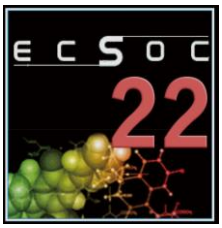
OPEN ACCESS



Conference Proceedings Paper – 22nd International Electronic Conference on Synthetic Organic Chemistry (November 2018) ECSOC-22

Insecticidal Activity Evaluation of Phenylazo and Dihydropyrrole-Fused Neonicotinoids Against Cowpea Aphids Using the MLR Approach

Simona Funar-Timofei * and Alina Bora

Institute of Chemistry Timisoara of the Romanian Academy, Bul. Mihai Viteazu 24, 300223 Timisoara, Romania; E-Mails: timofei@acad-icht.tm.edu.ro (S.F.T.); alina.bora@gmail.com (A.B.)

***** Author to whom correspondence should be addressed; E-Mail: timofei@acad-icht.tm.edu.ro (S.F.T.)
Tel.: +40-256-491-818; Fax: +40-256-491-824.

Abstract: This paper presents a QSAR study of a series of 24 dihydropyrrole-fused and phenylazo neonicotinoid derivatives, with insecticidal activity tested against Cowpea aphids (Aphis craccivora). In this regard, the conformational search ability of the OMEGA software was employed to model neonicotinoid conformer ensembles, using molecular mechanics calculations based on the MMFF94s (the 94s variant of the Merck Molecular) force field. The minimum energy conformers were used to calculate structural descriptors, which were further related to the insecticidal activity (pLC_{50} values), using the multiple linear regression approach. The genetic algorithm was used for variable selection and several criteria for internal and external model validation. A robust model (r^2 = 0.880, r^2_{adj} = 0.855, q^2_{LOO} = 0.827, s = 0.2098, F = 34.295) with predictive power (CCC_{ext} = 0.945, r^2_m= 0.824) was obtained, using the QSARINS software. The developed model can be confidently used for the prediction of the insecticidal activity of new chemicals, saving a substantial amount of time and money.

Keywords: neonicotinoid; MLR; Omega; cowpea aphids; QSARINS

---

1. Introduction

Neonicotinoids are considered to be one of the most important and relevant classes of insecticides used nowadays [1, 2].

Neonicotinoids, synthetic insecticides acting on the insect nicotinic acetylcholine receptor (nAChR), have been increasingly used to control various insects during recent decades since imidacloprid was introduced to the market [3].

The neonicotinoids success is, however, being provoked by the rapid development of resistance [2] and severe bee toxicity [4-6]. It is considered that neonicotinoid insecticides represent the most effective chemical class for the control of sucking insect pests (aphids, whiteflies, leaf- and planthoppers, thrips), some micro lepidoptera and a number of coleopteran pest species [7]. Neonicotinoids have the advantage of their plant systemicity over other insecticides. After application into the soil or the seed, these compounds are absorbed through the plant roots, where they are distributed and give therefore consistent and long-lasting control of sucking insects.

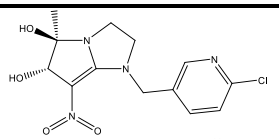
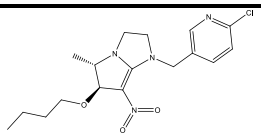
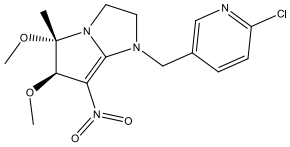
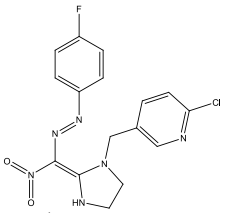
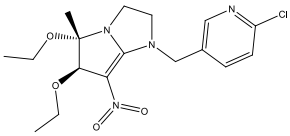
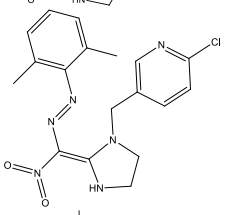
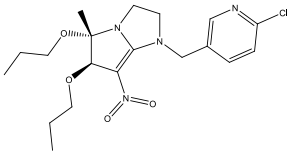
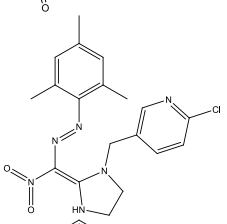
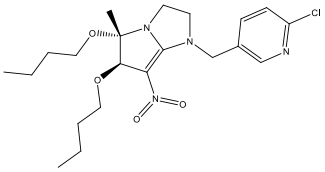
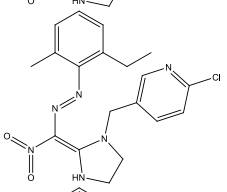
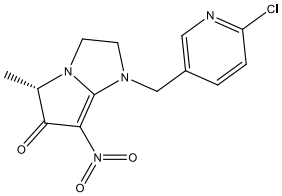
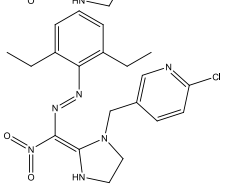
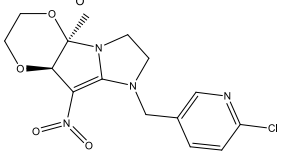
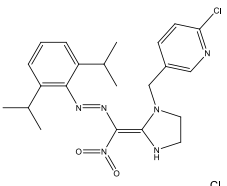
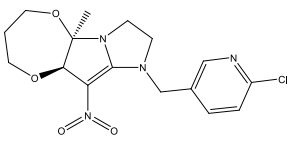
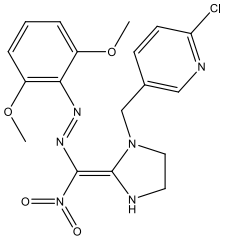
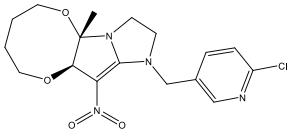
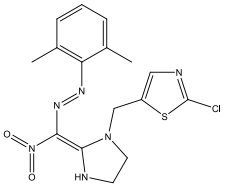
The coplanar segment between guanidine or amidine and pharmacophore in the neonicotinoids could create an electronic conjugation to facilitate the partial negative charge flow toward the tip atom and increase the binding affinity to the insect target [8]. Photostabilized compounds selective for insects relative to mammals have photolabile nithiazine with a nitromethylene moiety and no cationic substituent. [9].

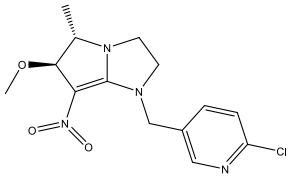
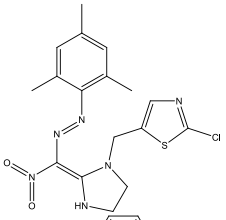
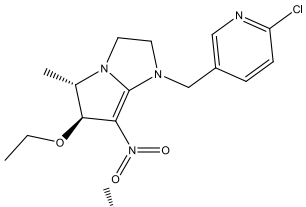
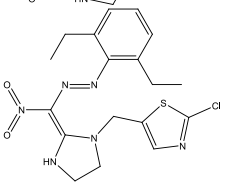
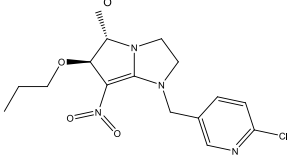
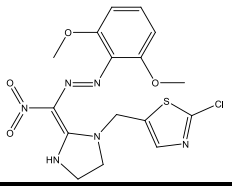
Quantitative Structure-Activity Relationship (QSAR) is the most commonly used method to understand how chemical structure features correlate with the toxicity of natural and/or synthetic chemicals like insecticides. This method offers the possibility of searching for new insecticides with enhanced activity against insects and pests. The urgent need for the development of new insecticide is related to the phenomenon of insecticide-resistant cases of pests. In this regard, several computational approaches were applied to study the insecticidal activity of neonicotinoids [10-15].

In this study, the QSAR model of 24 dihydropyrrole-fused and phenylazo neonicotinoid derivatives is derived from the data set of chemical structures and insecticidal activities tested against Cowpea aphids (*Aphis craccivora*) using multiple linear regression (MLR) approach.

Molecular mechanics calculations, using the MMFF94s force field, were used to model the neonicotinoid structures. Statistical analysis using several criteria was employed to find a robust and predictive MLR model. The best derived MLR model could be confidently used to predict the insecticidal activity of newly designed insecticides.

Table 1. Experimental (pLC_{50exp}) and predicted for the best MLR model (pLC_{50pred}) insecticidal activity values of neonicotinoids.

No	Structure	pLC_{50exp}	pLC_{50pred}	No	Structure	pLC_{50exp}	pLC_{50pred}
1		5.21	5.16	13*		3.97	4.04
2		5.70	5.57	14*		4.43	4.22
3*		5.80	5.59	15		5.37	5.49
4		5.71	5.61	16*		5.30	5.08
5		5.11	5.34	17		5.43	5.33
6		3.85	3.97	18		5.55	5.21
7		4.55	4.77	19		4.86	5.34
8		4.52	4.53	20		5.00	4.86
9		4.41	4.49	21		5.46	5.33

10		4.35	4.16	22		4.82	4.88
11*		3.96	4.23	23*		4.93	5.16
12		4.16	4.15	24		4.83	4.70

* Test compounds included in the best MLR model

2. Methods

2.1. Dataset and theoretical molecular descriptors calculation

A dataset of 24 phenylazo and dihydropyrrole-fused neonicotinoid derivatives (Table 1) having the insecticidal activity (LC_{50}, in mmol/L) against cowpea aphids (*Aphis craccivora*) [16, 17] was analyzed. pLC_{50} values were used as the 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. The Multiple Linear Regression (MLR) method

The multiple linear regression (MLR) approach [20] was employed to relate the pLC_{50} values with the calculated structural descriptors, using the QSARINS v. 2.2 program [21, 22]. The genetic algorithm with leave-one-out cross-validation correlation coefficient was used for variable selection, as constrained function to be optimized, a mutation rate of 20%, the population size of 10 and 500 iterations.

2.3. Model validation

The dataset was divided randomly into training and test (25% 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} [23], Q^2_{F2} [24], Q^2_{F3} [25], the concordance correlation coefficient (CCC) [26] (having the thresholds values higher than 0.85, [27]) and the predictive parameter r^2_m (with a lowest threshold value of 0.5) [28].

The model overfit was checked using the Y-randomization test [29] and by comparing the root-mean-square errors (RMSE) and the mean absolute error (MAE) of the training and validation sets [30].

Y-scrambling [31], the adjusted correlation coefficient (r^2_{adj}) and q^2 (leave-one-out, q^2_{LOO}, and leave-more-out, q^2_{LMO}) cross-validation coefficients were employed for internal model validation.

The Multi-Criteria Decision Making (MCDM) validation criterion [32] is used to summarize the performance of MLR models. To every validation criteria, a desirability function is associated, and MCDM have values between 0 (the worst) and 1 (the best).

3. Results and Discussion

The autoscaling method was employed for normalizing the data:

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

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.

The variables contained in the MLR models were selected using the genetic algorithm. The statistical (fitting and predictivity) results are included in Tables 2-4.

The ‚MCDM all' scores, based on the fitting, cross-validated and external criteria were considered for choosing the best MLR models.

Table 2. Fitting and cross-validation statistical results of the MLR models.*

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	0.880	0.827	0.806	0.855	0.185	0.147	0.936	0.176	-0.404	0.210	34.295
MLR2	0.865	0.793	0.774	0.837	0.196	0.164	0.928	0.174	-0.396	0.222	30.000
MLR3	0.854	0.777	0.755	0.822	0.205	0.172	0.921	0.178	-0.390	0.232	27.208
MLR4	0.854	0.790	0.772	0.823	0.204	0.161	0.921	0.177	-0.397	0.232	27.333

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

Table 3. The model predictivity results.*

Model	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	$RMSE_{ext}$	MAE_{ext}	CCC_{ext}
MLR1	0.904	0.844	0.945	0.211	0.202	0.945
MLR2	0.801	0.676	0.889	0.304	0.293	0.889
MLR3	0.818	0.704	0.896	0.291	0.281	0.896
MLR4	0.744	0.583	0.858	0.345	0.309	0.858

* 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. The 'MCDM all' score values, r^2_m predictivity parameter, and descriptors included in the MLR models.*

Model	r^2_m	MCDM all	Descriptors included in the MLR model*
MLR1	0.824	0.867	JGI2 HATSv R3m
MLR2	0.795	0.814	BEHp2 JGI2 R3m
MLR3	0.791	0.812	JGI2 Mor06m R3m
MLR4	0.720	0.786	JGI2 R3m R8m+

* JGI2 - mean topological charge index of order2 (Galvez topological charge index); HATSv - leverage-weighted total index / weighted by atomic van der Waals volumes (GETAWAY descriptor); R3m - R autocorrelation of lag 3 / weighted by atomic masses (GETAWAY descriptor); BEHp2 - highest eigenvalue n. 2 of Burden matrix / weighted by atomic polarizabilities (BCUT descriptor); Mor06m - 3D-MoRSE - signal 06 / weighted by atomic masses (3D-MoRSE descriptor); R8m+ - R maximal autocorrelation of lag 8 / weighted by atomic masses (GETAWAY descriptor).

For the reliability of the best MLR1 model, the experimental versus predicted pLC_{50} values, and Y-scramble plots for are presented in Figures 1 and 2, respectively.

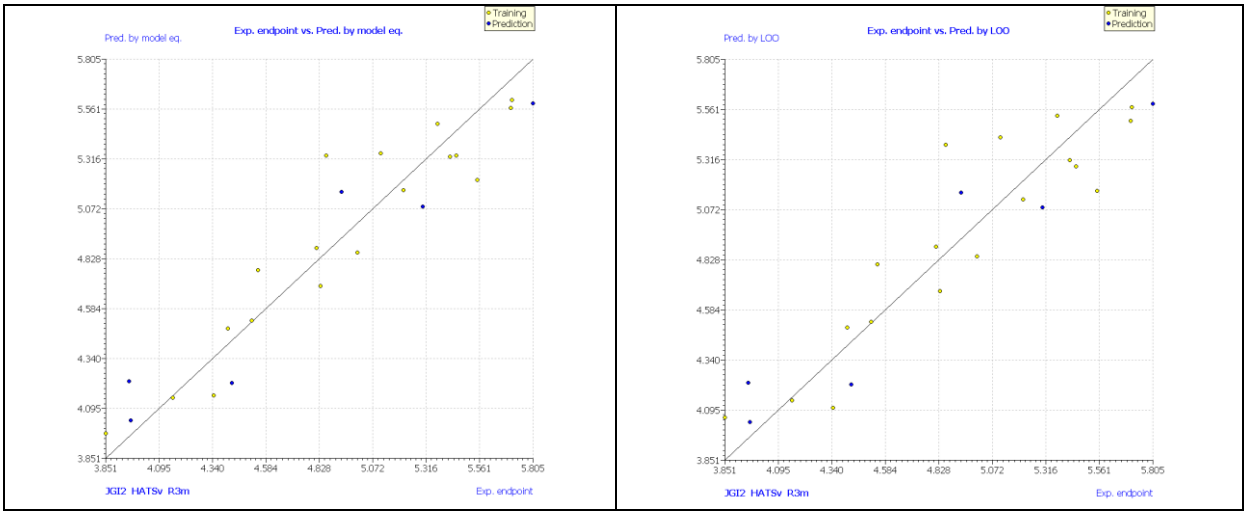


Figure. 1. Plots of experimental versus predicted pLC_{50} values for the MLR1 model predicted by the model (left) and by the leave-one-out (right) cross-validation approach (yellow circles-training compounds, blue circles-test compounds).

In the y-scrambling test performed for the MLR models, a significant low scrambled r^2 (r^2_{scr}) and cross-validated q^2 (q^2_{scr}) values were obtained for 2000 trials. Figure 2 shows that in case of all the randomized models, the values of r^2_{scr} and q^2_{scr} for the MLR1 model were < 0.5 (r^2_{scr}/q^2_{scr} of 0.1759/-0.4035). The low calculated r^2_{scr} and q^2_{scr} values indicate no chance correlation for all MLR chosen models (Table 2).

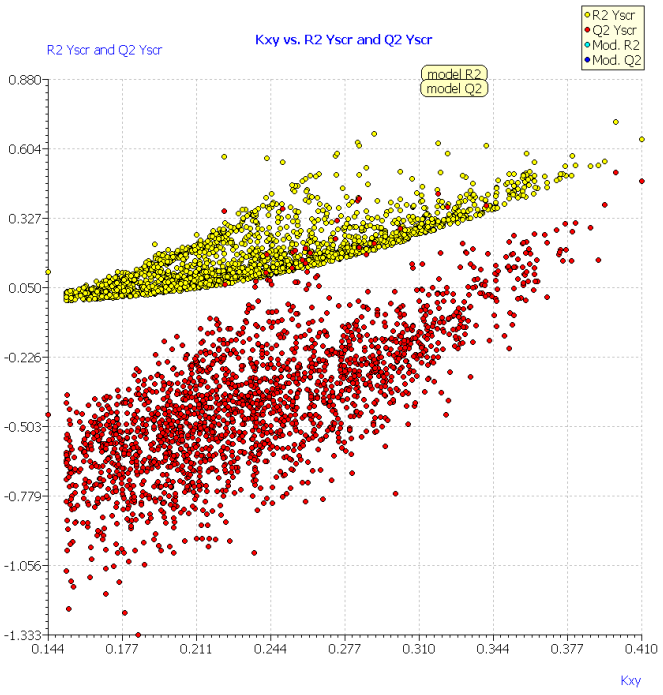


Figure. 2. Y-scramble plots for the MLR1 model.

The Williams plot (standardized residuals versus leverages, with the leverage threshold h* = 0.667 for the MLR1 model), in the range of ±2.5σ, was used to verify the domain applicability. All compounds in the dataset are within the applicability domain of the MLR1 model, as presented in Figure 3.

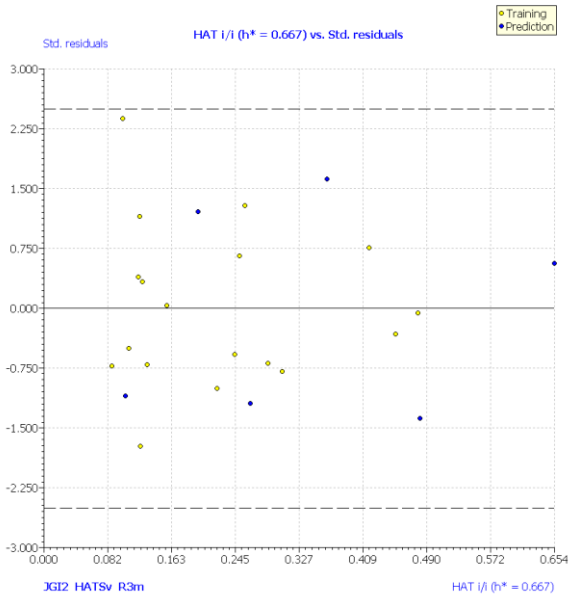


Figure. 3. Williams plot predicted by the MLR1 model (yellow circles-training compounds, blue circles-test compounds).

The selected descriptors included in the MLR1 best model are not intercorrelated, as presented in the correlation matrix from Table 5.

Table 5. Correlation matrix of the descriptors included in the best MLR1 model, and their standardized coefficients (Std. coeff.).

	JGI2	**HATSv**	**R3m**	**Std. coeff.**
JGI2	1			**0.967**
HATSv	-0.278	1		**0.321**
R3m	**-0.121**	**0.623**	**1**	**-0.617**

Good correlations with the insecticidal activity and predictive model power were notices for all MLR models. Model MLR4 is less predictive (in accordance to its Q^2_{F2} value), compared to the other MLR models. Closer values of the root-mean-square errors (RMSE) and the mean absolute error (MAE) of the training and validation sets were observed for the MLR2, MLR3, and MLR4 models. MLR1 model was considered being the best one according to several other statistical parameters of fitting and the 'MCDM all' score values.

The best MLR1 model has three descriptors: one Galvez topological charge index (JGI2, which means the topological charge index of order 2) and two GETAWAY descriptors (HATSv, which represents the leverage-weighted total index / weighted by atomic van der Waals volumes and R3m - R

autocorrelation of lag 3 / weighted by atomic masses). The increase of the JGI2 and HATSv descriptor values is favorable for high insecticidal activity. Lower values of R3m raise the insecticidal activity.

New neonicotinoid structures with insecticidal activity against the cowpea aphids can be designed based on the MLR models presented in this study.

4. Conclusions

Quantitative structure-insecticidal activity relationships were developed using the multiple linear regression approach for neonicotinoids with dihydropyrrole-fused and phenylazo moieties, active against the cowpea aphids (*Aphis craccivora*). Insecticide structures were modeled using the MMFF94s force field. Descriptors of the minimum conformers were related to the pLC50 values using the multiple linear regression approach. Good correlations and predictive models were obtained. Getaway and Galvez topological charge index descriptors included in the best MLR model can be used for prediction of new insecticides active against the cowpea aphids, saving experimental time and money.

Acknowledgments

This project was financially supported by Project 1.1 of the Institute of Chemistry of the Romanian Academy. Access to the Chemaxon Ltd., QSARINS and MOPAC 2016 software are greatly acknowledged by the authors.

Author Contributions

S.F.T. and A.B. 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. Ren, L.; Lou, Y.; Chen, N.; Xia, S.; Shao, X.; Xu, X.; Li, Z. Synthesis And Insecticidal Activities Of Tetrahydroimidazo[1,2-A]Pyridinones: Further Exploration On Cis-Neonicotinoids. *Synthetic Commun.* **2014**, *44*, 858–867.
2. Nauen, R.; Denholm, I. Resistance of Insect Pests to Neonicotinoid Insecticides: Current Status and Future Prospects. *Arch. Insect Biochem.* **2005**, *58*, 200–215.
3. Kagabu, S. Chloronicotinyl insecticides discovery, application and future perspective. *Rev. Toxicol.* **1997**, *1*, 75-129.
4. 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.

5. 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.

6. Chen, N.; Xia, S.; Zou, M.; Shao, X. Bridged heterocyclic neonicotinoid analogues: design, synthesis, and insecticidal activity. *Res. Chem. Intermed*. **2015**, *41 (8)*, 5293-5300.

7. 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.

8. 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.

9. Casida, J. E.; Durkin, K. A. Neuroactive Insecticides: Targets, Selectivity, Resistance, and Secondary Effects. *Annu. Rev. Entomol*. **2013**, *58*, 99–117.

10. 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.

11. Li, J.; Ju, X. L.; Jiang, F. C. Pharmacophore model for neonicotinoid insecticides. *Chinese Chem. Lett*. **2008**, *19*, 619–622.

12. Okazawa, A.; Akamatsu, M.; Nishiwaki, H.; Nakagawa, Y.; Miyagawa, H.; Nishimura, K.; Ueno, T. Three-dimensional quantitative structure–activity relationship analysis of acyclic and cyclic chloronicotinyl insecticides *Pest. Manag. Sci*. **2000**, *56*, 509-515.

13. Okazawa, A.; Akamatsu, M.; Ohoka, A.; Nishiwaki, H.; Cho, W.-J.; Nakagawa, Y.; Nishimura, K.; Ueno, T. Prediction of the Binding Mode of Imidacloprid and Related Compounds to House-Fly Head Acetylcholine Receptors Using Three-Dimensional QSAR Analysis. *Pestic. Sci*. **1998**, *54*, 134-144.

14. Nakayama, A., Sukekawa, M. Quantitative correlation between molecular similarity and receptor-binding activity of neonicotinoid insecticides. *Pestic. Sci*., **1998**, *52*, 104-110.

15. Sung, N.-D.; Jang, S.-C.; Choi, K.-S. CoMFA and CoMSIA on the Neuroblocking Activity of 1-(6-Chloro-3-pyridylmethyl)-2-nitroiminoimidazolidine Analogues. *Bull. Korean Chem. Soc.* **2006**, *27(11)*, 1741-1746.

16. 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.

17. 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.

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

21. Chirico, N.; Papa, E.; Kovarich, S.; Cassani, S.; Gramatica, P. 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.;
22. Gramatica, P.; Chirico, N.; Papa, E.; Cassani, S.; Kovarich, S. QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. *J. Comput. Chem*. **2013**, *34*, 2121–2132.
23. 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.
24. 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.
25. 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.
26. 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.
27. 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.
28. 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.
29. 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.
30. 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.
31. Todeschini, R.; Consonni, V.; Maiocchi, A. The K correlation index: Theory development and its application in chemometrics. *Chemometr. Intell. Lab*. **1999**, *46*, 13-29.
32. Keller, H.R.; Massart, D.L.; Brans, J.P. Multicriteria decision making: a case study. *Chemom. Intell. Lab. Syst*. **1991**, *11*, 175-189.

© 2018 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).