proceedings

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

* Correspondence : timofei@acad-icht.tm.edu.ro (S.F.T.); Tel.: +40-256-491-818; Fax: +40-256-491-824.

Abstract: A single 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 (pLC50 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 (CCC_{ext} = 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.

Keywords: neonicotinoid; PLS; Omega; Aphis craccivora; validation

1. Introduction

Neuroactive insecticides are the principal means of protecting crops, people, livestock, and pets from pest insect attack and disease transmission [1]. Most insecticides are nerve poisons and have been introduced since 1940, followed later by several classes of compounds, including neonicotinoids in the 1990s [2, 3]. Neurotoxicants are the major synthetic insecticides for several reasons [4]. They act rapidly to stop crop damage and disease transmission. There are many sensitive sites at which even a small disruption may ultimately prove to be lethal. A lipoidal sheath protects the insect nerve from ionized toxicants but not from lipophilic insecticides. Poor detoxification mechanisms in nerves provide prolonged toxicant effects.

Fundamental differences between the nicotinic acetylcholine receptor (nAChR) of insects on the one hand and mammals on the other confer remarkable selectivity to the neonicotinoids [5]. There is considerable inhibitor specificity between the acethylcolinesterase nerve target (AChE) of insects and mammals contributing to selective toxicity [4]. Photostabilized compounds selective for insects relative to mammals, commercialized as imidacloprid and six analogs designated neonicotinoids are highly insecticidal but the photolabile nithiazine with a nitromethylene moiety and no cationic substituent act as an nAChR agonist [1]. The guanidine or amidine unit is coplanar and electronically conjugated with the nitro or cyano substituent, which facilitates partial negative charge flow toward the tip. They act as agonists at multiple nAChR subtypes, with differential selectivity between insects and mammals conferred by only minor structural changes. Imidacloprid is not protonated, but its electronegative nitroimine extreme end may bind to a lysine or arginine

residue in a subsite of the insect nicotinic acetylcholine receptor [4]. Nicotine in contrast is protonated at physiological pH and undergoes cation-π interaction with the choline subsite of AChE for insect, at a nicotinic acetylcholine receptor subsite in mammals. Reactivation rates for inhibiting AChE may also vary for insects and mammals. Mammals have a second related enzyme, butyrylcholinesterase, abundant in blood plasma, which reacts with organosphophate and methylcarbamate toxicants and provides a protection not available to insects.

The coplanar segment between guanidine or amidine and pharmacophore in the neonicotinoids can create an electronic conjugation to facilitate the partial negative charge flow toward the tip atom and increase the binding affinity to the insect target [6].

Neuroactive compounds have been the dominant insecticides for 50 years, and this is not expected to change in the foreseeable future [7]. The early neuroactive insecticides were effective, inexpensive, and often persistent, ideal properties for the selection of resistant strains. Resistance can be conferred by minimal (or even a single) amino acid substitution at the target site. Resistance problems with this first generation of neuroactive insecticides prompted the search for new neurotoxicants acting at different sites to circumvent cross resistance. This was outstandingly successful in case of chloronicotinyl compounds or synthetic nicotinoids with outstanding systemic properties that are exemplified by imidacloprid acting at the nAChR nicotinic acetylcholine receptor. These nicotinoids act at the same site in insects as nicotine but with much higher effectiveness and safety. No target site–associated cross-resistance problems exist, and accordingly the synthetic nicotinoids are outstanding replacements for the insecticidal organophosphorus compounds and methylcarbamate insecticide systemics in many uses for sucking insect pests of crops.

One of the main success factors for neonicotinoids is their plant systemicity [5]. Applied into the soil or to the seed, the products are taken up via the roots, are distributed in the plant and give consistent and long-lasting control of sucking insects. Following foliar application, neonicotinoids penetrate into the leaf lamina and control pests on the lower side of the leaf owing to their good translaminar activity. Furthermore, they are distributed acropetally (xylem movement) and can protect new growing shoots.

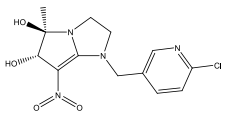
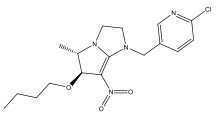
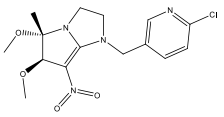
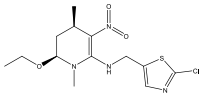
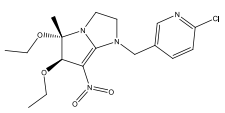
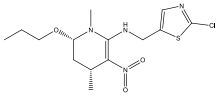
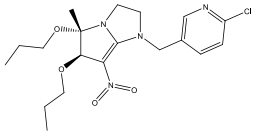
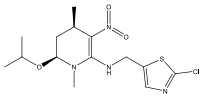
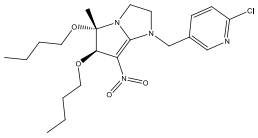
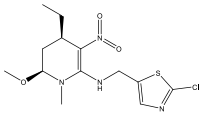
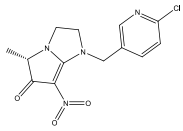
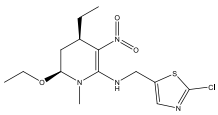
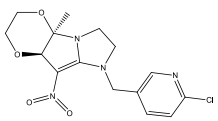
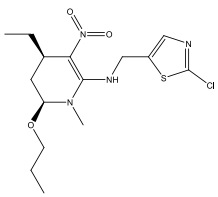
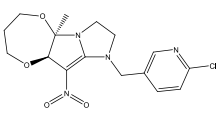
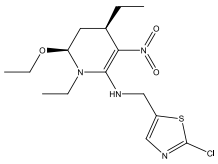
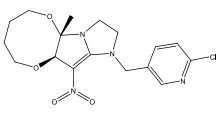
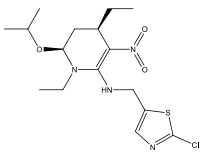
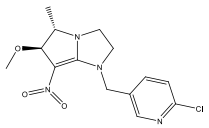
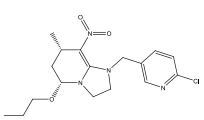
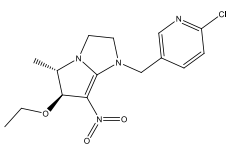
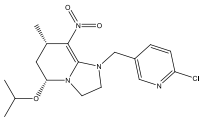
The development of neonicotinoids is provoked by the rapid resistance growing [8] and serious bee toxicity [9-11]. Thus, there is an urgent need for the development of novel, effective, neonicotinoid replacements.

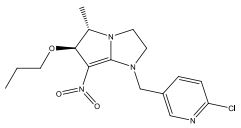
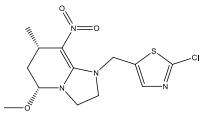
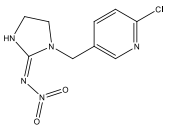
The target of the neonicotinoid insecticides is the nervous system of nAChR, as demonstrated by the quantitative relation between the neuroblocking activity and the insecticidal activity against American cockroaches [12]. In this study 23 variants of the key pharmacophore were built with the central ring conjugated to NCN, $CHNO_2$, or NNO_2 moieties, and demonstrated that the neuroblocking potency is proportional to the Mulliken charge on the nitro oxygen atom or cyano nitrogen atom.

A previous MLR analysis [13] of a series of 24 nitromethylene, pyrrole- and dihydropyrrole-fused neonicotinoids [14, 15] (Table 1) emphasized the importance of Galvez topological charge index (JGI2, which means the topological charge index of order 2) and two GETAWAY descriptors. The partial least squares approach was applied to the same series of neonicotinoids to find out the neonicotinoid structural features that influence the insecticidal activity [16]. Higher topological, edge adjacency indices, and 3D-MoRSE descriptor values were found to be favorable for the insecticidal potency.

This paper presents a 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. Based on the resulted PLS model new insecticide structures active against the cowpea aphids can be predicted.

Table 1. The neonicotinoid structures, the experimental (pLC_{50exp}) and predicted ($pLC_{50\ pred}$) insecticidal activity values obtained using the PLS model.

No	Structure	pLC_{50exp}	pLC_{50pred}	No	Structure	pLC_{50exp}	pLC_{50pred}
1		5.21	5.08	13		3.97	3.94
2		5.70	5.34	14		3.79	3.93
3*		5.80	5.42	15		4.25	4.40
4		5.71	5.35	16		4.07	3.86
5		5.11	5.33	17*		3.91	4.02
6		3.85	3.90	18		3.98	4.05
7		4.55	4.83	19		4.41	4.51
8*		4.52	4.85	20*		3.82	4.10
9		4.41	4.77	21		3.86	4.07
10		4.35	3.99	22		4.04	4.12
11		3.96	4.00	23*		3.58	4.09

12*		4.16	3.98	24		3.72	3.46
25		4.46	4.47				

* Test compounds included in the PLS model.

2. Methods

2.1. Definition of target property and molecular parameters

A dataset of 25 nitromethylene, pyrrole- and dihydropyrrole-fused neonicotinoids (Table 1) having the insecticidal activity (LC_{50}, in mmol/L) against cowpea aphids (*Aphis craccivora*) [15, 17] was analyzed. pLC_{50} values were used as dependant variable.

The neonicotinoid structures were pre-optimized using the MMFF94 molecular mechanics force field included in the Omega (Omega v.2.5.1.4, OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com) software [18, 19]. For conformer generation, the maximum number of conformers per compound set of 400 and an RMSD value of 0.5 Å were employed during the conformer ensemble generation.

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.

2.2. Partial Least Squares (PLS) method

The Partial Least Squares (PLS) approach [20] was employed to relate the pLC_{50} values with the calculated structural descriptors, using the SIMCA (SIMCA P+12 12.0.0.0 2008, Umetrics, Sweeden, www.umetrics.com) program. The PLS approach leads to stable, correct and highly predictive models even for correlated descriptors. The quality of the PLS model was verified using the squared correlation regression coefficient R^2(CUM), and the squared cross-validated correlation coefficient, Q^2(CUM). The Variables Importance in the Projection (VIP) values and the sign of the variables' coefficients were used to explain the activity mechanism. The leave-7-out crossvalidation procedure was employed to select the most significant principal components and to check the internal model validation.

The Y-randomization test was employed to test the model robustness and overfitting. In this procedure the dependent variable is randomly shuffled using the same descriptor matrix. The obtained PLS models (after 999 randomizations) must have the minimal r^2 and q^2 values [21].

2.3. Model validation

The dataset was divided randomly into training and test (24% of the total number of compounds) sets. Following compounds: **3**, **11**, **13**, **14**, **17** and **23** were included in the test set (Table 1).

Several criteria were used for testing the predictive model power: Q^2_{F1} [22], Q^2_{F2} [23], Q^2_{F3} [24], the concordance correlation coefficient (CCC) [25] (having the thresholds values higher than 0.85, [26]) and the predictive parameter r^2_m (with a lowest threshold value of 0.5) [27].

The model overfit was checked using the Y-randomization test [28] and by comparing the root-mean-square errors (RMSE) and the mean absolute error (MAE) of the training and validation sets [29].

3. Results and Discussion

The X matrix of neonicotinoid descriptors was analyzed using the PCA approach. A model with 5 significant components (N = 25 and X = 1426) was obtained; the first three components explain 75.7% of the information content.

The PLS model developed for the entire set of compounds gave the following statistical results: R^2_X(CUM) = 0.872, R^2_Y(CUM) = 0.990 and Q^2(CUM) = 0.731, obtained for seven 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). Therefore the noise variables (e.g. coefficient values insignificantly different from 0) were excluded from this model.

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^2_Y(CUM) = 0.87 and Q^2(CUM) = 0.796 was obtained. The descriptor coefficients and the VIP values included in the final PLS model are presented in Table 2.

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.

Highest contribution to the model is given by the 2D binary and frequency fingerprints descriptors.

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^2_m = 0.680, Q^2_{F1} = 0.805, Q^2_{F2} = 0.802, Q^2_{F3} = 0.704. They indicate the PLS model as having a good predictive power.

The coefficient and VIP plots are presented in Figure 1 and Figure 2, respectively.

Table 2. The PLS coefficients in descending order of the VIP values*.

No	Variable ID*	CoefCS[3]	VIP[3]
1	B03[O-O]	0.157	1.065
2	B05[O-O]	0.157	1.065
3	C-012	0.157	1.065
4	F03[O-O]	0.157	1.065
5	F05[O-O]	0.157	1.065
6	Mor15m	0.256	1.038
7	Xindex	0.193	0.914
8	B09[C-O]	0.367	0.902
9	Vindex	0.187	0.902
10	X1A	0.218	0.886

* B03[O-O] - presence/absence of O-O at topological distance 3 (2D binary fingerprints); B05[O-O] - presence/absence of O-O at topological distance 5 (2D binary fingerprints); C-012 – presence of the CR2X2 fragment (R: any group linked through carbon; X: halogen, atom-centred fragments); F03[O-O] - frequency of O-O at topological distance 3 (2D frequency fingerprints); F05[O-O] - frequency of O-O at topological distance 5 (2D frequency fingerprints); Mor15m - 3D-MoRSE - signal 15 / weighted by atomic masses (3D-MoRSE descriptors); Xindex - Balaban X index (information indices); B09[C-O] – presence/absence of O-O at topological distance 9 (2D binary fingerprints); Vindex - Balaban V index (information indices); X1A - average connectivity index chi-1 (connectivity indices).

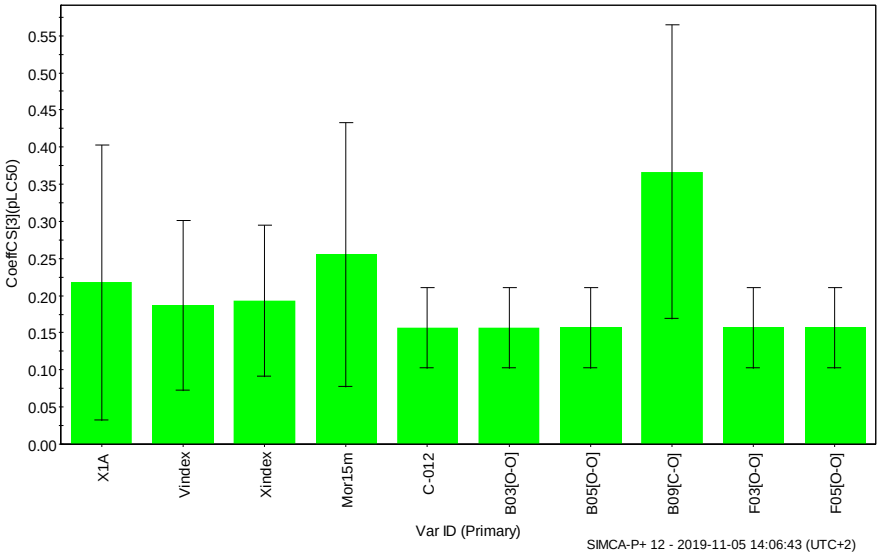


Figure 1. Coefficient plot of the final PLS model.

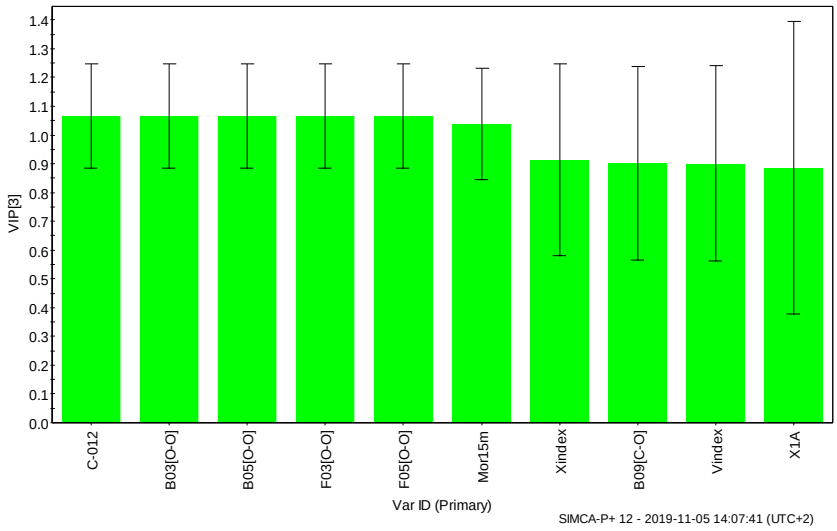


Figure 2. VIP plot for the final PLS model.

The Hotteling's T^2 range plot (Figure 3) confirms the absence of leverage compounds and outliers.

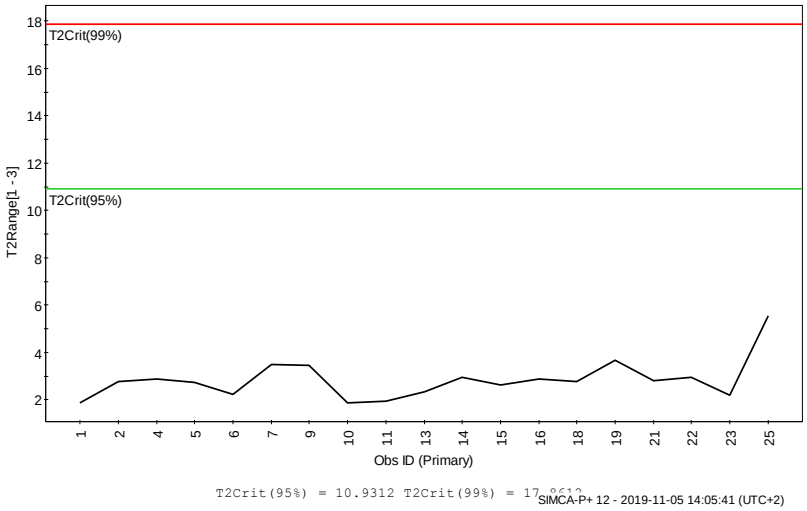


Figure 3. The Hotteling's T2 range plot of the final PLS model.

In the y-scrambling test performed for the PLS model, a significant low scrambled r^2 (r^2_{scr}) and cross-validated q^2 (q^2_{scr}) values were obtained for 999 trials. Figure 4 shows that in case of all the randomized models, the values of r^2_{scr} and q^2_{scr} for the PLS model were < 0.5 (r^2_{scr} / q^2_{scr} of 0.113/-0.428). The low calculated r^2_{scr} and q^2_{scr} values indicate no chance correlation for the PLS chosen model.

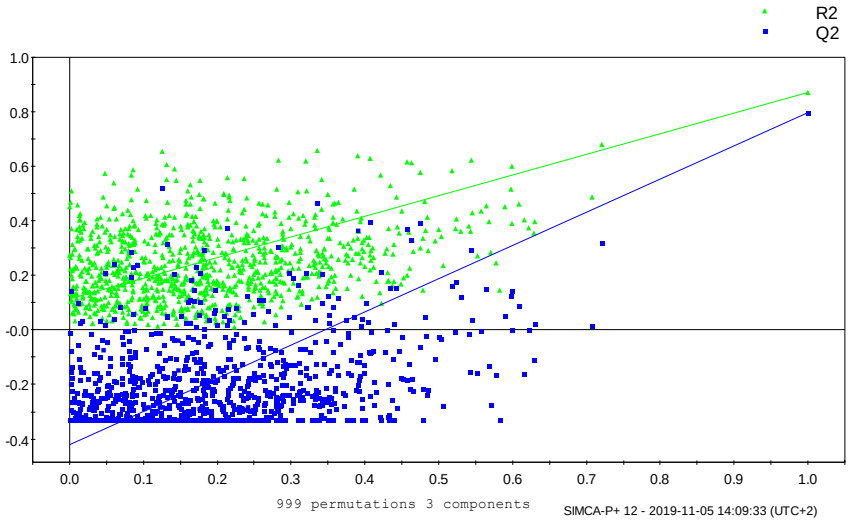


Figure. 4. Y-scramble plots for the PLS model.

New neonicotinoid structures with insecticidal activity against the cowpea aphids can be designed based on the final PLS model.

4. 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 descriptors 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.

Author Contributions

L.C. and A.B.1 analyzed the data; A.B. contributed to molecular modeling calculations; S.F.T. performed the statistical analysis and wrote the paper.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Casida, J. E.; Durkin, K. A. Neuroactive Insecticides: Targets, Selectivity, Resistance, and Secondary Effects. *Annu. Rev. Entomol*. **2013**, *58*, 99–117.
2. Casida JE, Quistad GB. 1998. Golden age of insecticide research: past, present or future? *Annu. Rev. Entomol.* 43:1–16.
3. Tomlin, C.D.S.: *The Pesticide Manual: A World Compendium*. Hampshire, UK: Br. Crop Prot. Counc. 15th ed. (2009).
4. Casida, J.E.; Quistad, G.B. Why insecticides are more toxic to insects than people: the unique toxicology of insects. *J. Pestic. Sci.* **2004**, *29*, 81–86.
5. Elbert, A.; Haas, M.; Springer, B.; Thielert, W.; Nauen, R. Applied aspects of neonicotinoid uses in crop protection. *Pest. Manag. Sci.* **2008**, *64*, 1099–1105.
6. Duan, H.; Zhang, W.; Zhao, J.; Liang, D.; Yang, X.; Jin, S. A novel halogen bond and a better-known hydrogen bond cooperation of neonicotinoid and insect nicotinic acetylcholine receptor recognition. *J. Mol. Model.* **2012**, *18*, 3867–3875.
7. Casida, J.E., Quistad, G.B. Golden age of insecticide research: past, present or future? *Annu. Rev. Entomol.* **1998**, *43*, 1–16.
8. Nauen, R.; Denholm, I. Resistance of Insect Pests to Neonicotinoid Insecticides: Current Status and Future Prospects. *Arch. Insect Biochem.* **2005**, *58*, 200–215.
9. Henry, M.; Béguin, M.; Requier F.; Rollin, O.; Odoux, J.F.; Aupinel, P.; Aptel, J.; Tchamitchian, S.; Decourtye, A. A common pesticide decreases foraging success and survival in honey bees. *Science* **2012**, *336*, 348–350.
10. Cameron, S.A.; Lozier, J.D.; Strange J.P.; Koch, J.B.; Cordes, N.; Solter, L.F.; Griswold, T.L.; Patterns of widespread decline in North American bumble bees. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 662–667.
11. Chen, N.; Xia, S.; Zou, M.; Shao, X. Bridged heterocyclic neonicotinoid analogues: design, synthesis, and insecticidal activity. *Res. Chem. Intermed.* **2015**, *41*, 5293-5300.
12. Kagabu, S.; Nishimura, K.; Naruse, Y.; Ohno, I. Insecticidal and neuroblocking potencies of variants of the thiazolidine moiety of thiacloprid and quantitative relationship study for the key neonicotinoid pharmacophore. *J. Pestic. Sci.*, **2008**, *33(1)*, 58–66.
13. Funar-Timofei, S.; Bora, A. Insecticidal Activity Evaluation of Phenylazo and Dihydropyrrole-Fused Neonicotinoids Against Cowpea Aphids Using the MLR Approach *Proceedings* **2019**, *9*, 18; doi:10.3390/ecsoc-22-05664
14. Ye, Z.; Xia, S.;Shao, X.; Cheng, J.;Xu, X.;Xu, Z.; Li, Z.; Qian, X. Design, synthesis, crystal structure analysis, and insecticidal evaluation of Phenylazoneonicotinoids. *J. Agric. Food Chem.* **2011**, *59*, 10615–10623;
15. Ye, Z.; Shi, L.; Shao, X.; Xu, X.; Xu, Z.; Li, Z. Pyrrole- and dihydropyrrole-fused Neonicotinoids: design, synthesis, and insecticidal evaluation. *J. Agric. Food Chem.* **2013**, 61, 312−319.

16. Suzuki, T.; Bora, A.; Funar-Timofei, S. PLS Evaluation of the Insecticidal Activity of Phenylazo, Pyrrole- and Dihydropyrrole-Fused Neonicotinoids, *Journal of Toyo University*, *Natural Science* **2019**, *63*, 101-112.
17. Lu, S.; Shao, X.; Li, Z.; Xu, Z.; Zhao, S.; Wu, Y.; Xu, X. Design, Synthesis, and Particular Biological Behaviors of Chain-Opening Nitromethylene Neonicotinoids with Cis Configuration. *J. Agric.Food Chem*. **2012**, *60*, 322−330.
18. Hawkins, P.C.D.; Skillman, A.G.; Warren, G.L.; Ellingson, B.A.; Stahl, M.T. Conformer generation with OMEGA: algorithm and validation using high quality structures from the Protein Databank and Cambridge Structural Database. *J. Chem. Inf. Model*. **2010**, *50*, 572–584.
19. Hawkins, P.C,D.; Nicholls, A. Conformer generation with OMEGA: learning from the data set and the analysis of failures. *J. Chem. Inf. Model*. **2012**, *52*, 2919-2936.
20. Wold, H., in: Kotz, S. and Johnson, N.L. (Eds.), Encyclopedia of Statistical Sciences, (Vol. 6), Wiley, New York, pp. 581 (1985).
21. Roy, P.P, Paul, S., Mitra, I., Roy, K. On two novel parameters for validation of predictive QSAR models. *Molecules* **2009**, *14*,1660-1701.
22. Shi, L.M.; Fang, H.; Tong, W.; Wu, J.; Perkins, R.; Blair, R.M.; Branham, W.S.; Dial, S.L.; Moland, C.L.; Sheehan, D.M. QSAR models using a large diverse set of estrogens. *J. Chem. Inf. Model*. **2001**, *41*, 186–195.
23. Schüürmann, G.; Ebert, R.U.; Chen, J.; Wang, B.; Kühne, R. External validation and prediction employing the predictive squared correlation coefficient test set activity mean vs training set activity mean. *J. Chem. Inf. Model*. **2008**, *48*, 2140–2145.
24. Consonni, V.; Ballabio, D.; Todeschini, R. Comments on the definition of the Q2 parameter for QSAR validation. *J. Chem. Inf. Model*. **2009**, *49*, 1669–1678.
25. Chirico, N.; Gramatica, P. 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.
26. Chirico, N.; Gramatica, P. Real External Predictivity of QSAR Models. Part 2. New Intercomparable Thresholds for Different Validation Criteria and the Need for Scatter Plot Inspection. *J. Chem. Inf. Model*. **2012**, *52*, 2044−2058.
27. Roy, K.; Mitra, I. On the Use of the Metric r_m^2 as an Effective Tool for Validation of QSAR Models in Computational Drug Design and Predictive Toxicology. *Mini-Rev. Med. Chem*. **2012**, *12*, 491−504.
28. Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Wold, S. *Multi and megavariate data analysis: principles and applications*. Umetrics AB, Umea, 2001, pp. 92–97, pp. 489–491.
29. Goodarzi, M.; Deshpande, S.; Murugesan, V.; Katti, S.B.; Prabhakar, Y.S. Is Feature Selection Essential for ANN Modeling? *QSAR Comb. Sci*. **2009**, *28*, 1487–1499.