21st International Electronic Conference on Synthetic Organic Chemistry (November 2017) ECSOC-21

QSAR Study of Neonicotinoid Insecticidal Activity Against Cowpea Aphids

Simona Funar-Timofei* and Alina Bora

Institute of Chemistry Timisoara of the Romanian Academy, 24 Mihai Viteazu, 300223 Timisoara, ROMANIA

E-mails: timofei@acad-icht.tm.edu.ro (S.F.T.); alina.bora@gmail.com (A.B.)

BACKGROUND

Neonicotinoids are considered to be one of the most important and relevant classes of insecticides used nowadays, accounting for over 10% of insecticidal market[1,2]. To date, there are eight insecticides commercialized with a neonicotinoid mode of action and others in development.

The neonicotinoids mode of action is similar to the natural insecticide nicotine. They are active on the insect postsynaptic nicotinic acetylcholine receptors (nAChRs) and still of current interest, despite their resistance and bee toxicity[3].

The basic neonicotinoid skeleton is composed of an amidine or a guanidine part conjugated to an electron-withdrawing group such as nitro or cyano. Every neonicotinoid poses two sites for binding to the nicotinic acetylcholine receptors: *(i)* a cationic site and *(ii)* a hydrogen acceptor site.

Several studies of computational chemistry and electrophysiology tried to model the neonicotinoid-receptor interactions. As outcomes, electrostatic interactions and possibly hydrogen bond formation were found to be important for the insecticidal activity[4].

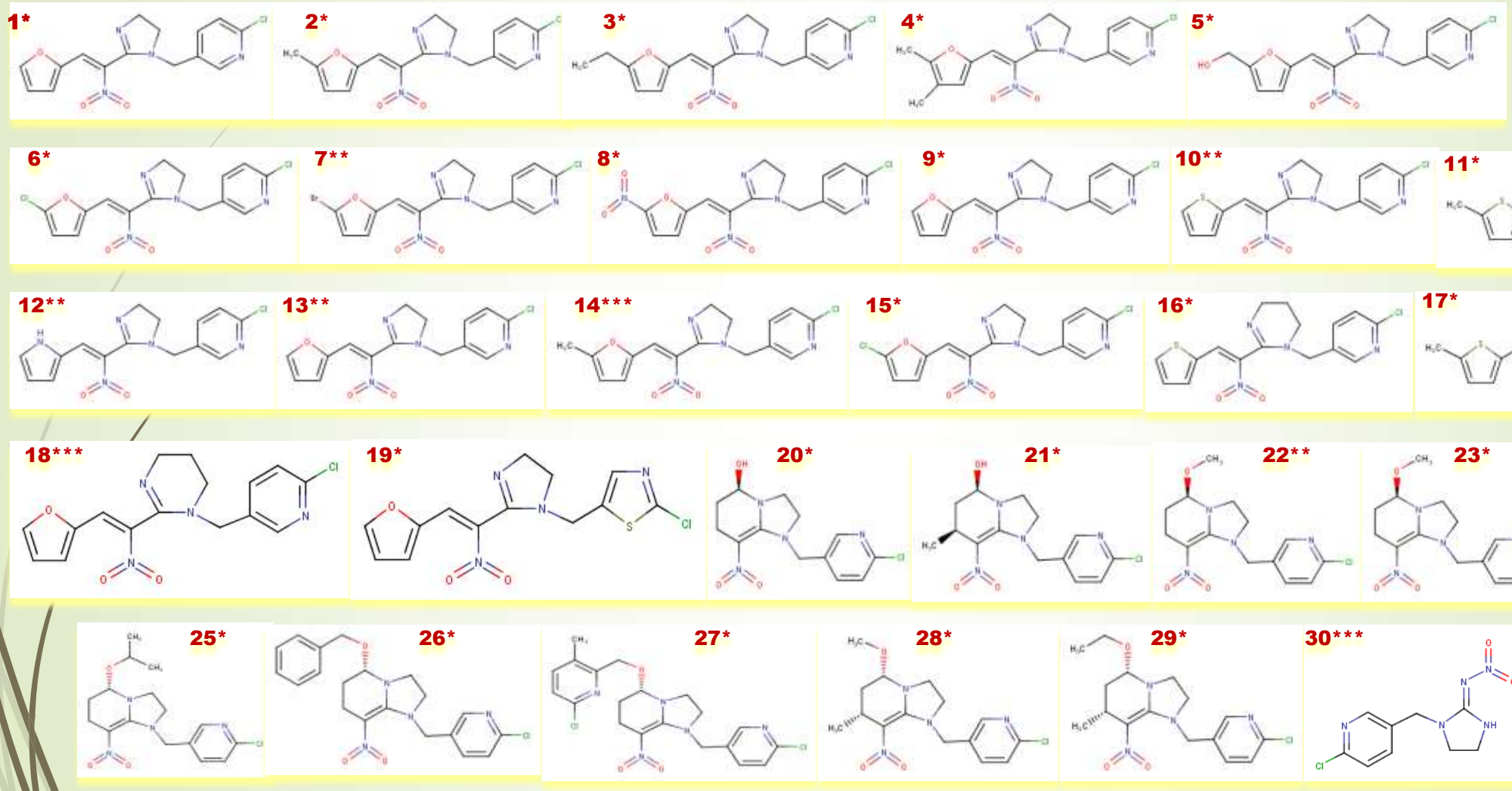
1. Ren L.; Lou Y.; Chen N.; Xia S.; Shao X.; Xu X.; Li Z. *Synthetic Commun*. 2014, 44, 858–867.
2. Nauen R.; Denholm I. *Arch. Insect Biochem*. 2005, 58, 200–215.
3. Matsuda K.; Kanaoka S.; Akamatsu M.; Sattelle D. B. *Mol. Pharmacol*. 2009, 76, 1–10.
4. Matsuda K.; Shimomura M.; Ihara M.; Akamatsu M.; Sattelle D.B. *Biosci. Biotechnol.Biochem.,* 2005, 69, 1442-1452.

AIM

A series of 30 neonicotinoid analogues tested against the cowpea aphids (Aphis craccivora) was modeled by molecular and quantum mechanics approaches.

Multiple linear regression (MLR) and genetic algorithm (GA) methods were used to simulate the relationship between pLC50 values and computed structural descriptors.

NEONICOTINOIDS CHEMICAL STRUCTURES[5,6]



5. Tian Z.; Shao X.; Li Z.; Qian X.; Huang Q. Synthesis, J. Agric. Food Chem. 2007, 55, 2288-2292.
6. Shao X.; Li Z.; Qian X.; Xu X. J. Agric. Food Chem. 2009, 57, 951–957.

*Training compounds included in the final MLR1 data set
**Test compounds included in the final MLR1 data set
***Compounds excluded from the final MLR1 model

METHODS

- *Definition of target property and molecular structures*

- **The insecticidal activity (expressed as pLC50 values) of 30 neonicotinoid analogues bearing nitroconjugated double bond and five-membered heterocycles and nitromethylene neonicotinoids containing a tetrahydropyridine ring with exo-ring ether modifications was used as dependent variable.**

- **The 30 neonicotinoid structures were pre-optimized using the conformer plugin of the MarvinSketch[7] package (with MMFF94 as molecular mechanics force field) and further the lowest energy conformers were refined using the semiempirical PM7 Hamiltonian of MOPAC[8] 2016 program .**

- **Structural 0D, 1D, 2D and 3D molecular descriptors were calculated for the lowest energy structures using the DRAGON[9] and InstanJChem[10] software.**

7. MarvinSketch 15.2.16.0, ChemAxon Ltd. http://chemaxon.com
8. MOPAC2016, James J. P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, HTTP://OpenMOPAC.net(2016)
9. Dragon Professional 5.5, 2007, Talete S.R.L., Milano, Italy
10. Instant JChem (2012) version 5.10.0, Chemaxon, http://www.chemaxon.com

METHODS

- The MLR calculations were performed using the QSARINS[11] v2.1 package.
- The high number of computed descriptors (N=1624) compared to the number of compounds (N = 30) imposed a proper variable selection method such as Genetic Algorithm (GA)[12].
- The QSARINS program uses GAs to select the meaningful descriptors that influence the biologic activity of the compounds. The following parameters were employed: the RQK fitness function with leave-one-out cross-validation 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.

11. Gramatica P.; Chirico N.; Papa E.; Cassani S.; Kovarich S. *J. Comput. Chem.* 2013, 34, 2121–2132.
12. Depczynski U.; Frost V.J.; Molt K., *Anal. Chim. Acta* 2000, 420, 217-227.

METHODS

- *Model validation*
- The neonicotinoid derivatives were randomly divided as fallows:
  - 18.5% of the total number of compounds (no. 7, 10, 12, 13, 22) as test set
  - 81.5% as training set
- The model's predictability was tested using the external validation parameters[13-15]:
  - Q^2_{F1} Q^2_{F2} Q^2_{F3}
  - the concordance correlation coefficient (CCC)
  - r^2_m (with a lowest threshold value of 0.5 to be accepted)
- For internal validation results, several measures of robustness were employed[16-18]:
  - Y-scrambling,
  - adjusted correlation coefficient (r^2_{adj})
  - q2 (leave-one-out, q^2_{LOO} and leave-more-out, q^2_{LMO}) cross-validation coefficient.
- The performance of the MLR models was tested by the Multi-Criteria Decision Making (MCDM) validation criteria (with values between 0 (the worst) and 1 (the best)).

13. Chirico N.; Gramatica P. J. Chem. Inf. Model. 2011, 51, 2320-2335.
14. Chirico N.; Gramatica P J. Chem. Inf. Model. 2012, 52, 2044−2058.
15. Roy K.; Mitra I. Mini-Rev. Med. Chem. 2012, 12, 491−504.
16. Eriksson L.; Johansson E.; Kettaneh-Wold N.; Wold S. Umetrics AB, Umea, 2001, pp. 92–97, pp. 489–491.
17. Todeschini R.; Consonni V.; Maiocchi A. Chemometr. Intell. Lab. 1999, 46, 13-29.
18. Keller H.R.; Massart D.L.; Brans J.P. Chemom. Intell. Lab. Syst. 1991, 11, 175-189.

RESULTS AND DISCUSSIONS

The statistical results for MLR model fitting and predictivity

Table 1. The fitting and cross-validation statistical results 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	0.896	0.853	0.845	0.885	0.261	0.216	0.945	0.095	-0.220	0.281	81.61
MLR2	0.887	0.851	0.841	0.876	0.271	0.220	0.940	0.095	-0.228	0.292	74.90
MLR3	0.808	0.770	0.763	0.799	0.354	0.302	0.894	0.045	-0.157	0.372	84.35
MLR4	0.824	0.786	0.779	0.815	0.340	0.294	0.904	0.049	-0.152	0.356	93.58

- $r^2_{training}$ 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 2. The MLR predictivity results (test set)*

	Model	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	$RMSE_{ext}$	MAE_{ext}	CCC_{ext}
1	MLR	0.851	0.840	0.916	0.235	0.179	0.907
2	MLR	0.805	0.790	0.890	0.269	0.244	0.913
3	MLR	0.876	0.867	0.930	0.214	0.207	0.934
4	MLR	0.820	0.806	0.898	0.258	0.236	0.921

* Q^2_{F1}; Q^2_{F2}; Q^2_{F3}-external validation parameters;
RMSEext-root-mean-square errors;
MAEext -mean absolute error;
CCCext-the concordance correlation coefficient

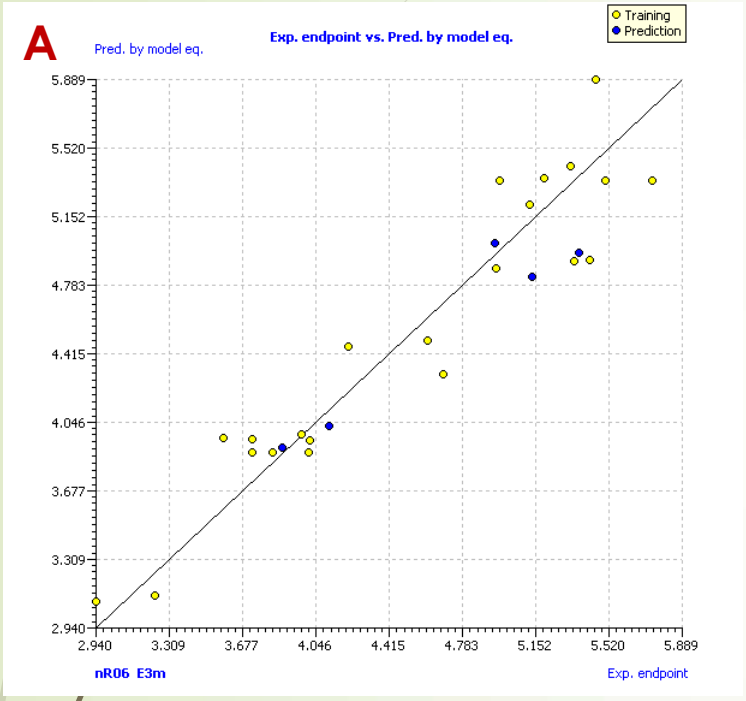
Table 3. The predictivity parameters, 'MCDM all' score values and descriptors in the final MLR models*

Model	r^2_m	MCDM all	Descriptors included in the model*
MLR1	0.810	0.878	nR06, E3m
MLR2	0.697	0.865	nCrs, C-003
MLR3	0.817	0.846	Strongest basic pKa
MLR4	0.656	0.840	nCrs

* nR06 – number of 6-membered rings, E3m- 3rd component accessibility directional WHIM index / weighted by atomic masses, nCrs- number of ring secondary C(sp3), C-003 - CHR3 (atom-centred fragments), strongest basic pKa- the basic pKa value for the first strength index.

RESULTS AND DISCUSSIONS

The reliability of the MLR model



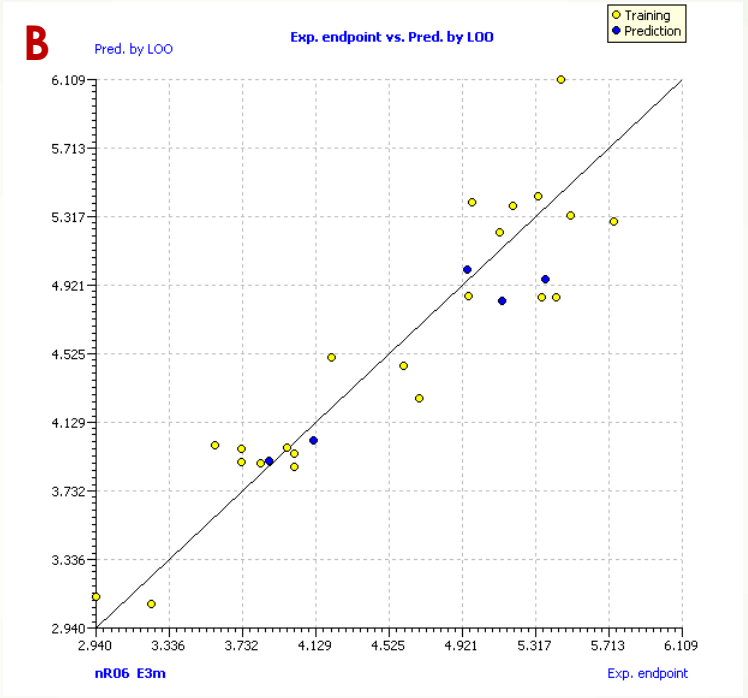


Figure 1. Plots of experimental versus predicted pLC50 values for the MLR1 model - predicted by the model (**A**) and by the leave-one-out (**B**) cross-validation approach (yellow circles-training compounds, blue circles-test compounds).

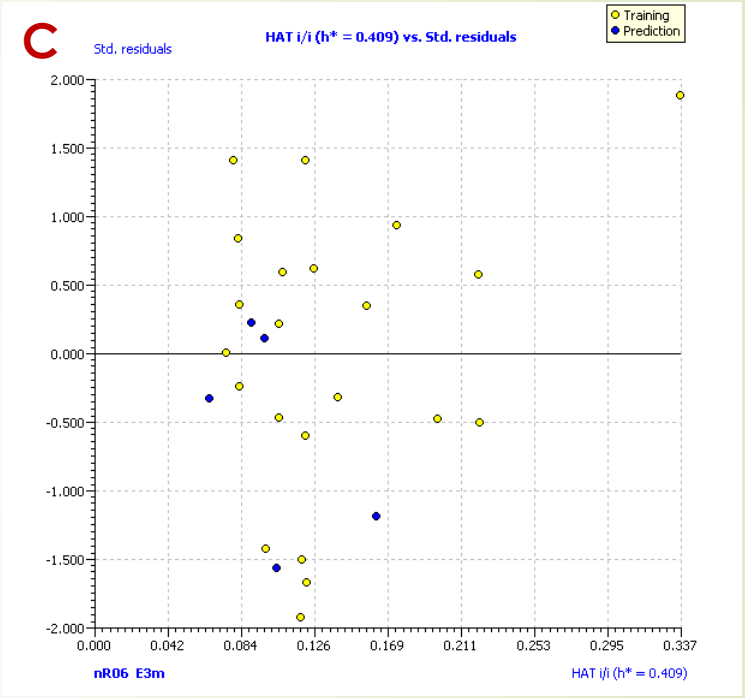


Figure 2. Williams plot predicted by the MLR1 model (**C**) (yellow circles-training compounds, blue circles-test compounds).

RESULTS AND DISCUSSIONS

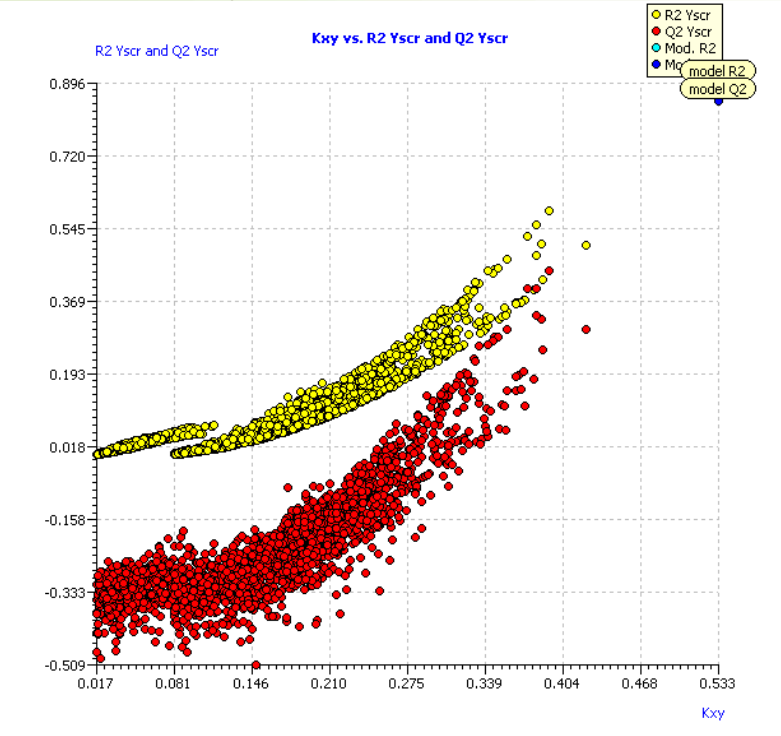


Figure 3. Y-scramble plots for the MLR1 model

The model robustness and predictive power

Table 4. Correlation matrix of the selected descriptors included in the best MLR1 model

	nR06	E3m
nR06	1	
E3m	0.247	1

The increases of E3m is beneficial for the insecticidal activity

The presence of more 6-membered rings in the structure decreases the insecticide action

CONCLUSIONS

Quantitative relationships between the molecular structure and cowpea aphids (Aphis craccivora) inhibitory activity of neonicotinoids analogues was verified by MLR approach.

The semiempirical quantum chemical PM7 method was employed for structure optimization and genetic algorithm for variable selection.

The final MLR models have good statistical parameters and predictive power.

Molecular descriptors related to the number of 6-membered rings in the structure, basic pKa capacity and the number of ring secondary C(sp3) have significant influence on the insecticidal activity.

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.