PLS Evaluation of the Insecticidal Activity of Phenylazo, Pyrrole- and Dihydropyrrole-Fused Neonicotinoids

Takahiro Suzuki[1] ・ Alina Bora[2] ・ Simona Funar-Timofei[2]

Abstract

In the present study the insecticidal activity of a series of 24 phenylazo, pyrrole- and dihydropyrrole-fused neonicotinoids was investigated using the partial least squares (PLS) approach. pLC_{50} values were used to express the activity against the cowpea aphid (*Aphis craccivora*). The insecticide structures were modeled using the MMFF94s force field included in the Omega software. Minimum energy conformers were used to calculate structural parameters, using the Dragon software. A PLS model with two components was obtained with acceptable statistical quality (R^2X(cum) = 0.830, R^2Y(cum) = 0.791 and Q^2(cum) = 0.749). Several internal and external criteria were checked to validate the model. The built PLS model is robust, with predictive power and can be used for the design of new insecticides, active against the cowpea aphids.

Keywords：Cowpea aphids, Neonicotinoids, PLS, Validation

1. Introduction

Since the introduction of imidacloprid in 1991 by Shinzo Kagabu (Kagabu, 2011), neonicotinoids became an important new class of insecticides, which act on the insect nicotinic acetylcholine receptor (nAChR). They are efficient since 50 years, being cheap and persistent, features which give them a long use in the future (Casida and Quistad, 1998; Tomlin, 2009; Casida and Quistad, 2004). The neonicotinoids success is, however, being provoked by the rapid development of resistance (Nauen and Denholm, 2005) and severe bee toxicity (Henry et al., 2012; Cameron et al., 2011; Chen et al., 2015).

Cross resistance problems, which appeared especially to the first generation of neuroactive insecticides, encouraged the search for new pesticides acting at different sites (Casida and Quistad, 1998). Synthetic nicotinoids (e.g. imidacloprid), active at the

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

nicotinic acetylcholine receptor (nAChR), were successfully employed with higher efficiency and safety to circumvent this inconvenience.

There are essential differences between the nAChR of insects and mammals, which give outstanding toxic selectivity to the neonicotinoids (Elbert et al., 2008)). The differential selectivity between insects and mammals is conferred by only minor structural changes (e.g. 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) (Casida and Quistad, 2004). In addition, mammals have a second related enzyme, butyrylcholinesterase, abundant in blood plasma, which reacts with organosphophate and methylcarbamates toxic compounds and offers a protection not available to insects. In mammals, nicotine is protonated at physiological pH and suffers π-cation interaction with choline subsite of acethylcolinesterase nerve target for insect at a nAChR subsite. Binding subsite specificity plays a major role in selective toxicity for electronegative neonicotinoids in insects and protonated nicotinoids in mammals.

A quantitative structure-insecticidal acivity was studied for 23 cyanoimine neonicotinoids active against *American cockroaches* (Kagabu et al., 2008). Variants of the key pharmacophore were constructed with the insecticide central ring conjugated to electron-withdrawing groups (e.g. NCN, $CHNO_2$, or NNO_2) and used to relate the insecticide structure to its neuroblocking potency. Using multiple linear regression it was found that the Mulliken charge on the nitro oxygen atom or cyano nitrogen atom influenced the insecticidal activity.

A pharmacophore model for neonicotinoid agonists of the nicotinic acetylcholine receptors was developed (Li et al., 2008). It included a hydrogen bonding acceptor, a hydrogen-bond donor, a hydrophobic aliphatic and a hydrophobic aromatic centre.

Comparative Molecular Field Analysis (CoMFA) was used to study the binding activity of chloronicotinyl insecticides to the nicotinic acetylcholine receptors (nAChR) of house flies (Okazawa et al., 2000). Steric interactions were found to be more significant for the acyclic than cyclic compounds and that the electrostatic property of the acyclic amino- and cyclic imdazolidine-moieties influenced their insecticidal activity.

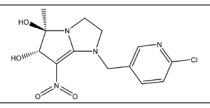
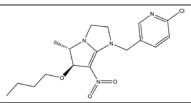
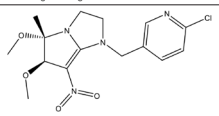
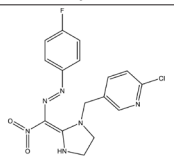
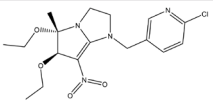
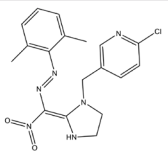
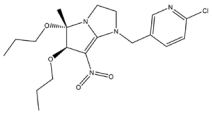
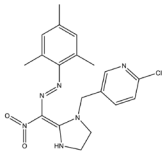
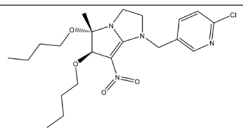
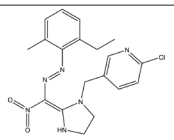
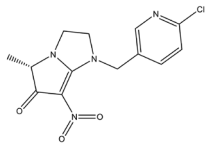
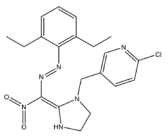
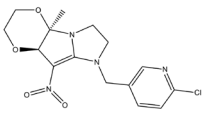
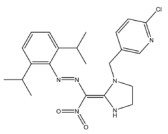
Nitromethylene compounds with insecticidal activity against house flies were studied using CoMFA. The nitroimino moiety and a portion of the imidazolidine ring were mainly surrounded by a sterically and electrostatically sensitive region of nAChR (Okazawa et al., 1998). Similar results were reported previously (Nakayama and Sukekawa, 1998). It was concluded that molecular similarity, expressed by electrostatic properties influenced the insecticidal activity, as well as superimposability, expressed by molecular shape.

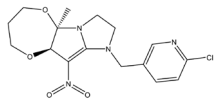
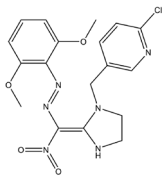
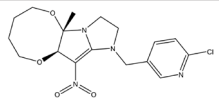
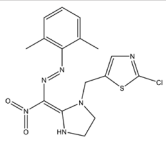
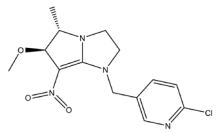
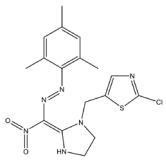
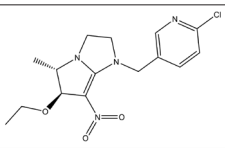
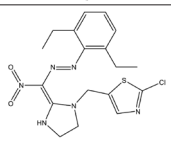
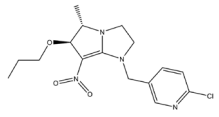
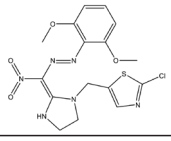
Favorable influence of steric and H-bond acceptor over the insecticidal activity was obtained using CoMFA and comparative molecular similarity indices analysis (CoMSIA) methodology applied to imidacloprid analogues (Sun et al., 2006).

A preliminary computer-aided molecular docking study was performed to investigate the interaction between receptor and phenylazoneonicotinoids (Ye et al., 2011). π - π stacking interaction and hydrogen bond were found between the studied neonicotinoids and the crystal structures of acetylcholine receptor binding protein-imidacloprid.

The purpose of this paper is to determine the structural features of a series of 24 neonicotinoids (Ye et al., 2011; Ye et al., 2013) (Table 1), which influence the cowpea aphids. The quantitative relationship between chemical features and the insecticide activity was determined by means of the partial least squares (PLS) approach.

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	5.23	13		3.97	3.98
2		5.70	5.53	14*		4.43	4.91
3		5.80	5.36	15		5.37	5.61
4		5.71	5.20	16		5.30	5.45
5		5.11	5.11	17*		5.43	5.59
6		3.85	4.12	18		5.55	5.61
7		4.55	4.62	19		4.86	5.33

8		4.52	4.68	20		5.00	5.21
9		4.41	4.61	21		5.46	5.04
10		4.35	4.01	22*		4.82	4.94
11*		3.96	4.00	23		4.93	4.98
12*		4.16	4.04	24		4.83	4.72

*Test compounds

2. Methods

Definition of target property and molecular structures

A dataset of 24 neonicotinoid derivatives (Table 1) with insecticidal activity (LC_{50}, in mmol/L, where LC_{50} represents the median lethal concentration of the chemical in air that kills 50% of the test animals during the observation period) against cowpea aphids was taken from the literature (Ye et al., 2011; Ye et al., 2013). The insecticidal activity against cowpea aphids (*Aphis craccivora*) activity data, expressed as pLC_{50} values was employed as dependent variable.

The neonicotinoid structures were built using the MarvinSketch (MarvinSketch 15.2.16.0, ChemAxon Ltd. http://chemaxon.com) package and were pre-optimized using the (MMFF94) molecular mechanics force field inside the Omega (Omega v.2.5.1.4, OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com) software (Hawkins et al., 2010; Hawkins and Nicholls, 2012). For conformer generation, the default parameters were employed excepting the maximum number of conformers per

compound set to 400. To avoid the presence of similarly shaped structures, any conformer having an RMSD fit outside 0.5 Å to another conformer was excluded.
Then, 22 types of structural 0D, 1D, 2D and 3D molecular descriptors were calculated for the lowest energy structures 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) method

Projections to latent structures (PLS) represent a regression technique for modeling the relationship between projections of dependent factors and independent responses. In this approach data analysis features link a block (or a column) of response variables to a block of explanatory variables (Wold, 1985). The relationship between the dependent and independent variables is described as a latent variable method (Eriksson et al, 2004). The PLS approach leads to stable, correct and highly predictive models even for correlated descriptors (Höskuldsson, 1998). PLS calculations were performed by the SIMCA package (SIMCA P+12.0.0.0, May 20 2008, Umetrics, Sweeden, http://www.umetrics.com/). The QSAR matrix (including the dependent and independent matrices) was analyzed in a first step by the principal component analysis (PCA) (Daszykowski et al., 2007), and subsequently by the partial least squares (PLS) approaches. The squared correlation regression coefficient R^2, and the squared cross-validated correlation coefficient, Q^2, are the most important statistical parameters that provide a measure of the quality and validity for the final PLS model, while the Variables Importance in the Projection (VIP) values and the sign of the variables' coefficients are more relevant in explaining the activity mechanism. The significant principal components were selected by 7 cross-validation groups.

The Y-randomization test is a widely used technique that indicates the robustness of a QSAR model and overfitting. The dependent variable (biological activity) is randomly shuffled and a QSAR model is built using the same descriptor matrix. The obtained PLS models (after 999 randomizations) must have the minimal r^2 and q^2 values (Pratim Roy et al., 2009).

Model validity

The dataset was divided into a random training set of 66.7% and a test set (no. 11, 12, 14, 17, 22) of 33.3% of the total number of compounds. Several criteria for model predictivity 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) (having the thresholds values higher than 0.85, as they have been rigorously determined by a simulation study (Chirico and Gramatica, 2012).

In addition, the predictive parameter r^2_m (Roy and Mitra, 2012) was used (a lowest

threshold value of 0.5 was used to be accepted).

The model robustness and overfitting were checked using the Y-randomization test. 999 randomizations were employed to develop PLS models with minimal r^2 and q^2 values (Eriksson et al., 2001).

The root-mean-square errors (RMSE) and the mean absolute error (MAE) of the training and validation sets were compared to check the model applicability and overfitting (Goodarzi et al., 2009).

3. Results and discussion

A statistical analysis of the phenylazo, pyrrole- and dihydropyrrole-fused neonicotinoids analogues was performed using the calculated variables. A PCA model was built for the whole X matrix (including N=24 compounds and X = 1478 descriptors). From the total of 5 significant principal components resulted from this analysis, the first three components already explained 69.1% of the information content of the descriptor matrix. A PLS model was built for the entire set of 24 compounds. The statistical results of the PLS model: R^2_X(cum) = 0.709 R^2_Y(cum) = 0.933 and Q^2(cum) = 0.597 obtained for four 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). This inconvenience was overstepped by excluding the noise variables from this model (e.g. coefficient values insignificantly different from 0). Thus, a robust model, M2 (N= 24 and X = 6) with two latent variables, which explains 83% of the information content of the descriptor matrix, with R^2_Y(cum) = 0.791 and Q^2(cum) = 0.749 was obtained. Table 2 presents the descriptor coefficients and the VIP values included in the final PLS model.

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]
1	EEig03r	0.239	1.164
2	JGI2	0.283	1.228
3	Mor11e	0.133	0.869
4	Mor11u	0.106	0.796
5	Mor13e	0.280	0.997
6	Rww	0.245	0.868

* Rww-reciprocal hyper-detour index (topological descriptor), EEig03r-Eigenvalue 03 from edge adj. matrix weighted by resonance integrals (edge adjacency index); JGI2-mean topological charge index of order 2 (topological charge index); Mor11u-3D-MoRSE - signal 11 / unweighted (3D-MoRSE descriptor); Mor11e-3D-MoRSE - signal 11 / weighted by atomic Sanderson electronegativities (3D-MoRSE descriptor); Mor13e-3D-MoRSE - signal 13 / weighted by atomic Sanderson electronegativities (3D-MoRSE descriptor)

The noise variables (with coefficient values insignificantly different from 0) were removed and the best resulted PLS model is robust and includes two principal components and 6 significant variables (Table 2). The coefficient and VIP plots are presented in Figure 1 and Figure 2, respectively.

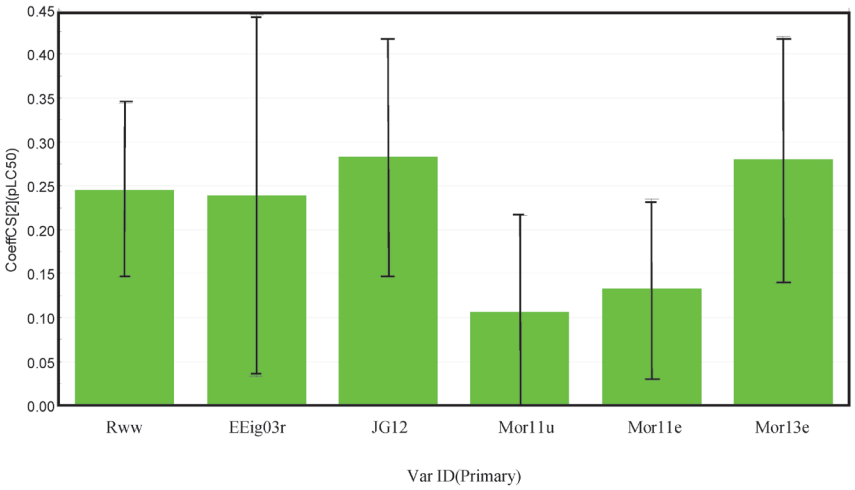


Fig. 1 Coefficient plot of the final PLS model.

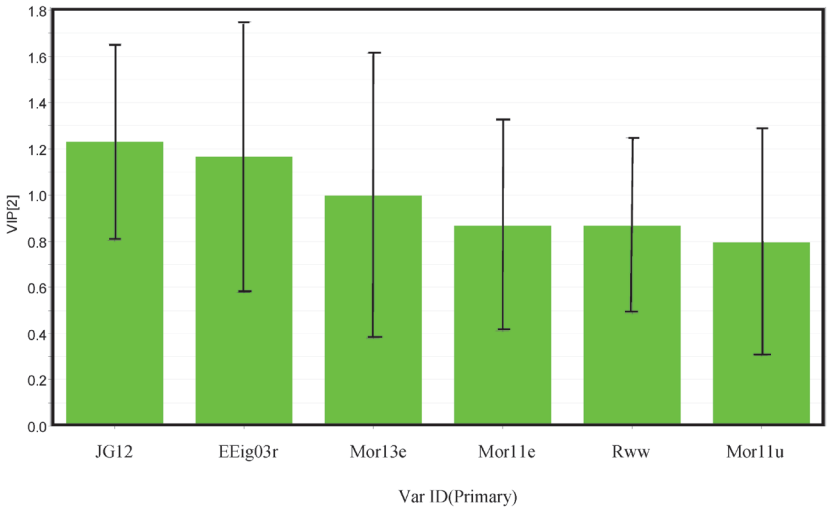


Fig. 2. VIP plot for the final PLS model.

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

The internal validation parameters (tr-for the training and CV-crossvalidation): CCC_{tr} =

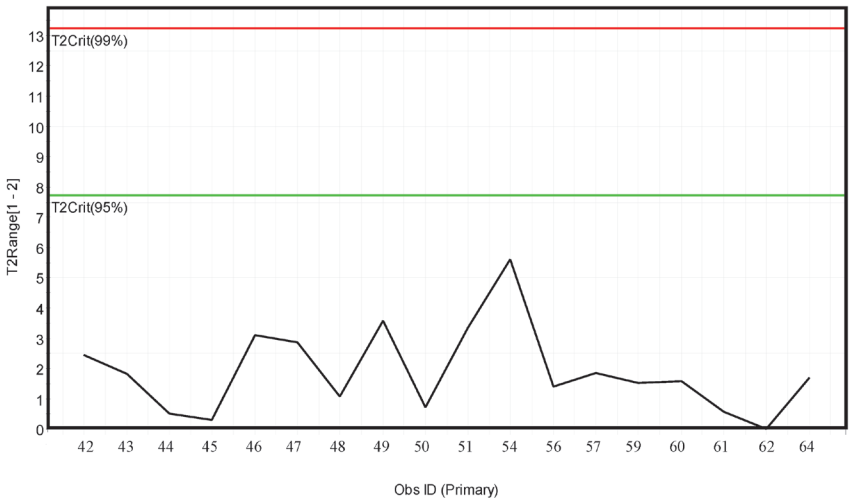


Fig. 3. The Hotteling's T 2 range plot of the final PLS model.

0.884, CCC_{CV} = 0.848, $RMSE_{tr}$ = 0.265, $RMSE_{CV}$ = 0.292, MAE_{tr} = 0.210, MAE_{CV} = 0.241 confirm the stability of the PLS model.

All selected variables in the final PLS model (Table 2) had VIP values greater than 1 and were considered to be the most relevant for the model. The Y-randomization procedure was applied using the SIMCA-P+ 12.0 software (for the final PLS model). It gave the following intercept (PLS) values 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, respectively: 0.146 for the R^2_Y line and -0.325 for the Q^2_Y line.

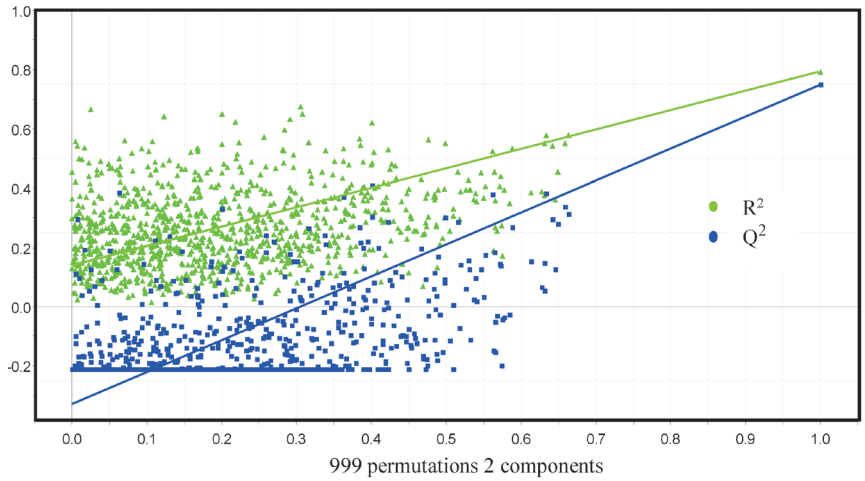


Fig. 4. Y-scrambling plot of the final PLS model.

The slope values close to zero indicate a stable model (see Figure 4).

The external validation parameters calculated for the test set: CCC_{ext} = 0.909, $RMSE_{ext}$ = 0.223, MAE_{ext} = 0.172, r^2_m = 0.718, Q^2_{F1} = 0.869, Q^2_{F2} = 0.791, Q^2_{F3} = 0.853 indicate the final PLS model has predictive power.

4. Conclusion

Quantitative structure activity relationships of a series of 24 neonicotinoid insecticides against the cowpea aphids (*Aphis craccivora*) was investigated using the partial least squares approach. Structure optimization modeling using the MMFF94s force field was used. Structural descriptors derived from the minimum energy structures were related to the insecticide activity. PLS models with good statistical results and predictive power were obtained. Higher topological, edge adjacency indices, 3 D-MoRSE descriptor values increase the insecticidal potency. The PLS model can be used for the prediction of new insecticides with high activity.

Acknowledgements

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

References

Cameron, S.A., Lozier, J.D., Strange, J.P. et al., Patterns of widespread decline in North American bumble bees. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 662–667 (2011).

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).

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

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)

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).

Daszykowski, M., Kaczmarek, K., Heyden, V., Walczak, B.: Robust statistics in data analysis-A review: Basic concepts. *Chemom. Intell .Lab. Syst.* **85**, 203-219(2007).

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).

Eriksson, L., Johansson, E., Kettaneh-Wold, N.,Wold, S.: Multi and megavariate data analysis:principles and applications, Umetrics AB, Umea, Sweden, pp.92-97; pp. 489-491(2001).

Eriksson, L., Gottfries, J., Johansson, E., Wold,, S.; Chemom. Time-resolved QSAR: an approach to PLS modelling of three-way biological data. *Intel. Lab. Syst.* **73**,73-84(2004).

Goodarzi, M., Deshpande, S., Murugesan, V., Katti, S.B., Prabhakar, Y.S.: Is feature selection essential for ANN modeling?, *QSAR Comb. Sci*., **28**, 1487-1499(2009).

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.* **50**, 572–584(2010).

Hawkins, P.C,D., Nicholls, A.:Conformer generation with OMEGA: learning from the data set and the analysis of failures, *J. Chem. Inf. Model.* **52**, 2919-2936(2012).

Henry, M., Béguin, M., Requier, F. *et al*., A common pesticide decreases foraging success and survival in honey bees. *Science* **336**, 348–350 (2012).

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

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.* **33**, 58–66 (2008).

Kagabu, S.: Discovery of Imidacloprid and Further Developments from Strategic Molecular Designs, *J. Agric. Food Chem.* **59**, 2887–2896(2011).

Keller, H.R., Massart, D.L., Brans, J.P.: Multicriteria decision making: a case study. *Chemom. Int. Lab. Syst.* **11**,175-189(1991).

Li, J., Ju, X.L., Jiang, F.C.:Pharmacophore model for neonicotinoid insecticides, *Chinese Chemical Letters* **19**, 619–622(2008).

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

Nauen, R., Denholm, I.: Resistance of insect pests to neonicotinoid insecticides: current status and future prospects. *Arch. Insect Biochem. Physiol.* **58**, 200–215 (2005).

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.* **54**, 134-144(1998).

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.* **56**, 509-515 (2000).

Pratim 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_m^2 as an effective tool for validation of QSAR models in computational drug design and predictive toxicology. *Mini-Rev. Med. Chem.* **12**, 491-504(2012).

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).

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.* **27**, 1741-1746 (2006).

Tomlin, C.D.S.: *The Pesticide Manual: A World Compendium*. Hampshire,UK: Br. Crop Prot. Counc. 15th ed(2009).

Ye, Z., Xia, S., Shao, X., Cheng, J., Xu, X., Xu, Z., Li, Z., Xuhong Qian, X.: Design, synthesis, crystal structure analysis, and insecticidal evaluation of Phenylazoneonicotinoids, *J. Agric. Food Chem.* **59**, 10615–10623(2011)

Ye, Z., Shi, L.,Shao,Xu, X., Xu, Z.,Li,Z.: Pyrrole- and dihydropyrrole-fused Neonicotinoids: design, synthesis, and insecticidal evaluation. *J. Agric. Food Chem.* **61**, 312－319(2013).

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

フェニルアゾ、ピロール、ジヒドロピロール系ネオニコチノイド農薬のPLSによる殺虫活性の検討

鈴 木 孝 弘, Alina Bora , Simona Funar-Timofei

ネオニコチノイド系農薬は、昆虫の神経伝達物質アセチルコリンの受容体に結合し、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効果がある。24種類のフェニルアゾ、ピロール、ジヒドロピロール系の縮合環構造を有するネオニコチノイド農薬の分子構造からDragon記述子を求め、ササゲアブラムシ（*Aphis craccivora*）に対する殺虫活性との定量的構造活性相関（QSAR）をPLSモデルにより検討した。その結果、新規の農薬分子の設計に有用である予測性のある統計的に有意なPLS回帰分析モデルを構築することができた。