# PLS Modeling of Neonicotinoid Insecticides Bearing Nitromethylene and Nitroconjugated Structural Moieties

Takahiro SUZUKI[1] · Alina BORA[2] · Luminita CRISAN[2] · Ana BOROTA[2] · Simona FUNAR-TIMOFEI[2]

**Abstract**

Neonicotinoids have a general use as potent agonists against a large spectrum of sucking and chewing pests. Their molecular target site is the insect nicotinic acetylcholine receptor. In this paper the partial least squares (PLS) approach is used to model the insecticidal activity against the cowpea aphids of a series of neonicotinoid derivatives bearing nitromethylene and nitroconjugated fragments. Structural descriptors were calculated from the energy optimized insecticide structures using the semiempirical quantum chemical PM7 approach and were related to the $pLC_{50}$ values. The resulted PLS model has good fitting results ($R^2X(Cum) = 0.968$, $R^2Y(cum) = 0.9$ and $Q^2(Cum) = 0.878$) and predictive power ($CCC_{tr} = 0.951$, $CCC_{ext} = 0.849$, $CCC_{CV} = 0.931$, $R^2_{ezt}$, $Q^2_{F1}$, $Q^2_{F2}$, $Q^2_{F3}$). New insecticides with improved activity can be designed based on the final PLS model.

**Keywords**：Neonicotinoids, *Aphis craccivora*, PLS, MOPAC

## 1. Introduction

Neonicotinoids or neonics are the most widely used insecticides introduced to the global market with a foremost impact in the economy and the ecosystem of any country (Jeschke et al., 2011; Casida and Durkin, 2013; Bonmatin et al., 2015). They are registered in more than 120 countries accounting over 26% of the global insecticide market (Jeschke and Nauen, 2008).

They have become the most successful class of insecticides with large-scale applications for plant protection (e.g. crops, vegetables, fruits), veterinary products,

[1] Natural Science Laboratory, Toyo University, 5-28-20 Hakusan, Bunkyo-ku, Tokyo 112-8606, Japan
[2] Coriolan Dragulescu Institute of Chemistry, Romanian Academy, 24 Mihai Viteazul Blvd., RO-300223 Timişoara, ROMANIA

animal health and biocides to invertebrate (e.g. insects such as aphids, whiteflies, nematodes, parasites) pest control since their first appearance in the mid-1990s (Casida, 2018). Due to their systemic nature, neonicotinoids are rapidly absorbed through the leaves or roots of the rising plant and subsequently transported through its tissues. This property confers them many advantages in pest control. Plant protection and controlling the action of pests are very important clues in order to improve the quality and quantity of the products (Elbert et al., 2008). Several studies revealed that minor structural changes influence neonicotinoids selectivity between insects and mammals (Casida and Quistad, 2004).

A large number of QSAR models which manage the acute toxicity of pesticides have been reported (Basant and Gupta, 2017; Hamadache et al., 2018; Wang, et al. 2017; Hamadache et al., 2016). Just limited scientific papers have been applied to neonicotinoids, which were tested against a broad array of insect species (Cronin et al., 2003; Shao et al., 2009; Shao et al., 2010; Xu et al., 2014; Lei et al., 2018). Neonicotinoids act on the insect nicotinic acetylcholine receptor (nAChR). Their development is provoked by the rapid development of resistance. Synthetic nicotinoids (e.g. imidacloprid) were successfully used with higher efficiency and safety to avoid this inconvenience (Casida and Quistad, 1998).

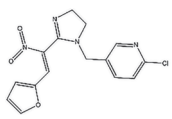
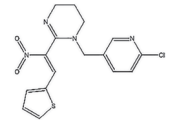
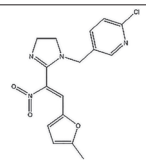
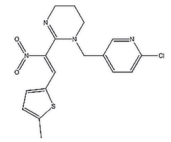
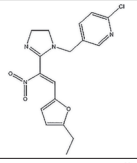
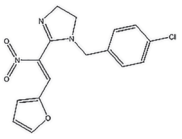
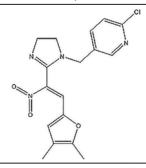
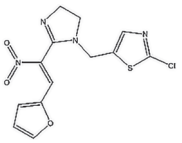
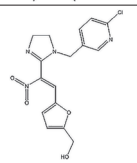
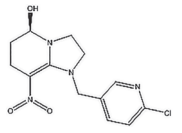
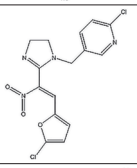
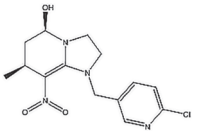
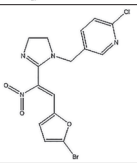
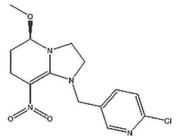
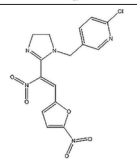
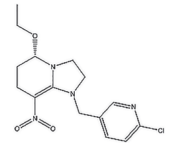
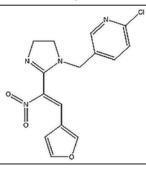
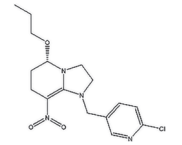
This paper presents the application of the partial least squares (PLS) approach to a series of 27 neonicotinoid derivatives bearing nitromethylene and nitroconjugated moieties (Tian et al., 2007; Shao et al., 2009). The insecticide structures were modeled using molecular mechanics calculations. Structural descriptors were calculated from the minimum energy structures and were related to the $pLC_{50}$ values using the partial lease squares (PLS) method. The best model was internally and externally validated using several criteria. Structural information of neonicotinoids which influence the cowpea aphid toxicity is obtained.

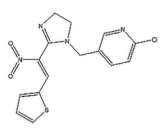
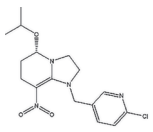
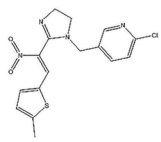
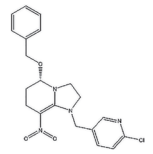
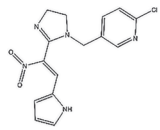
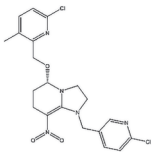
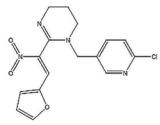
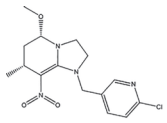
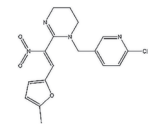
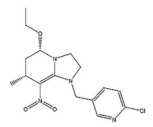
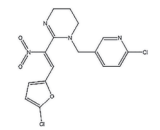
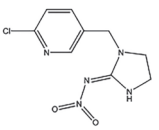
## 2. Methods

### Molecular structures and insecticidal activity

The insecticidal activity against cowpea aphids (*Aphis craccivora*) of a series of 27 neonicotinoids (Tian et al., 2007; Shao et al., 2009), expressed as dependent variable ($pLC_{50}$ values), is presented in Table 1.

**Table 1**. Structural neonicotinoid data, the experimental ($pLC_{50}$) and calculated ($pLC_{50}$ pred) insecticidal activity values obtained using the best PLS model

| No | Structure | $pLC_{50exp}$ | $pLC_{50pred}$ | No | Structure | $pLC_{50exp}$ | $pLC_{50pred}$ |
|---|---|---|---|---|---|---|---|
| 1 |  | 5.43 | 5.19 | 16* |  | 4.69 | 4.39 |
| 2 |  | 5.20 | 5.29 | 17* |  | 4.61 | 4.37 |
| 3 |  | 5.74 | 5.29 | 18** |  | 3.63 | |
| 4 |  | 5.33 | 5.25 | 19 |  | 5.46 | 5.86 |
| 5 * |  | 4.98 | 5.26 | 20 |  | 3.73 | 3.82 |
| 6 * |  | 5.12 | 5.29 | 21 |  | 4.01 | 3.70 |
| 7 |  | 5.14 | 5.21 | 22 |  | 3.88 | 3.81 |
| 8 |  | 4.96 | 5.22 | 23 |  | 4.02 | 3.75 |
| 9 |  | 5.35 | 5.18 | 24* |  | 3.98 | 3.79 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 10 |  | 5.37 | 5.16 | 25 |  | 3.59 | 3.75 |
| 11 |  | 5.51 | 5.31 | 26 |  | 3.24 | 3.79 |
| 12* |  | 4.95 | 5.14 | 27* |  | 2.94 | 3.69 |
| 13 |  | 4.12 | 4.31 | 28 |  | 3.83 | 3.67 |
| 14** |  | 3.16 | | 29 |  | 3.73 | 3.69 |
| 15 |  | 4.22 | 4.40 | 30** |  | 4.46 | |

*Test compounds
**Excluded compounds (Funar-Timofei and Bora, 2017).

The neonicotinoid structures were previously built and pre-optimized (Funar-Timofei and Bora, 2017) using the conformer plugin (with MMFF94 as molecular mechanics force field) inside the MarvinSketch (MarvinSketch 15.2.16.0, ChemAxon Ltd. http://chemaxon.com) package. The lowest energy conformers were refined further using the semiempirical PM 7 Hamiltonian (Stewart, 2013) implemented in MOPAC 2016 program (MOPAC2016, James J. P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, HTTP://OpenMOPAC.net (2016)).

Structural descriptors of the neonicotinoid derivatives were compute 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.

## The Partial Least Squares (PLS) approach

In the projections to latent structures (PLS) approach the projections of independent factors (the block of explanatory variables) are related to the dependent (block of)

responses by a linear relation (Wold, 1985). In the PLS approach stable, correct and highly predictive models are obtained, even for correlated descriptors (Höskuldsson, 1988). In this paper the SIMCA package (SIMCA P+12.0.0.0, May 20 2008, Umetrics, Sweeden, http://www.umetrics.com/) was used for PLS calculations.

As most important fitting parameters of the PLS model the following ones were used: the squared correlation regression coefficient ($R^2$), and the squared cross-validated correlation coefficient ($Q^2$). The Variables Importance in the Projection (VIP) values and the sign of the variable coefficients are very important in the elucidation of the activity mechanism. The leave-7-out cross-validation procedure was employed to select the significant principal components.

The robustness and overfitting of the PLS model were checked using the Y-randomization test, in which the insecticidal activity was randomly shuffled (999 randomizations). Minimal $R^2$ and $Q^2$ values are required for a good PLS model (Roy et al., 2009).

**Model validation**

Several criteria to test the model predictive power were used: $Q^2_{F1}$(Shi et al., 2001); $Q^2_{F2}$(Schüürmann et al., 2008); $Q^2_{F3}$(Consonni et al., 2009) and the concordance correlation coefficient (CCC) (Chirico and Gramatica, 2011), with the thresholds values higher than 0.85, as they have been rigorously determined by a simulation study (Chirico and Gramatica, 2012). The data set was splitted into training and validation sets. In the test set following compounds were randomly selected: 5, 6, 16, 17, 24, 27.

Other predictivity parameters were calculated: $r^2_m$(Roy and Mitra, 2012), with the lowest threshold value of 0.5 used to be accepted. The model applicability and overfitting were checked using the root-mean-square errors (RMSE) and the mean absolute error (MAE) of the training and validation sets.

## 3. Results and discussion

The matrix of the calculated independent variables was first studied using the principal component analysis for the series of 30 compounds and 1529 variables. A significant four principal component model was achieved (the first three components already explained 61.3% of the information content of the descriptor matrix). Compounds 14, 18 and 30 were excluded from the dataset, as outliers (Funar-Timofei and Bora, 2017).

For the entire set of 30 neonicotinoids following statistical results were obtained by the PLS model: $R^2_X$CUM) = 0.561 $R^2_Y$(CUM) = 0.859 and $Q^2$(CUM) = 0.55 ($R^2_{X(CUM)}$ and $R^2_{Y(CUM)}$ are the cumulative sum of squares of all the X and Y values), obtained for three

principal components, which demonstrated the model overfit. This problem was surpassed by excluding the noise variables from this model (e.g. coefficient values insignificantly different from 0).

PLS calculations were applied to the training set. Good statistical fitting results are obtained for the PLS model developed for the training set, having two latent variables (N= 21 and X = 15), which explain 96.8% of the information content of the descriptor matrix, and $R^2_Y$(CUM) = 0.9 and $Q^2$(CUM) = 0.878. The descriptor coefficients and the VIP values of the final PLS model are included in Table 2.

**Table 2.** The coefficients in descending order of VIP values for the two principal components of the final PLS model*.

| No | Variable ID* | CoefCS[2] | VIP[2] |
|---|---|---|---|
| 3 | SRW05 | 0.114 | 1.062 |
| 12 | stereoDoubleBondCount | 0.138 | 1.056 |
| 13 | Mor03u | 0.138 | 1.056 |
| 14 | Mor03e | 0.154 | 1.048 |
| 1 | D/Dr05 | 0.032 | 0.983 |
| 2 | nR05 | 0.032 | 0.983 |
| 4 | SRW07 | 0.032 | 0.983 |
| 8 | N-074 | 0.032 | 0.983 |
| 9 | nCconj | 0.032 | 0.983 |
| 10 | nN=C-N< | 0.032 | 0.983 |
| 11 | nR=Cs | 0.032 | 0.983 |
| 15 | MSD | 0.032 | 0.983 |
| 6 | C-040 | 0.065 | 0.977 |
| 5 | C-016 | 0.062 | 0.973 |
| 7 | doubleBondStereoisomerCount | 0.1 | 0.952 |

* D/Dr05—distance/detour ring index of order 5 (topological descriptor), nR05—number of 5-membered rings (constitutional descriptor), SRW05—self-returning walk count of order 05 (molecular walk counts descriptor), SRW07—self-returning walk count of order 07 (molecular walk counts descriptor), C-016— =CHR (atom-centred fragments descriptor), C-040—R-C (=X) -X / R-C#X / X=C=X (atom-centred fragments descriptor), doubleBondStereoisomerCount- the number of generated double bond stereoisomers, N-074—R#N / R=N- (atom-centred fragments descriptor), nCconj- number of non-aromatic conjugated C (sp 2) (functional group counts descriptor), nN=C-N<- number of amidine fragments (functional group counts descriptor), nR=Cs—number of aliphatic secondary C (sp 2) (functional group counts descriptor), stereoDoubleBondCount- the number of double bonds with defined stereochemistry, Mor03u- 3D-MoRSE—signal 03 / unweighted (3D-MoRSE descriptor), Mor03e- 3D-MoRSE—signal 13 / weighted by atomic Sanderson electronegativities (3D-MoRSE descriptor), MSD- mean square distance index (Balaban) (topological descriptor); # represents a triple bond; R represents any group linked through carbon; X represents any electronegative atom (O, N, S, P, Se, halogens).

The selected variables in the best PLS model (Table 2) had VIP values greater than 1, being considered to be the most relevant for the model. Coefficient plots and VIP of the final PLS model are presented in Figures 1 and 2, respectively.

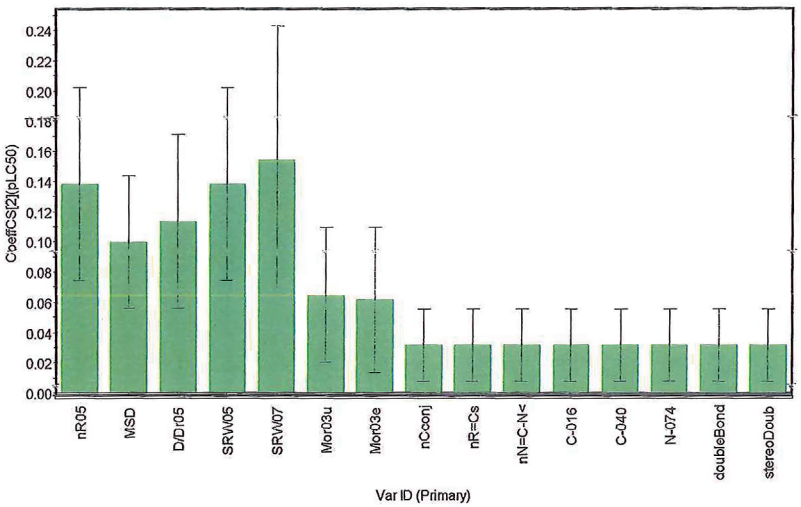



Fig. 1 Coefficient plot of the best PLS model.

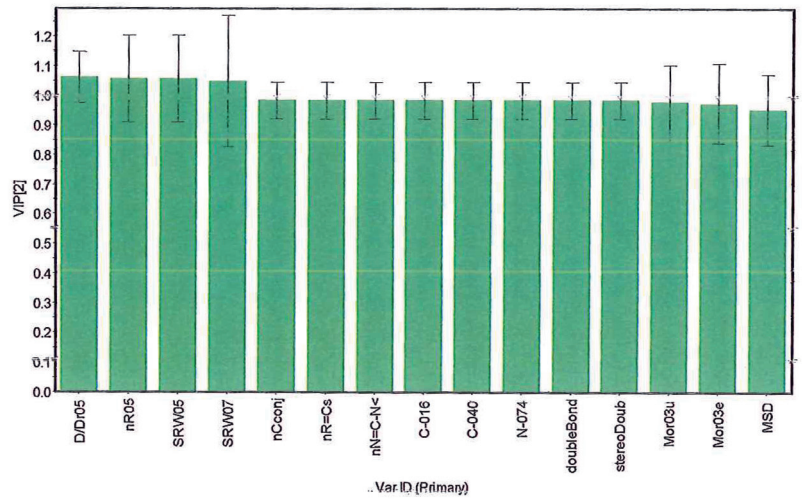



Fig. 2. VIP plot for the final PLS model.

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

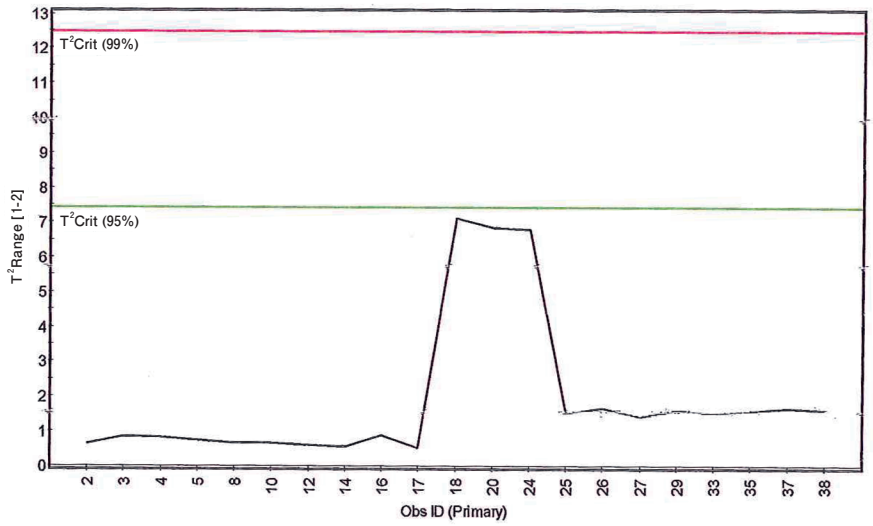



Fig. 3. The Hotteling's $T^2$ range plot of the final PLS model.

The stability of the final PLS model is confirmed by the internal validation parameters (tr-for the training and CV-crossvalidation): $CCC_{tr}$ = 0.951, $CCC_{CV}$ = 0.931, $RMSE_{tr}$ = 0.246, $RMSE_{CV}$ = 0.282, $MAE_{tr}$ = 0.210, $MAE_{CV}$ = 0.237. In addition, the Y-randomization procedure confirmed the model stability. The intercept values close to zero of the regression lines obtained by the correlation between the calculated $R^2$, respectively $Q^2$ values of the original Y-variable and the shuffled Y-variable were of 0.0158 for the $R^2Y$ line and -0.262 for the $Q^2_Y$ line, respectively (see Figure 4).

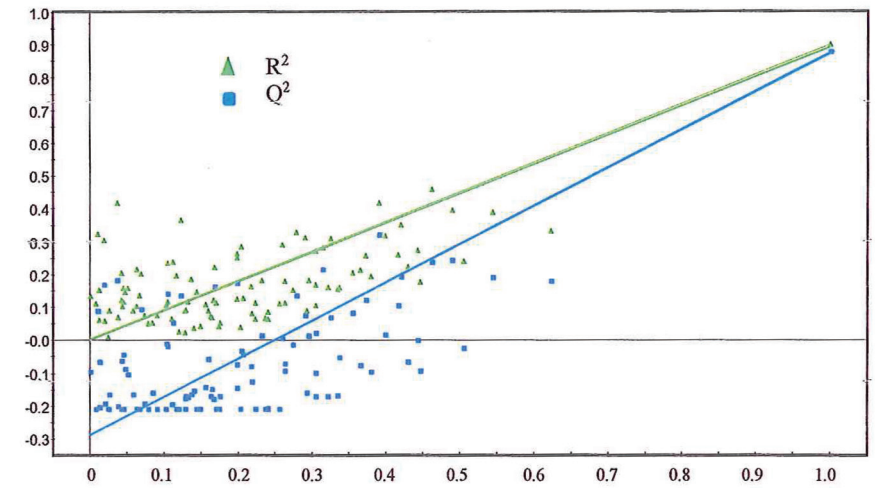



Fig. 4. Y-scrambling plot of the final PLS model.

The predictive power of the final PLS model was confirmed by the following external validation parameters, calculated for the test set: $CCC_{ext}$ = 0.849, $RMSE_{ext}$ = 0.379, $MAE_{ext}$ = 0.323, $r^2_m$= 0.728, $Q^2_{F1}$ = 0.759, $Q^2_{F2}$ = 0.738, $Q^2_{F3}$ = 0.762.

## 4. Conclusion

The PLS approach was applied to a series of 30 neonicotinoid insecticides which include nitromethylene and nitroconjugated structural fragments, active against the cowpea aphids (*Aphis craccivora*). Structural descriptors calculated previously from the minimum energy structures optimized using the PM 7 semiempirical Hamiltonian were related to the insecticide activity. Good statistical results and predictive power model were obtained. Higher number of 5-membered rings, of double bonds with defined stereochemistry, of amidine fragments, of non-aromatic conjugated C(sp 2) are favorable for an increased insecticidal activity. The PLS model can be used for the design of new insecticides from this class of neonicotinoids, with improved activity.

## Acknowledgements

This project was financially supported by Project 1.1 of the Coriolan Dragulescu Institute of Chemistry of the Romanian Academy. Access to the MOPAC2016 and Chemaxon Ltd. software are greatly acknowledged by the authors.

## References

Basant N., Gupta, S.: QSAR modeling for predicting mutagenic toxicity of diverse chemicals
for regulatory purposes. *Environ Sci Pollut Res* **24**, 14430–14444 (2017).

Bonmatin, J.M., Giorio, C., Girolami, V., Goulson, D., Kreutzweiser, D.P., Krupke, C., Liess, M., Long, E., Marzaro, M., Mitchell, E.A.D., Noome, D.A., Simon-Delso, N., Tapparo, A.: Environmental fate and exposure; neonicotinoids and fipronil. *Environ. Sci. Pollut. Res.* 22, 35–67 (2015).

Casida, J.E., Quistad, G.B.: Golden age of insecticide research: past, present or future? *Annu. Rev. Entomol.* **43**, 1–16 (1998).

Casida, J.E., Quistad, G.B.: Why insecticides are more toxic to insects than people: the unique toxicology of insects. *J. Pestic. Sci.* **29**, 81–86 (2004).

Casida, J.E., Durkin, K.A.: Neuroactive insecticides: targets, selectivity, resistance, and

secondary effects, *Annu. Rev. Entomol.* **58**, 99–117 (2013).

Casida, J.E.: Neonicotinoids and other insect nicotinic receptor competitive modulators: progress and prospects, *Annu. Rev. Entomol.* **63**, 125–144 (2018).

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.* **51**, 2320–2335 (2011).

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.* **52**, 2044–2058 (2012)

Cronin, M.T.D., Walker, J.D., Jaworska, J.S, Comber, M.H.I., Watts, C.D., Worth, A.P.;

Use of QSARs in international decision-making frameworks to predict ecologic effects and environmental fate of chemical substances. *Environ Health Perspect,* **111**, 1376–1390 (2003).

Consonni, V., Ballabio, D., Todeschini, R.: Comments on the definition of the Q2 parameter for QSAR validation, *J. Chem. Inf. Model.* **49**, 1669–1678 (2009).

Elbert, A., Haas,M., Springer, B., Thielert, W., Nauen, R.: Applied aspects of neonicotinoid uses in crop protection. *Pest. Manag. Sci.* **64**, 1099–1105 (2008).

Funar-Timofei S., Bora A., QSAR Study of Neonicotinoid Insecticidal Activity Against Cowpea Aphids by the MLR Approach, *Proceedings of the The 21st International Electronic Conference on Synthetic Organic Chemistry*, 1–30 November 2017; Sciforum Electronic Conference Series, Vol. 21, **2017**; doi:10.3390/ecsoc-21-04727

Hamadache, M., Benkortbi, O., Hanini, S., Amrane, A., Khaouane, L., Si Moussa, C.: A quantitative structure activity relationship for acute oral toxicity of pesticides on rats: validation, domain of application and prediction. *J Hazard Mater* **303**, 28–40 (2016).

Hamadache, M., Benkortbi, O., Hanini, S., Amrane, A.: QSAR modeling in ecotoxicological risk assessment: application to the prediction of acute contact toxicity of pesticides on bees (*Apis mellifera L*). *Environ Sci Pollut Res*, **25**, 896–907 (2018).

Höskuldsson, A.: PLS regression methods, *J. Chemometrics* **2**, 211–228 (1988).

Jeschke, P., Nauen, R.: Neonicotinoids - from zero to hero insecticide chemistry. *Pest Manag. Sci.* **64**, 1084–1098 (2008).

Jeschke, P., Nauen, R., Schindler, M., Elbert, A.: Overview of the status and global strategy for neonicotinoids, *J. Agric. Food Chem.* **59**, 2897–2908 (2011).

Lei, C., Geng, L., Xu, X., Shao, X., Li, Z.: Isoxazole-containing neonicotinoids: Design, synthesis, and insecticidal evaluation, *Bioorg Med Chem Lett.* **28**, 831–833(2018).

Roy, P.P, Paul, S., Mitra, I., Roy, K.: On two novel parameters for validation of predictive QSAR models. *Molecules* **14**, 1660–1701 (2009).

Roy, K., Mitra, I.: On the use of the metric $r^2_m$ as an effective tool for validation of QSAR models in computational drug design and predictive toxicology. *Mini-Rev. Med. Chem.* **12**, 491–504 (2012).

Shao, X., Li, Z., Qian, X., Xu, X.: Design, synthesis, and insecticidal activities of novel analogues of neonicotinoids: Replacement of nitromethylene with nitroconjugated system, *J Agric Food Chem*., **57**, 951–957 (2009)

Shao, X., Fu, H., Xu, X., Xu, X., Liu, Z., Li, Z., Qian, X.: Divalent and oxabridged neonicotinoids constructed by dialdehydes and nitromethylene analogues of imidacloprid: design, synthesis, crystal structure, and insecticidal activities. *J Agric Food Chem.*, **58**, 2696–702 (2010)

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.* **41**, 186–195 (2001).

Schüürmann, G., Ebert, RU., 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.* **48**, 2140–2145 (2008).

Stewart, J.J.P.:Optimization of parameters for semiempirical methods VI: More modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* **19**, 1–32 (2013).

Tian, Z., Shao, X., Li, Z., Qian, X., Huang, Q.: Synthesis, insecticidal activity, and QSAR of novel nitromethylene neonicotinoids with tetrahydropyridine fixed cis configuration and exo-ring ether modification, *J. Agric. Food Chem*. **55**, 2288–2292 (2007). This is the correct abbreviation of the authors.

Wang, X., Chu, Z., Yang, J., Li, Y.: Pentachlorophenol molecule design with lower bioconcentration through 3D-QSAR associated with molecule docking. *Environ Sci Pollut Res*, **24**, 25114–25125 (2017);

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

Xu, R., Xia, R., Luo, M., Xu, X., Cheng, J., Shao, X., Li, Z.: Design, synthesis, crystal structures, and insecticidal activities of eight-membered azabridge neonicotinoid analogues. *J Agric Food Chem*., **62**, 381–90 (2014).

# ニトロメチレンとニトロ基を包含した構造を有するネオニコチノイド系殺虫剤のPLSモデリング

**鈴 木 孝 弘, Alina Bora, Luminita Crisan, Ana Borota, Simona Funar-Timofei**

ネオニコチノイド系農薬は、昆虫のニコチン酸アセチルコリン受容体に結合し、神経を興奮状態にして方向感覚を狂わせ、筋肉を収縮させて死に至らしめる。それらの分子標的サイトは、昆虫のニコチン性アセチルコリン受容体である。27種類のニトロメチレンとニトロ基を包含した構造を有するネオニコチノイド系農薬のササゲアブラムシ（*Aphis craccivora*）に対する殺虫活性の定量的構造活性相関（QSAR）をPLSモデルにより検討した。構造記述子は、半経験的量子化学的手法PM 7 を使用してエネルギー最適化殺虫剤構造から計算され、pLC50値との相関を検討した。得られたPLSモデルによる殺虫活性の計算値と実測値は良好な相関を示し（$R^2X$(Cum) = 0.968, $R^2Y$(cum) = 0.9 and $Q^2$(Cum) = 0.878)、また十分な予測機能を有することが確認できた（$CCC_{tr}$ = 0.9506, $CCC_{ext}$ = 0.849, $CCC_{CV}$ = 0.931, $R^2_{ext}$ = 0.738, $Q^2_{F1}$ = 0.759, $Q^2_{F2}$ = 0.738, $Q^2_{F3}$ = 0.763).