessment of metal mixtures in aquatic systems. Environ. Toxicol. Chem. 10.1002/etc.4039.

OECD, 1998. Test No. 213: Honeybees, acute oral toxicity test. OECD Guidel. Test. Chem. 10.1787/9789264070165-en.

OECD, 2007. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models, ENV/JM/MONO(2007)2. OECD Publishing, Paris.

Qin, L.T., Chen, Y.H., Zhang, X., Mo, L.Y., Zeng, H.H., Liang, Y.P., 2018. QSAR prediction of additive and non-additive mixture toxicities of antibiotics and pesticide. Chemosphere. https://doi.org/10.1016/j.chemosphere.2018.01.142.

Rortais, A., Arnold, G., Lou Dorne, J., More, S.J., Sperandio, G., Streissl, F., Szentes, C., Verdonck, F., 2017. Risk assessment of pesticides and other stressors in bees: Principles, data gaps and perspectives from the European Food Safety Authority. Sci. Total Environ. 587–588, 524–537. https://doi.org/10.1016/j.scitotenv.2016.09.127.

Roy, K., Mitra, I., Kar, S., Ojha, P.K., Das, R.N., Kabir, H., 2012. Comparative studies on some metrics for external validation of QSPR models. J. Chem. Inf. Model. https://doi.org/10.1021/ci200520g.

Singh, K.P., Gupta, S., Basant, N., Mohan, D., 2014. QSTR modeling for qualitative and quantitative toxicity predictions of diverse chemical pesticides in honey bee for regulatory purposes. Chem. Res. Toxicol. 10.1021/tx500100m.

Skorbiłowicz, M., Skorbiłowicz, E., Cieśluk, I., 2018. Bees as bioindicators of environmental pollution with metals in an urban area. J. Ecol. Eng. 19, 229–234. https://doi.org/10.12911/22998993/85738.

Spurgeon, D., Hesketh, H., Lahive, E., Svendsen, C., Baas, J., Robinson, A., Horton, A., Heard, M., 2017. Chronic oral lethal and sub-lethal toxicities of different binary mixtures of pesticides and contaminants in bees (Apis mellifera, Osmia bicornis and Bombus terrestris). EFSA Support. Publ. 13. https://doi.org/10.2903/sp.efsa.2016.en-1076.

Thoreau, F., 2016. 'A mechanistic interpretation, if possible': how does predictive modelling causality affect the regulation of chemicals? Big Data Soc. https://doi.org/10.1177/2053951716670189.

Tian, D., Lin, Z., Yin, D., 2013. Quantitative structure activity relationships (QSAR) for binary mixtures at non-equitoxic ratios based on toxic ratios-effects curves. Dose-Response. 10.2203/dose-response.11-042.Lin.

Toropov, A.A., Toropova, A.P., Lombardo, A., Roncaglioni, A., De Brita, N., Stella, G., Benfenati, E., 2012a. CORAL: the prediction of biodegradation of organic compounds with optimal SMILES-based descriptors. Cent. Eur. J. Chem. 10.2478/s11532-012-0031-4.

Toropov, A.A., Raška, I., Toropova, A.P., Raškova, M., Veselinović, A.M., Veselinović, J. B., 2019. The study of the index of ideality of correlation as a new criterion of predictive potential of QSPR/QSAR-models. Sci. Total Environ. 659, 1387–1394. https://doi.org/10.1016/j.scitotenv.2018.12.439.

Toropov, A.A., Toropova, A.P., Benfenati, E., Diomede, L., Salmona, M., 2018. Use of quasi-SMILES to model biological activity of "micelle–polymer" samples. Struct. Chem. 10.1007/s11224-018-1115-3.

Toropov, A.A., Toropova, A.P., Rasulev, B.F., Benfenati, E., Gini, G., Leszczynska, D., Leszczynski, J., 2012b. CORAL: binary classifications (active/inactive) for liver-related adverse effects of drugs. Curr. Drug Saf. 7, 257–261.

Toropova, A.P., Toropov, A.A., Benfenati, E., Gini, G., Leszczynska, D., Leszczynski, J., 2012. CORAL: Models of toxicity of binary mixtures. Chemom. Intell. Lab. Syst. 10.1016/j.chemolab.2012.10.001.

Toropova, A.P., Toropov, A.A., Carnesecchi, E., Benfenati, E., Dorne, J.L., 2019a. The index of ideality of correlation: models for flammability of binary liquid mixtures. Chem. Pap. Doi: 10.1007/s11696-019-00903-w.

Toropova, A.P., Toropov, A.A., 2017. CORAL: Binary classifications (active/inactive) for drug-induced liver injury. Toxicol. Lett. 268, 51–57. https://doi.org/10.1016/j.toxlet.2017.01.011.

Toropova, A.P., Toropov, A.A., 2018. Use of the index of ideality of correlation to improve models of eco-toxicity. Environ. Sci. Pollut. Res. https://doi.org/10.1007/s11356-018-3291-5.

Toropova, A.P., Toropov, A.A., Benfenati, E., Castiglioni, S., Bagnati, R., Passoni, A., Zuccato, E., Fanelli, R., 2018. Quasi-SMILES as a tool to predict removal rates of pharmaceuticals and dyes in sewage. Process Saf. Environ. Prot. 10.1016/j.psep.2018.07.003.

Toropova, A.P., Toropov, A.A., Benfenati, E., 2019b. Semi-correlations as a tool to build up categorical (active/inactive) model of GABA A receptor modulator activity. Struct. Chem. 30, 853–861. https://doi.org/10.1007/s11224-018-1226-x.

Tosi, S., Costa, C., Vesco, U., Quaglia, G., Guido, G., 2018. A 3-year survey of Italian honey bee-collected pollen reveals widespread contamination by agricultural pesticides. Sci. Total Environ. 615, 208–218. https://doi.org/10.1016/j.scitotenv.2017.09.226.

U.S. EPA (Environmental Protection Agency), 2014. Guidance for assessing pesticide risks to bees. Office of Research and Development, National Center for Environmental Assessment, Washington, DC 20460. Available at: https://www.epa.gov/sites/production/files/2014-06/documents/pollinator_risk_assessment_guidance_06_19_14.pdf.

U.S. EPA (Environmental Protection Agency), 2003. Framework for Cumulative Risk Assessment. Office of Research and Development, National Center for Environmental Assessment, Washington, DC 20460. EPA/630/P-02/001F. Available at: https://www.epa.gov/sites/production/files/2014-11/documents/frmwrk_cum_risk_assmnt.pdf.

VanEngelsdorp, D., Meixner, M.D., 2010. A historical review of managed honey bee populations in Europe and the United States and the factors that may affect them. J. Invertebr. Pathol. 103, S80–S95. https://doi.org/10.1016/j.jip.2009.06.011.

Venko, K., Drgan, V., Župerl, Š., Vračko, M., Novič, M., 2017. In silico evaluation of toxicity towards honey bees with QSAR models. Toxicol. Lett. https://doi.org/10.1016/j.toxlet.2017.07.787.

Wade, A., Lin, C.-H., Kurkul, C., Regan, E.R., Johnson, R.M., 2019. Combined toxicity of insecticides and fungicides applied to California almond orchards to honey bee larvae and adults. Insects 10, 20. https://doi.org/10.3390/insects10010020.

Wang, D., Wu, X., Lin, Z., Ding, Y., 2018a. A comparative study on the binary and ternary mixture toxicity of antibiotics towards three bacteria based on QSAR investigation. Environmental Research. https://doi.org/10.1016/j.envres.2017.12.015.

Wang, T., Tang, L., Luan, F., Cordeiro, M.N., 2018b. Prediction of the toxicity of binary mixtures by qsar approach using the hypothetical descriptors. International Journal of Molecular Sciences. https://doi.org/10.3390/ijms19113423.