# STRUCTURE-TOXICITY STUDY OF SOME PYRETHROIDAL ESTER INSECTICIDES

Simona Funar-Timofei, Sorin Avram, Ana Borota
Institute of Chemistry of the Romanian Academy, Bv. Mihai Viteazu 24, 300223 Timisoara, Romania

## ABSTRACT

Mortality $LD_{50}$ values of a series of 37 pyrethroidal esters have been previously reported against a susceptible strain of housefly (*Musca domestica*). The aim of the study is to correlate the structure features of these compounds to the logarithm of $LD_{50}$ values by multiple linear regression (MLR). Pyrethroidal esters structures were first built by the Marwin Sketch software and then conformational analysis was performed by the OMEGA program. Structural descriptors were calculated for the active stereoisomers mentioned in the literature by the Dragon, Instant JChem and EPI Suite™ software. Genetic algorithm was applied for descriptor selection. The dataset was divided into training and test (obtained randomly by taking out 25% from the entire series) sets. MLR analysis has been applied after variable selection performed by the genetic algorithm included in the QSARINS v. 1.2 program. The analysis was carried out using the RQK fitness function, with leave-one-out cross-validation correlation coefficient as constrained function to be optimized. Structural features which influence insecticide toxicity were discussed.

**Keywords**: pyrethroidal esters, insecticides, $LD_{50}$, MLR, genetic algorithm

## INTRODUCTION

The housefly is an important medical pest because it carries a large number of pathogens of human and animal diseases [1]. The use of insecticides is an effective method for the control of housefly populations.

Pyrethroids constitute one of the most widely used classes of insecticides worldwide [2]. They are synthetic structural derivatives of natural pyrethrins present in the pyrethrum extract of Chrysanthemum species.

Pyrethroids show the following characteristics [3]: quick knock-down effect against insects, efficacy against insects with organophosphorus and/or carbamate-resistant strains, easy decomposition in the environment and low mammalian toxicity [3-5]. Such features make

pyrethroids useful for household insecticides, and those with improved chemical stability, for example, to light and in the air, are valuable agrochemicals [3]. The favourable selective toxicity of pyrethroids can be explained by the fact that pyrethroids which act on the nervous system are metabolised and excreted by mammals before reaching the central nervous system. On the other hand, pyrethroids do reach the nervous system in insects, causing such symptoms as excitement and paralysis, eventually leading to knock-down or death of the insects.

Pyrethroids act primarily on the nervous system, although the specific mechanism of activity is uncertain [6]. Several mechanisms of action have been proposed, including alterations in sodium channel dynamics in nerve tissues, which polarise membranes and result in abnormal discharge in targeted neurons. Pyrethroids are composed of several structural groups, including an acid moiety, a central ester bond and an alcohol moiety. The acid moiety contains two chiral carbons, meaning that the pyrethroids typically exist as stereoisomeric compounds. Furthermore, some compounds also contain a chiral carbon on the alcohol moiety, which allows for three chiral carbons and a total of eight different stereoenantiomers. All pyrethroids can therefore exist in at least four stereoisomeric forms, each with different biological activities. They may be formulated as racemic mixtures or as single isomers (e.g. deltamethrin) and different isomers may have individual common names. The mechanisms by which pyrethrins and pyrethroids alone are toxic are complex. The cis isomers are usually more toxic than the trans isomers [6]. For example, the 1R and 1S cis isomers bind competitively to one site, and the 1R and 1S trans isomers bind non-competitively to another. In mammals the 1R isomers are active and the 1S isomers inactive, making the 1S isomers non-toxic.

Although this classification system is widely employed, it has several shortcomings for the identification of common toxic effects [7]. In particular, it does not reflect the diversity of intoxication signs found following oral administration of various pyrethroids. Pyrethroids act in vitro on a variety of putative biochemical and physiological target sites, four of which merit consideration as sites of toxic action: voltage-sensitive sodium, calcium and chloride channels, and peripheral-type benzodiazepine receptors.

There are two basic signs of pyrethroid toxicity in laboratory rodents [4]. The first sign is Type I or T (tremor) syndrome, which is characterized by whole-body tremor, aggressive behavior, hyperexcitation and ataxia. The second sign is Type II, or CS (choreoathetosis with salivation) syndrome, which is characterized by choreoathetosis and profuse salivation. In general, pyrethroids without an α-cyano group (e.g., bifenthrin and permethrin) produce Type I symptoms, and pyrethroids with this group (e.g., deltamethrin) produce Type II symptoms. In humans, the symptoms that may arise from acute oral exposure to pyrethroids include dizziness, headache,

nausea, anorexia, fatigue, vomiting, mild disturbance of consciousness, or muscular fasciculation in limbs.

Toxicity among the various pyrethroids varies greatly, as is evidenced by the wide range in $LD_{50}$ values (concentrations or doses that result in 50% mortality in exposed laboratory animals) [8]. These differences are related to several factors, including specific pyrethroid, ratios of stereo and optical isomers within a given pyrethroid formulation, and vehicle. Acute oral $LD_{50}$ values are generally lower in Type II than Type I pyrethroids, indicating a greater degree of toxicity for Type II pyrethroids. In the case of tetramethrin, like all other Type I pyrethroids, isomers of the 1R conformation are considerably more toxic than those of the 1S conformation. The 1S isomer can also inhibit toxicity by competitive inhibition at a number of stereospecific pyrethroid binding sites, thus preventing binding of the more toxic 1R isomer. Furthermore, it has been observed that the cis isomers possess greater mammalian toxicity than the trans isomers. For Type II pyrethroids, the S conformation at the alpha carbon adjacent to the cyano group is considerably more toxic than the R conformation.

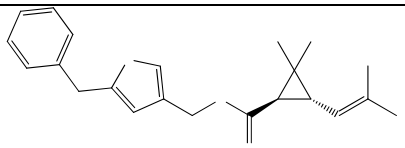
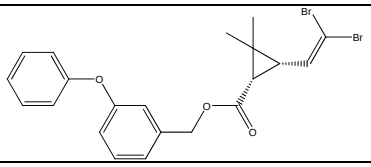
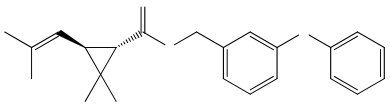
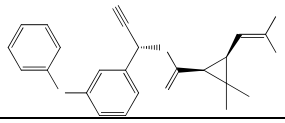
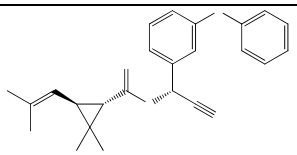
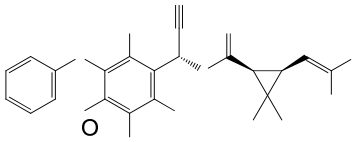
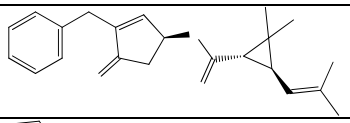
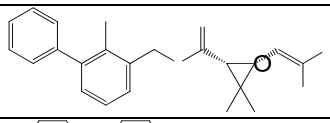
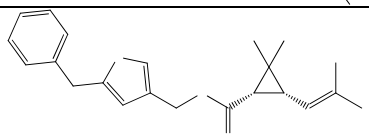
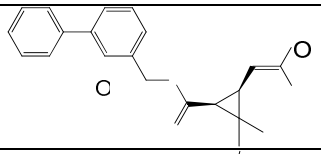
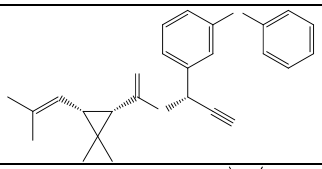
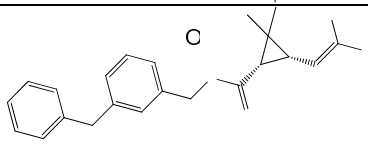
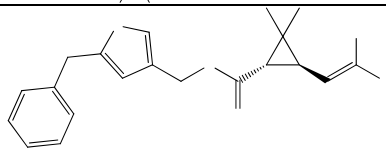
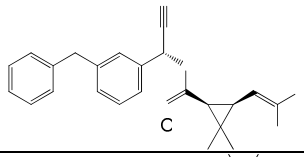
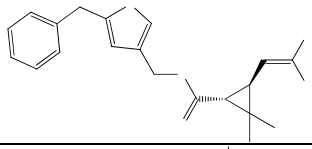
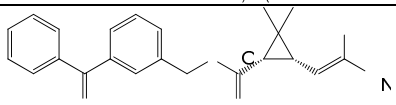
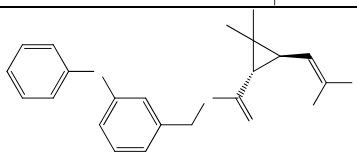
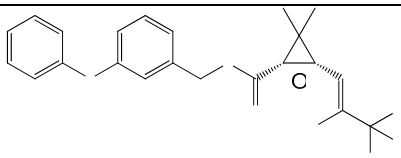
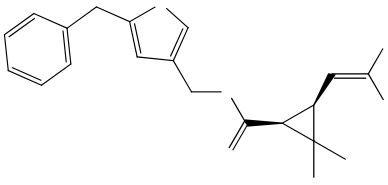
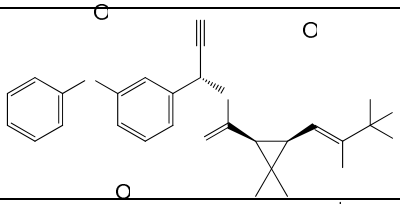
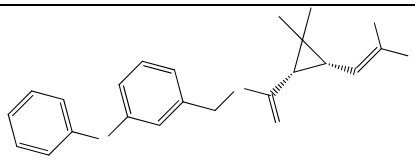
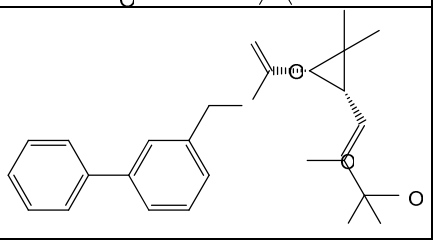
This paper studies the toxicity of 37 pyrethroidal esters (Table 1), expressed by the logarithm of $LD_{50}$ values measured against a susceptible strain of housefly (*Musca domestica*), by multiple linear regression (MLR). Stereoisomers selected according to the literature [9] were modeled by conformational analysis performed by molecular mechanics calculations. Structural descriptors of the title compounds calculated for these isomers were correlated to the logarithm of $LD_{50}$ values.

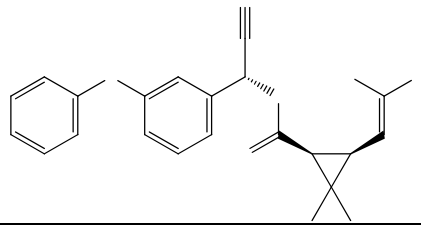
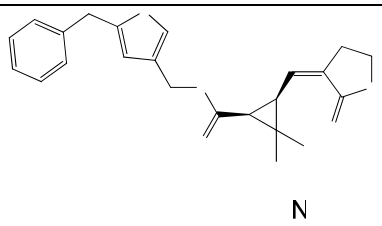
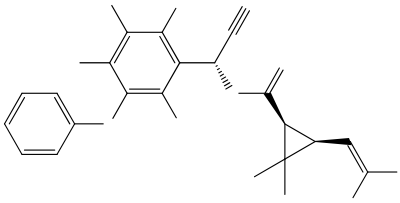
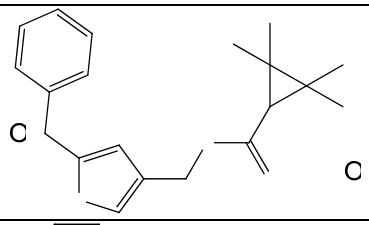
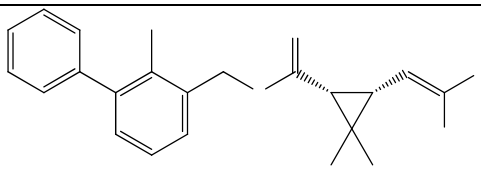
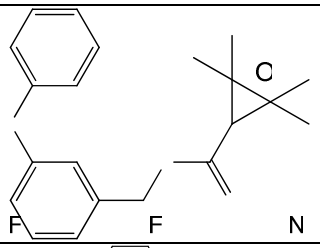
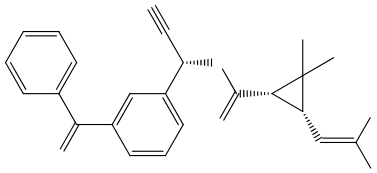
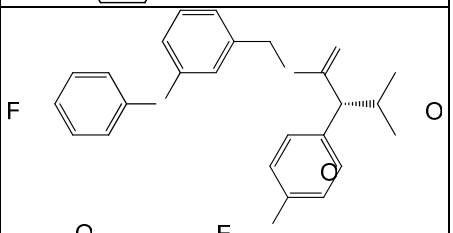
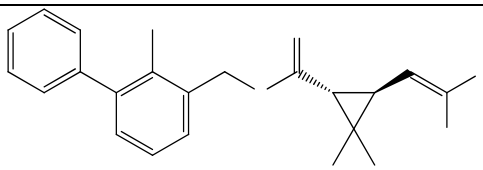
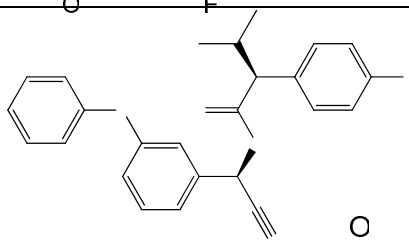
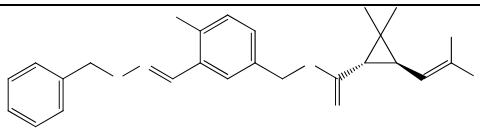
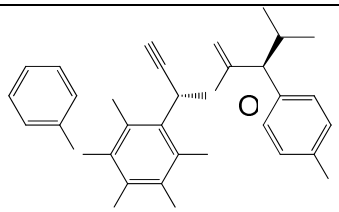
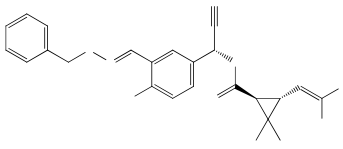
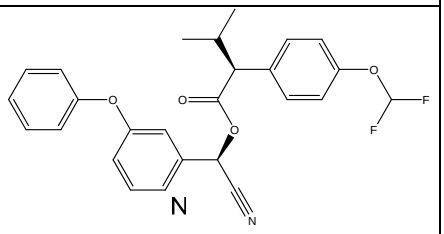
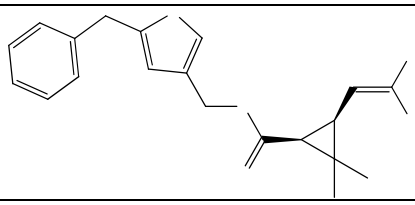
## METHODS

### Definition of target property and molecular structures

The experimental $LD_{50}$ values of 37 pyrethroidal ester derivatives have been previously [9] measured against a susceptible strain of housefly (*Musca domestica*). Starting structures were first built by the Marwin Sketch (Marwin Sketch 6.0, 2013, ChemAxon, http://www.chemaxon.com) software and then conformational analysis was performed by the OMEGA (Omega v.2.4.6 (OpenEye Scientific Software, Santa Fe, NM., http://www.eyesopen.com, 2010) program [10]. Conformers' generation was performed using the default parameters except the maximum number of conformations to be generated that was set to 400. MMFF94s was used as force field. Active stereoisomers mentioned in the literature [9] were considered in further calculations.

**Table 1**. Pyrethroidal ester structure

| No | Structure | No | Structure |
|---|---|---|---|
| A1 |  | G2 |  |
| A2 |  | G3 |  |
| A3 |  | G4 |  |
| A11 |  | G5 |  |
| B1 |  | G6 |  |
| B3 |  | G7 |  |
| C1 |  | G8 |  |
| D1 |  | G9 |  |
| D2 |  | H2 |  |
| E1 |  | H3 |  |
| E2 |  | H5 |  |

| | | | |
|---|---|---|---|
| E3 |  | I1 |  |
| E4 |  | J1 |  |
| E5 |  | J2 |  |
| E10 |  | K2 |  |
| F5 |  | K3 |  |
| F12 |  | K4 |  |
| F13 |  | L3 |  |
| G1 |  | | |

Twenty-two types of molecular descriptors calculated by the Dragon software (Dragon Professional 5.5/2007, Talete S.R.L., Milano, Italy) were derived from molecular structures: constitutional, functional groups counts, topological descriptors, Burden eigenvalues, eigenvalue-based indices, Galvez descriptors (topological charge indicies), Randic descriptors, RDF descriptors, MWC (Molecular walk counts path counts – atomic and molecular descriptors) and 3D-MoRSE, atom-centred fragments, information indices, edge adjacency indices, topological charge indices, connectivity indices, 2D-autocorrelations, molecular properties, 2D binary fingerprints, and 2D frequency fingerprints descriptors.

Instant JChem (Instant JChem v. 6.0, Chemaxon Ltd., Budapest, Hungary) was used for structure database management, search and prediction. By this software additional structural parameters were calculated.

EPI Suite™ (US EPA. [2012]. Estimation Programs Interface Suite™ for Microsoft® Windows, v. 4.11. United States Environmental Protection Agency, Washington, DC, USA.) software was employed to calculate other descriptors.

The data are normalized based on the autoscaling method, which can be described as:

$$XT_{mj} = \frac{X_{mj} - \overline{X}_m}{S_m} \qquad (1)$$

where for each variable *m*, *XTmj* and *Xmj* are the values *j* for the variable *m* after and before scaling respectively, is the mean and *Sm* the standard deviation of the variable.

Multiple linear regression (MLR) analysis [11] has been applied after variable selection carried out by the genetic algorithm included in the QSARINS v. 1.2 program [12, 13], using the RQK fitness function, with leave-one-out cross-validation correlation coefficient as constrained function to be optimized. The dataset was divided in training and a randomly selected (25% of the total number of compounds) test set. Compounds: C1, D1, G1, H3 and L3 were included in the test set.

**Model validation**

All the statistical tests were performed at a significance level of 5 %. In MLR models, outliers were detected by a value of residual greater than 2.5 times the value of standard error in calculation.

The leave-one-out cross-validation procedure was employed for internal validation. The over fitting of data and model applicability was controlled by comparing the root-mean-square errors (RMSE) of training and validation sets. To test the predictive power of the model, the concordance correlation coefficient (CCC) [14] (which simply verifies how small the differences

are between experimental data and external data set predictions, independently of their range) was used.

Y-randomization (also known as Y-scrambling) testing is a technique for checking the robustness of a QSAR model and the statistical significance of the estimated predicted power. In this test, the dependent variable vector, Y-vector, is randomly shuffled and a new QSAR model is developed using the original independent variable matrix. The process was repeated 2000 times. It is expected that the resulting QSAR models will generally have low $R^2$ and low $Q^2$ (leave-one-out, LOO) values. If the new models developed from the data set with randomised responses have significantly lower $R^2$ and $Q^2$ than the original model, then this is strong evidence that the proposed model is well founded, and not just the result of chance correlation. Satisfactory leave-one-out cross-validation values are stable and predictive if validated by the leave-more-out procedure.

Models based on chance correlation can be detected using the QUIK rule [15], a simple criterion that allows the rejection of models with high predictor collinearity, which could lead to chance correlation. The QUIK rule is based on the K multivariate correlation index (Table 2) that measures the total correlation of a set of variables. The rule is derived from the assumption that the total correlation in the set given by the model predictors X plus the response Y ($K_{XY}$) should always be greater than that the one measured only in the set of predictors ($K_{XX}$). Therefore, according to the QUIK rule, only models with the $K_{XY}$ correlation among the [X + Y] variables greater than the $K_{XX}$ correlation among the [X] variables can be accepted.

A QSAR model can be considered robust when its performance remains satisfactory and stable when heavy perturbations (for instance by leave-many-out) in the training composition is made. For internal validation results several measures of robustness were employed: $Q^2$ cross-validation, Y-scrambling (performed for 2000 iterations) and for internal predictivity $Q^2_{LMO}$ leave-more-out (LMO) cross-validation (performed for 2000 iterations).

The Multi-Criteria Decision Making (MCDM) [16] is a technique that summarizes the performances of a certain number of criteria simultaneously, as a single number (score) between 0 and 1. This is done associating to every validation criteria a desirability function which values range from 0 to 1 (where 0 represents the worst validation criteria value and 1 the best). The geometric average of all the values obtained from the desirability functions gives the MCDM value. The ‚MCDM all' scores were calculated using all the criteria: fitting, cross validated and external and were used to choose the best MLR models.

## RESULTS AND DISCUSSION

Several MLR models were built after variable selection carried out by the genetic algorithm. Seven outliers (compounds A1, B1, B3, F12, F13, G2 and H5) were found and removed from the final MLR models included in Table 2. This fact can be related to possible experimental errors.

The applicability domain of the selected models was evaluated by leverage analysis expressed as Williams plot (Figures 1 and 2 for model 1), in which the jackknifed (Studentized) residuals and the leverage values were plotted. These plots confirm the absence of outliers and influential points in the final selected MLR models.

The MLR models included in Table 2 are completely satisfactory in the fitting, but have modest predictive power. They have been assessed by internal (LOO and LMO) cross-validation, Y-scrambling. From this table it is evident that the models, even with good fitting performances (high values of $R^2$ and $R^2_{adj}$), have low predictive power, verified by different validations.

The LOO validation highlights that the model is stable, not obtained by chance, in fact the difference between $R^2$ and $Q^2$ is small: 6.8 % in case of model 1, considered best. Model 1 is stable and internally predictive having less difference (of -5.2%) between $Q^2_{LMO}$ and $Q^2$ and of 0.3% between $R^2_{LMO}$ and $R^2$.

The descriptors are not very correlated (medium Kxx: 20.9 %) and the difference in the correlation between the block of X variables plus the response Y (Kxy) and that of X (Kxx) is sufficiently high: medium delta = 15.05.

The risk of chance correlation was verified also by the Y-scrambling procedure. The extremely low calculated $R^2$Yscrambling and $Q^2$Yscrambling values (Table 2) indicate no chance correlation for the chosen models.

The RMSE values for the training and validation sets are similar. The chosen models demonstrate a satisfactory stability in internal validation, have high fitting and internal predictivity, but a modest predictive power (see the $CCC_{ext}$ values in Table 2). The difference of CCC values between the training and test sets of 11.1% (model 1), 5.5% (model 2), 14% (model 3), demonstrates that these models are not able to predict the response for chemicals not used in the model development (validation set) just as they do for chemicals used to find the relationship (training set).

**Table 2**. MLR statistical results for the training , cross-validated and test sets*

| Model | Equation | $R^2$ | $Q^2$ | $R^2_{adj}$ | SEE | RMSEtr | RMSEex | $K_{XX}$ | ΔK | CCCtr | CCCex | MCDM all | $R^2_{LMO}$ | $Q^2_{LMO}$ | $R^2Y_{scr}$ | $Q^2Y_{scr}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | $\log LD_{50} = -0.35(\pm 0.22) - 0.99(\pm 0.47)EEig02d - 0.55(\pm 0.31)BEHm3 + 0.62(\pm 0.32)BELm8 - 1.92(\pm 0.76)KOAWIN\log Kaw$ | 0.857 | 0.789 | 0.828 | 0.252 | 0.225 | 0.231 | 0.209 | 0.151 | 0.923 | 0.812 | 0.794 | 0.860 | 0.860 | 0.17 | -0.32 |
| 2 | $\log LD_{50} = -0.54(\pm 0.24) - 0.97(\pm 0.44)EEig04d + 0.44(\pm 0.30)nCp - 2.06(\pm 0.82)KOAWIN\log Kaw$ | 0.759 | 0.668 | 0.724 | 0.319 | 0.292 | 0.222 | 0.377 | 0.101 | 0.863 | 0.808 | 0.742 | 0.761 | 0.761 | 0.13 | -0.25 |
| 3 | $\log LD_{50} = -0.41(\pm 0.25) - 0.79(\pm 0.37)EEig02d - 0.74(\pm 0.30)MW + 0.52(\pm 0.27)BELm8 - 1.83(\pm 0.72)KOAWIN\log Kaw$ | 0.834 | 0.756 | 0.800 | 0.271 | 0.243 | 0.258 | 0.261 | 0.148 | 0.909 | 0.769 | 0.754 | 0.840 | 0.840 | 0.17 | -0.33 |

* $R^2$ – correlation coefficient, $Q^2$ – leave-one-out 'crossvalidated $r^2$', $R^2_{adj}$- adjusted $R^2$, SEE – standard error of estimates, RMSE - root mean squared error, MAE - mean absolute error, CCC - concordance correlation coefficient, for the training (tr), and test (ex) sets; MCDM all - Multi-Criteria Decision Making calculated for fitting cross-validation and external validation; $R^2_{LMO}$ and $Q^2_{LMO}$ – leave many-out correlation coefficient and cross-validation coefficients; $R^2Y_{scr}$ and $Q^2Y_{scr}$-Y scramble correlation and cross-validation coefficients; EEig02d-Eigenvalue 02 from edge adj. matrix weighted by dipole moments; BEHm3-highest eigenvalue n. 3 of Burden matrix / weighted by atomic masses; BELm8-lowest eigenvalue n. 8 of Burden matrix / weighted by atomic masses; KOAWIN Log Kaw–air-water partition coefficients; nCp-number of terminal primary C(sp3); MW-molecular weight

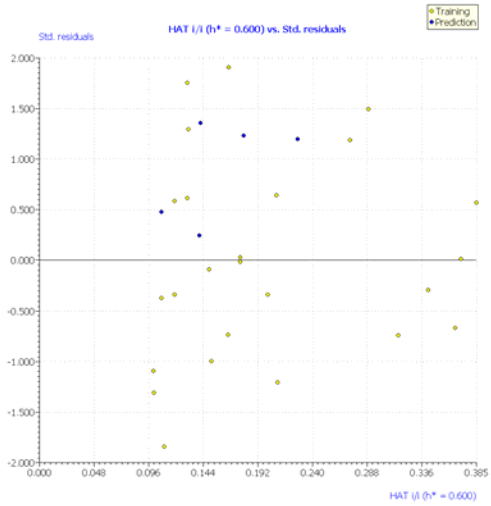



**Figure 1.** Williams plot – predicted by fitting for model 1

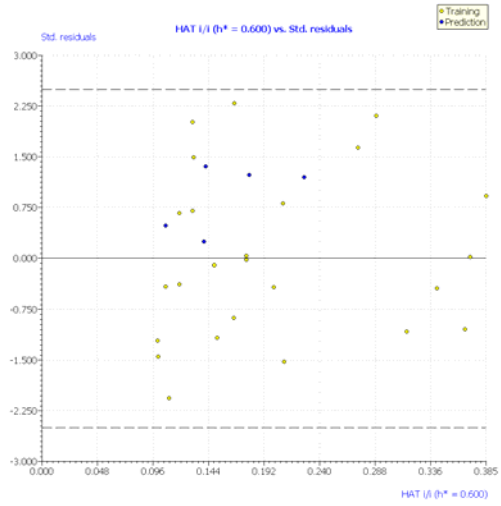



**Figure 2.** Williams plot – predicted by leave-one-out (LOO) for model 1

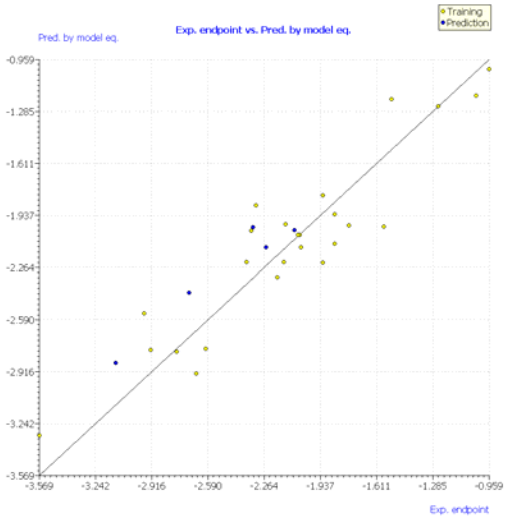



**Figure 3.** Experimental versus $\log LD_{50}$ values predicted by fitting for model 1

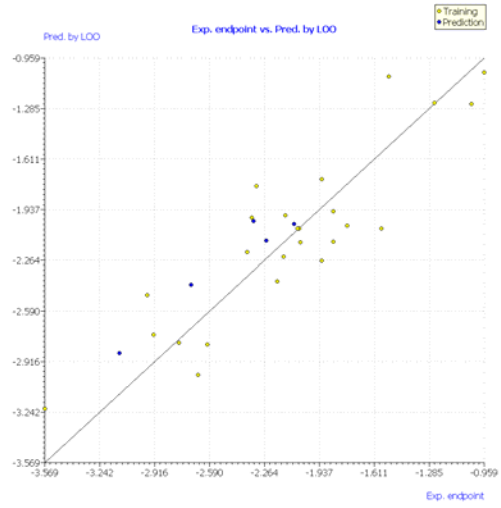



**Figure 4.** Experimental versus $\log LD_{50}$ values predicted by LOO for model

Edge adjacency indices descriptors are indices that show good structural selectivity [17]. The presence of the EEig02d (Eigenvalue 02 from edge adj. matrix weighted by dipole moments) descriptor indicates non-toxic effect of the title compounds, but EEig04d seems to induce high toxicity.

The Burden eigenvalue descriptors are important in capturing structural information important for understanding polar intermolecular interactions [18]. The presence of the BELm8 (lowest eigenvalue n. 8 of Burden matrix / weighted by atomic masses) descriptor suggests non-toxic tendency of pyrethroidal esters, opposite to the BEHm3 (highest eigenvalue n. 3 of Burden matrix / weighted by atomic masses) descriptor.

The presence of the number of terminal primary C(sp3) group is favorable for low toxicity. A high molecular weight leads to high toxicity. The KOAWIN Log Kaw – air-water partition coefficients is related to the waste water treatment because solubility affects volatilization of toxic compounds into the air; higher descriptor values are related to high toxicity.

## CONCLUSIONS

In order to study the toxicity of 37 pyrethroidal esters, the logarithm of $LD_{50}$ values measured against a susceptible strain of housefly (*Musca domestica*) were related by multiple linear regression to their structural descriptors. Stereoisomers selected according to the literature [9] were modeled by conformational analysis performed by molecular mechanics calculations. Several criteria for internal and external validation were applied. The obtained MLR models are satisfactory in the fitting, but have modest predictive power. They indicate structural features of the title compounds which contribute to the death of housefly (*Musca domestica*). The presence number of terminal primary C(sp3) group is favorable for low toxicity. High values of air-water partition coefficients and of molecular weight can be associated with high toxicity of the title compounds.

## ACKNOWLEDGEMENTS

The authors are indebted to Chemaxon Ltd., OpenEye Scientific Software Inc. and Prof. for Prof. Paola Gramatica from The University of Insubria (Varese, Italy) for giving access to their programs. This project was financially supported by Project 1.1 of the Institute of Chemistry Timisoara of the Romanian Academy.

## REFERENCES

1. Q. Gao, M. Li, C. Sheng, J. G. Scott, X. Qiu, Multiple cytochrome P450s overexpressed in pyrethroid resistant house flies (*Musca domestica*), *Pestic. Biochem. Phys.*, 2012, *104*, 252–260.

2. Y. Du, Y. Nomura, N. Luo, Z. Liu, J. –E. Lee, B. Khambay, K. Dong, Molecular determinants on the insect sodium channel for the specific action of type II pyrethroid insecticides, *Toxicol. Appl. Pharm.*, 2009, *234*, 266–272.

3. Y. Katsuda, Development of and future prospects for pyrethroid chemistry, *Pestic. Sci*., 1999, *55*, 775-782.

4. M. F. Hughes, B. C. Edwards, In vitro dermal absorption of pyrethroid pesticides in human and rat skin, *Toxicol. Appl. Pharm.*, 2010, *246*, 29–37.

5. F. D. Rinkevich, C. A. Leichter, T. A. Lazo, M. C. Hardstone, J. G. Scott, Variable fitness costs for pyrethroid resistance alleles in the house fly, *Musca domestica*, in the absence of insecticide pressure, *Pestic. Biochem. Phys.*, 2013, *105*, 161–168.

6. A. Anadón, M.R. Martínez-Larranãaga, M.A. Martínez, Use and abuse of pyrethrins and synthetic pyrethroids in veterinary medicine, *Vet. J.*, 2009, *182*, 7–20.

7. D. M. Soderlund, J. M Clark, L. P. Sheets, L. S. Mullin, V. J. Piccirillo, D. Sargent, J. T. Stevens, M. L. Weiner, Mechanisms of pyrethroid neurotoxicity: implications for cumulative risk assessment, *Toxicology*, 2002, *171(1)*, 3–59.

8. Toxicological Profile for Pyrethrins and Pyrethroids, Agency for Toxic Substances and Disease Registry (ATSDR), based in Atlanta, Georgia, is a federal public health agency of the U.S, September 2003, http://www.atsdr.cdc.gov/toxprofiles/tp.asp?id=787&tid=153.

9. A. W. Farnham, B. P. S. Khambay, The Pyrethrins and Related Compounds. Part XXXIX*-Structure-Activity Relationships of Pyrethroidal Esters with Cyclic Side Chains in the Alcohol Component against Resistant Strains of Housefly *(Musca domestica)*, *Pestic. Sci*. 1995, *44*, 269-275.

10. P.C.D. Hawkins, A. Nicholls, Conformer Generation with OMEGA: Learning from the Data Set and the Analysis of Failures, *J. Chem. Inf. Model*. 2012, *52*, 2919−2936.

11. S. Wold, W.J. Dunn III, Multivariate quantitative structure activity relationships (QSAR): Conditions for their applicability, *J. Chem..Inf. Comput. Sci.* 1983, *23*, 6-13.

12. N. Chirico, E. Papa, S. Kovarich, S. Cassani, P. Gramatica, *QSARINS, software for QSAR MLR model development and validation. 2012, QSAR Res. Unit in Environ. Chem. and Ecotox., DiSTA, University of Insubria, Varese, Italy. http://www.qsar.it.*

13. P. Gramatica, N. Chirico, E. Papa, S. Cassani, S. Kovarich, QSARINS: A new software for the development, analysis, and validation of QSAR MLR models, *J. Comput. Chem.*, 2013, *34(24)*, 2121 – 2132.

14. N. Chirico, P. Gramatica, Real External Predictivity of QSAR Models: How To Evaluate It? Comparison of Different Validation Criteria and Proposal of Using the Concordance Correlation Coefficient, *J. Chem. Inf. Model.* 2011, *51*, 2320-2335.

15. R. Todeschini, V. Consonni, A. Maiocchi, The K correlation index: theory development and its application in chemometrics. *Chemom. Int. Lab. Syst.* 1999, *46*, 13–29.

16. H.R. Keller, D.L. Massart, J.P. Brans, Multicriteria decision making: a case study. *Chemom. Int.Lab. Syst*. 1991, *11*, 175-189.

17. E. Estrada, A. Ramírez, Edge Adjacency Relationships and Molecular Topographic Descriptors. Definition and QSAR Applications, *J. Chem. Inf. Comput. Sci*., 1996, *36* (4), 837–843.

18. D. T. Stanton, Evaluation and Use of BCUT Descriptors in QSAR and QSPR Studies**,** *J. Chem. Inf. Comput. Sci*. 1999**,** *39*, 11-20.