The 23rd International Electronic Conference on Synthetic Organic Chemistry

15 Nov – 15 Dec 2019

PLS Structure-Insecticidal Activity Relationships of Nitromethylene, Pyrrole- and Dihydropyrrole-Fused Neonicotinoids

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

Coriolan Dragulescu Institute of Chemistry, Romanian Academy,

Bul. Mihai Viteazu 24, 300223 Timisoara, Romania

timofei@acad-icht.tm.edu.ro

The 23rd International Electronic Conference on Synthetic Organic Chemistry

15 Nov – 15 Dec 2019

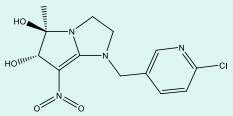
ABSTRACT

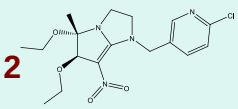
Neonicotinoids are considered to be one of the important classes of insecticides used at the present time. In this study the partial least squares (PLS) approach is applied to a series of nitromethylene, pyrrole- and dihydropyrrole-fused neonicotinoids to model their insecticidal activity (pLC_{50} values) against the cowpea aphids. The structures were modeled using the MMFF94s force field. A robust PLS model (3PCs, $R^2X(Cum) = 0.963$, $R^2Y(cum) = 0.870$ and $Q^2(Cum) = 0.796$) with predictive power (CCCext = 0.873, $r_m^2 = 0.680$, $Q_{F1}^2 = 0.805$, $Q_{F2}^2 = 0.802$, $Q_{F3}^2 = 0.704$) is obtained. New insecticides active against the cowpea aphids can be predicted.

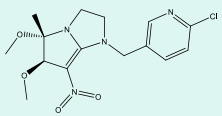
AIM

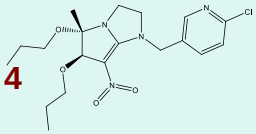
- **Quantitative Structure-Activity Relationship (QSAR) study of a series of nitromethylene, pyrrole- and dihydropyrrole-fused neonicotinoids, active against the cowpea aphids, using the partial least squares (PLS) method. Structural insecticide parameters calculated for the minimum energy conformers are related to the insecticidal activity, expressed as pLC_{50} values.**

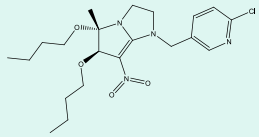
Neonicotinoid structures[1,2]

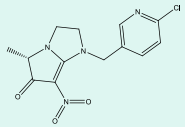
1 

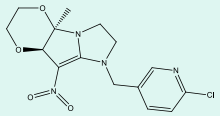
2 

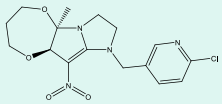
3* 

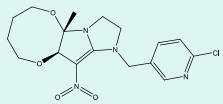
4 

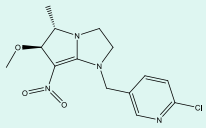
5 

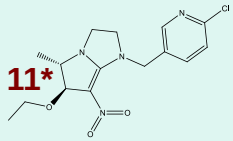
6 

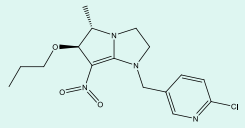
7 

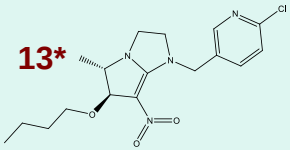
8 

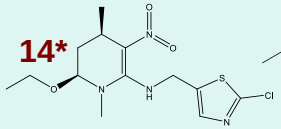
9 

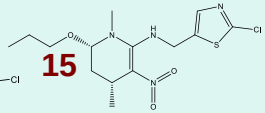
10 

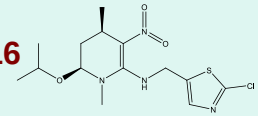
11* 

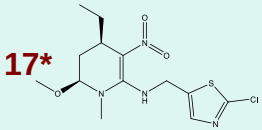
12 

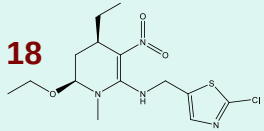
13* 

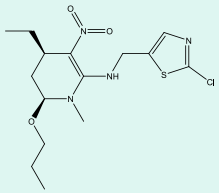
14* 

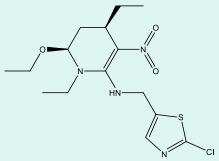
15 

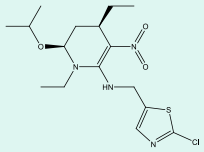
16 

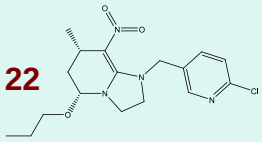
17* 

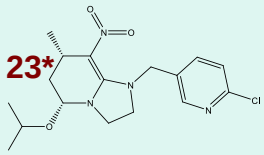
18 

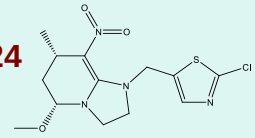
19 

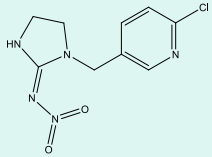
20 

21 

22 

23* 

24 

25 

*** test compounds**

1. Ye, Z.; Shi, L.; Shao, X.; Xu, X.; Xu, Z.; Li, Z. *J. Agric. Food Chem.* 2013, 61, 312−319.
2. Lu, S.; Shao, X.; Li, Z.; Xu, Z.; Zhao, S.; Wu, Y.; Xu, X. *J. Agric.Food Chem*. 2012, *60*, 322−330.

METHODS

Dataset

- The insecticide derivatives were taken from literature.[1,2]
- The lethal concentration values of the 25 neonicotinoid derivatives, expressed as pLC_{50} values, were used as the dependent variable.
- The insecticide structures were pre-optimized using the MMFF94 molecular mechanics force field of OMEGA v.2.5.1.4 (OpenEye) tool.[3,4] The default parameters were employed excepting the maximum number of conformers per compound set to 400 and an RMSD value of 0.5 Å.

Molecular descriptors

- The conformers of minimum energy were then used to calculate the structural parameters, using the DRAGON (Dragon Professional 5.5, 2007, Talete S.R.L., Milano, Italy) and InstanJChem (Instant JChem (2012) version 5.10.0, Chemaxon, http://www.chemaxon.com) software.

3. Hawkins, P.C.D.; Skillman, A.G.; Warren, G.L.; Ellingson, B.A.; Stahl, M.T. *J. Chem. Inf. Model.* 2010, 50, 572–584.
4. Hawkins, P.C,D.; Nicholls, A. *J. Chem. Inf. Model.* 2012, 52, 2919-2936.

METHODS

The Partial Least Squares (PLS) approach[5] was employed to relate the pLC_{50} values to the calculated structural descriptors, using the SIMCA (SIMCA P+12 12.0.0.0 2008, Umetrics, Sweeden, www.umetrics.com) program. The Variables Importance in the Projection (VIP) values and the sign of the variables' coefficients were used to explain the activity mechanism.

Internal validation criteria	*Parameter details*
$R^2X(CUM)$; $R^2Y(CUM)$; $Q^2(CUM)$	$r^2=1$ – strong relationship between the dependent and the independent variables
q^2_{L7O} (leave - 7 - out cross - validation coefficient)	q^2_{LOO} >0.7 – acceptable threshold
RMSE $_{tr}$ (root - mean - square errors for the training set)	q^2_{LMO} >0.7 – acceptable threshold
MAE $_{tr}$ (mean absolute error for the training set)	The reliability of the model fitness was tested by comparing the RMSE and the MAE of the training and the validation sets.
CCC $_{tr}$ (the concordance correlation coefficient for the training set)	CCC $_{tr}$ > 0.85 – acceptable threshold
Y - scrambling	Used to verify the robustness and chance correlation of a QSAR model and the statistical significance of the estimated predicted power

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

METHODS

External validation criteria	*Parameter details*
Q^2_{F1}	Q^2_{F1} >0.7 – acceptable threshold
Q^2_{F2}	Q^2_{F2} >0.7 – acceptable threshold
Q^2_{F3}	Q^2_{F3} >0.7 – acceptable threshold
$RMSE_{ext}$ (root-mean-square errors for the test set) **MAE_{ext} (mean absolute error for the test set)**	**The reliability of the model fitness was tested by comparing the RMSE and the MAE of the training and the validation sets.**
CCC_{ext} (the concordance correlation coefficient for the test set)	**CCC_{tr} > 0.85 – acceptable threshold**
r^2_m **metric**	r^2_m **> 0.5 – acceptable threshold**

RESULTS AND DISCUSSION

- **A robust model with three significant principal components which explains 96.3% of the information content of the descriptor matrix (for 10 structural descriptors), with $R^2Y(CUM) = 0.87$ and $Q^2(CUM) = 0.796$ was obtained.**
- **The PLS model was validated by the following internal validation parameters: $CCC_{tr} = 0.930$, $CCC_{CV} = 0.892$, $RMSE_{tr} = 0.216$, $RMSE_{CV} = 0.269$, $MAE_{tr} = 0.180$, $MAE_{CV} = 0.221$, (*tr* is the notation for training and *CV* for crossvalidation). A stable PLS model was obtained.**
- **The prediction model power was checked using external validation parameters, calculated for the test set: $CCC_{ext} = 0.873$, $RMSE_{ext} = 0.327$, $MAE_{ext} = 0.299$, $r_m^2 = 0.680$, $Q_{F1}^2 = 0.805$, $Q_{F2}^2 = 0.802$, $Q_{F3}^2 = 0.704$. They indicate the PLS model as having predictive power.**
- **Highest contribution to the model is given by the 2D binary and frequency fingerprints.**

RESULTS AND DISCUSSION

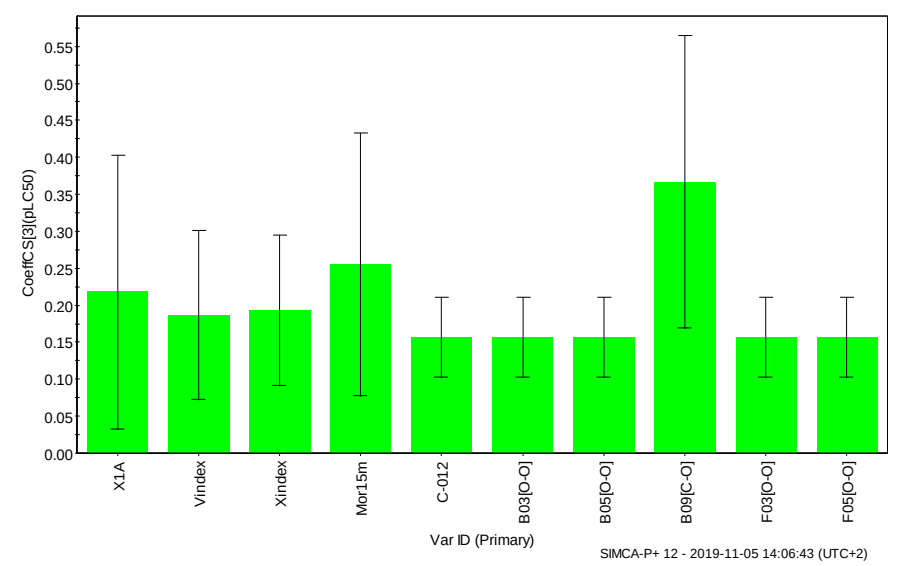


Figure 1. Coefficient plot of the final PLS model.

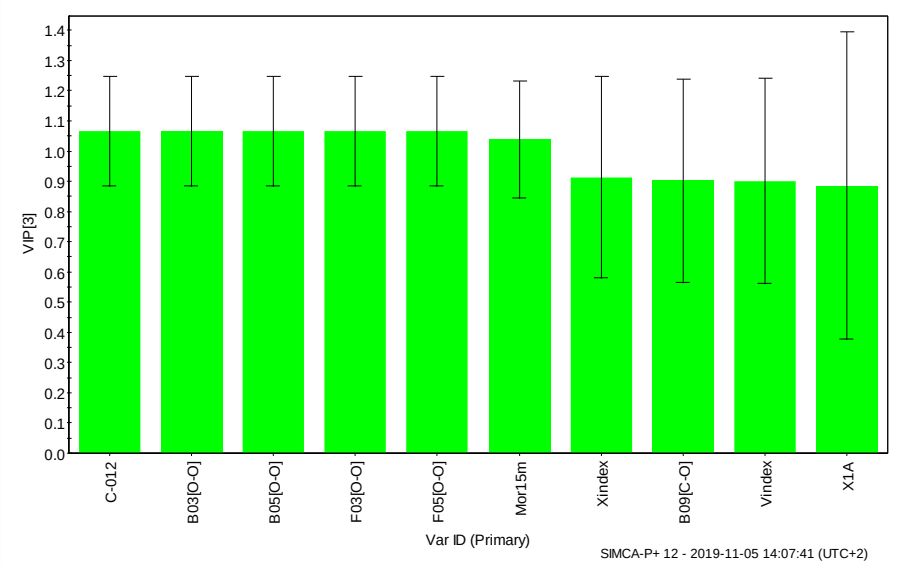


Figure 2. VIP plot for the final PLS model.

RESULTS AND DISCUSSION

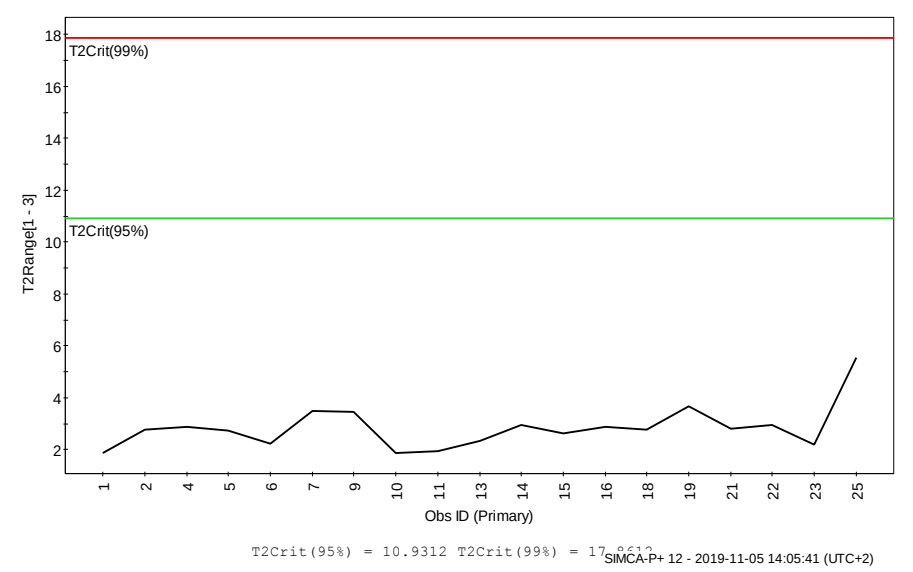


Figure 3. The Hotteling's T^2 range plot of the PLS model.

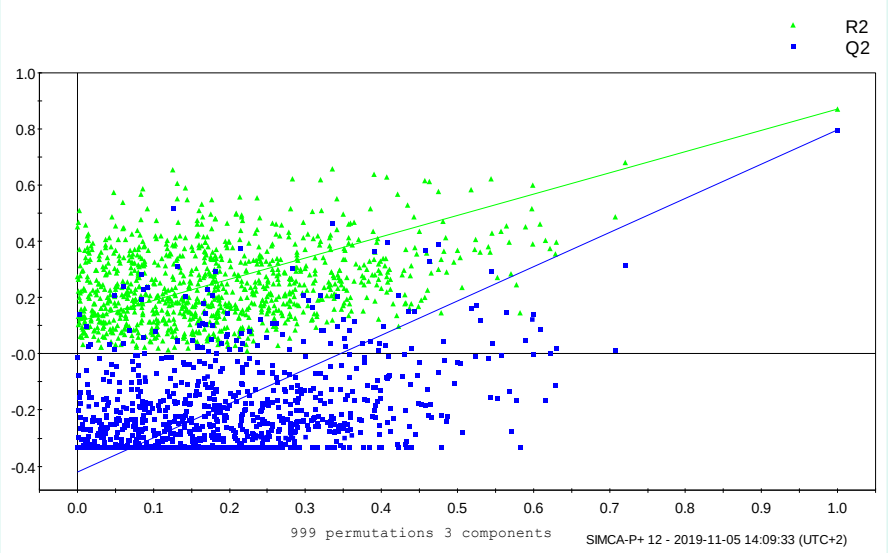


Figure 4. Y-scramble plots for the PLS model.

RESULTS AND DISCUSSION

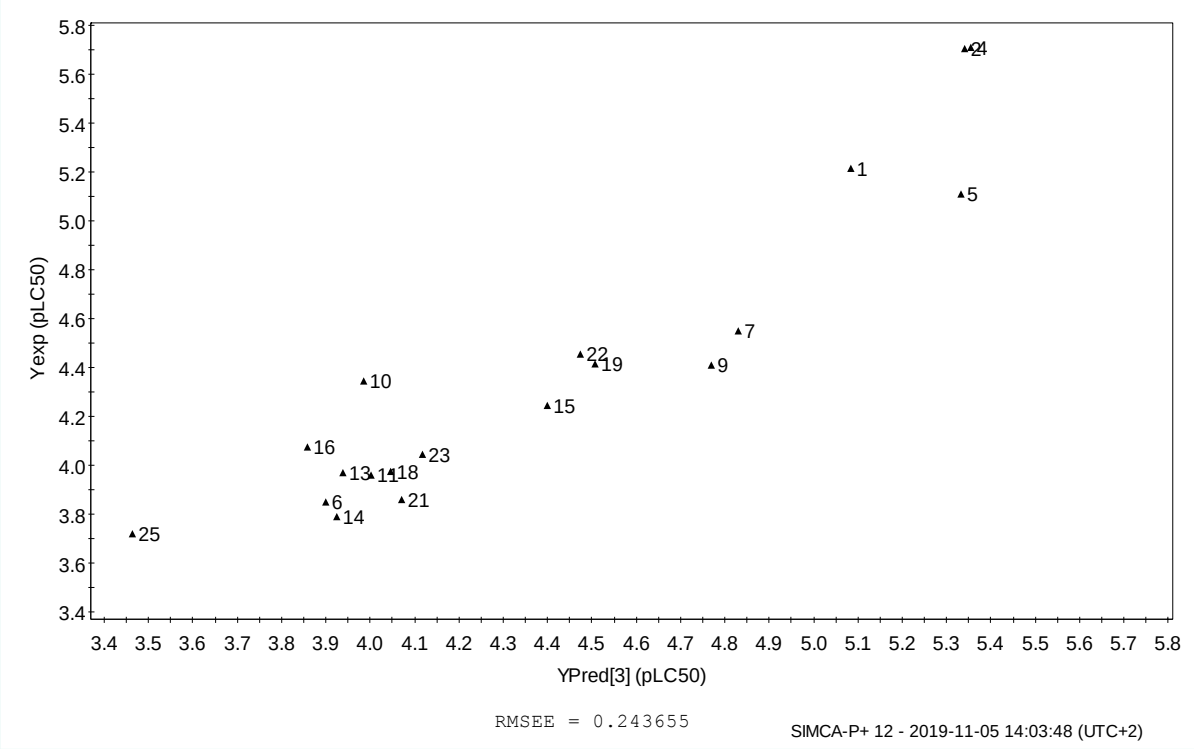


Figure 5. Experimental versus calculated pLC_{50} values.

CONCLUSIONS

- **The partial least squares (PLS) approach was used to study the insecticide activity against the cowpea aphids (*Aphis craccivora*) of a series of nitromethylene, pyrrole- and dihydropyrrole-fused neonicotinoids.**

- **The structures were pre-optimized using the MMFF94 molecular mechanics force field.**

- **Structural descriptors were derived from the minimum energy conformers and were related to the pLC_{50} values.**

- **2D binary and frequency fingerprints had highest contribution to the PLS model.**

- **The resulted PLS model, having good statistical results and predictive power can be used for the design of new insecticides active against the cowpea aphids.**

Acknowledgments

This project was financially supported by Project 1.1 of the Coriolan Dragulescu Institute of Chemistry of the Romanian Academy. Access to the OpenEye Ltd., and Chemaxon Ltd. software are greatly acknowledged by the authors.