Scientific paper

Application of Inverse QSAR/QSPR Analysis for Pesticides Structures Generation

Belgacem Souyei,[1,2] Abdelkader Hadj Seyd,[3,*] Faouzi Zaiz[4] and Abdelkrim Rebiai[5]

[1] *Faculty of Applied Sciences, University of Kasdi Merbah-Ouargla, 30000, Algeria*

[2] *Chemistry Department, Echahid Hamma Lakhdar University of El Oued.*

[3] *Faculty of Renewable Energy, Hydrocarbons, Earth and Universe Sciences, University of Kasdi Merbah- Ouargla*

[4] *Informatics department, Echahid Hamma Lakhdar University of El Oued.*

[5] *VTRS Laboratory, University of El Oued, P.O. Box 789, 39000, El-Oued, Algeria.*

* Corresponding author: E-mail: seydtg@gmail.com
Tel: +213 662 726024

Abstract

The present work has focused on the application of the inverse-QSAR/QSPR problem for generating new structures of pesticides; this is in view of its extremely important and widespread use in several areas, particularly the agricultural field. For this reason, we implemented a methodology containing nine detailed successive steps that include a quantitative structure–activity/property relationship (QSAR/QSPR) study performed to develop a model that relates the structures of 190 pesticides compounds to their n-octanol–water partition coefficients ($\mathbf{logk_{ow}}$). We used the unique atomic signatures which represent the structures and acts as independent variables while the property ($\mathbf{logk_{ow}}$) as the dependent variable. The model was constructed using 130 molecules as training set, and predictive ability tested using 60 compounds. Modeling of $\mathbf{logk_{ow}}$ of these compounds as a function of the signatures descriptors was established by multiple linear regression (MLR) using (LOO) cross-validation. As a result, a QSAR/QSPR equation with 14 atomic signatures was hereby obtained with a $R^2 = 0.659273$, $Q^2 = 0.65617$ and $RMSE_{training} = 0.930192$, s = 1.37297 for the training set and in leave-one-out (LOO) cross-validation experiment set value, $q^2 = 0.605676$, $RMSE_{LOO} = 1.0936$ respectively. In addition to all of the above, new structures have been generated for a range of pesticides that can be included as future search topics.

Keywords: Atomic Signatures; I-QSPR algorithm; multiple linear regression (MLR); n-Octanol–water partition coefficients; pesticides.

1. Introduction

Pesticides are a large group of substances used to kill insects. These substances are mainly used to control pests that infest cultivated plants and crops or to eliminate disease-carrying insects in specific areas.[1] The definition of pesticides according to Food Agriculture Organization of the United Nations (FAO 1989), a pesticide is any substance or mixture of substances intended for preventing, destroying, or controlling any pest including vectors of human or animal diseases, unwanted species of plants or animals causing harm during, or otherwise interfering with, the production, processing, storage, or marketing of food, agricultural commodities, wood and wood products, or animal feedstuffs, or which may be administered to animals for the control of insects, arachnids or other pests in or on their bodies.[2]

Pesticides and agrochemicals, in general, became an important component of worldwide agriculture systems during the last century, allowing for a noticeable increase in crop yields and food production.[3]

Poisoning from pesticides is a global public health problem and accounts for nearly 300,000 deaths worldwide every year.[4]

Pesticides have numerous beneficial effects. These include the protection of crop, preservation of food and ma-

terials and prevention of vector-borne diseases. For example pesticides may be used in the prevention of malaria, which kills up to 1 million children per year,[5] and for preventing other vector-borne diseases such as dengue, leishmaniosis and Japanese encephalitis. Sorption, volatilization, solubility in water, hydrolysis or oxidation, photo degradation and biodegradation are some of the important factors dealing with the fate of OPPs in the environment.[6] Pesticides vary by source, structure and usage, for example, we find Botanical Ps and Neonicotinoid pesticides. Botanical Ps are naturally occurring chemical compounds extracted or derived from plants to manage field and storage crop pests.[7] The BPs can easily degrade in the environment, and they are easily available, less toxic to human and non-targeted organisms and are compatible with different human cultures.[8,9] Studies have shown that, plants are very good source of crop protectants against pests.[10,11]

Neonicotinoid pesticides were first introduced in the mid-1990s and since then their use has grown rapidly so that they have become the most widely used class of insecticides in the world, with the majority being used as seed coatings.[12] As for the distribution of pesticides, it is concluded that lipophilicity is the chief determinant of pesticide distribution in sediment/water systems.[13] Accordingly lipophilicity k_{ow} (n-octanol–water partition coefficients) is a physico-chemical property that characterizes the ability of a chemical compound to dissolve in fats (lipids) and non-polar solvents.[13] Lipophilicity plays an important role in the development of drugs and pesticides, since this parameter affects the pharmacokinetic and pharmacodynamic behavior of a biologically active substance.[15,16] According to IUPAC, lipophilicity reflects the affinity of a molecule or a fragment thereof with a lipophilic medium.[17]

Due to the importance of lipophilicity parameter k_{ow} in the distribution of these compounds (pesticides) between the water and organic phases in the organism, numerous studies report k_{ow} values for ionisable compounds.[18–20]

However, most of them determined a single k_{ow} value, reflecting the lipophilicity of the neutral species only. Lipophilicity is expressed by the octanol–water partition coefficient (k_{ow}), estimates the solubility in both aqueous and organic phases.[21] The values of k_{ow} generated using these various methods may vary by several orders of magnitude hence k_{ow} is usually expressed in the logarithmic form (**log-k_{ow}**).[22] Given all the above mentioned importance of the **logk_{ow}**, it is necessary to study the shape and characteristics of the relationship between this important property of (**log-k_{ow}**) and the molecular structure of these compounds.

The aim of this work is the application of the Inverse-quantitative structure-property relationship (I-QSPR) study. This method is based on a nine-step methodology. The first is a selection of database compound. The second is the generation of the 2D structures, while the third step is the QSPR analysis after translation of the database compounds into unique atomic signatures. In the fourth one, we construct constraint equations, specifically the graphicality and consistency equations, which facilitate the reconstruction of the solution compounds directly from the signatures. Fifth, we solve constraint equations while the sixth step is the database solutions checking. The seventh, eighth and ninth steps are respectively, keeping solutions with desired range, new structures generating, and finally, the database focused.
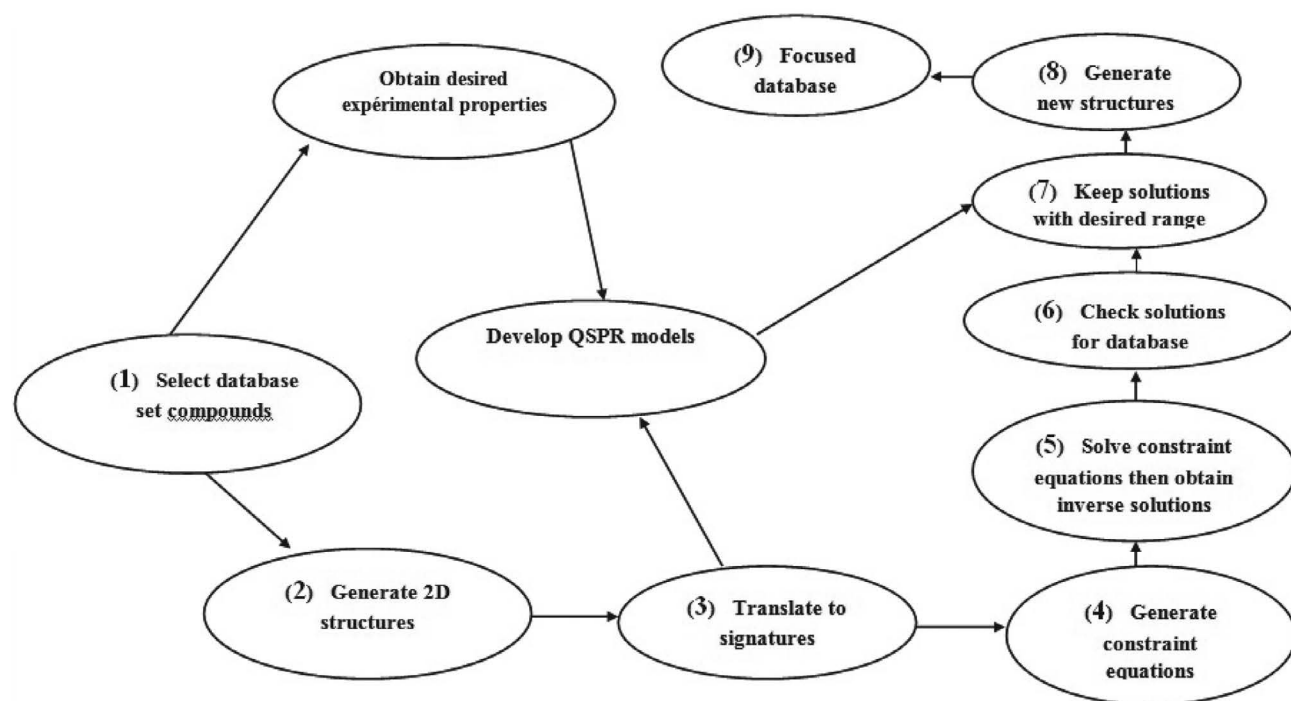

Figure 1. Outline of the I-QSPR Algorithm.[24]

As for important uses of the QSPR analysis step is to develop a QSPR model that relates the structures of 190 pesticides compounds to their n-octanol–water partition coefficients using multiple linear regression technique and to generate new pesticides structures with novel physico-chemical and QSAR properties.

2. Methodology

Inverse-QSAR/QSPR is known as the technical uses values for the independent variables of a particular compound in the QSAR /QSPR to solve for the activity /property of that compound (the dependent variable). In contrast, the goal of the inverse-QSAR/QSPR problem is to determine values for the independent variables given a desired activity /property.[23]

An inverse-QSPR(I-QSPR) problem is a signature-based CAMD (Computer Aided Molecular Design) algorithm that identifies compounds possessing a certain performance (or property) of interest predicted using a developed QSPR model.[24] The I-QSPR technique is interchangeable with the molecular signature descriptor CAMD algorithm.

To achieve the I- QSPR algorithm in our study, we provided the steps detailed in **Figure 1** below, which explains how this algorithm is performed in nine fundamental stages:

(1) The selection of database compounds ; (2) Generation of the 2D structures; (3) Translation of the database compounds into signatures in addition to QSPR analysis; (4) Generation of constraint equations ; (5) Constraint equations solving then inverse solutions obtaining;(6) Check solutions for database; (7)Store solutions within desired range ; (8) New structure generation;(9) Focused database.

2. 1. Step 1: Selection of Database Compounds

In this work the database contains 190 compounds (Pesticides) of different classes, which have an important role in human life. The corresponding experimental data (*n*-octanol/water partition coefficients **logk_{ow}**) are obtained from the literature (www.chemspider.com chemical structures and www.pubchem.com).

2. 2. Step 2 : Generation of the 2D Structures

We have developed a code that allows us to calculate the atomic signatures of molecules after generating 2D structures.

2. 3. Step 3 : Translation of the Database Compounds into Signatures in Addition to QSPR Analysis

The structural representation of the studied compounds is of great importance for describing, circulating and explaining the significant structural information depending on their characteristics. Based on this representation, the extent to which this structure is related to the activity /physochemical properties of the studied molecule. The structural information of a molecule is evaluated by entities called molecular descriptors. The descriptor which is distinctive and in accordance with the applicable conditions to this technique (I-QSPR) is called signature.[25,26]

a) Signature

The signature is a fragment based descriptor that encodes the local topology of an atom in a molecule.[27] Degeneracy, when using signature, is controlled by the height of the signature, which represents the level of branching in a structure. Signature at height-1 or height-2 has lower degeneracy than height-0, and shows high correlation ability for atomic signatures of a molecule to its corresponding property of interest.[28]

b) Definition of the Atomic Signature

Signature, which has its origins in structural elucidation studies of Faulon, [29] is based on the molecular graph of a molecule, $G = (V_G, E_G)$, where the elements in V_G denote the atoms in the molecule, and the edges of E_G correspond to the bonds between those atoms. We define an atomic signature, $^h\sigma_G(x)$, as the canonical sub-graph of G consisting of all atoms a distance h from the root x.[29]

Once a signature height is specified, the molecular signature of each of the N compounds identified in step 1 is calculated using an in-house translator program.

c) Definition of the Molecular Signature Descriptors

Descriptors encoding significant structural information are used to present the physicochemical characteristics of compounds to build the relationship between structure and property in this study. The molecular descriptor used in this project was the molecular descriptor called signature because of its success to address the I-QSAR problem. The success of signature is threefold, First, signature performs the QSAR analysis as well as conventional molecular descriptors.[30,31] Second, signature has a lower degeneracy than other molecular descriptors and can be controlled by the user by a variable termed height. The molecular signature for a compound is the sum of each atomic signature multiplied by the occurrence vector of that atomic signature in the given compound and it can be calculated using the following equation.[32]

$$^h\sigma(G)= \sum_{x \in V_G} {}^h\sigma_G(x) = {}^h\alpha_G \sum^{h} \qquad (1)$$

Where the elements of V_G (matrix of the vertices) are the atoms (X), $\sum^h i$ is the basis set of all atomic signatures of height h and $^h\alpha_G$ is the vector of occurrence number of atomic h-signatures of the graph G. Example of molecular signatures for Ethephon is given in **Figure2**.
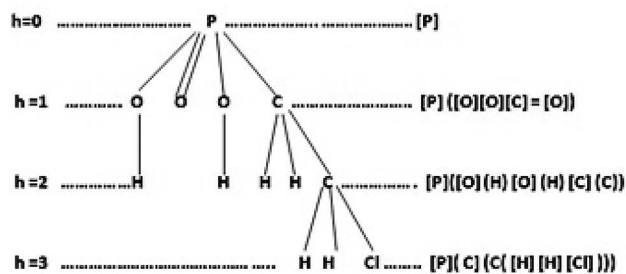
Molecular Signatures for height-1:

$$\sigma^1 = [P]([O][O][C]=[O]) + =[O]([P]) + 2[O]([P][H]) + [C]([P][C][H][H])$$
$$+[C]([C][H][H][Cl]) + 2[H]([O]) + 4[H]([C]) + [Cl]([C])$$

Figure 2. Illustration of the atomic and molecular signatures of Ethephon ($C_2H_6ClO_3P$)

d) QSAR/QSPR analysis:

The QSPR analysis was performed according to the organizational chart below:

Quantitative structure-activity / property relationship (QSAR/QSPR) as an important area of chemometrics has been the subject of a series of investigations.[32,33] The main aim of (QSAR/QSPR) studies is to establish an empirical rule or function relating the structural descriptors of compounds under investigation to properties. This rule or function is then utilized to predict the same properties of the compounds not involved in the training set from their structural descriptors.
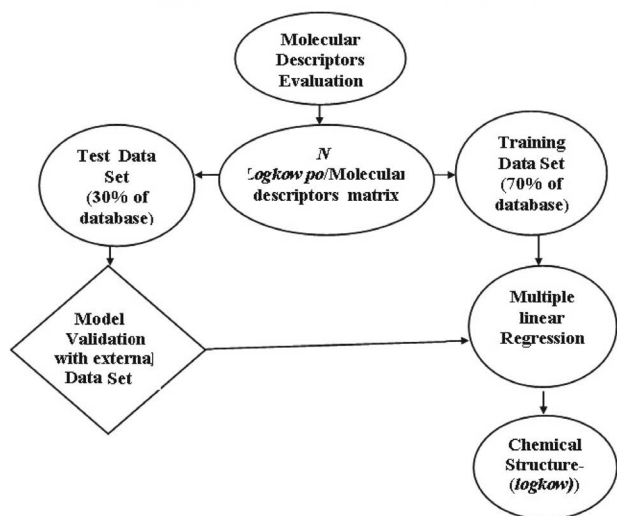

Figure 3: The Steps of QSPR proposed methodology

2. 4. Step 4: Generation of Constraint Equations

In this step, we constructed the constraint equations, which serve in the construction of new compounds by reconnecting atomic signatures into molecular signatures with desired properties determined by the QSPR equa-

tion.[34] Constraint Equations are generated from the atomic signature database. Graphicality Equation is developed from the height 0 atomic signatures, and is a necessary condition for a connected graph.[35]

2. 5. Step 5: Solve Constraint Equations

Since the space of solutions is infinite we limit the range that these solutions (independent variables) can take, based on their range in the original training set, minimum and maximum value (per atomic signatures) provides the additional constraints necessary to solve the system. Due to the large number of equations, we have used the min/max values in the Particle Swarm Optimization algorithm called PSO. This algorithm seems to satisfy the constraint equations in a step-wise manner such that the iterations involving those variables which occur in the equations go from least to most iterations.

2. 6. Step 6: Check Solutions for Database

Since the constraint equations are derived from the number of used compounds, this number should represent solutions to the constraint equations, these are evaluated according to the bonding of the atoms within the molecule.

2. 7. Step 7 : Keep Solutions With Desired Range

In this step the solutions must be scored for fitness relative to a desired property value. The solutions which have the desired fitness are kept, while the unsuitable atomic signature must be removed from the solution. It is at this stage where various heuristics can be applied to focus the solution space based on expert knowledge or other means.[26]

2. 8. Step 8: Generate New Structures

The molecular Signatures (solutions) which emerge from Step 7 are the molecular signatures from which structure generation will occur. Structure generation is performed using an algorithm developed by Faulon and Coworkers,[36] which is based on an earlier isomer enumeration algorithm developed by Faulon (**Appendix 1** in supplementary files). In this step we have selected suitable structures after various filters to remove those undesirable candidate structures.

2. 9. Step 9: Focused Database

The structures which have survived until this point become part of the focused database. These are the high-quality structures which are worthy of further investigation. It is here where experiments run on a select num-

ber of compounds to verify the predictions of the algorithm would be employed. Often, the results of the experimentation can be used to refine the QSPRs and the focused database itself.[26]

# 3. Results and Discussion

3. 1. Atomic Signatures Calculation

As the first step in our I-QSPR analysis, is the translation of the 2D molecular structures into atomic signatures illustrated in **Table 1** bellow**.**

Table 1. The unique height-1 atomic signatures used in the QSPR analysis.

Variable	Height-1 Atomic Signature	Occurrence [Min, max]
X_2	[H]([N])	[0,5]
X_5	**[O](=[C])**	[0,3]
X_6	**[C]([C] [H] [H] H])**	[0,5]
X_8	[S]([P] [C])	[0,3]
X_9	**[O]([P][C])**	[0,4]
X_{11}	[C] ([O] [H] [H] [H])	[0, 4]
X_{12}	**[H] ([C])**	[0,29]
X_{14}	**p[C] ([N] p[C] p[C])**	[0,4]
X_{18}	p[C] (p[C] p[C] [C])	[0,5]
X_{19}	**[Cl] ([C])**	[0,4]
X_{20}	**p[C] (p[C] p[C] [H])**	[0,10]
X_{21}	**[H](p[C])**	[0,11]
X_{23}	pC(p[N] p[N] [N])	[0,3]
X_{24}	**p[N] (p[C] p[C])**	[0,3]
X_{33}	[O] (p[C] [C])	[0,3]
X_{35}	[O] (=[S])	[0,4]
X_{37}	C(N] H] H] H])	[0,4]
X_{39}	[N] ([C] [C] [C])	[0,3]
X_{43}	[C](p[C][H][H][H])	[0,4]
X_{44}	**[pN] (pC] pN])**	[0,4]
X_{49}	[C]([C] [H] [H]S])	[0,3]
X_{52}	**p[C](p[C]p[C]Cl])**	[0,5]
X_{53}	**[Cl](p[C])**	[0,5]
X_{57}	p[C]([pC] p[C][O])	[0,3]
X_{64}	[C]([O][C][H][H])	[0,4]
X_{74}	**[C]([C][C][H][H])**	[0,7]
X_{75}	**[F] ([C])**	[0,6]
X_{89}	[H] ([O])	[0,3]
X_{104}	[H](=[C])	[0,3]

The QSPR analysis was calculated on the basis of a descriptor matrix. The descriptor matrix for the height -1 atomic signature contained 190 rows and 253 columns, one column for the **logk_ow** and 252 columns for the unique atomic signatures. The QSPR equation, however, was only calculated on the basis of 29 atomic signatures, hereby removing 223 atomic signatures, in order to perform the LOOCV.[37] For performing the forward stepping MLR 223 unique atomic signatures removed and we leaved 29 with

occurrence numbers greater than or equal to 3 in order to perform the LOOCV analysis. The most significant atomic signatures were then added one at a time, on the basis of the R^2 and Q^2 values were calculated for each step resulting in **Figure 4**, which depicts the R^2 and Q^2 values as function of the number of independent variables, i.e. atomic signatures.
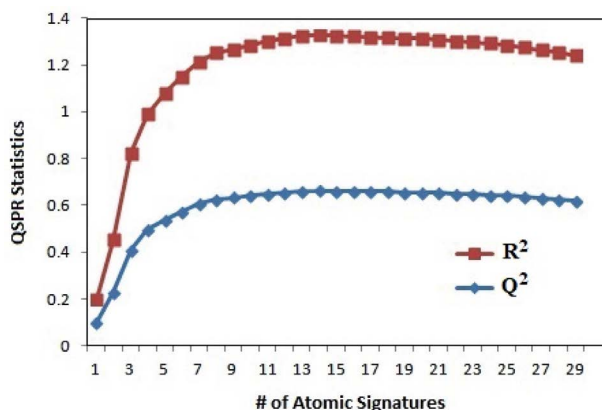
Figure 4. Impact of pesticides height-1 atomic Signatures on the QSPR statistics, which is plotted as function of the number of independent variables

QSPR statistics consists in analyzing the improvement of the correlation with the increase of the number of variables of the model. The representation of the values of R^2 and Q^2 as a function of the number of descriptors (Fig.4) brings out an asymptotic behavior, the model is considered optimal when the improvement of the correlation becomes maximal, that is to say representing the better compromise between correlation and parameterization.

The calculations were terminated at **14** atomic signatures, thus the 15th atomic signature was insignificant. Statistically QSPR model using **MLR** was obtained, the QSPR equation was chosen on the basis of the best predicting model, i.e. highest Q^2 value. A QSPR equation with **14** atomic signatures was hereby obtained with a **R^2 = 0.659273, Q^2 =0.65617 and RMSE_training = 0.930192, s = 1.37297** for the training set, and in leave-one-out (LOO) cross-validation experiment set value, **q² = 0.605676, RMSE_LOO = 1.0936** respectively. Where R^2 represents the determination coefficient, Q^2: square validation coefficient, s: standard deviation, RMSE: the root square error.
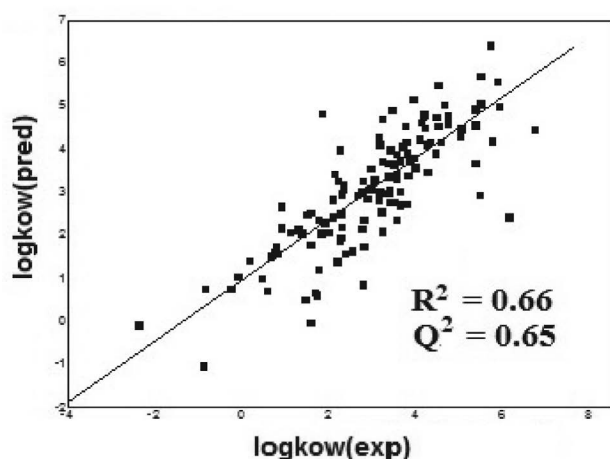
3. 2. The model Equation

The model equation can be written as fellow:

logk_ow = - 0.167497 + 0.444669 * **X_{52}** + 0.417366 * **X_6** + 0.785521 * **X_{20}** + 0.461849 * **X_{75}** + 0.0716288 * **X_{12}** - 0.419674 * **X_5** - 0.409265 * **X_{21}** + 0.528737 * **X_{19}** - 0.220096 * **X_{44}** + 0.251859 * **X_{74}** + 0.365269 * **X_9** + 0.226272 * **X_{14}** + 0.445708 * **X_{53}** + 0.0357723 * **X_{24}**

The **14** atomic signatures included in the QSPR equation are marked with bold in **Table 1.** Statistical results and significance of this final model, illustrate that the positive high value coefficient is for atomic signature X_{20}. It was also suggested from this model that the atomic signature X_{24} was necessary contributor to **logk$_{ow}$**, the atomic signature X_5 was assigned as an effective variable on **logk$_{ow}$**, but with a negative coefficient. Using the QSPR equation to predict the **logk$_{ow}$** of the pesticides in the same training set and plotting these values against the experimental data, resulted in **Figure 5.**

The plot shows the predicted **logk$_{ow}$** values based on the model equation which is validated to be statistically significant by the leave-one-out cross-validation versus experimental ones. Obviously, the predicted **logk$_{ow}$** values are in a good agreement with experimental ones. The 14-parameters of model provide a high statistical quality: $R^2 = 0,66$ and $Q^2 = 0.65$, and this shows that the condition of predictability according to the consideration of R. Veerasamy,[37] and A. Golbraikh,[38] is satisfied.

3. 3. The Constraint Equations

The following constraint equations are written in the order of the smallest parameters number to the greatest, thus in the order they were solved: (i) Consistency equations which ensure the alignment of atoms in the construction of molecular signatures. (ii) Graphicality equation which represents the valence of each atom. To solve these equations, a method developed by Weis and Visco.[28] was adopted, and because of the wide database and the large number of constraint equations, we have used a program based on PSO algorithm.

3. 4. Generating of new Pesticides Structures

Base on the inverse solutions obtained from solving the system of constraint equations, new structures can be constructed from molecular signatures. It is worth mentioning that for the same molecular signature there are multiple structures.
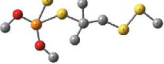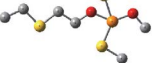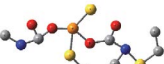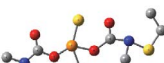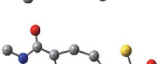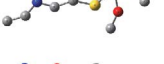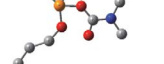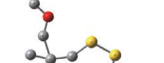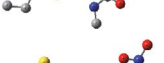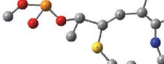
Solving the constraint equations (**Table 2**) a total number of 5500 solutions (new molecular signatures) which will be a new structures. Since it would be difficult to examine over 5500 structures, the newly generated structures were refined according to the different chemical

Figure 5: The experimental- versus predicted values for the QSPR equation based on the **logk$_{ow}$** with 29 height-1 atomic Signatures (*logKow(Pred)=0.909412+0.704451.logKow(exp)*)

Table 2. Constraint equations

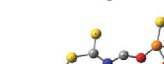| N°	Constraint equation
Eq.1	**Mod(+X_{217},2) = 0**
Eq.2	**Mod(+X_{242},2) = 0**
Eq.3	**+X_{44}+X_{169} = 2**
Eq.4	**-X_{45}+X_{48} = 0**
Eq.5	**-X_{91}+X_{92} = 0**
Eq.6	**-2X_{136}+X_{137} = 0**
Eq.7	**-X_{138}+X_{139} = 0**
Eq.8	**-X_{144}+X_{145} = 0**
Eq.9	**-X_{191}+X_{193} = 0**
Eq.10	**Mod(+X_{213}+X_{215},2) = 0**
Eq.11	**-X_{243}+X_{244} = 0**
Eq.12	**-X_{41}+X_{42}-X_{63} = 0**
Eq.13	**-X_{70}+X_{72}-X_{185} = 0**
Eq.14	**-X_{70}+X_{73}-X_{185} = 0**
Eq.15	**-X_{93}+X_{95}+X_{159} = 0**
Eq.16	**-X_{113}+X_{114}-X_{208} = 0**
Eq.17	**-X_{173}+2X_{174}-X_{175} = 0**
Eq.18	**-X_{206}+X_{207}+X_{209} = 0**
Eq.19	**-X_{52}+X_{53}-X_{119}-X_{187} = 0**
Eq.20	**-X_{88}+X_{89}-X_{153}-X_{195} = 0**

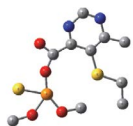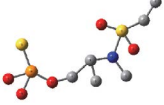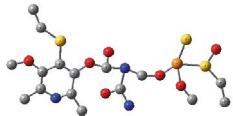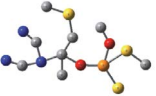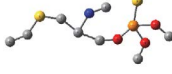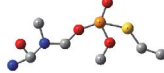| N°	Constraint equation
Eq.21	$-X_{149}+X_{150}+X_{226}+X_{250} = 0$
Eq.22	$-X_{168}+X_{169}-X_{186}-X_{246} = 0$
Eq.23	$-X_{121}+2X_{122}-X_{123}-X_{234}-X_{235} = 0$
Eq.24	$-X_{151}+X_{152}+X_{170}-X_{176}-X_{252} = 0$
Eq.25	$-X_{1}+X_{3}+X_{160}-X_{172}+X_{188}-X_{205} = 0$
Eq.26	$-X_{20}+X_{21}-X_{47}-X_{58}-X_{123}-X_{175} = 0$
Eq.27	$-X_{50}+X_{51}-X_{116}-X_{144}-X_{170}-X_{188} = 0$
Eq.28	$-X_{62}+X_{63}+X_{198}-X_{199}-X_{212}-X_{247} = 0$
Eq.29	$-X_{101}+X_{104}-X_{150}-2X_{155}-X_{165}-X_{226} = 0$
Eq.30	$-3X_{71}+X_{75}-X_{133}-3X_{167}-2X_{219}-X_{220}-2X_{239} = 0$
Eq.31	$-X_{109}+X_{112}-X_{142}-X_{207}-X_{215}-X_{250}+X_{251} = 0$
Eq.32	$-X_{3}+X_{7}-X_{59}-X_{96}-X_{134}-X_{152}-X_{160}-X_{232} = 0$
Eq.33	$-X_{17}+X_{19}-2X_{80}-3X_{103}-X_{115}-3X_{117}-2X_{133}-X_{245} = 0$
Eq.34	$-X_{32}+X_{33}-X_{57}+2X_{85}+X_{93}+X_{118}-X_{124}+X_{195} = 0$
Eq.35	$-X_{3}+X_{8}-X_{50}-X_{59}-2X_{96}+X_{146}-X_{170}-3X_{232}+X_{241} = 0$
Eq.36	$-X_{27}+X_{28}-X_{67}+X_{120}+X_{146}-X_{218}+X_{230}+2X_{236}+2X_{240} = 0$
Eq.37	$\mathbf{Mod}(+X_{60}+X_{62}+X_{140}+X_{141}+X_{184}+X_{185}+X_{201}+X_{202}+X_{238},2)= 0$
Eq.38	$-2X_{34}+X_{35}-2X_{38}-2X_{95}-2X_{120}-2X_{159}-2X_{223}-X_{224}-2X_{230}-2X_{240}-2X_{241} = 0$
Eq.39	$-X_{4}+X_{5}-X_{30}-X_{56}-X_{61}-X_{65}-X_{66}-X_{82}-X_{163}-X_{171}-X_{177}-X_{222} = 0$
Eq. 40	$-X_{31}+2X_{34}-2X_{36}+X_{38}-X_{83}-X_{84}+X_{95}+X_{99}+2X_{107}+X_{120}-2X_{132}-X_{221} =0$
Eq.41	$-X_{1}+X_{2}-X_{25}-X_{31}-2X_{46}-X_{69}-X_{79}-X_{83}-X_{162}-X_{184}-2X_{202}-2X_{205}-2X_{221}-X_{238} =0$
Eq.42	$-X_{13}+X_{14}+X_{23}-X_{25}-X_{46}+X_{55}-X_{70}-X_{76}-X_{77}-X_{83}-X_{132}-X_{140}-2X_{162}+X_{234}-X_{238} = 0$
Eq.43	$-X_{3}+X_{9}-2X_{50}-2X_{59}-X_{96}-3X_{116}+X_{118}-3X_{134}-2X_{144}-2X_{152}+X_{153}-2X_{160}-X_{170}-2X_{188}+2X_{231}+X_{251} = 0$
Eq.44	$-2X_{23}+2X_{24}-2X_{27}-X_{32}+X_{44}+2X_{45}-2X_{47}-X_{55}-X_{58}-X_{94}-X_{119}-2X_{124}-2X_{125}-2X_{126}+X_{169}-2X_{187}-X_{218}-X_{234}-X_{235} = 0$
Eq.45	$-X_{40}+X_{41}-X_{105}+X_{149}-X_{177}+2X_{178}-X_{181}-X_{182}-X_{183}-X_{192}+X_{193}-X_{196}-X_{197}+X_{198}-X_{210}-X_{216}+X_{217}-X_{228}-2X_{248}-X_{249} = 0$
Eq.46	$-X_{76}-X_{77}+3X_{78}-X_{79}-X_{141}+X_{142}+X_{150}+X_{165}-2X_{166}-X_{180}+X_{194}-X_{199}+2X_{200}-X_{201}+X_{207}-X_{208}+X_{209}-X_{211}-X_{212}-2X_{214}+X_{215}+X_{250}$ $= 0$
Eq.47	$+2X_{14}+2X_{18}+2X_{20}+X_{32}+2X_{42}+2X_{52}+X_{55}+2X_{57}+X_{58}+2X_{67}+3X_{86}+2X_{91}+X_{94}+X_{119}+X_{121}+X_{123}+X_{126}+2X_{138}+2X_{145}+X_{173}+X_{175}+X_{218} = 2$
Eq.48	$-X_{18}+X_{22}+X_{43}+X_{61}+X_{71}+X_{81}+X_{82}+2X_{87}-X_{94}-X_{121}-X_{125}+X_{127}+2X_{130}+X_{131}+X_{147}+X_{148}+X_{157}+3X_{161}+X_{164}-X_{173}+X_{179}+X_{197}-X_{235}= 0$
Eq.49	$-X_{8}+X_{10}-X_{28}-X_{38}+X_{49}+X_{54}+X_{66}+X_{81}+X_{97}-X_{99}+X_{103}+X_{108}-2X_{111}+X_{133}+2X_{156}-X_{159}+X_{179}+X_{182}+X_{196}-X_{206}+X_{210}-2X_{223}-2X_{224}-X_{230}+X_{237}-X_{241}-X_{242}+X_{243}= 0$
Eq.50	$-X_{100}+2X_{101}-X_{102}-X_{108}+2X_{109}-X_{110}-X_{135}+X_{136}+X_{142}-X_{143}-X_{154}+X_{155}-X_{164}+X_{165}-X_{181}-X_{189}+2X_{190}-X_{192}+2X_{194}+X_{200}-2X_{204}+X_{209}+X_{213}-X_{222}-X_{225}+X_{226}-X_{227}-X_{228}-2X_{229}-X_{245} = 0$
Eq. 51	$-X_{9}+X_{11}+X_{15}-2X_{16}-X_{33}+X_{56}+X_{64}+X_{65}+2X_{68}+X_{82}+X_{87}-X_{88}+X_{90}+X_{106}-X_{112}\ X_{114}+X_{115}+X_{135}+X_{147}+X_{148}+2X_{158}+X_{161}+X_{167}+X_{189}-X_{191}+X_{196}+X_{219}+X_{220}+X_{228}+X_{246}+X_{252} = 0$
Eq.52	$-X_{1}+X_{4}-2X_{13}+X_{15}-X_{25}+X_{26}+X_{29}+2X_{30}-X_{31}-X_{36}+X_{37}-3X_{39}+X_{40}+X_{54}+X_{56}-2X_{60}+X_{61}+X_{66}-2X_{69}-X_{76}-X_{79}-2X_{84}+2X_{105}-2X_{113}+X_{127}+X_{129}-X_{140}-X_{166}+X_{171}-2X_{172}+X_{176}+X_{177}-2X_{180}-X_{184}+X_{189}-X_{201}+2X_{203}+X_{210}-X_{211}-X_{212}+X_{222}+X_{225}+2X_{227}+X_{233}+X_{243}-2X_{247}+X_{249} = 0$
Eq.53	$-3X_{6}-3X_{10}-3X_{11}+X_{12}-2X_{15}-2X_{17}-2X_{22}-2X_{26}-X_{29}-3X_{37}-X_{40}-3X_{43}-2X_{49}-2X_{54}-2X_{64}-2X_{74}-X_{80}-2X_{81}-X_{90}-X_{97}-X_{98}-X_{100}-2X_{102}-3X_{110}-X_{115}-2X_{127}-X_{130}-X_{131}-X_{135}-X_{148}-2X_{151}-2X_{156}-X_{158}-2X_{168}-X_{171}-2X_{176}-X_{179}-X_{181}-2X_{183}-X_{186}-2X_{203}-X_{204}-X_{220}-X_{225}-X_{239}-X_{246}-2X_{249}-X_{252} = 0$
Eq.54	$+X_{4}+X_{6}+X_{17}+X_{22}+X_{26}+2X_{29}+X_{49}+X_{64}+X_{65}+2X_{68}+2X_{74}+X_{80}+X_{87}+2X_{90}+2X_{97}+3X_{98}+X_{100}+X_{102}+3X_{106}+X_{108}+X_{115}+X_{117}+4X_{128}+2X_{129}+X_{130}+2X_{131}+3X_{143}+2X_{147}+X_{148}+X_{151}+2X_{154}+3X_{157}+X_{158}+2X_{163}+X_{164}+X_{168}+X_{179}+X_{182}+X_{183}+2X_{186}+X_{192}+X_{197}+2X_{216}+X_{219}+X_{220}+X_{229}+3X_{233}+3X_{237}+X_{239}+X_{245}+X_{246}+X_{248}+X_{252} = 2$
Eq.55	$\mathbf{Mod}(\, X_{1}-X_{2}+2X_{3}+X_{4}-X_{5}+2X_{6}-X_{7}+X_{8}+X_{9}+2X_{10}+2X_{11}-X_{12}+X_{13}+X_{14}+2X_{15}+X_{16}+2X_{17}+X_{18}-X_{19}+X_{20}-X_{21}+2X_{22}+X_{23}+X_{24}+X_{25}+2X_{26}+X_{27}+X_{28}+2X_{29}+X_{30}+X_{31}+X_{32}+X_{33}+2X_{34}-X_{35}+X_{36}+2X_{37}+2X_{38}+X_{39}+X_{40}+X_{41}+X_{42}+2X_{43}+X_{44}+X_{45}+X_{46}+X_{47}-X_{48}+2X_{49}+2X_{50}-X_{51}+X_{52}X_{53}+2X_{54}+X_{55}+X_{56}+X_{57}+X_{58}+2X_{59}+X_{60}+X_{61}+X_{62}+X_{63}+2X_{64}+X_{65}+X_{66}+X_{67}+2X_{68}+X_{69}+X_{70}+2X_{71}-X_{72}-X_{73}+2X_{74}-X_{75}+X_{76}+X_{77}+X_{78}+X_{79}+2X_{80}+2X_{81}+X_{82}+X_{83}+X_{84}+X_{85}+X_{86}+2X_{87}+X_{88}-1X_{89}+2X_{90}+X_{91}-X_{92}+X_{93}+X_{94}+2X_{95}+2X_{96}+2X_{97}+2X_{98}+X_{99}+X_{100}+X_{101}+2X_{102}+X_{103}-X_{104}+X_{105}+2X_{106}+X_{107}+X_{108}+X_{109}+2X_{110}+X_{111}+X_{112}+X_{113}+X_{114}+2X_{115}+2X_{116}+2X_{117}+X_{118}+X_{119}+2X_{120}+X_{121}+X_{122}+X_{123}+X_{124}+X_{125}+X_{126}+2X_{127}+2X_{128}+X_{129}+2X_{130}+2X_{131}+X_{132}+2X_{133}+2X_{134}+X_{135}+X_{136}-X_{137}+X_{138}-X_{139}+X_{140}+X_{141}+X_{142}+2X_{143}+2X_{144}+X_{145}+X_{146}+2X_{147}+2X_{148}+X_{149}+X_{150}+2X_{151}+2X_{152}+X_{153}+X_{154}+X_{155}+2X_{156}+2X_{157}+2X_{158}+2X_{159}+2X_{160}+2X_{161}+X_{162}+X_{163}+X_{164}+X_{165}+X_{166}+2X_{167}+2X_{168}+X_{169}+2X_{170}+X_{171}+X_{172}+X_{173}+X_{174}+X_{175}+2X_{176}+X_{177}+X_{178}+2X_{179}+X_{180}+X_{181}+X_{182}+2X_{183}+X_{184}+X_{185}+2X_{186}+X_{187}+2X_{188}+X_{189}+X_{190}+X_{191}+X_{192}+X_{193}+X_{194}+X_{195}+X_{196}+X_{197}+X_{198}+X_{199}+X_{200}+X_{201}+X_{202}+2X_{203}+X_{204}+X_{205}+X_{206}+X_{207}+X_{208}+X_{209}+X_{210}+X_{211}+X_{212}+X_{213}+X_{214}+X_{215}+X_{216}+X_{217}+X_{218}+2X_{219}+2X_{220}+X_{221}+X_{222}+2X_{223}+X_{224}+X_{225}+X_{226}+X_{227}+X_{228}+X_{229}+2X_{230}+X_{231}+2X_{232}+2X_{233}+X_{234}+X_{235}+X_{236}+2X_{237}+X_{238}+2X_{239}+2X_{240}+2X_{241}+X_{242}+X_{243}-X_{244}+X_{245}+2X_{246}+X_{247}+X_{248}+2X_{249}+X_{250}+X_{251}+2X_{252}\,,2) = 0$

Table 3. Example of a solved molecular Signature, note that only atomic Signatures with occurrence numbers greater than 0 are depicted.

Molecular Signature no. 1	X_6	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{49}	X_{50}	X_{51}	X_{64}
	1	1	2	1	1	15	1	1	1	1

Table 4. Values of logk$_{ow}$ of the 20 new structures calculated by I-QSPR theory and those predicted by Hyperchem software with other QSAR properties simulated by molecular mechanic MM$^+$ and semi-empirical PM3 calculations.

Formula	3D Structure [a]	logk$_{ow}$[b]	logk$_{ow}$[c]	Surface area	Volume	Refractivity	Mass	MM+ (KJ/mol)	PM3 (KJ/mol)
$C_7H_{17}O_2PS_4$		2.93	2.62	536.71	760.29	76.60	292.42	22.67	−2641.14
$C_6H_{15}O_2PS_3$		2.17	2.61	531.90	730.74	64.50	246.34	24.518	−2302.97
$C_7H_{15}N_2O_4PS_3$		6.05	2.19	580.95	882.26	77.57	318.36	146.488	−2905.23
$C_8H_{17}N_2O_4PS_3$		6.74	2.20	577.82	882.76	81.80	332.39	131.31	−3174.98
$C_9H_{17}Cl_2N_2O_3PS_2$		6.30	2.19	564.48	874.47	77.39	353.22	26.47	−2922.12
$C_7H_{18}NO_2PS_2$		1.87	1.36	573.67	483.36	65.30	243.32	54.01	−2703.03
$C_5H_{12}NO_5PS$		2.65	2.61	451.91	657.32	52.07	229.19	21.28	−2288.44
$C_7H_{16}NO_4PS$		4.56	2.20	510.26	729.30	59.29	241.24	19.674	−2737
$C_6H_{14}OS_2$		1.46	1.88	410.80	566.15	47.31	166.30	-0.46	−1997.52
$C_9H_{19}N_2O_4PS_4$		6.00	2.20	651.1	992.59	95.46	378.47	32.8	−3488
$C_{13}H_{23}N_4O_5PS_2$		3.34	2.61	600.25	1074.63	106.29	366.51	44.67	−4527.12
$C_6H_{16}NO_2PS_2$		1.55	2.61	606.55	978.56	81.77	347.45	31.30	3549.116
$C_9H_{22}N_3O_3PS_3$		5.02	2.62	651.7	982.59	95.42	346.48	32.8	−3458
$C_6H_{14}NO_3PS_4$		4.71	2.19	499.29	791.32	73.09	307.40	14.102	2539.87

Formula	3D Structure [a]	$logk_{ow}$[b]	$logk_{ow}$[c]	Surface area	Volume	Refractivity	Mass	MM+ (KJ/mol)	PM3 (KJ/mol)
$C_{10}H_{15}N_2O_4PS_2$		5.2	2.20	488.23	853.74	76.81	322.33	27.922	–3329.07
$C_6H_{16}NO_5PS_2$		4.27	2.61	493.77	731.	63.56	277.23	49.9	–2743.26
$C_{16}H_{26}N_3O_7PS_3$		8.66	1.47	693	1249	115.85	499.5	46.36	–5356.36
$C_8H_{18}N_3O_2PS_3$		2.59	2.61	506.16	858.53	74.30	315.40	43.37	–3075.79
$C_8H_{20}NO_3PS_2$		1.51	1.46	543.54	803.91	70.45	273.35	26.001	–3074.00
$C_5H_{15}N_2O_4PS$		5 .52	2 .20	492.04	72.69	57.54	242.23	40.15	–26.21

[a]: ● : Carbon (C); ● : Chlorine (Cl) ; ● : Nitrogen (N); ● : Oxygen (O); ● : Phosphorus; ● : Sulfur (S) ; (the H atoms are hidden)
[b]: logkow predicted by Hyperchem software ; [c]: logkow calculated by I-QSPR theory

structures of all existing pesticides. All newly generated structures were passed through the ChemSpider, PUB Chem and LookChem (structure search) which are reliable database to identify commercially available compounds. After all these steps, a set of 20 samples was selected (Appendix 2 in supplementary files) according to their **$logk_{ow}$**, which is close to the average value of the database of the 190 pesticides studied (**$logk_{ow}$ = 2.94**).

The 20 identified compounds became part of the focused database. In order to assess the diversity level among the newly generated structures, they were compared to the training set structures.In addition to the work done, we compared the values of the **$logk_{ow}$** predicted by our model to those calculated by Hyperchem software (ver. 8), it was concluded that the values for the two results (Table 4) are close in most cases.

4. Conclusion

The high interest in pesticides and their uses in diverse fields, especially in agriculture, requires us to study these pesticides extensively and in depth . This is done by identification and focus on the characteristics including the physico-chemical properties, then the attempt to establish new chemical structures.

The identification of new pesticides with desired properties was done by developing an inverse-quantitative structure-property relationship on the basis of octanol-water partition coefficient (**$logk_{ow}$**).

We processed a database of 190 pesticide compounds, after developing molecular signatures calculated from atomic signatures. And in order to perform LOOCV. Only 29 atomic signatures from out of 252 are used as independent variables and **$logk_{ow}$** as a dependent variable in the QSPR realization, then the resolution of the constraint equations to the number of 55 by a computation code developed for this purpose, based on the successful PSO method to find 5500 solutions which represent new structures.

Based on the goal of the inverse-QSPR method was to predict, if any, novel compounds structures possessing a **$logk_{ow}$** values are close to those in the training set. There were 20 new compounds classified as pesticides.

We have presented and studied these new structures that do not yet exist in the databases of chemical compounds based on our search of reliable databases for this purpose.

This work indicates that the inverse-QSPR method can be used as a reliable approach to generate new compound structures, since, on the one hand, the coefficient of determination R^2 of the model is greater than 60%, and on the other hand, the predicted results are close to the values calculated by other software such as Hyprchem. This research is envisaged to serve as a base for further studies.

5. Acknowledgments

6. References

1. A. Sharma, P. J. Gadi , V. P. Reddy, *Science of The Total Environment*, **2018**, *643*, 1522–1532.
   DOI:10.1016/j.scitotenv.2018.06.312

2. S. Knillmanna, P. Orlinskiyabc, O. Kaskea, K. Foita, M. Liessad, *Science of The Total Environment*, **2018**, *630*, 1619–1627.
   DOI:10.1016/j.scitotenv.2018.02.056

3. R. Münze, P. Orlinskiy, R. Gunold, A. Paschke, O. Kaske, M. A. Beketov, M. Hundt, C. Bauer, G. Schüürmann, M. Möder, M. Liess, *Science of The Total Environment*, **2015**, *537*, 69–80.
   DOI:10.1016/j.scitotenv.2015.07.012

4. A. Sabarwal, K. Kumar, R. P. Singh, *Environmental Toxicology and Pharmacology*, **2018**, *63*, 103–114.
   DOI:10.1016/j.etap.2018.08.018

5. V. Sujitha, K. Murugan, D. Dinesh, A. Pandiyan, R. Aruliah, J.Hwang, K. Kalimuthu, C. Panneerselvam, A. Higuchi, A. T. Aziz, S. Kumar, A. A. Alarfaj, B. Vaseeharan, A. Canale, G. Benelli, *Aquatic Toxicology*, **2017**, *188*, 100–108
   DOI:10.1016/j.aquatox.2017.04.015

6. K. V. Ragnarsdottir, *Journal of the Geological Society*, **2000**, *157*, 859–876. **DOI:**10.1144/jgs.157.4.859

7. N. E. El-Wakeil, *Gesunde Pflanzen*, **2013**, *65*, 125–149
   DOI:10.1007/s10343-013-0308-3

8. E. V. R. Campos, P. L. F.Proença, J. L. Oliveira, M. Bakshi, P. C. Abhilash, L. F. Fraceto, *Ecological Indicators*, (in press), Available online 27 April **2018**. **DOI:**10.1016/j.ecolind.2018.04.038

9. Y. Huang, L. Li , J. Liu, W. Lin, *Algal Research*, **2014**, *4*, 62–69.
   DOI:10.1016/j.algal.2013.08.001

10. M. L. Umpiérrez, J. Paullier, M. Porrini, M. Garrido, E. Santos, C. Rossini, *Industrial Crops and Products*, **2017**, *109*, 686–692. **DOI:**10.1016/j.indcrop.2017.09.025

11. B. I. Murray, *Pest Management Sci.* **2008**, *64*, 8–11.
    DOI:10.1002/ps.1470

12. R. Potts, M. R. Clarke, S. E. Oldfield, K. L. Wood, N. H. de Ibarra, J. E. Cresswell, *Journal of Insect Physiology,* **2018**, *104*, 33–39. **DOI:**10.1016/j.jinsphys.2017.11.006.

13. O. Ogbeide, A. Chukwuka, I. Tongo, L. Ezemonye, *Journal of Environmental Management*, **2018**, *217*, 23–37.
    DOI:10.1016/j.jenvman.2018.03.065

14. M. Kah, C. D. Brown, *Chemosphere*, **2008**, *10*,1401–1408, https: //doi.org/10.1016/j. chemosphere. 2008.04.074.

15. F. Tsopelas, C.Giaginis, A. Tsantili-Kakoulidou, *Expert Opinion on Drug Discovery*, **2017**, *12*, 885–896.
    DOI:10.1080/17460441.2017.1344210

16. C. Vraka, L. Nics, K. Wagner, M. Hacker, W. Wadsak, M. Mitterhauser, *Nuclear Medicine and Biology*. **2017**, *50*, 1–10.
    DOI:10.1016/j.nucmedbio.2017.03.003

17. V. S. Talismanov, S. V. Popkov, O. G. Karmanova, S.S. Zykova, A. P., Chernobrovkina, *J. Pharm. Sci. & Res.* **2017**, *9*, 2372–2375.

18. E. Benfenati, G. Gini, N. Piclin, A. Roncaglioni, M. R. Varı, *Chemosphere*, **2003**, *53*, 1155–1164.
    DOI:10.1016/S0045-6535(03)00609-X

19. R. N. Waterhouse, *Molecular Imaging & Biology*, **2003**, *5*, 376–389. **DOI:** 0.1016/ j.mibio. 2003.09.014

20. A. Finizio, M. Vighi, D. Sandroni, *Chemosphere*, **1997**, *34*, 131–161. **DOI:**10.1016/S0045-6535(96)00355-4

21. S. G. Machatha, S. H. Yalkowsky, *Int. J. Pharm*, **2005**, *294*, 185–192. **DOI:**10.1016/j.ijpharm.2005.01.023

22. Y. P. Chin, W. J Weber, T. C. Voice, *Water Research*, **1986**, *20*, 1443–1450. **DOI:**10.1016/0043-1354(86)90144-2

23. T. Ferrari, A. Lombardo, E. Benfenati, *Science of The Total Environment*, **2018**, *638*, 1158–1165.
    DOI:10.1016/j.scitotenv.2018.05.072

24. C. J. Churchwell, M. D. Rintoul, S. Martin, D. P. Jr Visco, A. Kotu, R. Slarson, L. O. Sillerud, D. C. Brown, J. L. Faulon, *J Mol Graph Model*, **2004**, *22*, 263–273.
    DOI:10.1016/j.jmgm.2003.10.002

25. J. L. Faulon, D. P. Jr.Visco, R. S.Pophale, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 707–720. **DOI:**10.1021/ci020345w

26. T. Miyao, H. Kaneko, K. Funatsu, *J Chem Inf Model*. **2016**, *56*, 286–299. **DOI:**10.1021/acs.jcim.5b00628

27. T. Miyao, M. Arakawa, K. Funatsu, *Mol Inform*. **2010**, *29*, 111–125. **DOI:**10.1002/minf.200900038

28. C. D. Weis, D. P. Visco, *Computers & Chemical Engineering*, **2010**, *34*, 1018–1029.
    DOI:10.1016/j.compchemeng.2009.10.017

29. J. L. Faulon, *J. Chem. Info. Comput. Sci.* **1994**, *34*, 1204–1218.
    DOI:10.1021/ci00021a031

30. V. A. Dev, N. G. Chemmangattuvalappil, M. R. Eden, *Computer Aided Chemical Engineering*, **2014**, *33*, 151–156.

31. N. Brown, B. McKay, J. Gasteiger, *J. Comput. Aided Mol. Des.* **2006**, *20*, 333–341. **DOI:**10.1007/s10822-006-9063-1

32. C. Hansch, R. M. Muir, T. Fujita, P. P. Maloney, F. Geiger, M. Streich, *j. Am. Chem. Soc.* **1963**, *85*, 2817–2825.
    DOI:10.1021/ja00901a033

33. E. X. Esposito, A. J. Hopfinger, J. D. Madura, *Methods. Mol. Biol.* **2004**, *275*, 131–214.
    DOI:10.1385/1-59259-802-1:131

34. W. M. Brown, S. Martin, M. D. Rintoul, J. L. Faulon, *J. Chem. Inf. Model.*, **2006**, *46*, 826–835. **DOI:**10.1021/ci0504521

35. P. Donald, J. R. Visco, R. S. Pophale, M. D. Rintoul, and J. L. Faulon, *Journal of Molecular Graphics and Modelling*, **2002**, *20*, 429–438.

36. J. L. Faulon, C. J. Churchwell, D. P. Jr. Visco, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 721–734. **DOI:**10.1021/ci020346o

37. R. Veerasamy, H. Rajak, A.Jain, S. Sivadasan, C. P. Varghese, R.K. Agrawal , *International Journal of Drug Design and Discovery* . **2011**, *2*, 511–519

38. A. Golbraikh, A. Tropsha, *J. Mol. Graphics Mod.* **2002**, *20*, 269–276.
    DOI:10.1016/S1093-3263(01)00123-1

Povzetek

Predstavljena raziskava preučuje uporabo inverznega-QSAR/QSPR pristopa za generiranje novih struktur pesticidov, kar je izredno pomembno v luči njihove široke uporabe, še posebej na področju kmetijstva. S tem namenom smo uporabili metodologijo devetih zaporednih korakov, ki vključujejo kvantitativno študijo relacije med strukturo in aktivnostjo/ lastnostmi (QSAR/QSPR) s ciljem razviti model, ki povezuje strukturo 190 pesticidov z njihovim porazdelitvenim koeficientov za sistem n-oktanol-voda ($\log k_{ow}$). Uporabili smo enolične atomske deskriptorje, ki predstavljajo strukture in nastopajo kot neodvisne spremenljivke, medtem kot je vrednost $\log k_{ow}$ odvisna spremenljivka. Model smo razvili na učenju z nizom 130 molekul, njegovo sposobnost napovedovanja pa smo preverili na ostalih 60 komponentah. Modeliranje $\log k_{ow}$ vrednosti teh komponent kot funkcije deskriptorjev smo izvedli z večkratno linearno regresijo (MLR) z uporabo pristopa izpusti-enega (LOO) navzkrižne validacije. Rezultat je QSAR/QSPR enačba s 14 atomskimi deskriptorji z $R^2 = 0.659273$, $Q^2 = 0.65617$ in $RMSE_{training} = 0.930192$, $s = 1.37297$ za niz na katerem smo model učili ter $q^2 = 0.605676$, $RMSE_{LOO} = 1.0936$ s pristopom izpusti-enega (LOO) navzkrižne validacije na testiranem nizu. Generirali smo tudi nove strukture pesticidov, ki bi bili lahko vključeni v nadaljnje študije.