Journal of Environmental Science and Health, Part B
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597269

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

To cite this Article Slavov, Svetoslav, Gini, Giuseppina and Benfenati, Emilio(2008)'QSAR trout toxicity models on aromatic pesticides',Journal of Environmental Science and Health, Part B,43:8,633 — 637
To link to this Article: DOI: 10.1080/03601230802352658
URL: http://dx.doi.org/10.1080/03601230802352658

Journal of Environmental Science and Health Part B (2008) **43**, 633–637

ISSN: 0360-1234 (Print); 1532-4109 (Online)
DOI: 10.1080/03601230802352658

Taylor & Francis
Taylor & Francis Group

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:

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

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.

Address corespondence to Emilio Benfenati, Laboratory of Environmental Chemistry and Toxicology Istituto di Ricerche Farmacologiche "Mario Negri", Via La Masa 19, 20156 Milano, Italy; E-mail: benfenati@marionegri.it
Received April 23, 2008.

Table 1. Literature data for the biological activity. Chemical Abstracts Service (CAS) number was used for compound identification.

CAS	*LOG(1/LC$_{50}$)*	*CAS*	*LOG(1/LC$_{50}$)*	*CAS*	*LOG(1/LC$_{50}$)*	*CAS*	*LOG(1/LC$_{50}$)*	*CAS*	*LOG(1/LC$_{50}$)*
50293	1.9368	709988	−1.0233	3691358	0.1721	25311711	−0.7175	52918635	2.7441
55389	−0.2964	957517	−2.6084	3861414	1.0777	26002802	1.3214	53404221	−3.301
63252	−0.8429	1214397	−1.9782	5234684	−0.93	26225796	−0.4187	53404312	−0.9063
72435	0.4179	1320189	−0.4527	5259881	−1.8723	27314132	−1.4263	55283686	0.9542
76879	1.1047	1563662	−0.2354	7085190	−2.7648	28249776	−0.6104	58138082	−0.2184
83261	0.0395	1582098	1.1826	7166190	−0.5231	28434017	2.7365	60168889	−0.8024
88040	−0.6863	1689845	−0.878	7745893	−1.836	28772567	−0.4244	60207901	−0.466
94746	−2.6569	1689992	0.6051	10007859	−2.7715	33089611	−1.9865	62476599	−1.6735
94757	−3.2095	1918009	−2.1028	10453868	2.2736	33245395	1.2498	62924703	1.2446
94804	−0.4606	1918167	0.0949	10605217	−0.0807	33629479	−0.0984	66332965	−1.2232
94826	−1.7591	1929733	−0.8136	13684565	−0.7533	33820530	−0.4148	66441234	−0.1046
99309	−0.8881	1982496	−1.7485	13684634	−0.6721	35367385	−2.8881	67485294	0.5176
100027	−1.9709	1982690	−3.4042	15299997	−1.5714	35400432	−1.8539	68359375	−0.195
101053	0.294	2008391	−3.0555	15972608	−1.1376	36734197	−1.1047	69409945	2.2387
101213	−1.4265	2032657	−0.2873	17804352	−0.1504	39300453	1.385	74051802	−0.425
114261	−1.2481	2039465	−2.7682	19044883	−0.9741	39515418	2.1811	76578148	−0.3684
120321	−0.5179	2122705	−1.8157	19666309	−1.233	40487421	−0.5513	77501634	−0.904
120365	−1.0603	2164172	−1.7959	21564170	0.7234	41198087	1.2499	81335377	−2.9544
122145	−0.6715	2312358	0.4722	22224926	−0.2691	43222486	−3.4557	81777891	−1.8994
133073	0.2049	2439012	−0.0458	22248799	−0.07	50471448	−0.997	86209510	−2.652
134623	−2.5718	2491385	−0.7817	22781233	−0.7309	51218452	−1.1385	88671890	−1.1631
140567	−2.958	2492264	−0.6432	23422539	−1.2991	51338273	0.171	90982324	−1.3068
148798	−0.9521	2536314	−0.6259	23564058	−1.8674	51630581	2.5436	110488705	−1.2041
330541	−1.9249	2675776	−1.2522	23950585	−2.4491	52315078	−1.5048	112410238	−1.1842
330552	−1.0809	3380345	0.0023	25168267	−2.1851	52645531	2.1298	116255482	−0.8709

Material and methods

Toxicity data

Toxicity data are based on the U.S. Environmental Protection Agency (EPA) database of ecotoxicological data.[10] Only pesticides with measured at 96h lethal concentration killing 50% of the trout population (LC_{50} (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 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 mutlilinear 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

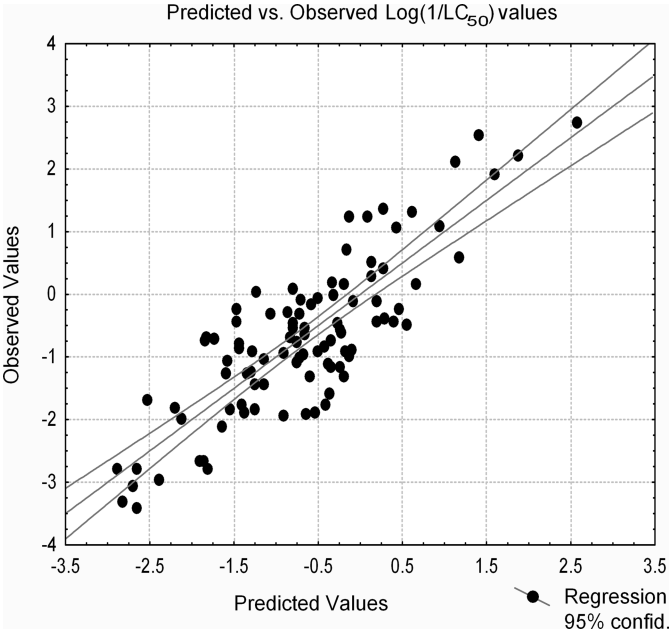

Fig. 1. Predicted vs. observed $Log(1/LC_{50})$ values.

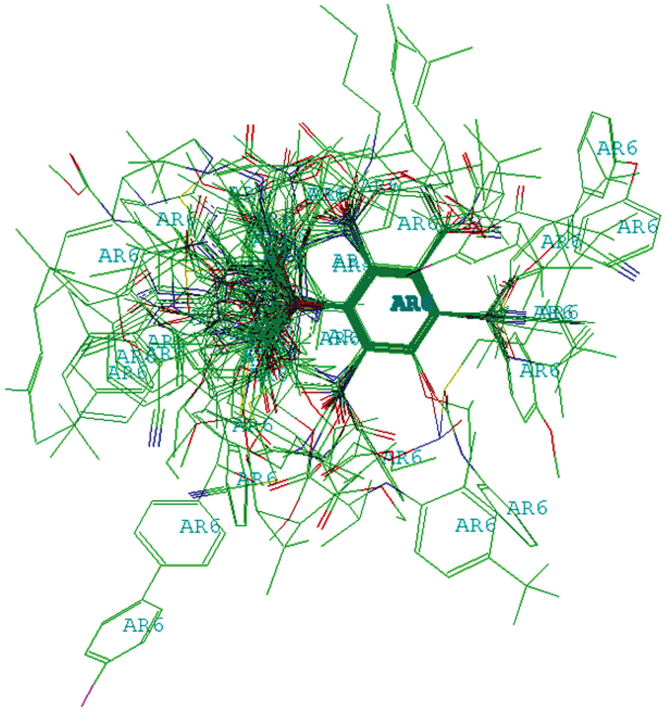

Fig. 2. Aligned molecules from the series.

having positive regression coefficients.

$$Log(1/LC_{50}) = 0.010CODE\_MID + 0.008MWEIGHT + 0.008HEATOFFORMATION - 3.412 \quad (Eq.1)$$

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

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 ($Q^2_E = 0.32$) 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 $R^2 = 0.90$; $F = 310.73$; $Q^2_S = 0.75$ for the training set and $R^2_{pred} = 0.89$ 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

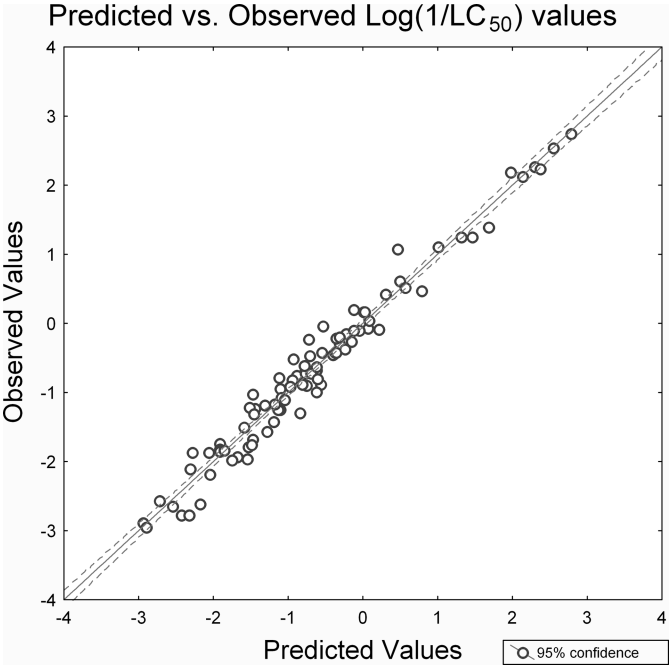

Fig. 3. CoMFA results for the training set.

positions 3 and 4 of the aromatic ring and near the heteroatoms 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;

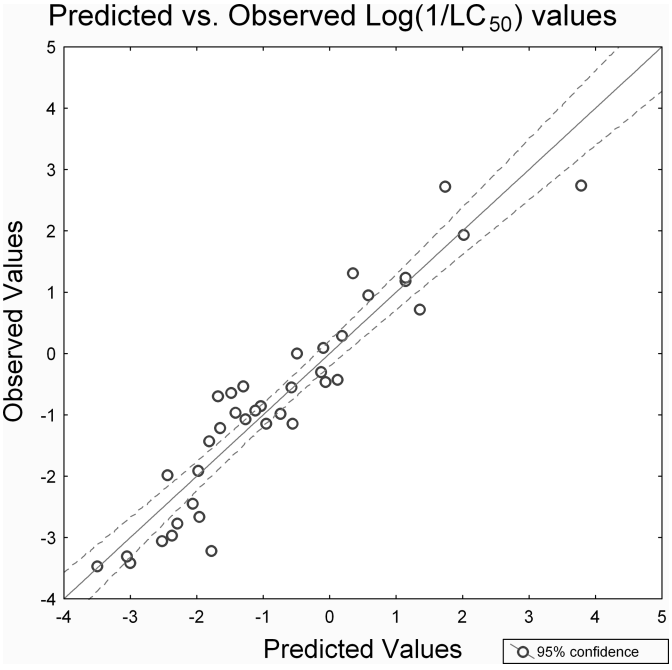

Fig. 4. CoMFA results for the test set.

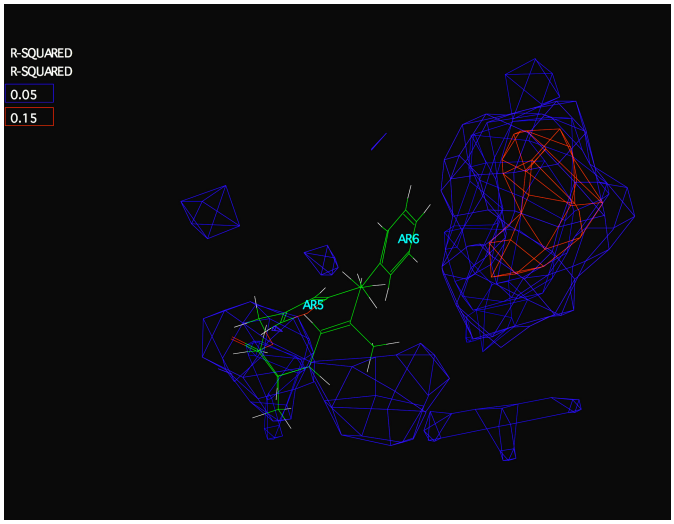

Fig. 5. 3D-WLS map of the steric interactions.

(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 features 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 typically 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 towards trout was presented. The application of CoMFA led to a satisfactory QSAR model, demonstrating the importance of the steric interactions. The CoMFA model developed 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 Evaluation of Toxicities (IMAGETOX).

References

[1] Bradbury, S. P. Predicting modes of toxic actions from chemical structure: An overview. SAR and QSAR Environ. Res. **1994**, *2*, 89–104.

[2] Vehraar, H.J.; Van Leeuwen, C.J.; Hermens, J.L.M. Classifying environmental pollutants. 1: Structure-activity relationships for prediction of aquatic toxicity. Chemosphere. **1992**, *25*, 471–491.

[3] Russom, C.L.; Bradbury, S.P.; Broderius, S.J.; Hammermeister, D.E.; Drummond, R.A. Predicting modes of action from chemical structure: Acute toxicity in the Fathead Minnow (Pimephales Promelas). Environ. Toxicol. Chem. **1997**, *16*, 948- 957.

[4] McFarland, J.W. On the Parabolic Relationship Between Drug Potency and Hydrophobicity. J. Med. Chem. **1970**, *13*, 1092–1099.

[5] Pintore, M.; Piclin, N.; Benfenati, E.; Gini, G.; Chrétien, J.R. Predicting toxicity against the fathead minnow by adaptative fuzzy partition. QSAR. **2003**, 22, 210–219.

[6] Pasha, F.A.; Muddassar, M.; Chung, H.W.; Cho, S.J.; Cho, H. Hologram and 3D-quantitative structure toxicity relationship studies of azo dyes. Journal of Molecular Modeling. **2008**, *14*(4), 293–302.

[7] Fratev, F.; Mihayova, E.; Tadjer, A.; Benfenati, E. Study of species-specific carcinogenicity of benzene derivatives. 1. Combination of CoMFA and GRID analysis, Oxidation Communications. **2007**, *30*(4), 891–911.

[8] Lo Piparo, E.; Smiesko, M.; Mazzatorta, P.; Benfenati, E.; Idinger, J.; Blümel, S. Preliminary analysis of toxicity of benzoxazinones and their metabolites for Folsomia candida. Journal of Agricultural and Food Chemistry. **2006**, *54*(4), 1099–1104.

[9] Liu, X.; Yang, Z.; Wang, L. Three-dimensional quantitative structure-activity relationship study for phenylsulfonyl carboxylates using CoMFA and CoMSIA, Chemosphere **2003**, *53*(8), 945–952.

[10] Mid-Continent Ecology Division, Environmental Protection Agency, Duluth, Minn. ECOTOX (ECOTOXicology) database. Release 4.0. http://cfpub.epa.gov/ecotox/

[11] Benfenati, E. Quantitative Structure-Activity Relationships (QSAR) for Pesticide Regulatory Purposes. Elsevier: Amsterdam, 2007; 510 pp.

[12] Porcelli, C.; Boriani, E.; Roncaglioni, A.; Chana, A.; Benfenati, E. Regulatory perspectives in the use and validation of QSAR. A case study: DEMETRA model for daphnia toxicity. Environ. Sci. Technol. **2008**, *42*, 491–496.