Journal of *Xenobiotics*

MDPI

Article

Simulation of the Impact of Pesticides on Pollinators Under Different Conditions Using Correlation Weighting of Quasi-SMILES Components Together with the Index of Ideality of Correlation (IIC)

Alla P. Toropova *, Andrey A. Toropov, Sofia Mescieri, Alessandra Roncaglioni and Emilio Benfenati

Department of Environmental Health Sciences, Istituto di Ricerche Farmacologiche Mario Negri IRCCS, Via Mario Negri 2, 20156 Milan, Italy; andrey.toropov@marionegri.it (A.A.T.); sofia.mescieri@marionegri.it (S.M.); alessandra.roncaglioni@marionegri.it (A.R.); emilio.benfenati@marionegri.it (E.B.)

* Correspondence: alla.toropova@marionegri.it; Tel.: +39-02-3901-4595

Abstract

Background: Pesticide toxicity to insects is an important adverse effect with a potentially large ecological impact when considering the effect on beneficial insects, as pollinators. The assessment of this endpoint is necessary to avoid applying ecologically dangerous pesticides. *Aim of the study:* Assessment of the availability of the Monte Carlo method for the development of a model for toxicity (pLD50) towards bees and other pollinators. In addition, the index of ideality of correlation is examined as a possibility to increase the statistical quality of quantitative structure–activity relationships (QSARs) for the toxicity of pesticides to pollinators. *Main results and novelty:* models with good performance on the toxic effect of pesticides towards different pollinators, wrapping acute and chronic effects, using the Monte Carlo method for QSAR analysis.

Keywords: pollinators; CORAL software; Las Vegas algorithm; Monte Carlo method; pesticide toxicity; QSAR

1. Introduction

Pollinators are essential components of terrestrial ecosystems and global agriculture, as a large proportion of crop species depend on insect-mediated pollination [1–8]. By ensuring the reproduction of wild and cultivated plants, they sustain biodiversity, ecosystem functioning and food security [2,4,5]. In Europe, 80% of the crops depend on pollinators, with an economic output of at least 5 billion euros [9]. Despite their ecological and economic importance, pollinators are increasingly exposed to multiple stressors, including habitat loss, pathogens, climate change and agrochemical use, among which pesticide exposure is recognized as a major driver of decline worldwide [8]. For these reasons, the protection of pollinators has become a growing priority in environmental risk assessment and the European Commission established an Action Plan to protect pollinators [9]. Honey bees are commonly used as sentinel organisms in studies of pesticide effects on pollinators [10,11], because they are widely distributed, relatively simple to monitor and have a concise and well-understood biological cycle [8,12,13]. However, the pollinator community is much broader, and other wild pollinators, such as *Bombus terrestris* spp., *Osmia* spp. and *Megachile rotundata*, play an equally crucial role in the maintenance of biodiversity, ecosystem functioning and agricultural productivity. Recent studies have emphasized

Check for updates

Academic Editors: Mohamed A. Ghorab and Shaohua Chen

Received: 25 November 2025
Revised: 2 January 2026
Accepted: 6 January 2026
Published: 8 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

the importance of extending pesticide toxicity assessments beyond honey bees to other species, in order to reduce existing data gaps and improve the ecological relevance of risk assessments [8,10,11,14].

Since experimental toxicity testing is costly and requires in vivo procedures, in silico approaches, such as quantitative structure–activity relationship (QSAR) models, have been recommended to estimate pesticide toxicity, and indeed QSAR models have been developed for honey bees, in particular addressing contact exposure [15–23]. The regression models [24–27] and models that include classification [28], distinguishing substances as toxic and non-toxic [29–33] using a threshold value, for instance, can predict the dose producing the adverse effect. The models used a range of molecular descriptors and algorithms. However, the studies on pollinators other than honey bees are very limited [20,33,34]. This scarcity of studies may be partly due to the fact that OECD test guidelines for bumblebees (adopted 2017) [35,36] and solitary bees such as Mason Bee (*Osmia* sp.) (adopted 2025) [37] are relatively recent and also to the limited availability of toxicity data for these pollinators compared to *Apis mellifera* [10,11,14]. Therefore, predictive models should not be limited to honey bees but should also include other pollinators to ensure a more effective pesticide risk assessment. The objective of this study is therefore to develop a QSAR model using toxicological data from different pollinator species by integrating experimental toxicity values from the databases with predicted environmental fate characteristics in order to improve ecological risk assessment. Furthermore, we addressed not only acute toxicity and effects towards adults and larvae, in order to have better tools to protect pollinators.

2. Materials and Methods

2.1. Data

A dataset containing 541 data points of chemicals tested on pollinators was compiled from the ECOTOX database [38] and the EFSA OpenFoodTox database [39]. Only studies reporting food intake (oral exposure) of active ingredients or mono-constituents with a purity of at least 80% were included. Preference was given to studies conducted following standardized OECD guidelines for oral acute [36,40] and chronic tests [41]: acute tests with exposure durations between 24 and 96 h and chronic tests of 10 days. In addition, subchronic tests of 5–8 days and chronic tests extending up to 14 days were also included to increase dataset coverage, particularly for non-Apis pollinators, for which experimental information is limited. All values were harmonized in µg/organism. The dataset includes acute and chronic oral toxicity values for adult and larval pollinators (*Apis mellifera*, *Apis mellifera linguistica*, *Apis mellifera carnica*, *Bombus terrestris*, *Bombus terrestris audax*, *Megachile rotundata*, *Osmia excavata*, *Osmia lignaria* and *Osmia rufa*), with exposure durations ranging from 1 to 14 days, expressed in µg/organism. Values indicated by the qualifiers ">" or "<" were used simply by considering their numerical value, in order to maximize the number of available data for non-Apis pollinators, where experimental data are generally scarce. Among these, 540 entries correspond to LD_{50} values and 1 to an LC_{50} value. We used canonical SMILES to verify the presence of duplicates. We removed a total of 159 duplicates, obtaining a dataset of 382 substances. The single LC_{50} data point in our dataset was not retained for model construction, because during the duplicate-removal step for the same molecule, a lower LD_{50} value was available.

Toxicity values (LD_{50} expressed in µg/organism) were converted into negative decimal logarithms (pLD_{50}).

Physicochemical and environmental properties were predicted using VEGA (version 1.2.4) and JANUS (version 1.0.3) software, both available from VEGAHUB [42], including persistence in water, soil and sediment, expressed in days.

In this way, we obtained a dataset and generated quasi-SMILES for the substances [43,44]. The SMILES (Simplified Molecular-Input Line-Entry System) is a widely used format for processing chemical substances in QSAR. Quasi-SMILES contains the SMILES (preprocessed with VEGA, to obtain canonical, standardized SMILES) with additional labels, relative to the features as in Table 1. These features contain a label to specify the persistence of the substance in different compartments, the pollinator species, the life stage and the exposure duration.

Table 1. Features of quasi-SMILES used for modelling.

CODE	COMMENT
WT	Persistence in water [days]
SS	Persistence in sediment [days]
SO	Persistence in soil [days]
SP	Species
LS	Life stage
OB	Observation duration [days]

Figure 1 contains an example of quasi-SMILES construction. This is quite a clear action which is aimed at providing a simulation system by adding information about phenomena considered. In fact, the majority of phenomena in general and QSAR phenomena in particular are characterized by the Hilbert space of experimental (observed) conditions. In practice, it is impossible to separate the important and non-important conditions and circumstances. However, it is clear that the more circumstances involved in the simulation process, the higher the probability of obtaining more robust results.

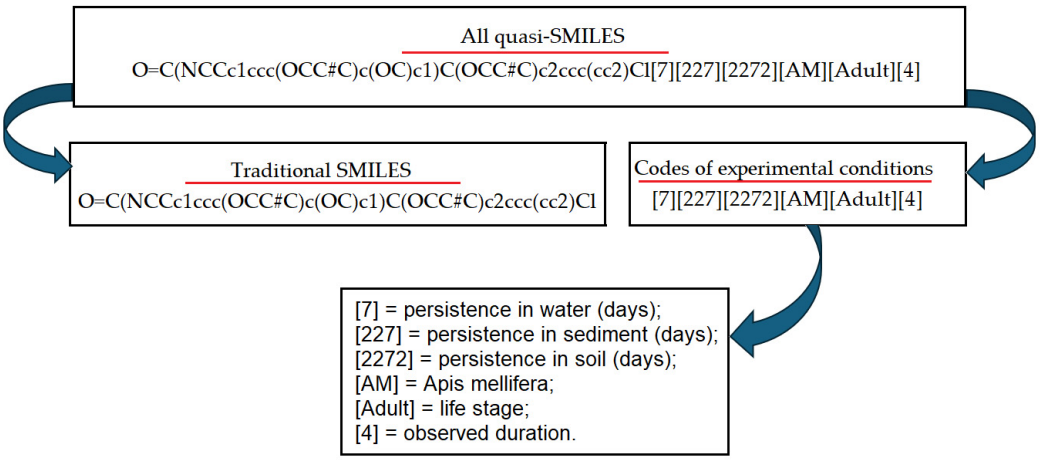

Figure 1. The scheme of quasi-SMILES construction. Underlines demonstrate the logic of quasi SMILES.

But, on the other side, the large number of circumstances involved in the simulation process can lead to the extraordinary complexity of a model. In order to avoid this situation, it is necessary to limit the system at the basis of the simulation to a reasonable number of considered conditions. Table 1 contains the list of experimental and available considerations in the process of simulating the model for pesticide toxicity towards pollinators. Preliminary computational experiments have indicated that various kinds of persistence (in water, sediment and soil) are able to serve as a reliable basis for developing corresponding models for the toxicity of pesticides.

2.2. Simulation Scheme

The basic steps of building a model within the framework of the used method are as follows: (1) Splitting into active and passive training sets, calibration set, and validation

set. (2) Optimizing the correlation weights of molecular features extracted from quasi-SMILES. (3) Building a regression model linking the descriptor calculated by the correlation weights with the endpoint under study (toxicity for bee and bumblebee). (4) Validating the predictive potential of the model.

Within step (1), the set of all compounds was randomly divided into four subsets of approximately the same number of quasi-SMILES: (i) an active training (≈25%), (ii) a passive training (≈25%), (iii) a calibration (≈25%) and (iv) a validation (≈25%). It should be noted that the active and passive training sets, together with the calibration set, are the training set in the traditional approach. However, in practical terms, the calibration set was used to generate the final, optimized model, while the active and passive training sets were used only in the initial steps, and thus, the models obtained with these subsets are preliminary ones. In step (2), the active training set was used to calculate the correlation weights, and then the passive training set was used as an inspector of the correlation weights supplied by the active training set. In step (3), the calibration set was used to check for overfitting. Finally, in step (4), the external validation set was used to evaluate the predictive potential of the resulting model. Since the validation set contains substances that are not visible in the calculation of the correlation weights, this step is useful to estimate the observed predictive potential. This process was repeated five times to obtain a reliable statistical basis, creating five random splits [45].

We clarify that, as described above, the set of quasi-SMILES has been randomly split into 4 subsets: active training set, passive training set, calibration set, and the validation set. None of the quasi-SMILES is present in more than one of these four subsets.

It should be noted that the approach under consideration, which involves considering multiple groups of splits for training and validation sets, requires a tool to verify that the splits under consideration are distinct. For this purpose, data on the identity of the splits under consideration is used. Table 2 presents the information, separated into the training and validation parts of the list, which is fully available for analyzing data on the used organic compounds. This means that the same quasi-SMILES can be present in both splits 1 and 2, for instance, but it cannot be present twice in the same split, as we wrote.

Table 2. Percentage of identity for the five studied splits into the training and validation set.

	1	**2**	**3**	**4**	**5**
1	100	39.4 *	34.4	44.8	36.1
2	40.2	100	42.9	38.7	37.3
3	32.5	36.5	100	34.7	45.8
4	33.5	44.8	41.2	100	38.5
5	45.3	42.9	35.2	37.3	100

* If i > j, then the matrix element [i,j] means the percentage of identity for the active training sets; if i < j, then the matrix element [i,j] means the percentage of identity for the validation sets (external sets). The i (i.e., for the columns) and j (for the rows) mean the numbering of the 5 splits examined.

2.3. Optimal Descriptors

The optimal descriptors are the correlation weights for all non-rare quasi-SMILES attributes, summed within a single equation. The correlation weights are (i) attributes of traditional SMILES, and (ii) attributes indicated experimental conditions collected in Table 1. The non-rare quasi-SMILES are identified within the modelling development, as described below.

2.4. Optimization of Correlation Weights

The correlation weights of quasi-SMILES attributes are calculated using the CORAL software (version CORALSEA-2025) developed by Istituto Mario Negri, Milano, Italy [46].

The optimization process used here includes a special algorithm called Index of Ideality of Correlation (IIC) [47,48]. The algorithm avoids "local" situations, not valid from a general point of view. The initial modelling steps, as implemented using the active and passive sets, start from specific cases and can be tailored to the local situation. To obtain more general lessons, the prediction is improved using the calibration set, even if the statistics on the calibration set may be lower than those on the active and passive training sets (which may show higher values for overfitting). The overfitting can be controlled by using the weighting coefficients for the IIC. The choice of the coefficient is made empirically, based on the results of preliminary observations of the stochastic optimization system with different weights for the IIC.

Having the numerical data on the correlation weights, the considered toxicity to pollinators can be calculated using Equation (1):

$$pLD_{50} = C_0 + C_1 \times DCW(T,N) \quad (1)$$

C_0 and C_1 are regression coefficients; T is the threshold used to define rare and non-rare SMILES attributes. The SMILES attribute is considered rare if its frequency in the active training set is less than T (rare attributes are not considered; their correlation weights are equal to zero). N is the number of epochs of Monte Carlo optimization. In this study, T = 3 and N = 15 were obtained from computational experiments within the modelling optimization, as in Equation (2).

$$DCW(T,N) = \sum CW(S_k) + \sum CW(SS_k) \quad (2)$$

For the calculation of the correlation weights necessary for calculating the descriptors by Equation (1), the following target functions were used, and, in particular, IIC was used in Equation (4):

$$TF_0 = R_{AT} + R_{PT} - |R_{AT} - R_{PT}| \times 0.1 \quad (3)$$

$$TF_1 = TF_0 + IIC \times 0.3 \quad (4)$$

R_{AT} and R_{PT} are correlation coefficients between observed and calculated values of pLD_{50} observed for the active training and passive training sets, respectively.

A clear advantage of the Monte Carlo method for constructing the models is its ability to build a prediction based on a simple representation of the molecular structure by SMILES, without the need to take into account their spatial structure and complex descriptors based on representations of atom–atom potentials and quantum mechanical interpretations of charges and electronic density distributions.

Figure 2 shows the scheme of the optimization used in this study.

Figure 3 shows the sequence of the process of the Monte Carlo optimization. One can see that the selection of the threshold 3 (used in Equation (2)) gives better statistical results for the calibration set for both TF0- and TF1-optimization. In the case of TF0, the number of epochs of the optimization selected is equal to six (Figure 3). In the case of TF1, the results slightly improved with the number of epochs, and we stopped the model at 15 epochs. The Monte Carlo process is a statistical procedure, and thus, results may vary. The statistical parameters shown in Figure 3 exhibit clear dispersion, and there may be different ideal values across several runs of the optimization under the same split in active, passive training, calibration, and validation sets. We have empirically selected reasonable T and N values, also to limit the time needed for the process, observing that the modifications on the different epochs were minimal with a high number of epochs.

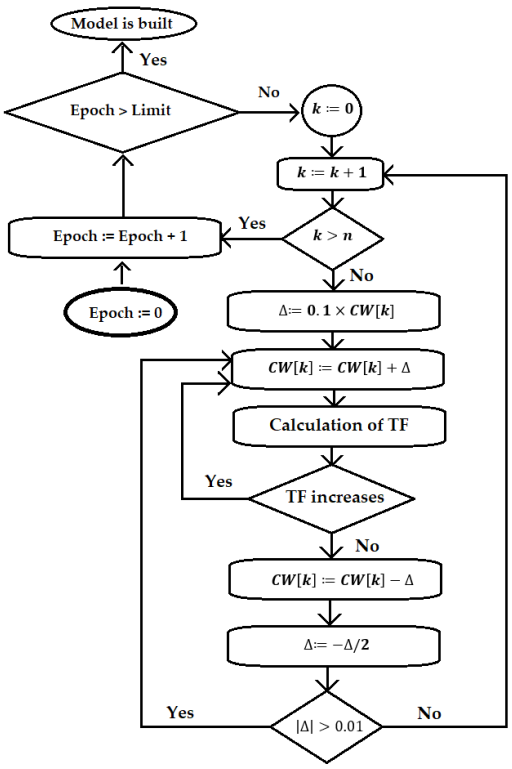

Figure 2. The block scheme of the Monte Carlo optimization.

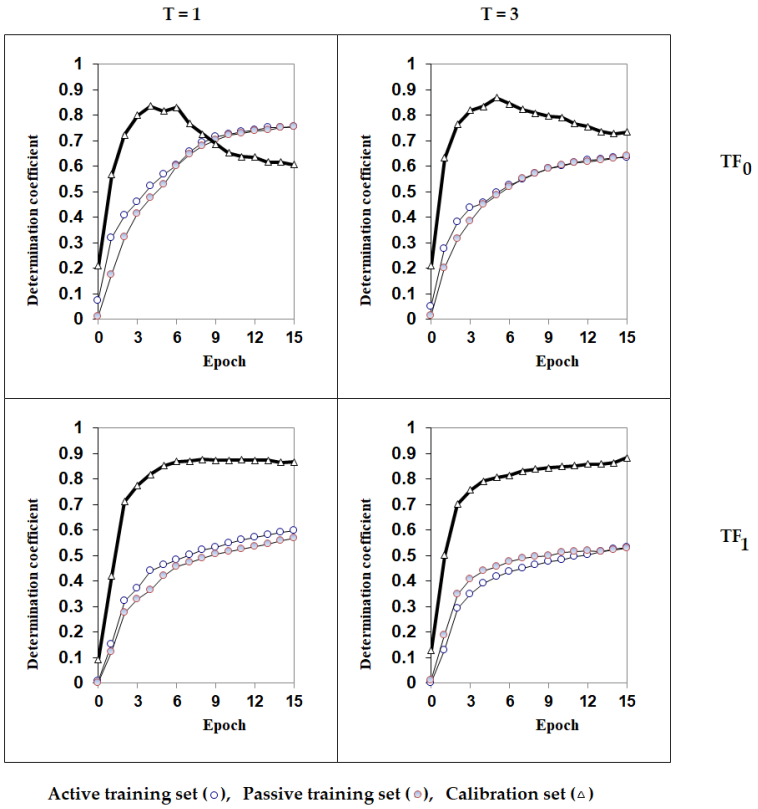

Figure 3. The results of the modelling process performed within the progressive epochs of the Monte Carlo optimization with different thresholds (T) and target functions.

2.5. Applicability Domain

The applicability domain for the described model, calculated with Equation (1), defines the so-called statistical defects of SMILES attributes. These defects can be calculated as follows:

$$d_k = \frac{|P(A_k) - P\prime(A_k)|}{N(A_k) + N\prime(A_k)} + \frac{|P(A_k) - P''(A_k)|}{N(A_k) + N''(A_k)} + \frac{|P\prime(A_k) - P''(A_k)|}{N\prime(A_k) + N''(A_k)} \quad (5)$$

where $P(A_k)$, $P'(A_k)$ and $P''(A_k)$ are the probability of A_k in the active training set, passive training set, and calibration set, respectively; $N(A_k)$, $N'(A_k)$ and $N''(A_k)$ are frequencies of A_k in the active training set, passive training set, and calibration set, respectively. The statistical SMILES-defects (D_j) are calculated as

$$D_j = \sum_{k=1}^{NA} d_k \quad (6)$$

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

A SMILES falls in the domain of applicability if

$$D_J < 2 \times \overline{D} \quad (7)$$

The results of the process related to the applicability domain are reported in Table S1 in Supplementary Materials, showing the outliers on split-1. The results presented in Section 3 consider all substances, without excluding outliers. The outliers for splits 1, 2, 3, 4 and 5, respectively, are 11, 14, 12, 16 and 13.

2.6. Mechanistic Interpretation

Several runs of the CORAL program can provide the basis for a mechanistic interpretation of quasi-SMILES codes. Using numerical data on the correlation weights of features obtained in several Monte Carlo optimization runs; two types of these features can be distinguished: (1) features with a positive correlation weight for all program runs. Such features can be considered as promoters of the growth of the values of the endpoint under study. (2) Features with only a negative correlation weight in all runs. These features can be classified as promoters of a decrease in the value of the endpoint under study. All other features cannot be considered as promoters of an increase or decrease for the endpoint under study.

3. Results

QSAR models have been obtained with the CORAL software [46], using quasi-SMILES [43,44] that represent in the same format the information on the substance, and the result of the experiment about toxicity towards different pollinators with different exposure durations; further data on the environmental persistence are codified within the quasi-SMILES too. The equation has been optimized using two target functions, TF_0 and TF_1. Table 3 contains the results for the training and validation set for five different splits in the case of the TF_0 and TF_1. One can see that in the case of TF_1, the predictive potential of the models is better in comparison with the case using TF_0. The results are presented for the five splits to evaluate if there is consistency among splits and thus sound results. The results are shown for the different sets of substances. The active and passive training sets are those used in the initial phases of the modelling process. In these phases, the final model is not yet developed. Indeed, the model is optimized using the calibration set. Thus, the statistical values related to the final model are those indicated for the calibration model. The validation set contains substances not used for the model development. Table 3

provides the results of the different statistical evaluation checks that we did. All these parameters show that the models using TF_1 are better. Let us consider the values of the determination coefficient D obtained with the TF_0. They are summarized in Table 4.

Table 3. The statistical characteristics of models observed for five splits into the training (active-, passive-, and calibration) and validation sets.

Target Function	Split	Set *	n	D	CCC	IIC	Q^2	$<R_m^2>$	MAE	F	Na
TF_0	1	A	97	0.6157	0.7621	0.7376	0.5987		0.848	152	
		P	94	0.6157	0.7732	0.7058	0.6000		0.728	147	
		C	97	0.8028	0.8949	0.7521	0.7946	0.7564	0.552	387	
		V	94	0.8570	-	-	-	-	0.51	-	100
	2	A	96	0.6629	0.7973	0.8142	0.6502		0.770	185	
		P	96	0.6659	0.7941	0.7818	0.6501		0.777	187	
		C	95	0.7192	0.8471	0.8106	0.7096	0.6700	0.655	238	
		V	95	0.7725	-	-	-	-	0.56	-	96
	3	A	95	0.6607	0.7957	0.6172	0.6451		0.762	181	
		P	96	0.6548	0.7969	0.7190	0.6415		0.720	178	
		C	94	0.8728	0.9226	0.6259	0.8677	0.8498	0.569	631	
		V	97	0.8096	-	-	-	-	0.61	-	101
	4	A	95	0.6993	0.8230	0.7526	0.6887		0.698	216	
		P	95	0.7285	0.8134	0.5765	0.7169		0.846	249	
		C	95	0.6479	0.7966	0.7356	0.6350	0.5732	0.815	171	
		V	97	0.6871	-	-	-	-	0.77	-	99
	5	A	97	0.7316	0.8450	0.8040	0.7180		0.609	259	
		P	94	0.7313	0.8487	0.7829	0.7196		0.662	250	
		C	95	0.7106	0.8343	0.7874	0.6977	0.6425	0.731	228	
		V	96	0.6644	-	-	-	-	0.77	-	100
TF_1	1	A	97	0.4855	0.6537	0.4890	0.4603		1.01	90	
		P	94	0.4798	0.6635	0.5363	0.4566		0.932	85	
		C	97	0.9102	0.9491	0.9537	0.9069	0.8380	0.383	963	
		V	94	0.9405	-	-	-	-	0.32	-	100
	2	A	96	0.5303	0.6930	0.5908	0.5127		0.957	106	
		P	96	0.5274	0.7047	0.5737	0.5042		0.959	105	
		C	95	0.8832	0.9344	0.9394	0.8792	0.8589	0.416	703	
		V	95	0.9025	-	-	-	-	0.35	-	96
	3	A	95	0.5090	0.6746	0.4756	0.4842		0.949	96	
		P	96	0.4869	0.6608	0.4876	0.4665		0.981	89	
		C	94	0.8452	0.9064	0.9193	0.8392	0.7715	0.509	502	
		V	97	0.8335	-	-	-	-	0.46	-	101
	4	A	95	0.5211	0.6851	0.6228	0.5015		0.975	101	
		P	95	0.6771	0.7500	0.4740	0.6627		0.920	195	
		C	95	0.8356	0.9102	0.9139	0.8294	0.7929	0.487	473	
		V	97	0.8177	-	-	-	-	0.48	-	99
	5	A	97	0.5875	0.7402	0.6363	0.5670		0.864	135	
		P	94	0.5875	0.7530	0.7039	0.5690		0.903	131	
		C	95	0.8492	0.9171	0.9182	0.8423	0.7722	0.488	524	
		V	96	0.8507	-	-	-	-	0.50	-	100

* A = active training set; P = passive training set; C = calibration set; and V = validation set; *n* = the number of quasi-SMILES in a set; D = determination coefficient; CCC = concordance correlation coefficient; IIC = index of ideality of correlation; Q^2 = cross validated correlation coefficient (should be >0.6); $<R_m^2>$ = Roy and Kar metric [49] (should be >0.6); MAE = mean absolute error; F = Fischer F-ration; Na = the number of active quasi-SMILES attributes.

Table 4. The average values of the determination coefficient for the calibration and validation sets for the TF_0 and TF_1 (in parentheses, the range of the values in the five splits).

	TF_0	TF_1
Calibration set	0.75 (0.65–0.87)	0.86 (0.84–0.91)
Validation set	0.76 (0.66–0.86)	0.87 (0.82–0.94)

Figure 4 contains the plots of experimental vs. calculated values of the endpoint for the five splits.

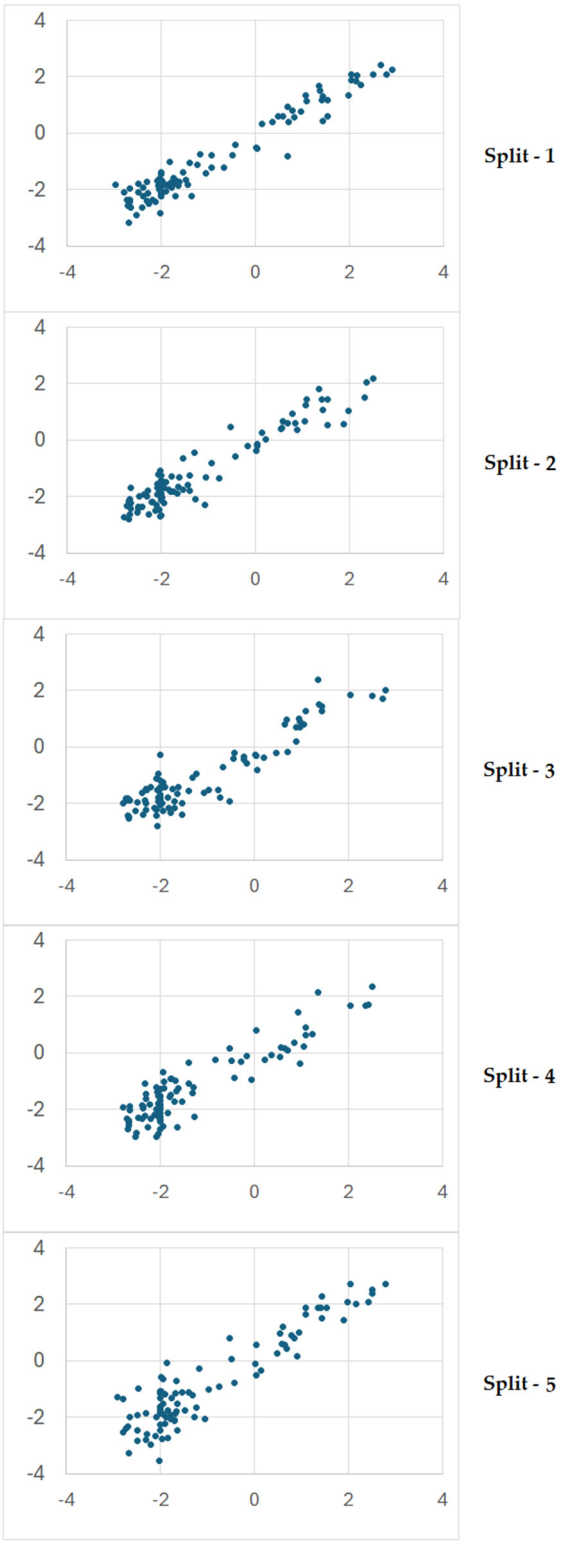

Figure 4. Plots of experimental vs. calculated values of the endpoint for splits 1–5. On the *x*-axis, there are the experimental values. On the *y*-axis, there are the predicted values. Dots represent the toxicity values.

Table 4 shows the average values of the determination coefficient for the calibration and validation sets for the TF_0 and TF_1.

Table 4 represents the results on the calibration set, which are those of the final model developed. The validation set contains substances not used to develop the model. Thus, the results on the validation set provide an indication of the expected results when the model is used for new substances. Table 4 clearly shows the improvement obtained using TF_1, which uses IIC too. There is an improvement of 0.11 for both the calibration and validation sets. The results of the validation and calibration sets are very similar, indicating the model is predictive also towards new substances and is quite robust. In addition to the improvement of the average value, Table 4 shows that the range of the values obtained replicating the models using the five different splits is much smaller in the case of the TF_1, which indicates that the models are more stable and reproducible.

Although the model was developed using a pooled dataset, each data point is explicitly associated with a specific pollinator species, life stage (adult or larva) and exposure duration. To assess the ecological and regulatory relevance of the model, we evaluated its predictive performance across the most representative data subsets. In particular, model performance was analyzed separately for major taxonomic groups (*Apis* spp., *Bombus terrestris* spp. and *Osmia* spp.), life stages (adults and larvae) and exposure durations (acute, subchronic and chronic). The results of this subgroup analysis for the calibration and validation sets are summarized in Table 5.

Table 5. Model performance by species, life stage and exposure duration on calibration and validation sets.

Subgroup	N	R^2	RMSE
Apis spp.	175	0.91	0.46
Bombus terrestris spp.	9	0.97	0.25
Megachile rotundata	-	-	-
Osmia spp.	7	0.92	0.36
Adults	123	0.94	0.43
Larvae	68	0.84	0.47
Acute	143	0.93	0.44
Subchronic	45	0.70	0.46
Chronic	2	0.78	0.23

As shown in Table 5, the model exhibited consistently high predictive performance across the most represented pollinator taxa. Performance for *Apis* spp., which constitutes the largest subset, was strong (R^2 = 0.91, RMSE = 0.46), supporting the robustness of the model for honey bees. Despite the smaller number of data points, the model also performed well for *Bombus terrestris* spp. (R^2 = 0.97, RMSE = 0.25) and *Osmia* spp. (R^2 = 0.92, RMSE = 0.36), suggesting good generalization.

When stratified by life stage, predictive performance was slightly higher for adults (R^2 = 0.94) than for larvae (R^2 = 0.84), likely reflecting greater biological variability and higher uncertainty in larval toxicity data. Performance statistics could not be computed for *Megachile rotundata*, as no data points for this species were included in the calibration or validation sets.

Regarding exposure duration, the model showed high performance for acute exposures (R^2 = 0.93), while performance decreased for subchronic exposures (R^2 = 0.70), consistent with the smaller size and higher heterogeneity of this subset. The chronic subset contained too few data points to allow robust conclusions, although the reported metrics are provided for completeness.

Table S1 in the Supplementary Material contains the results of the model observed for split 1 using TF_1 optimization, together with data on numerical values of the descriptor as well as the quasi-SMILES configurations.

4. Discussion

The new approach applied within this study is to address simultaneously multiple pollinators, with data across different exposure and life stages. To achieve this, on a methodological point of view, we applied (1) the quasi-SMILES technology and (2) the index of ideality of correlation. The quasi-SMILES, as shown in Table 1 and Figure 1 in this specific case, contains information on the chemical substance through the SMILES structure. The use of SMILES for QSAR is very common. What is new here is that the quasi-SMILES includes in the code additional pieces of information. In this case, information on the different species of pollinators is present, extending the models not only to honey bees. Furthermore, the duration of exposure is taken into account within the quasi-SMILES string. Finally, the environmental behaviour of the substance, in particular persistence in water, soil and sediment, is used in the quasi-SMILES. In this way, the model largely extends its applicability, covering a range of situations, and exploits environmental data that may play a role in the overall impact of the pesticides. In this way, the model applies to multiple species and covers multiple exposure durations. The fixed exposure duration, used to define the different kinds of protocols for toxicity assessment towards bees, is useful to gather standardized data in a reproducible way. Conversely, the adverse effects of the same substance in different pollinators may be observed at different times, and thus the model presented here is effective in covering effects appearing at different times depending on the substance and species. This approach is aligned with the framework of the Dynamic Energy Budget (DEB), which explores the complexity of the phenomena leading to toxic effects occurring in different taxa and at different times [50]. From a methodological point of view, the approach we present is convenient because it allows us to exploit data that are sparse and poorly represented in some circumstances.

The development of multiple individual models specific to each species and at different times would be impossible due to the lack of sufficient data suitable to represent all cases. Conversely, merging all data and organizing it into a coherent scheme, as in the present case, allows extending the modelling approach to situations that take advantage of associated experimental observations on related substances or species or durations. We clarify that the lack of data is always an issue, and this approach cannot cover all situations. We expect that there is less uncertainty for predictions related to honey bees, which is more broadly represented in the training set, compared with other species, and the same applies for the evaluation at different durations. Further data are needed to better verify the performance and the different pollinators for chronic exposure.

Another limitation regards the interpretation related to the biological relevance and the differences among species. Further information related to the traits of the different pollinators may be used for this purpose, when available.

Regarding the algorithm, without quasi-SMILES or without supplementing traditional SMILES with codes of experimental condition, the set of objects under consideration exhibits a high degree of degeneracy (identical, indistinguishable traditional SMILES). Without the use of the IIC, the statistical quality of the forecast is significantly lower (Tables 3 and 4). The IIC implements an algorithm able to focus the attention of the model towards substances representing the “correct” correlation between experimental and calculated values, and vice versa, able to “avoid” signals derived from uncommon substances, which will be outliers. There are many studies in the direction of algorithms for attention, also in other areas, such as generative artificial intelligence. Our algorithm

proved to be effective in prioritizing the most relevant parts of the quasi-SMILES. Thus, the above ideas can be considered as useful tools for the QSAR analysis of toxicity towards pollinators. It is to be noted that the basic idea of the approach considered is the Monte Carlo method.

As shown in Table 5, the subgroup analysis supports the applicability of the quasi-SMILES-based approach across different pollinator taxa, life stages and exposure durations, while also revealing clear data limitations for non-Apis species and longer exposure durations that constrain the statistical evaluation of model performance.

Table 6 contains attributes of quasi-SMILES that are promoters of an increase or decrease in pesticide toxicity in pollinator insects. One can see that the presence of branching in the molecular structure (denoted by brackets), carbon in the sp^2 state, double bonds, chlorine and nitrogen are promoters of toxicity increase. This is explainable by the fact that, for instance, branching of the carbon structure is often associated with higher hydrophobicity. The carbon in the sp^2 state and double bonds are often associated with higher reactivity of the molecule. Chlorinated pesticides are often more toxic than others. Nitrogen, and in particular substances with nitro groups, are often more toxic than others. Conversely, the presence of cycles (denoted by digits), pairs of connected carbon atoms in the sp^3 state, as well as the presence of sulfur atoms, are promoters of toxicity decrease. This series of attributes has a statistical meaning and it may help in the identification of potential mechanisms. We clarify that the proposed assumptions regarding the decrease and increase in toxicity towards pollinators are justified on a probabilistic point of view, because they have been identified by the software considering the population of substances in the training and validation sets. It is useful to make reasoning about the possible mechanisms and associations, as exemplified above. However, the overall process codified by the software is more complex, and it cannot be simplified with a few linear relationships with individual features. Given that most natural phenomena are complex, it is appropriate to address them in a probabilistic way. At the same time, this property of this method conceals a drawback: this approach lacks robust universality and the ability to standardize, because it is strictly associated with the specific set of substances and features used. New substances and new experimental conditions may modify the population of items and conditions, with the possibility of further improvements.

Table 6. List of promoters for the increase and decrease in pesticide toxicity. Promoters of toxicity decrease are indicated by grey.

Attribute of Quasi-SMILES	CWs * Run 1	CWs Run 2	CWs Run 3	CWs Run 4	CWs Run 5	NA	NP	NC	S_k
C...(.....	0.0779	0.2255	0.1014	0.0347	0.1229	91	89	93	0.0002
c...c.....	0.0766	0.0026	0.0844	0.3675	0.2334	78	68	77	0.0007
Cl..(.....	0.1406	0.2713	0.2941	0.2074	0.4830	58	55	51	0.0009
N...(.....	0.1919	0.0678	0.2758	0.1823	0.2387	56	58	61	0.0006
=...(.....	0.5996	0.6721	0.4614	0.1333	0.5304	55	52	59	0.0007
n...c.....	0.3686	0.1722	0.3637	0.2684	0.6192	46	40	45	0.0007
N...C.....	0.7570	0.9766	0.9795	0.6388	0.7433	42	38	43	0.0006
c...C.....	1.4670	1.9743	1.8103	1.3863	2.0857	32	29	28	0.0009
c...O.....	0.2966	0.3994	0.6405	0.1038	0.4649	24	16	20	0.0026
1.....	−0.1586	−0.3426	−0.3408	−0.3667	−0.1228	95	83	89	0.0007
1...(.....	−0.2410	−0.4791	−0.4058	−0.2481	−0.1908	49	38	42	0.0016
C...C.....	−0.3833	−0.1924	−0.2086	−0.2670	−0.3992	45	47	49	0.0006
S...(.....	−0.1742	−0.2165	−0.1575	−0.2682	−0.4748	14	22	23	0.0031

* CW = correlation weights. The attributes associated with a decrease in toxicity are in a grey background. NA, NP and NC are frequencies of the quasi-SMILES code in active training, passive training and calibration sets, respectively.

Table 7 contains a comparison of the models for bee acute toxicity (oral exposure) suggested in the literature.

Table 7. Comparison of the results in validation of QSAR models of pollinators' toxicity (oral exposure).

Number of Compounds in Validation Set	Determination Coefficient	Root Mean Squared Error	Reference
28	0.75	0.68	[29]
94	0.94	0.43	Best model in this work

One can see that the statistical quality of the models suggested here is better. The advantages are in the highest number of substances (reflected also in the higher number of substances in the validation set), in the higher value of the determination coefficients, and in the lower value of the root mean square error. In addition, this model works not only on acute toxicity and not only on honey bees, thus these two are other advantages, not represented in Table 7. There are other advantages, compared with the model published by Moreira-Filho et al. [29]: that model relies on feed-forward neural networks (FNN) and requires molecular descriptors (and fingerprints for classification), making it dependent on feature generation and preprocessing. In contrast, the quasi-SMILES CORAL regression developed here requires only SMILES strings and no descriptor or fingerprint calculation, while directly encoding species identity, exposure duration and physicochemical persistence within the same quasi-SMILES representation. Overall, the present approach offers a simpler, descriptor-free and computationally lighter alternative to existing neural network models, while expanding prediction capability beyond *Apis mellifera* and beyond a single oral toxicity endpoint.

It is worth noting that only a limited number of studies have developed regression-based QSAR models for acute oral endpoints. As we wrote, there are more studies on contact exposure. If we compare the results we obtained with those from published QSAR models on contact toxicity (thus through a different exposure route) [19,31,32], the set of substances used in the other studies is much lower (in the range of 6 to 17 substances), with determination coefficients in the range of 0.84–0.96. Thus, the performance of our model is quite similar to or better than what was achieved in other studies on contact exposure, but the advantages that we mentioned above remain. For instance, Hamadache et al. [19] used 1666 molecular descriptors starting from the 3D molecular structure and then used neural networks. This requires much more effort to achieve the 3D structure, introduces uncertainty regarding the 3D optimization, needs preprocessing of the molecules and complex molecular descriptors, and sophisticated algorithms. Some models are focused on specific chemical classes. In another study, Dulin et al. only addressed carbamates and organophosphorus substances [31]. They used 3D structures, 2489 molecular descriptors, and Genetic Function Approximation (GFA), producing 500 equations. Thus, also in this case, it is clear that a major effort is needed to generate models. We developed models for honey bees too [25,26,30,32]. In some cases, we used a previous version of CORAL and fewer substances [25,30]. In another study, a complex approach was performed using hybrid models, thus developing two models based on molecular descriptors, which were then combined into a single model [32].

Overall, the current approach provides one of the few available oral toxicity QSARs in pollinators, while being simpler to implement, trained on a richer dataset, applicable to multiple species (*Apis mellifera* spp., *Bombus terrestris* spp., *Osmia* spp. and *Megachile rotundata*), two different life stages (adult and larva), and exposure durations (acute, subchronic, and chronic), and capable of delivering predictive performance comparable or superior

to existing models. By requiring only SMILES as input and avoiding complex descriptor calculations, this framework represents a robust and versatile predictive tool for pollinator risk assessment.

Last, but not least, the approach described has been applied in different QSPR/QSAR studies in the aspect of using the quasi-SMILES methodology [51–64] as well for the investigation of the IIC as a tool to improve the predictive potential [65–83].

5. Conclusions

We introduced a general model approach to the toxicity of pesticides towards pollinators, not focused on honey bees. Results were good for different pollinators, exposure time and life stage. This is relevant to introduce tools useful for the protection of wild pollinators, accounting for their important ecological and agricultural role. The novel approach also goes beyond the exposure times with a fixed granularity, which may fail to capture the effects occurring at times different from the official ones. The quasi-SMILES technique gives models of pesticide toxicity to pollinators of quite satisfactory quality. IIC improved the predictive potential of models. Heuristically, the approach proposed here represents a convenient palette of possibilities for solving similar problems related to QSAR analysis. In purely computational terms, the above-mentioned correlation ideality index represents an opportunity to examine previously unapplied correlation quality. Our study provides recommendations for effectively managing the stochastic selection process of correlation weights, which have a significant influence on the mechanistic interpretation. Due to its statistical basis, the uncertainty of the model is lower for conditions that are better represented, such as honey bees and acute exposure; further data may increase the confidence (or possibly allow the development of better models in the future) on less represented conditions, such as chronic toxicity or *Osmia* species.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jox16010010/s1, Table S1: Technical details for split 1.

Author Contributions: A.P.T.: conceptualization, methodology, resources; writing—original draft preparation; visualization; formal analysis; writing—review and editing; A.A.T.: conceptualization, methodology, resources; writing—original draft preparation; visualization; formal analysis; S.M.: conceptualization, methodology, resources; writing—original draft preparation; visualization; formal analysis; writing—review and editing; A.R.: project administration; writing—review and editing; E.B.: project administration; writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Funding: We acknowledge EFSA for the financial contribution within the project sOFT-ERA, OC/EFSA/IDATA/2022/02.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in this study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Tosi, S.; Démares, F.J.; Nicolson, S.W.; Medrzycki, P.; Pirk, C.W.W.; Human, H. Effects of a neonicotinoid pesticide on thermoregulation of African honey bees (*Apis mellifera scutellata*). *J. Insect. Physiol.* **2016**, *93–94*, 56–63. [CrossRef] [PubMed]
2. Baas, J.; Goussen, B.; Miles, M.; Preuss, T.G.; Roessink, I. BeeGUTS—A toxicokinetic–toxicodynamic model for the interpretation and integration of acute and chronic honey bee tests. *Environ. Toxicol. Chem.* **2022**, *41*, 2193–2201. [CrossRef]
3. Gisder, S.; Genersch, E. Viruses of commercialized insect pollinators. *J. Invertebr. Pathol.* **2017**, *147*, 51–59. [CrossRef]

4. Stanley, D.A.; Raine, N.E. Chronic exposure to a neonicotinoid pesticide alters the interactions between bumblebees and wild plants. *Funct. Ecol.* **2016**, *30*, 1132–1139. [CrossRef] [PubMed]
5. Jumarie, C.; Aras, P.; Boily, M. Mixture of herbicides and metals affect the redox system of honey bees. *Chemosphere* **2017**, *168*, 163–170. [CrossRef]
6. Charreton, M.; Decourtye, A.; Henry, M.; Rodet, G.; Sandoz, J.C.; Charnet, P.; Collet, C. A locomotor deficit induced by sublethal doses of pyrethroid and neonicotinoid insecticides in the honeybee Apis mellifera. *PLoS ONE* **2015**, *10*, e0144879. [CrossRef]
7. Parmentier, L.; Meeus, I.; Cheroutre, L.; Mommaerts, V.; Louwye, S.; Smagghe, G. Commercial bumblebee hives to assess an anthropogenic environment for pollinator support: A case study in the region of Ghent (Belgium). *Environ. Monit. Assess.* **2014**, *186*, 2357–2367. [CrossRef] [PubMed]
8. Potts, S.G.; Biesmeijer, J.C.; Kremen, C.; Neumann, P.; Schweiger, O.; Kunin, W.E. Global pollinator declines: Trends, impacts and drivers. *Trends Ecol. Evol.* **2010**, *25*, 345–353. [CrossRef]
9. European Commission. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Revision of the EU Pollinators Initiative—A New Deal for Pollinators; COM(2023) 35 Final; Brussels. 24 January 2023. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52023DC0035 (accessed on 30 December 2025).
10. European Food Safety Authority (EFSA). EFSA Guidance Document on the risk assessment of plant protection products on bees (*Apis mellifera*, *Bombus* spp. and solitary bees). *EFSA J.* **2013**, *11*, 3295. [CrossRef]
11. United States Environmental Protection Agency (US EPA). Guidance for assessing pesticide risks to bees. In *Office of Research and Development, National Center for Environmental Assessment*; United States Environmental Protection Agency: Washington, DC, USA, 2014. Available online: https://www.epa.gov/sites/production/files/2014-06/documents/pollinator_risk_assessment_guidance_06_19_14.pdf (accessed on 30 December 2025).
12. Tosi, S.; Sfeir, C.; Carnesecchi, E.; van Engelsdorp, D.; Chauzat, M.-P. Lethal, sublethal, and combined effects of pesticides on bees: A meta-analysis and new risk assessment tools. *Sci. Total Environ.* **2022**, *844*, 156857. [CrossRef]
13. Devillers, J.; Pham-Delègue, M.H. (Eds.) *Honey Bees: Estimating the Environmental Impact of Chemicals*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2002; pp. 1–352. [CrossRef]
14. European Food Safety Authority (EFSA). Towards an integrated environmental risk assessment of multiple stressors on bees: Review of research projects in Europe, knowledge gaps and recommendations. *EFSA J.* **2014**, *12*, 3594. [CrossRef]
15. Como, F.; Carnesecchi, E.; Volani, S.; Dorne, J.L.; Richardson, J.; Bassan, A.; Pavan, M.; Benfenati, E. Predicting acute contact toxicity of pesticides in honeybees (*Apis mellifera*) through a k-nearest neighbor model. *Chemosphere* **2017**, *166*, 438–444. [CrossRef]
16. Carnesecchi, E.; Toropov, A.A.; Toropova, A.P.; Kramer, N.; Svendsen, C.; Dorne, J.L.; Benfenati, E. Predicting acute contact toxicity of organic binary mixtures in honey bees (*A. mellifera*) through innovative QSAR models. *Sci. Total Environ.* **2020**, *704*, 135302. [CrossRef]
17. Carnesecchi, E.; Toma, C.; Roncaglioni, A.; Kramer, N.; Benfenati, E.; Dorne, J.L.C.M. Integrating QSAR models predicting acute contact toxicity and mode of action profiling in honey bees (*A. mellifera*): Data curation using open source databases, performance testing and validation. *Sci. Total Environ.* **2020**, *735*, 139243. [CrossRef]
18. Devillers, J. (Ed.) Chapter 7—QSAR Modeling of Pesticide Toxicity to Bees. In *Silico Bees*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2014; p. 314. [CrossRef]
19. Hamadache, M.; Benkortbi, O.; Hanini, S.; Amrane, A. 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.* **2018**, *25*, 896–907. [CrossRef]
20. Zhao, X.; Li, H.; Mo, Q.; Cui, J.; Zhang, L. Toxicity prediction of pesticide to bumblebee and honey bee based on machine learning methods. *Chin. J. Pestic. Sci.* **2020**, *22*, 933–941. [CrossRef]
21. Damião, T.C.; de Oliveira Neto, R.F.; de Alencar Filho, E.B. New generation of QSAR modeling for bee safety: Predicting toxicity using graph neural networks and apistox data. *Ecotoxicology* **2025**, *34*, 2028–2039. [CrossRef]
22. Sharifi, M.; Harwood, G.P.; Harris, M.; Patel, D.M.; Collison, E.; Lunsman, T. Leveraging in silico structure-activity models to predict acute honey bee (*Apis mellifera*) toxicity for agrochemicals. *J. Agric. Food Chem.* **2024**, *72*, 20775–20782. [CrossRef] [PubMed]
23. Chatterjee, M.; Banerjee, A.; Tosi, S.; Carnesecchi, E.; Benfenati, E.; Roy, K. Machine learning—Based q-RASAR modeling to predict acute contact toxicity of binary organic pesticide mixtures in honey bees. *J. Hazard. Mater.* **2023**, *460*, 132358. [CrossRef]
24. Devillers, J.; Pham-Delègue, M.H.; Decourtye, A.; Budzinski, H.; Cluzeau, S.; Maurin, G. Structure-toxicity modeling of pesticides to honey bees. *SAR QSAR Environ Res.* **2002**, *13*, 641–648. [CrossRef] [PubMed]
25. Toropov, A.A.; Benfenati, E. SMILES as an alternative to the graph in QSAR modeling of bee toxicity. *Comput. Biol. Chem.* **2007**, *31*, 57–60. [CrossRef] [PubMed]
26. Toropov, A.A.; Benfenati, E. Additive SMILES-based optimal descriptors in QSAR modelling bee toxicity: Using rare SMILES attributes to define the applicability domain. *Bioorg. Med. Chem.* **2008**, *16*, 4801–4809. [CrossRef] [PubMed]
27. Singh, K.P.; Gupta, S.; Basant, N.; Mohan, D. QSTR modeling for qualitative and quantitative toxicity predictions of diverse chemical pesticides in honey bee for regulatory purposes. *Chem. Res. Toxicol.* **2014**, *27*, 1504–1515. [CrossRef]

28. Xu, X.; Zhao, P.; Wang, Z.; Zhang, X.; Wu, Z.; Li, W.; Tang, Y.; Liu, G. In silico prediction of chemical acute contact toxicity on honey bees via machine learning methods. *Toxicol. Vitr.* **2021**, *72*, 105089. [CrossRef]
29. Moreira-Filho, J.T.; Braga, R.C.; Milhomem Lemos, J.; Alves, V.M.; Borba, J.V.V.B.; Costa, W.S.; Kleinstreuer, N.; Muratov, E.N.; Horta Andrade, C.; Neves, B.J. BeeToxAI: An artificial intelligence-based web app to assess acute toxicity of chemicals to honey bees. *Artif. Intell. Life Sci.* **2021**, *1*, 100013. [CrossRef]
30. Toropov, A.A.; Toropova, A.P.; Como, F.; Benfenati, E. Quantitative structure–activity relationship models for bee toxicity. *Tox. Environ. Chem.* **2017**, *99*, 1117–1128. [CrossRef]
31. Dulin, F.; Halm-Lemeille, M.P.; Lozano, S.; Lepailleur, A.; Sopkova-de Oliveira Santos, J.; Rault, S.; Bureau, R. Interpretation of honey bees contact toxicity associated to acetylcholinesterase inhibitors. *Ecotoxicol. Environ. Saf.* **2012**, *79*, 13–21. [CrossRef]
32. Amaury, N.; Benfenati, E.; Boriani, E.; Casalegno, M.; Chana, A.; Chaudhry, Q.; Chrétien, J.R.; Cotterill, J.; Lemke, F.; Piclin, N.; et al. Results of DEMETRA models. In *Quantitative Structure-Activity Relationships (QSAR) for Pesticide Regulatory Purposes*; Benfenati, E., Ed.; Elsevier Science: Amsterdam, The Netherlands, 2007; Chapter 7; pp. 201–281. [CrossRef]
33. Devillers, J.; Decourtye, A.; Budzinskid, H.; Pham-Delègue, M.H.; Cluzeau, S.; Maurin, G. Comparative toxicity and hazards of pesticides to Apis and non-Apis bees. A chemometrical study. *SAR QSAR Environ. Res.* **2003**, *14*, 389–403. [CrossRef] [PubMed]
34. Lewis, K.A.; Tzilivakis, J. Wild bee toxicity data for pesticide risk assessments. *Data* **2019**, *4*, 98. [CrossRef]
35. OECD. OECD Guidelines for the Testing of Chemicals, Section 2. In *Test No. 246: Bumblebee, Acute Contact Toxicity Test*; OECD Publishing: Paris, France, 2017. [CrossRef]
36. OECD. OECD Guidelines for the Testing of Chemicals, Section 2. In *Test No. 247: Bumblebee, Acute Oral Toxicity Test*; OECD Publishing: Paris, France, 2017. [CrossRef]
37. OECD. OECD Guidelines for the Testing of Chemicals, Section 2. In *Test No. 254: Mason bees (Osmia sp.), Acute Contact Toxicity Test*; OECD Publishing: Paris, France, 2025. [CrossRef]
38. U.S. Environmental Protection Agency. *ECOTOX Database*; U.S. Environmental Protection Agency: Washington, DC, USA, 2025. Available online: https://cfpub.epa.gov/ecotox/index.cfm (accessed on 30 December 2025).
39. European Food Safety Authority. *OpenFoodTox Chemical Hazards Database*; European Food Safety Authority: Parma, Italy, 2025; Available online: https://www.efsa.europa.eu/en/data-report/chemical-hazards-database-openfoodtox (accessed on 30 December 2025).
40. OECD. OECD Guidelines for the Testing of Chemicals, Section 2. In *Test No. 213: Honeybees, Acute Oral Toxicity Test*; OECD Publishing: Paris, France, 1998. [CrossRef]
41. OECD. OECD Guidelines for the Testing of Chemicals, Section 2. In *Test No. 245: Honey Bee (Apis mellifera L.), Chronic Oral Toxicity Test (10-Day Feeding)*; OECD Publishing: Paris, France, 2017. [CrossRef]
42. VEGA-Software. Available online: https://www.vegahub.eu/ (accessed on 30 December 2025).
43. Toropova, A.P.; Toropov, A.A. (Eds.) QSPR/QSAR analysis using SMILES and Quasi-SMILES. In *Challenges and Advances in Computational Chemistry and Physics*; Springer: Cham, Switzerland, 2023; Volume 33, pp. 241–268. [CrossRef]
44. Toropov, A.A.; Di Nicola, M.R.; Toropova, A.P.; Roncaglioni, A.; Dorne, J.L.C.M.; Benfenati, E. Quasi-SMILES: Self-consistent models for toxicity of organic chemicals to tadpoles. *Chemosphere* **2023**, *312*, 137224. [CrossRef]
45. Toropova, A.P.; Toropov, A.A.; Benfenati, E.; Leszczynska, D.; Leszczynski, J. QSAR model as a random event: A case of rat toxicity. *Bioorganic. Med. Chem.* **2015**, *23*, 1223–1230. [CrossRef]
46. CORAL Software. Available online: https://www.insilico.eu/coral (accessed on 30 November 2025).
47. Toropov, A.A.; Toropova, A.P. The index of ideality of correlation: A criterion of predictive potential of QSPR/QSAR models? *Mutat. Res. Genet. Toxicol. Environ. Mutagen.* **2017**, *819*, 31–37. [CrossRef] [PubMed]
48. Toropova, A.P.; Toropov, A.A.; Roncaglioni, A.; Benfenati, E. The index of ideality of correlation improves the predictive potential of models of the antioxidant activity of tripeptides from frog skin (*Litoria rubella*). *Comput. Biol. Med.* **2021**, *133*, 104370. [CrossRef]
49. Roy, K.; Kar, S. The rm2 metrics and regression through origin approach: Reliable and useful validation tools for predictive QSAR models (Commentary on 'Is regression through origin useful in external validation of QSAR models?'). *Eur. J. Pharm. Sci.* **2014**, *62*, 111–114. [CrossRef] [PubMed]
50. Baas, J.; Augustine, S.; Marques, G.M.; Dorne, J.L. Dynamic energy budget models in ecological risk assessment: From principles to applications. *Sci. Total Environ.* **2018**, *628–629*, 249–260. [CrossRef]
51. Hassan, R.; Baghban, A. Beyond the surface: Quasi-SMILES machine learning approaches for precise estimation of organic sorption. *Mater. Today Commun.* **2025**, *49*, 114126. [CrossRef]
52. He, Y.; Liu, F.; Min, W.; Liu, G.; Wu, Y.; Wang, Y.; Yan, X.; Yan, B. De novo design of biocompatible nanomaterials using quasi-SMILES and recurrent neural networks. *ACS Appl. Mater. Interfaces* **2024**, *16*, 66367–66376. [CrossRef]
53. Hamidi, E.; Hossein Fatemi, M.; Jafari, K. Thermal conductivity of carbon-based nanofluids; a theoretical modeling using nano-quantitative structure–property relationships. *Chem. Phys. Lett.* **2024**, *846*, 141344. [CrossRef]
54. Azimi, A.; Ahmadi, S.; Jebeli Javan, M.J.; Rouhani, M.; Mirjafary, Z. QSAR models for the ozonation of diverse volatile organic compounds at different temperatures. *RSC Adv.* **2024**, *14*, 8041–8052. [CrossRef]

55. Cheng, K.; Pan, Y.; Yuan, B. Cytotoxicity prediction of nano metal oxides on different lung cells via Nano-QSAR. *Environ. Pollut.* **2024**, *344*, 123405. [CrossRef]
56. Pang, Y.; Li, R.; Zhang, Z.; Ying, J.; Li, M.; Li, F.; Zhang, T. Based on the Nano-QSAR model: Prediction of factors influencing damage to C. elegans caused by metal oxide nanomaterials and validation of toxic effects. *Nano Today* **2023**, *52*, 101967. [CrossRef]
57. Kumar, P.; Kumar, A.; Sindhu, J.; Lal, S. Quasi-SMILES as a basis for the development of QSPR models to predict the CO_2 capture capacity of deep eutectic solvents using correlation intensity index and consensus modelling. *Fuel* **2023**, *345*, 128237. [CrossRef]
58. Ahmadi, S.; Ketabi, S.; Qomi, M. CO_2 uptake prediction of metal-organic frameworks using quasi-SMILES and Monte Carlo optimization. *New J. Chem.* **2022**, *46*, 8827–8837. [CrossRef]
59. Ahmadi, S.; Aghabeygi, S.; Farahmandjou, M.; Azimi, N. The predictive model for band gap prediction of metal oxide nanoparticles based on quasi-SMILES. *Struct. Chem.* **2021**, *32*, 1893–1905. [CrossRef]
60. Jafari, K.; Fatemi, M.H. Application of nano-quantitative structure–property relationship paradigm to develop predictive models for thermal conductivity of metal oxide-based ethylene glycol nanofluids. *J. Therm. Anal. Calorim.* **2020**, *142*, 1335–1344. [CrossRef]
61. Jafari, K.; Fatemi, M.H. A new approach to model isobaric heat capacity and density of some nitride-based nanofluids using Monte Carlo method. *Adv. Powder Technol.* **2020**, *31*, 3018–3027. [CrossRef]
62. Ahmadi, S. Mathematical modeling of cytotoxicity of metal oxide nanoparticles using the index of ideality correlation criteria. *Chemosphere* **2020**, *242*, 125192. [CrossRef]
63. Choi, J.-S.; Trinh, T.X.; Yoon, T.-H.; Kim, J.; Byun, H.-G. Quasi-QSAR for predicting the cell viability of human lung and skin cells exposed to different metal oxide nanomaterials. *Chemosphere* **2019**, *217*, 243–249. [CrossRef]
64. Trinh, T.X.; Choi, J.-S.; Jeon, H.; Byun, H.-G.; Yoon, T.-H.; Kim, J. Quasi-SMILES-based nano-Quantitative Structure-Activity Relationship model to predict the cytotoxicity of multiwalled carbon nanotubes to human lung cells. *Chem. Res. Toxicol.* **2018**, *31*, 183–190. [CrossRef]
65. Bhawna, B.; Kumar, S.; Kumar, P.; Kumar, A. Correlation intensity index-index of ideality of correlation: A hyphenated target function for furtherance of MAO-B inhibitory activity assessment. *Comput. Biol. Chem.* **2024**, *108*, 107975. [CrossRef] [PubMed]
66. Goyal, S.; Rani, P.; Chahar, M.; Hussain, K.; Kumar, P.; Sindhu, J. Quantitative structure activity relationship studies of androgen receptor binding affinity of endocrine disruptor chemicals with index of ideality of correlation, their molecular docking, molecular dynamics and ADME studies. *J. Biomol. Struct. Dyn.* **2023**, *41*, 13616–13631. [CrossRef] [PubMed]
67. Kumar, A.; Kumar, P.; Singh, D. QSRR modelling for the investigation of gas chromatography retention indices of flavour and fragrance compounds on Carbowax 20 M glass capillary column with the index of ideality of correlation and the consensus modelling. *Chemometr. Intell. Lab. Syst.* **2022**, *224*, 104552. [CrossRef]
68. Ahmadi, S.; Lotfi, S.; Kumar, P. Quantitative structure–toxicity relationship models for predication of toxicity of ionic liquids toward leukemia rat cell line IPC-81 based on index of ideality of correlation. *Toxicol. Mech. Methods* **2022**, *32*, 302–312. [CrossRef]
69. Duhan, M.; Sindhu, J.; Kumar, P.; Devi, M.; Singh, R.; Kumar, R.; Lal, S.; Kumar, A.; Kumar, S.; Hussain, K. Quantitative structure activity relationship studies of novel hydrazone derivatives as α-amylase inhibitors with index of ideality of correlation. *J. Biomol. Struct. Dyn.* **2022**, *40*, 4933–4953. [CrossRef]
70. Kumar, A.; Kumar, P. Cytotoxicity of quantum dots: Use of quasi SMILES in development of reliable models with index of ideality of correlation and the consensus modelling. *J. Hazard. Mater.* **2021**, *402*, 123777. [CrossRef] [PubMed]
71. Kumar, P.; Kumar, A. Unswerving modeling of hepatotoxicity of cadmium containing quantum dots using amalgamation of quasi SMILES, index of ideality of correlation, and consensus modeling. *Nanotoxicology* **2021**, *15*, 1199–1214. [CrossRef]
72. Kumar, A.; Kumar, P. Prediction of power conversion efficiency of phenothiazine-based dye-sensitized solar cells using Monte Carlo method with index of ideality of correlation. *SAR QSAR Environ. Res.* **2021**, *32*, 817–834. [CrossRef]
73. Ghiasi, T.; Ahmadi, S.; Ahmadi, E.; Talei Bavil Olyai, M.R.; Khodadadi, Z. The index of ideality of correlation: QSAR studies of hepatitis C virus NS3/4A protease inhibitors using SMILES descriptors. *SAR QSAR Environ. Res.* **2021**, *32*, 495–520. [CrossRef]
74. Kumar, A.; Sindhu, J.; Kumar, P. In-silico identification of fingerprint of pyrazolyl sulfonamide responsible for inhibition of N-myristoyltransferase using Monte Carlo method with index of ideality of correlation. *J. Biomol. Struct. Dyn.* **2021**, *39*, 5014–5025. [CrossRef]
75. Kumar, A.; Kumar, P. Quantitative structure toxicity analysis of ionic liquids toward acetylcholinesterase enzyme using novel QSTR models with index of ideality of correlation and correlation contradiction index. *J. Mol. Liq.* **2020**, *318*, 114055. [CrossRef]
76. Javidfar, M.; Ahmadi, S. QSAR modelling of larvicidal phytocompounds against Aedes aegypti using index of ideality of correlation. *SAR QSAR Environ. Res.* **2020**, *31*, 717–739. [CrossRef] [PubMed]
77. Kumar, P.; Kumar, A. In silico enhancement of azo dye adsorption affinity for cellulose fibre through mechanistic interpretation under guidance of QSPR models using Monte Carlo method with index of ideality correlation. *SAR QSAR Environ. Res.* **2020**, *31*, 697–715. [CrossRef]
78. Kumar, A.; Kumar, P. Construction of pioneering quantitative structure activity relationship screening models for abuse potential of designer drugs using index of ideality of correlation in Monte Carlo optimization. *Arch. Toxicol.* **2020**, *94*, 3069–3086. [CrossRef]

79. Kumar, P.; Kumar, A. Nucleobase sequence based building up of reliable QSAR models with the index of ideality correlation using Monte Carlo method. *J. Biomol. Struct. Dyn.* **2020**, *38*, 3296–3306. [CrossRef]
80. Bagri, K.; Kumar, A.; Nimbhal, M.; Kumar, P. Index of ideality of correlation and correlation contradiction index: A confluent perusal on acetylcholinesterase inhibitors. *Mol. Simul.* **2020**, *46*, 777–786. [CrossRef]
81. Kumar, P.; Kumar, A. CORAL: QSAR models of CB1 cannabinoid receptor inhibitors based on local and global SMILES attributes with the index of ideality of correlation and the correlation contradiction index. *Chemom. Intell. Lab. Syst.* **2020**, *200*, 103982. [CrossRef]
82. Nimbhal, M.; Bagri, K.; Kumar, P.; Kumar, A. The index of ideality of correlation: A statistical yardstick for better QSAR modeling of glucokinase activators. *Struct. Chem.* **2020**, *31*, 831–839. [CrossRef]
83. Kumar, P.; Kumar, A.; Sindhu, J. Design and development of novel focal adhesion kinase (FAK) inhibitors using Monte Carlo method with index of ideality of correlation to validate QSAR. *SAR QSAR Environ. Res.* **2019**, *30*, 63–80. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.