LINEAR REGRESSION MODELS OF MOULTING ACCELERATING COMPOUNDS WITH INSECTICIDE ACTIVITY AGAINST SILKWORM *BOMBYX MORI* L.[1]

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

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

**e-mail: timofei@acad-icht.tm.edu.ro*

[1]**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].**

[1]. L. Swevers, T. Soin, H. Mosallanejad, K. Iatrou, G. Smagghe, *Insect Biochem.* 38 (2008) 825
[2]. T. Harada, Y. Nakagawa, M. Akamatsu, H. Miyagawa, *Bioorgan. Med. Chem.* 17 (2009) 5868.

AIM:

- **The ecdysone agonistic activity of dibenzoylhydrazine insecticides (Table 1), expressed by pEC_{50} values (where EC_{50} represents the concentration at which 50% of the maximum response is achieved) was studied by multiple linear regression (MLR) partial least squares (PLS).**

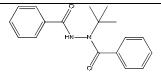
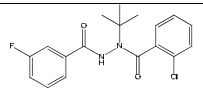
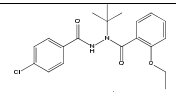
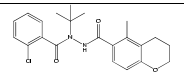
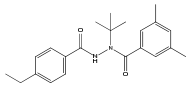
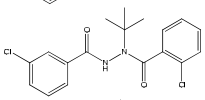
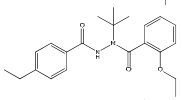
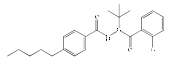
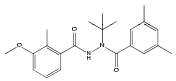
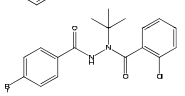
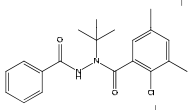
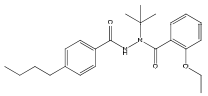
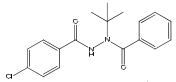
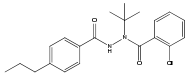
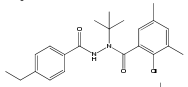
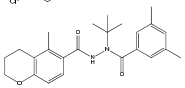
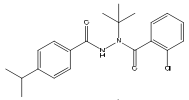
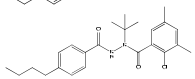
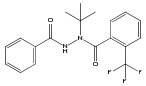
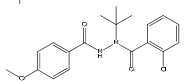
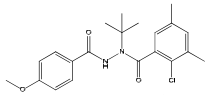
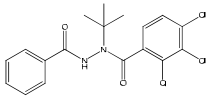
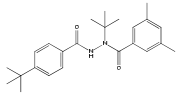
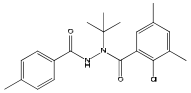
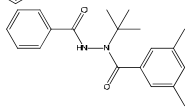
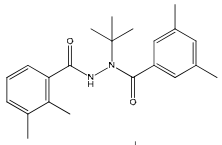
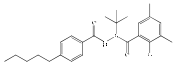
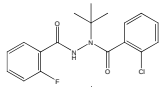
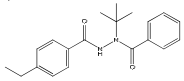
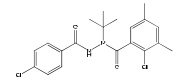
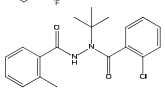
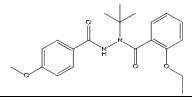
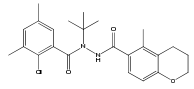
- **These insecticides were energy optimized using the MMFF94 force field (included in the Marvin Sketch MarvinSketch 15.2.16.0, ChemAxon Ltd.**

  http://chemaxon.com) and the PM7 semiempirical quantum chemical approach, using the MOPAC 2016 program (MOPAC2016, James J. P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA,

  HTTP://OpenMOPAC.net (2016)) Structural descriptors of these compounds were correlated to the pEC_{50} values.

METHODS

Table 1. The dibenzoylhydrazine structures

No	Structure	No	Structure	No	Structure	No	Structure
1		11*		21		31	
2		12		22		32	
3*		13		23		33	
4*		14		24			
5		15		25			
6		16*		26			
7		17		27			
8		18*		28*			
9*		19		29*			
10*		20		30			

*** Test compounds**

METHODS

- ***Definition of target property and molecular structures***

- **A series of 33 dibenzoylhydrazine structures was used, having the insecticide activity (pEC_{50} values) [3], as dependent variable.**

- **These structures 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 and further optimized using the PM7 semiempirical quantum chemical approach [4] included in the MOPAC2016 program.**

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

[3]. 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.

[4]. J.J.P. Stewart, *J. Mol. Modeling* 19 (2013) 1.

METHODS

- **Multiple linear regression (MLR) combined with genetic algorithm for variable selection was applied to the series of dibenzoylhydrazines, using the QSARINS [5] software.**

- **Partial Least Squares (PLS) [6] was employed to relate the structural descriptors to the ecdysone agonistic activity measured in the silkworm *Bombyx Mori* lepidopteran species cell lines. The PLS calculations were performed using the SIMCA (SIMCA P+12.0.0.0, May 20 2008, Umetrics, Sweeden, http://www.umetrics.com/) package.**

[5]. P. Gramatica, N. Chirico, E. Papa, S. Cassani, S. Kovarich, *J. Comput. Chem.* 34 (2013) 2121.

[6]. H. Wold, Partial Least Squares, in: S. Kotz and N. L. Johnson (Eds.), *Encyclopedia of Statistical Sciences* (Vol. 6), Wiley, New York, 1985, pp. 581-591.

METHODS

- ***Model validation***
- **The leave-seven-out cross-validation procedure was employed for internal validation, the data over fit and model applicability was controlled by comparing the root-mean-square errors (RMSE) and the mean absolute error (MAE) of training and validation sets and the predictive power of the model by the concordance correlation coefficient (CCC) [6].**
- **Y-scrambling was used to check the model robustness.**
- **T test the predictive power of the model, several external prediction parameters were employed (Q^2_{F1} [7]; Q^2_{F2} [8]; Q^2_{F3} [9] and r^2_m [10].**

[6]. N. Chirico, P. Gramatica, *J. Chem. Inf. Model.* 2011, *51*, 2320-2335.
[7]. 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.
[8]. G. Schüürmann, R.U. Ebert, J. Chen, B. Wang, R. Kuhne, *J. Chem. Inf. Model.* 48 (2008) 2140.
[9]. V. Consonni, D. Ballabio, R. Todeschini. *J. Chem. Inf. Model.* 49 (2009) 1669.
[10]. K. Roy, I. Mitra. *Mini-Rev .Med. Chem.* 12 (2012) 491.

RESULTS AND DISCUSSION

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.

RESULTS AND DISCUSSION

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

RESULTS AND DISCUSSION

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

RESULTS AND DISCUSSION

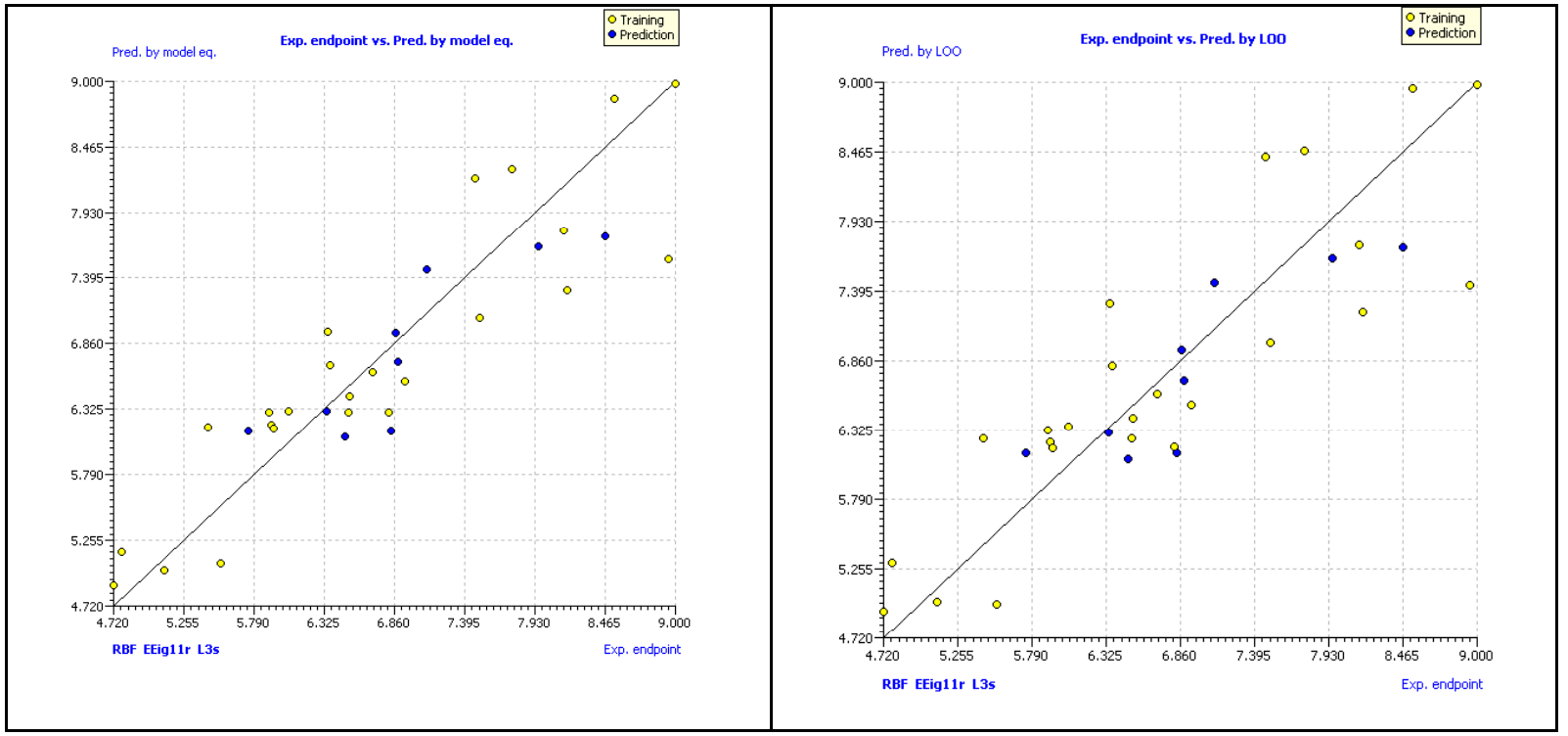


Figure. 1. Experimental versus predicted pEC_{50} values for the MLR1 model predicted by the model (left) and by the leave-one-out (right) crosvalidation approach (yellow circles-training compounds, blue circles-test compounds).

RESULTS AND DISCUSSION

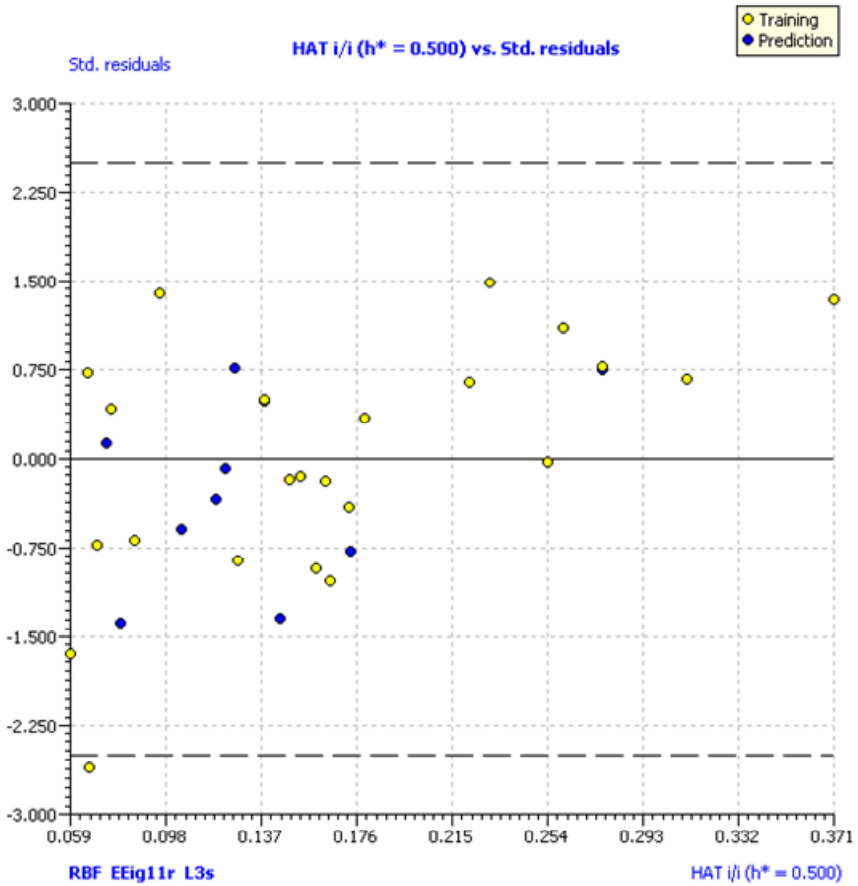


Figure. 2. Williams plot predicted by the final MLR1 model (yellow circles-training compounds, blue circles-test compounds).

RESULTS AND DISCUSSION

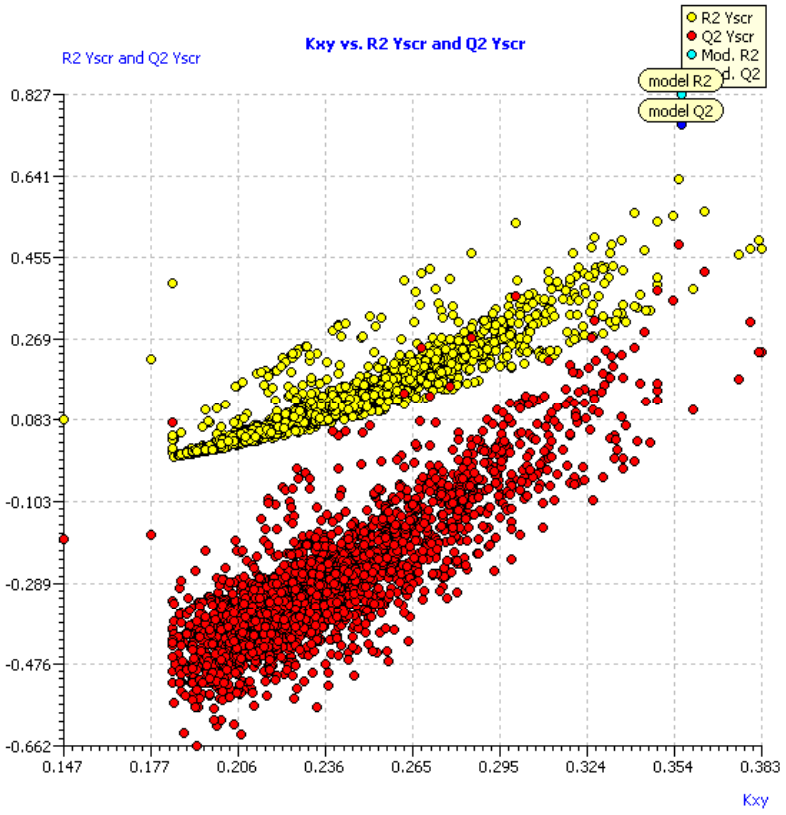


Figure. 3. Y-scramble plots for the MLR1 model.

RESULTS AND DISCUSSION

PLS results

- A two-components PLS model with satisfactory statistical quality was obtained: $R^2X(Cum) = 0.723$, $R^2Y(cum) = 0.780$, $Q^2(Cum) = 0.717$.

- Y-randomization test and leave-seven-out crossvalidation runs were performed to check the robustness and internal predictive ability of the PLS models. The Y-scrambling procedure, which was repeated 999 times. The extremely low calculated scrambled R^2 (0.204) and Q^2 (-0.289) values indicate no chance correlation for the chosen model.

- The PLS model has poorer statistical results and predictive power compared to the MLR1 best model .

CONCLUSIONS

- **MLR and PLS approaches were used to model the ecdysone agonistic activity of a series dibenzoylhydrazine insecticides.**
- **Better statistical results were obtained by the MLR1 model, which is satisfactory in the fitting and has predictive power, compared to the final PLS model.**
- **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.**

ACKNOWLEDGEMENTS

- **Access to the Chemaxon Ltd., QSARINS and MOPAC 2016 software are greatly acknowledged by the authors.**

- **This project was financially supported by Project 1.1 of the Institute of Chemistry Timisoara of the Romanian Academy.**