QSAR trout toxicity models on aromatic pesticides

Svetoslav Slavov a, Giuseppina Gini a, Emilio Benfenati b

a Department of Electronics and Information, Politecnico di Milano, Milano, Italy
b Istituto di Ricerche Farmacologiche “Mario Negri”, Milano, Italy

Online Publication Date: 01 November 2008
QSAR trout toxicity models on aromatic pesticides

SVETOSLAV SLAVOV\(^1\), GIUSEPPINA GINI\(^1\) and EMILIO BENFENATI\(^2\)

\(^1\)Department of Electronics and Information, Politecnico di Milano, Milano, Italy
\(^2\)Istituto di Ricerche Farmacologiche “Mario Negri,” Milano, Italy

The pesticides originally designed to kill target organisms are dangerous for many other wild species. Since they are applied directly to the environment, they can easily reach the water basins and the topsoil. A dataset of 125 aromatic pesticides with well-expressed aquatic toxicity towards trout was subjected to quantitative structure activity relationships (QSAR) analysis aimed to establish the relationship between their molecular structure and biological activity. A literature data for LC\(_{50}\) concentration killing 50% of fish was used. In addition to the standard 2D-QSAR analysis, a comparative molecular field analysis (CoMFA) analysis considering the electrostatic and steric properties of the molecules was also performed. The CoMFA analysis helped the recognition of the steric interactions as playing an important role for aquatic toxicity. In addition, the transport properties and the stability of the compounds studied were also identified as important for their biological activity.

Keywords: Pesticides; QSAR; COMFA; modelling; aquatic toxicity.

Introduction

The toxicological profile of aromatic compounds has been under investigation since the 1970s. The chemical interaction of the toxicant with the organism results in a number of biochemical and thus physiological effects. The interaction of toxicants described at the molecular level is termed the mechanism of toxic action of the chemical. To understand the toxic behavior of chemicals and for classification purposes toxic modes of action (MOA) have been identified in aquatic species, such as non-polar narcosis, polar narcosis, uncoupling of oxidative phosphorylation, respiratory membrane irritation, acetylcholinesterase inhibition, central nervous system seizure, inhibition of photosynthesis, and alkylation, but different classifications have been proposed.\(^1\)−\(^3\) Classification of chemicals into appropriate MOA is particularly complicated with variety of functional groups involved in one compound and in cases of metabolic rearrangements in the cell environment.

For diverse sets of chemicals, where specific mechanism of action is assumed, the toxicity effect can be expressed as a combination of penetration into or through biological membranes and the interaction of the toxicant with the site of action. McFarland \(^4\) represented this principle mathematically by the following generic QSAR equation:

\[
\log(\text{toxicity})^{-1} = A(\log \text{of penetration}) + B(\log \text{of interaction}) + C
\]

One issue in toxicity prediction and modelling is how to describe the chemical information. Russom et al. used the presence of fragments in the molecule and developed an expert system for prediction of the MOA and then toxicity.\(^3\)

Chemical descriptors of 2- and 3-dimensional structures have been used in many cases. For instance, Pintore et al. developed a model for toxicity prediction and compared 2- and 3-D descriptors.\(^5\) A quite rich chemical description is likely important considering complex structures, where specific reactivity is involved; conversely, if the main mechanism is narcosis, quite general and unspecific descriptors, such as logP, can be suitable. Pesticides, by their definition and chemical complexity, act through a series of mechanisms involving specific reactivity.

In recent years CoMFA has been widely used in toxicity prediction.\(^6\)−\(^9\) Its basic assumption is that at the molecular level the steric and electrostatic interactions produce an observable biological effect. The analysis of the data sampled at the intersections of a 3D-lattice, by partial least squares (PLS), using cross-validation to maximize the likelihood of predictive power. Thus, the relationships between the fields and the activities can be built without knowledge of the 3D-structures of the receptors. The graphical representation of the results provides a basis for the mechanism study and new molecular understanding of toxicity. In addition to the standard 2D-QSAR approach, the use of CoMFA for QSAR modelling of pesticides was also evaluated.
Material and methods

Toxicity data

Toxicity data are based on the U.S. Environmental Protection Agency (EPA) database of ecotoxicological data. Only pesticides with measured at 96h lethal concentration killing 50% of the trout population (LC50 (mmol/l)) were selected. A logarithmic transformation function was applied to obtain data distribution closer to Gaussian (normal). When more than one toxicity value was reported, all selected. When more than one toxicity value was reported, all values above or below the average for 25%, were discarded and then the average of the remaining values was calculated (Table 1).

Computational procedure

Different types of electronic (total energy, core-core repulsion energy, electronic energy, HOMO, LUMO, Dipole moments and their components, etc.), physico-chemical (heat of formation, etc.), geometrical (CODE_POS, CODE_NEG, CODE_MID, representing the projections of the isoelectrostatic potential surfaces over the Van der Waals surface of the molecule for regions where the energy is higher than 10 kcal/mol, lower than −10 kcal/mol, and between −10 kcal/mol and 10 kcal/mol, respectively. Van der Waals volume, etc.), and lipophilic descriptors (LogP) were calculated using Chem-X (version 1999.1, Oxford Molecular Ltd., Oxford, UK) and HyperChem (version 6.0, Hypercube, Inc., Gainesville, USA) programs. All descriptors were evaluated using the PM3 semi-empirical quantum mechanical method. The computational procedure involved the following steps: i) theoretical determination of geometrical parameters and assessment of the conformational isomerism of the compounds from the series; ii) alignment of the structures using Chem-X flexyfit 3D-search method in case of CoMFA analysis; iii) evaluation of the steric and electrostatic energies of interaction; iv) partial least squares (PLS) and weighted least squares (WLS) for important region mapping and iv) a forward stepwise multilinear regression procedure, as implemented in STATISTICA program (Version 6.1, StatSoft Italia, Italy) for 2D-QSAR model generation.

Statistical parameters

Conventional ($R^2$) correlation coefficient, cross-validated correlation coefficient ($Q^2$), predictive correlation coefficient ($R^2_{pred}$), standard deviation (SD) and the Fischer criterion were used as criteria for statistical significance and predictive ability of the QSAR models reported. An external test set of 37 compounds was used for validation.

Results and discussion

2D-QSAR analysis

The best multilinear QSAR equation obtained (see Figure 1 and Eq. 1) involved three independent variables: CODE_MID, molecular weight and heat of formation, all
having positive regression coefficients.

\[
\log(1/LC_{50}) = 0.010\text{CODE\_MID} + 0.008\text{MWEIGHT} + 0.008\text{HEATOFFORMATION} - 3.412 \quad (\text{Eq.1})
\]

\[
n = 96; R^2 = 0.70; F(3, 92) = 71.21; SD = 0.70
\]

The CODE\_MID descriptor represents the areas of the Van der Waals surface where the projections of the isoelectrostatic potential are in the range between – 10 kcal/mol and 10 kcal/mol. Due to the positive regression coefficient sign, larger CODE\_MID values will lead to an increased toxicity effect of pesticides. The remaining two descriptors, i.e. the molecular weight and the heat of formation could be related to the transport properties and the thermodynamic stability of the compounds, respectively. The pesticides characterized by larger heats of formation are more stable and thus the probability to reach the target site unchanged is higher. Since no charge distribution related descriptors (such as CODE\_NEG or CODE\_POS) were involved into the model, it can be concluded that the electrostatic interactions are of much lesser importance for the aquatic toxicity than the steric interactions. This conclusion is fully supported by the CoMFA results obtained.

However, due to the moderate quality of the 2D-QSAR analysis results, we decided to explore the structure-activity relationship applying the methods of 3D-QSAR.

3D-QSAR analysis

As a common structural feature for all compounds the presence of an aromatic ring was selected as a searching criterion for the alignment (Fig. 1). In addition, this is well-known as one of the most important pharmacophore structures. Aiming to avoid bad Van der Waals contacts between the atoms within the molecules the “bump check” option was selected. For the purposes of CoMFA, the initial dataset was randomly split into two subsets: training (87 compounds) and test (38 compounds).

At the next stage, all aligned ligands (see Figure 2) were placed in a 3D-lattice with 2.0 angstrom grid spacing along all cartesian directions. Using a fictitious hydrogen probe atom with a charge of +1 the steric and electrostatic fields of the ligands at various grid points of the lattice were calculated. The resulting field matrix was then analyzed by the partial least squares (PLS) method. The WLS method was further used for 3D-mapping of the important for the steric and electrostatic interactions regions.

Due to their low squared cross-validated correlation coefficient \(Q^2_\text{E} = 0.32\) the contribution of the electrostatic interactions to the explanation of the data variance was considered insufficient. However, the PLS analysis conducted for the steric interactions resulted in \(R^2 = 0.90; F = 310.73; Q^2_\text{S} = 0.75\) for the training set and \(R^2_{\text{pred}} = 0.89\) for the test set (see Figures 3 and 4), respectively. On the basis of its deviation from the regression line one compound (CAS number 94757) from the test series was identified as an outlier and was therefore removed.

From all the results outlined above, it can be concluded that the steric interactions play a much more important role for the aquatic toxicity than the electrostatic. The visual examination of the steric interactions map (see Figure 5) showed that the presence of bulky substituents around
positions 3 and 4 of the aromatic ring and near the het-
eratoms of the side chain will lead to an increased toxicity
effect.

The quality of the CoMFA analysis results clearly
demonstrates the following advantages:

(i) It provides highly predictive models for a data sets of
large number of similar compounds;

(ii) The calculation process is relatively fast;

(iii) The derived results are easily interpretable and provide
high predictive abilities.

However, the requirement for common substructural fea-
tures somehow narrows the diversity of the compounds
within the dataset (in our case the presence of an aromatic
ring in the structure). The results on pesticides currently
reported were found superior to those obtained for other
pesticide data sets, such as those we recently presented using
the DEMETRA software.\[11\]

As is it well-known, the pesticides are chemicals typi-
cally containing a variety of functional groups, and thus,
the modelling of their toxic effect is a challenging task.
For instance, in similar case of toxicity prediction for
another aquatic organism, Daphnia magna, the DEMETRA
program produced results superior to those by other
software.\[12\]

Conclusion

A new 3D-QSAR model for the acute aquatic toxicity to-
wars trout was presented. The application of CoMFA led
to a satisfactory QSAR model, demonstrating the impor-
tance of the steric interactions. The CoMFA model devel-
oped has been validated using an external set of chemicals,
which confirms the model robustness.

The advantages and limitations of the models have been
discussed, in respect to the modelling scheme applicability.

Acknowledgments

We acknowledge the European Commission (EC) funded
project Intelligent Modeling Algorithms for General Eval-
uation of Toxicities (IMAGETOX).
References