LINEAR REGRESSION MODELS OF MOULTING ACCELERATING COMPOUNDS WITH INSECTICIDE ACTIVITY AGAINST SILKWORM *BOMBYX MORI* L.[†]

Simona Funar-Timofei[*], Alina Bora, Luminita Crisan, Ana Borota

Institute of Chemistry Timisoara of the Romanian Academy, Bul. Mihai Viteazu 24, 300223 Timisoara, Romania

ABSTRACT

Dibenzoylhydrazine derivatives are used as insect growth regulators that act through the induction of a lethal larval molting process in insects that belong to the species of Lepidoptera and Coleoptera. This paper presents linear regression models of ecdysone agonistic activity of dibenzoylhydrazine insecticides measured in the silkworm *Bombyx Mori* lepidopteran species cell lines. These structures were modeled through the PM7 semiempirical quantum chemical method using the MOPAC 2016 software. Several structural descriptors were derived from the energy optimized structures and were related to the insecticidal activity, expressed as pEC_{50} values, using the multiple linear regression (MLR) and the partial least squares (PLS) methods. The dataset was divided into training and test (30% of the total number of compounds, chosen randomly) sets to test the model predictive power by several parameters. According to the squared correlation coefficients values, of 0.827 and 0.78 for the MLR and PLS models, respectively, and other statistical tests, the MLR model had better fitting results and good predictive ability compared to the PLS one. Structural features which influence the ecdysone agonistic activity of dibenzoylhydrazine insecticides encode chemical information on molecular flexibility, are related to sigma and pi bonding patterns in molecules and to geometrical descriptors invariant to translation and rotation, which contain electronic and topological information.

Keywords: ecdysone agonistic activity, dibenzoylhydrazines, insecticide, MLR, PLS, PM7

[†] Dedicated to the 150th anniversary of the Romanian Academy

INTRODUCTION

Dibenzoylhydrazine compounds are insect growth regulators that act through the induction of an early and lethal larval molting process in vulnerable insects that belong to the species of Lepidoptera and Coleoptera [1]. These compounds activate the steroid receptor complex of ecdysone type at lower concentrations than the natural hormone. The insect cannot remove them efficiently from its body and as consequence a constant state of ecdysteroid signaling is displayed in the insect, which avoids it to complete the molting process. Because the insect stays permanently trapped in the molting process and is unable to feed, it dies in the period of a few days from desiccation and starvation.

The activity of ecdysteroids is mediated by a heterodimer protein complex composed of ecdysone receptor and ultraspiracle, which activates the translation of the associated genes after the trigger caused by the binding of the corresponding ligand molecule [2].

The molecular mechanism of action of ecdysteroids is still unknown because one of the three interaction sites of the hormone-receptor model is not present in some active compounds [3].

The objective of our study is to estimate the ecdysone agonistic activity of dibenzoylhydrazine insecticides [4] measured in the silkworm *Bombyx Mori* lepidopteran species cell lines by linear regression techniques (multiple linear regression (MLR) and the partial least quares (PLS) methods).

Table 1. The smiles notation of dibenzoylhydrazine structures and their experimental insecticidal (pEC_{50}) and predicted ($pEC_{50\ pred}$) activity values obtained using the MLR/PLS methods

No	Smiles	pEC_{50}	pEC_{50pred_MLR}	pEC_{50pred_PLS}
1	CC(C)(C)N(NC(=O)C1=CC=CC=C1)C(=O)C1=CC=CC=C1	6.36	6.95	6.87
2	CCC1=CC=C(C=C1)C(=O)NN(C(=O)C1=CC(C)=CC(C)=C1)C(C)(C)C	8.95	7.55	7.88
3*	COC1=CC=CC(C(=O)NN(C(=O)C2=CC(C)=CC(C)=C2)C(C)(C)C)=C1C	8.47	7.73	7.72
4*	CC(C)(C)N(NC(=O)C1=CC=C(Cl)C=C1)C(=O)C1=CC=CC=C1	7.11	7.47	7.29
5	CC1=CC(=CC(C)=C1)C(=O)N(NC(=O)C1=C(C)C2=C(OCCC2)C=C1)C(C)(C)C	9	8.99	8.69
6	CC(C)(C)N(NC(=O)C1=CC=CC=C1)C(=O)C1=C(C=CC=C1)C(F)(F)F	6.06	6.31	6.28
7	CC(C)(C)N(NC(=O)C1=CC=CC=C1)C(=O)C1=C(Cl)C(Cl)=C(Cl)C=C1	5.44	6.18	6.17
8	CC(C)(C)N(NC(=O)C1=CC=CC=C1)C(=O)C1=C(Cl)C(Cl)=C(Cl)C=C1	6.7	6.63	7.42
9*	CC(C)(C)N(NC(=O)C1=C(F)C=CC=C1)C(=O)C1=C(Cl)C=CC=C1	6.84	6.15	5.40
10*	CC1=C(C=CC=C1)C(=O)NN(C(=O)C1=C(Cl)C=CC=C1)C(C)(C)C	6.89	6.71	5.94
11*	CC(C)(C)N(NC(=O)C1=CC(F)=CC=C1)C(=O)C1=C(Cl)C=CC=C1	6.49	6.10	5.68
12	CC(C)(C)N(NC(=O)C1=CC(Cl)=CC=C1)C(=O)C1=C(Cl)C=CC=C1	6.51	6.31	6.66

13	CC(C)(C)N(NC(=O)C1=CC=C(Br)C=C1)C(=O)C1=C(Cl)C=CC=C1	6.82	6.30	6.38
14	CCCC1=CC=C(C=C1)C(=O)NN(C(=O)C1=C(Cl)C=CC=C1)C(C)(C)C	7.51	7.07	6.69
15	CC(C)C1=CC=C(C(NN(C(C)(C)C)C(C2=C(Cl)C=CC=C2)=O)=O)C=C1	8.18	7.29	6.93
16*	COC1=CC=C(C=C1)C(=O)NN(C(=O)C1=C(Cl)C=CC=C1)C(C)(C)C	6.87	6.95	6.33
17	CC(C)(C)N(NC(=O)C1=CC=CC=C1)C(=O)C1=CC=CC=C1	8.15	7.79	8.75
18*	CC1=CC(=CC(C)=C1)C(=O)N(NC(=O)C1=CC=C(C=C1)C(C)(C)C)C(C)(C)C	7.96	7.65	7.22
19	CC1=CC=C(CC(C)=C1)C(=O)N(NC(=O)C1=C(C)C(C)=CC=C1)C(C)(C)C	7.76	8.29	7.48
20	CCC1=CC=C(C=C1)C(=O)NN(C(=O)C1=CC=CC=C1)C(C)(C)C	4.72	4.90	4.95
21	CCOC1=C(C=CC=C1)C(=O)N(NC(=O)C1=CC=C(OC)C=C1)C(C)(C)C	5.11	5.02	4.75
22	CCOC1=C(C=CC=C1)C(=O)N(NC(=O)C1=CC=C(Cl)C=C1)C(C)(C)C	5.54	5.07	5.31
23	CCOC1=C(C=CC=C1)C(=O)N(NC(=O)C1=CC=C(CC)C=C1)C(C)(C)C	5.93	6.19	5.87
24	CC1=CC(C(=O)N(NC(=O)C2=CC=CC=C2)C(C)(C)C)=C(Cl)C(C)=C1	6.94	6.55	6.36
25	CCC1=CC=C(C=C1)C(=O)NN(C(=O)C1=C(Cl)C(C)=CC(C)=C1)C(C)(C)C	5.94	6.17	6.69
26	CCCCC1=CC=C(C=C1)C(=O)NN(C(=O)C1=C(Cl)C(C)=CC(C)=C1)C(C)(C)C	5.91	6.30	5.91
27	COC1=CC=C(C=C1)C(=O)NN(C(=O)C1=C(Cl)C(C)=CC(C)=C1)C(C)(C)C	6.52	6.43	6.23
28*	CC1=CC=C(C=C1)C(=O)NN(C(=O)C1=C(Cl)C(C)=CC(C)=C1)C(C)(C)C	5.75	6.15	7.02
29*	CCCCCC1=CC=C(C=C1)C(=O)NN(C(=O)C1=C(Cl)C(C)=CC(C)=C1)C(C)(C)C	6.35	6.31	6.22
30	CC1=CC(C(=O)N(NC(=O)C2=CC=C(Cl)C=C2)C(C)(C)C)=C(Cl)C(C)=C1	7.48	8.21	8.09
31	CC1=CC(C(=O)N(NC(=O)C2=CC=C3OCCCC3=C2C)C(C)(C)C)=C(Cl)C(C)=C1	8.54	8.86	8.41
32	CC1=C2CCCOC2=CC=C1C(=O)NN(C(=O)C1=C(Cl)C=CC=C1)C(C)(C)C	6.37	6.69	6.93
33	CCCCCC1=CC=C(C=C1)C(=O)NN(C(=O)C1=C(Cl)C=CC=C1)C(C)(C)C	4.79	5.16	5.51

* test compounds

METHODS

Definition of target property and molecular structures

A set of 33 dibenzoylhydrazine ecdysone agonists with known biological activity was analyzed in this study. The ecdysone agonistic activity data [4], expressed as pEC_{50} values (where EC_{50} represents the concentration at which 50% of the maximum response is achieved) was used as dependent variable.

In the first step, the structures of the investigated molecules were pre-optimized using the (MMFF94) molecular mechanics force field included in the MarvinSketch (MarvinSketch 15.2.16.0, ChemAxon Ltd. http://chemaxon.com) package. In the next step, the minimized structures were refined using the semiempirical PM7 Hamiltonian [5] implemented in MOPAC 2016 program (MOPAC2016, James J. P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, HTTP://OpenMOPAC.net (2016).). For the geometry optimization a gradient norm limit of 0.01kcal/Å was set. For the conformations of minimum energy several quantum chemical descriptors were calculated: electronegativity,

hardness, chemical potential, electrophilicity, HOMO and LUMO molecular orbital energies, heat of formation, dipole moment, molecular surface area, softness.

Structural 0D, 1D, 2D and 3D molecular descriptors were calculated for the lowest energy structures using the DRAGON (Dragon Professional 5.5, 2007, Talete S.R.L., Milano, Italy) software.

The Multiple Linear Regression (MLR) method

Because the number of 1412 of computed descriptors is too high compared to the number of compounds (N = 33), a proper variable selection method was mandatory. The Genetic Algorithm (GA) is a trustworthy and extensively used variable selection method [6]. GA uses a stochastic algorithm that elucidates the optimization issues illustrated by fitness criteria, implying the evolution assumption of Darwin and various genetic functions, including crossover and mutation. The QSARINS v. 2.1 program [7] uses GAs to choose the meaningful descriptors that influence the variation of biologic activity of the compounds. The following parameters were employed: the RQK fitness function [8] with leave-one-out cross-validation [9] correlation coefficient as constrained function to be optimized, a crossover/mutation trade-off parameter of T = 0.5 and a model population size of P = 50.

The Partial Least Squares (PLS) method

Projections to latent structures (PLS) represent a regression technique for modeling the relationship between projections of dependent factors and independent responses. In this approach a block (or a column) of response variables is linked to a block of explanatory variables [10]. The relationship between the dependent and independent variables is described as a latent variable approach [11]. In the PLS approach stable, correct and highly predictive models are obtained even for correlated descriptors [12]. In this work PLS calculations were performed using the SIMCA (SIMCA P+12.0.0.0, May 20 2008, Umetrics, Sweeden, http://www.umetrics.com/) package. The QSAR matrix (of dependent and independent variables) was analyzed in a first step by the principal component analysis (PCA) [10], and subsequently by the partial least squares (PLS) approach. The squared correlation regression coefficient r^2, and the squared cross-validated correlation coefficient, q^2, are the most eloquent statistical parameters that ensure a measure of the quality and validity of the final PLS model, while the Variables Importance in the Projection (VIP) values and the sign of the variables' coefficients are more relevant in explaining the activity mechanism. The significant principal components were selected by 7 cross-validation groups.

Model validity

The dibenzoylhydrazine derivatives were divided into training and test sets by random split, taking out 27% of the total number of compounds (no. 3, 4, 9, 10, 11, 16, 18, 28, 29), while the remaining 73% were used as training set. The model's predictability was tested using the Q^2_{F1} [13]; Q^2_{F2} [14]; Q^2_{F3} [15] and the concordance correlation coefficient (CCC) [16] (having the thresholds values higher than 0.85, as they have been rigorously determined by a simulation study [17])-external validation parameters.

The predictive power of the QSAR models was, also, evaluated based on the predictive parameter r^2_m (with a lowest threshold value of 0.5 to be accepted) [18].

The Y-randomization test is a usually used technique that exhibits the robustness of a QSAR model, being a measure of model overfit. The dependent variable (biological activity) is arbitrarily mixed and a QSAR model is built using the same X matrix of molecular descriptors. The obtained MLR and PLS models (after 999 randomizations) must have minimal r^2 and q^2 values [19].

The data over fitting and model applicability was controlled by comparing the root-mean-square errors (RMSE) and the mean absolute error (MAE) of the training and validation sets [20].

For internal validation results several measures of robustness were employed: Y-scrambling [21], adjusted correlation coefficient (r^2_{adj}) and q^2 (leave-one-out, q^2_{LOO}, and leave-more-out, q^2_{LMO}) cross-validation coefficient.

Results and discussion

MLR analysis

The data was normalized using the autoscaling method:

$$XT_{mj} = \frac{X_{mj} - \overline{X}_m}{S_m} \tag{1}$$

where for each variable m, XT_{mj} and X_{mj} are the j values for the m variable after and before scaling, respectively, $\overline{X}_m$ is the mean, and S_m is the standard deviation of the variable.

Several MLR models were built after variable selection, which was carried out by genetic algorithm. The fitting and predictivity criteria for these models are presented in Table 2-4.

Table 2 Fitting and cross-validation parameters of the MLR models (training set)*

Model	$r^2_{training}$	q^2_{LOO}	q^2_{LMO}	r^2_{adj}	$RMSE_{tr}$	MAE_{tr}	CCC_{tr}	r^2_{scr}	q^2_{scr}	SEE	F
MLR1	0.827	0.760	0.736	0.801	0.509	0.411	0.906	0.130	-0.266	0.558	31.924
MLR2	0.785	0.687	0.652	0.753	0.568	0.441	0.880	0.129	-0.267	0.622	24.320
MLR3	0.799	0.714	0.688	0.768	0.550	0.460	0.888	0.131	-0.259	0.602	26.433
MLR4	0.808	0.736	0.712	0.779	0.537	0.403	0.894	0.132	-0.258	0.588	28.001
MLR5	0.774	0.682	0.640	0.740	0.582	0.429	0.873	0.131	-0.266	0.638	22.862
PLS-M2	0.780	-	0.717	-	0.575	0.485	0.876	0.204	-0.289	-	-

* $r^2_{training}$ -correlation coefficient; q^2_{LOO} - leave-one-out correlation coefficient; q^2_{LMO} leave-more-out correlation coefficient; r^2_{adj} -adjusted correlation coefficient; $RMSE_{tr}$-root-mean-square errors; MAE_{tr}-mean absolute error; CCC_{tr}-the concordance correlation coefficient; r^2_{scr} and q^2_{scr} -Y-scrambling parameters; SEE-standard error of estimates; F-Fischer test.

Table 3 Predictivity criteria calculated for the MLR models (test set)*

| Model	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	$RMSE_{ext}$	MAE_{ext}	CCC_{ext}
MLR1	0.734	0.705	0.883	0.420	0.352	0.829
MLR2	0.733	0.705	0.882	0.420	0.343	0.834
MLR3	0.612	0.571	0.829	0.507	0.407	0.730
MLR4	0.540	0.491	0.797	0.552	0.465	0.744
MLR5	0.627	0.588	0.836	0.497	0.417	0.741
PLS-M2	-0.121	-0.240	0.732	0.862	0.755	0.455

* Q^2_{F1}; Q^2_{F2} ; Q^2_{F3} -external validation parameters; $RMSE_{ext}$-root-mean-square errors; MAE_{ext} -mean absolute error; CCC_{ext}-the concordance correlation coefficient

Table 4 Other predictivity parameters (r_m^2) and final descriptors selected in the MLR/PLS models*

Model	r_m^2	Descriptors included in the model*
MLR1	0.734	RBF, EEig11r, L3s
MLR2	0.677	RBF, BEHv8, L3s
MLR3	0.569	RBF, Mor02p, L3s
MLR4	0.518	RBF, BEHe5, L3s
MLR5	0.594	X1A, BEHv8, L3s
PLS-M2	0.136	BEHp2, BELe1, BELm1, BELp1, BELv1, EEig04r, EEig04x, F02[C-C], F03[C-C], F09[C-C], HATS4e, HATS4u, Mor02m, Mor02p, Mor02v, Mor11e, Mor11m, Mor11p, Mor11u, Mor11v, Mor24m, Mor24p, Mor24v, RDF025m, RDF025v, SPH, VEA2

* RBF – rotatable bond fraction; EEig11r – Eigenvalue 11 from edge adj. matrix weighted by resonance integrals; L3s - 3rd component size directional WHIM index / weighted by atomic electrotopological states; BEHv8 - highest eigenvalue n. 8 of Burden matrix / weighted by atomic van der Waals volumes; Mor02p - 3D-MoRSE - signal 02 / weighted by atomic polarizabilities; BEHe5 - highest eigenvalue n. 5 of Burden matrix / weighted by atomic Sanderson electronegativities; X1A - average connectivity index chi-1; BEHp2- highest eigenvalue n. 2 of Burden matrix / weighted by atomic polarizabilities; BELe1- lowest eigenvalue n. 1 of Burden matrix / weighted by atomic Sanderson electronegativities; BELm1-; BELp1- lowest eigenvalue n. 1 of Burden matrix / weighted by atomic masses; BELv1- lowest eigenvalue n. 1 of Burden matrix / weighted by atomic van der Waals volumes; EEig04r- Eigenvalue 04 from edge adj. matrix weighted by resonance integrals; EEig04x- Eigenvalue 04from edge adj. matrix weighted by edge degrees; F02[C-C]- frequency of C-C at topological distance 2; F03[C-C]- frequency of C-C at topological distance 3; F09[C-C]- frequency of C-C at topological distance 9; HATS4e- leverage-weighted autocorrelation of lag 4 / weighted by atomic Sanderson electronegativities; HATS4u- leverage-weighted autocorrelation of lag 4 / unweighted; Mor02m- 3D-MoRSE - signal 02 / weighted by atomic masses; Mor02v- 3D-MoRSE - signal 02 / weighted by atomic van der Waals volumes; Mor11e- 3D-MoRSE - signal 11 / weighted by atomic Sanderson electronegativities; Mor11m- 3D-MoRSE - signal 11 / weighted by atomic masses; Mor11p- 3D-MoRSE - signal 11 / weighted by atomic polarizabilities; Mor11u- 3D-MoRSE - signal 11 / unweighted; Mor11v- 3D-MoRSE - signal 11 / weighted by atomic van der Waals volumes; Mor24m- 3D-MoRSE - signal 24 / weighted by atomic masses; Mor24p- 3D-MoRSE - signal 24 / weighted by atomic polarizabilities; Mor24v- 3D-MoRSE - signal 24 / weighted by atomic van der Waals volumes; RDF025m- Radial Distribution Function - 2.5 / weighted by atomic masses; RDF025v- Radial Distribution Function - 2.5 / weighted by atomic van der Waals volumes; SPH- spherosity; VEA2- average eigenvector coefficient sum from adjacency matrix.

In order to verify the reliability of the developed equations, experimental versus predicted pEC_{50} values, Williams plots and Y-scramble plots for the MLR1 best model are presented in Figure 1, 2 and 3, respectively.
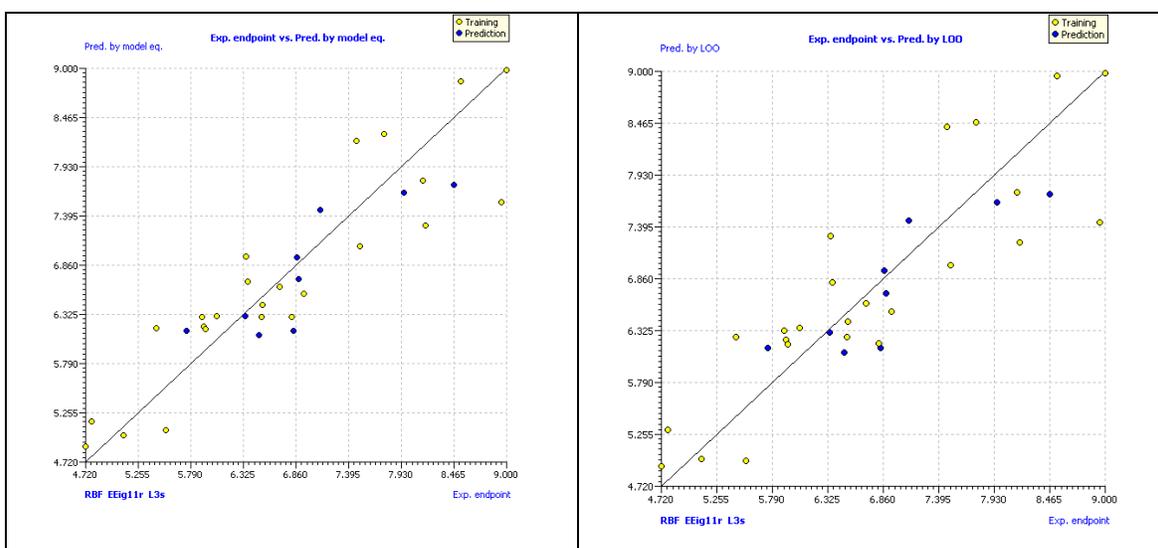
Figure. 1. Experimental versus predicted pEC_{50} values for the MLR1 model predicted by the model (left) and by the leave-one-out (right) crosvalidation approach (yellow circles-training compounds, blue circles-test compounds).

The Williams plot is used to identify compounds with the greatest structural influence ($h_i > h^*$; h_i =leverage of a given chemical; h^* = the warning leverage) in the QSAR model. The Williams plot for the training sets presented in Figure 2 (for the MLR1 model), establishes the applicability domain of the models within ±2.5σ and a leverage threshold h* of 0.500. The analysis of Figure 2 suggests that all the compounds in the dataset are within the applicability domain of the models.

The y-scrambling test indicates the robustness of a QSAR model, being a measure of the model overfit. The robustness of the developed models is confirmed by a significant low scrambled r^2 (r_{scr}^2) and cross-validated q^2 (q_{scr}^2) values obtained for 999 trials. Figure 3 suggest that in case of all the randomized models, the values of r_{scr}^2 and q_{scr}^2 were < 0.5 (r_{scr}^2/q_{scr}^2 of 0.130/-0.266). The low calculated r_{scr}^2 and q_{scr}^2 values indicate no chance correlation for the chosen models.
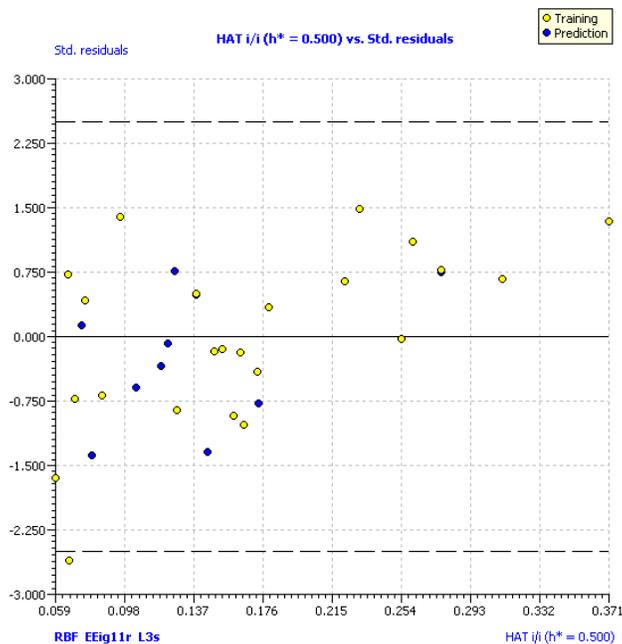
Figure. 2. Williams plot predicted by the final MLR1 model (yellow circles-training compounds, blue circles-test compounds).
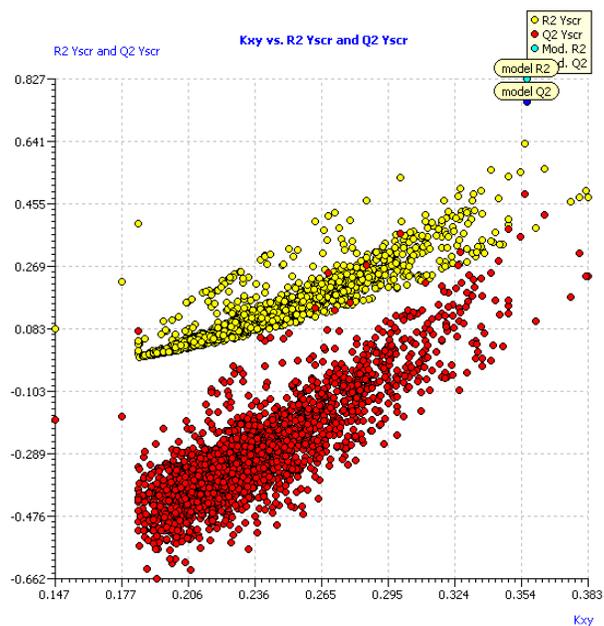
Figure. 3. Y-scramble plots for the MLR1 model.

In the present study, the best MLR1 model has three parameters. A higher or lower number of molecular descriptors does not have any significant effect on the model's accuracy. Additionally, the predictive r^2 (leave-one-out, q^2_{LOO}, and leave-more-out, q^2_{LMO}, cross-

validation parameters) were calculated and presented in Table 2. The most important descriptors, selected by genetic algorithm, which influence the ecdysone agonistic activity are: the constitutional RBF descriptor (which describes the rotatable bond fraction), the EEig11r (Eigenvalue 11 from edge adj. matrix weighted by resonance integrals) edge adjacency index, the L3s (3rd component size directional WHIM index / weighted by atomic electrotopological states) WHIM descriptors.

Increase of the EEig11r values yield increased the insecticide activity. Higher values of the RBF and L3s descriptors diminish the insecticide ability of the compounds.

An intercorrelation analysis of the selected molecular descriptors from the MLR1 model is presented in Table 5. The three selected descriptors are not intercorrelated.

Table 5. Correlation matrix of the selected descriptors included in the MLR1 model

	RBF	EEig11r	L3s
RBF	1.0000		
EEig11r	0.3144	1.0000	
L3s	0.2600	0.3715	1.0000

The statistical results and intercorrelation coefficients presented above confirm that the MLR method associated with a proper variable selection procedure generates an efficient QSAR model for predicting the ecdysone agonistic activity of dibenzoylhydrazine insecticides.

PLS analysis

A PCA model was built using the SIMCA-P+ version 12.0 software for the entire X matrix which include N=33 compounds and X = 1412 molecular descriptors. From the total of 7 significant principal components resulted from this analysis, we observed that the first three components already explained 65.5% of the information content of the descriptor matrix. PLS calculations were, as well, performed using the same training and test sets, as in case of MLR models. The statistical results of the PLS model: $R^2_{X(CUM)} = 0.47$, $R^2_{Y(CUM)} = 0.806$ and $Q^2_{(CUM)} = 0.619$ obtained for two principal components demonstrated the model overfit ($R^2_{X(CUM)}$ and $R^2_{Y(CUM)}$ are the cumulative sum of squares of all the X and Y values, respectively, explained by all extracted principal components; $Q^2_{(CUM)}$ is the fraction of the

total variation of the Y values that can be predicted for all the A extracted principal components in the cross-validation procedure (7 rounds) used to establish the number of significant principal components, A). The noise variables from this model was excluded and a robust PLS-M2 model (N= 24 and X= 27) with two latent variables (Tables 2, 3 and 4) was obtained. Although the PLS-M2 model contains only the descriptors significantly different from zero it has poorer statistical results and predictive power compared to the MLR1 model.

Conclusion

A series dibenzoylhydrazine insecticide with ecdysone agonistic activity measured in the silkworm *Bombyx Mori* lepidopteran species cell lines was investigated using linear regression methods. After structure optimization modeling using the semiempirical quantum chemical PM7 approach, calculated descriptors were related to the insecticide activity using the multiple linear regression and partial least squares approaches. The final model of dibenzoylhydrazine non-steroidal ecdysone agonists obtained using the MLR method have good statistical parameters. Molecular descriptors related to molecular flexibility, to sigma and pi bonding patterns in molecules and to geometrical descriptors invariant to translation and rotation, which contain electronic and topological information influenced the insecticidal activity. PLS modeling of the same data gave worser statistical results and a less predictive model compared to the MLR one.

References

1. L. Swevers, T. Soin, H. Mosallanejad, K. Iatrou, G. Smagghe, Insect Biochem. 38 (2008) 825

2. Toshiyuki Harada, Yoshiaki Nakagawa, Miki Akamatsu, Hisashi Miyagawa, Evaluation of hydrogen bonds of ecdysteroids in the ligand–receptor interactions using a protein modeling system, Bioorganic & Medicinal Chemistry 17 (2009) 5868–5873.

3. N. Ferro, J.E. Tacoronte, T. Reinard, P. Bultinck, L.A. Montero, J. Mol. Struct.-THEOCHEM 758 (2006) 263.

4. T. Soin, E. De Geyter, H. Mosallanejad, M. Iga, D. Martín, S. Ozaki, S. Kitsuda, T. Harada,

H. Miyagawa, D. Stefanou, G. Kotzia, R. Efrose, V. Labropoulou, D. Geelen, K. Iatrou, Y. Nakagawa, C.R. Janssen, G. Smagghe, L. Swevers, Pest. Manag. Sci. 66 (2010) 526.

5. J.J.P. Stewart, J. Mol. Modeling 19 (2013) 1-32.

6. U. Depczynski, V. J. Frost, K. Molt, Anal. Chim. Acta 420 (2000) 217.

7. P. Gramatica, N. Chirico, E. Papa, S. Cassani, S. Kovarich, J. Comput. Chem. 34 (2013) 2121–2132.

8. R. Todeschini, V. Consonni, A. Mauri, M. Pavan, Anal. Chim. Acta 515 (2004) 199.

9. D.M. Hawkins, S.C. Basak, D. Mills, J. Chem. Inf. Comput. Sci. 43(2) (2003) 579-86.

10. H. Wold, in: S. Kotz and N.L. Johnson (Eds.), Encyclopedia of Statistical Sciences, (Vol. 6), Wiley, New York, 1985, pp. 581.

11. L. Eriksson, J. Gottfries, E. Johansson, S. Wold, Chemom. Intel. Lab. Syst. 73 (2004) 73.

12. A. Höskuldsson, J. Chemometrics 2(3) (1998) 211.

13. L.M. Shi, H. Fang, W. Tong, J. Wu, R. Perkins, R.M. Blair, W.S. Branham, S.L. Dial, C.L. Moland, D.M. Sheehan. J. Chem. Inf. Model. 41 (2001) 186–195.

14. G. Schüürmann, R.U. Ebert, J. Chen, B. Wang, R. Kuhne, J. Chem. Inf. Model. 48 (2008) 2140–2145.

15. V. Consonni, D. Ballabio, R. Todeschini. J. Chem. Inf. Model. **49** (2009) 1669-1678.

16. N. Chirico, P. Gramatica. J. Chem. Inf. Model. **51**, (2011)2320–2335.

17. N. Chirico, P. Gramatica, J. Chem. Inf .Model. 52 (2012) 2044−2058.

18. K. Roy, I. Mitra. Mini-Rev .Med. Chem. 12 (2012) 491−504.

19. P.P. Roy, S. Paul, I. Mitra, K. Roy, Molecules 14 (2009) 1660.

20. P. Gramatica, In: Reisfeld B, Mayeno AN, editors. Computational Toxicology, Volume II, *Methods in Molecular Biology*, Vol. 930, "On the Development and Validation of QSAR Models", Springer, 2013, pp. 499-526.

21. R. Todeschini, V. Consonni, A. Maiocchi, Chemometr. Intell. Lab. 46 (1999) 13-29.