SAR Simulation for Insecticidal Activity Against the Cowpea Aphids (*Aphis craccivora*) of Neonicotinoids

Takahiro SUZUKI[1], Alina BORA[2], Simona FUNAR-TIMOFEI[2]

Abstract

Neonicotinoids represent one of the most important classes of insecticides on the market. In this study a series of 27 neonicotinoid derivatives bearing five-membered heterocycles and cis-nitromethylene neonicotinoids with insecticidal activity active against the cowpea aphids (*Aphis craccivora*) was investigated using molecular modeling and linear regression techniques. Neonicotinoid structures were modeled using the semiempirical quantum chemical PM7 approach. The conformers of minimum energy were further used to calculate structural descriptors, which were related to the insecticidal activity by applying a traditional QSAR modeling approach, using the multiple linear regression (MLR). The genetic algorithm was employed for variable selection and several mono and bi-parametric models were selected. The final MLR model has good fitting statistical parameters ($r^2_{training}$ = 0.913; q^2_{LOO} = 0.880; $RMSE_{tr}$ = 0.211; SEE = 0.235; F = 56.14) and predictive power. Higher values of 3D-MoRSE and 2D autocorrelation descriptors and the presence of significant hydrophobic molecular moieties and aromaticity influence the pesticide activity.

Keywords : Cowpea aphids, Neonicotinoid, MLR, QSARINS, PM7

1. Introduction

Neonicotinoids represent an important and actual class of insecticides, which act on the insect nicotinic acetylcholine receptors (nAChRs) (Nauen and Denholm, 2005; Ren et al., 2014). Research on the neonicotinoid design faced new challenges related to limit adverse effects on bees, elucidation of mechanisms of action and metabolism, resistance monitoring and management.

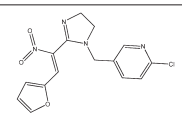
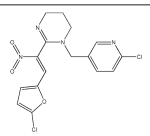
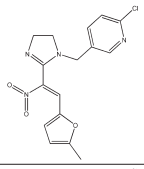
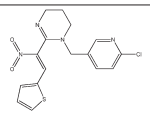
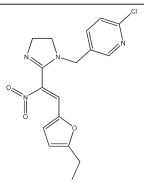
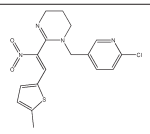
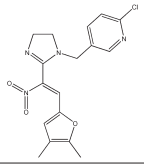
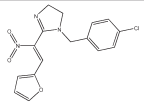
Synthetic nitromethylene heterocyles are among the major commercial insecticides

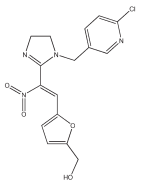
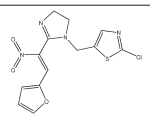
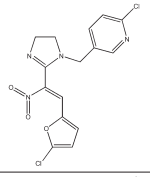
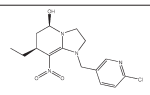
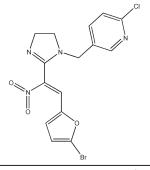
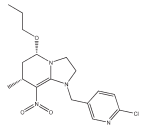
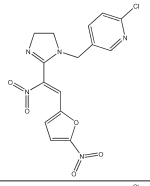
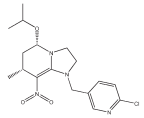
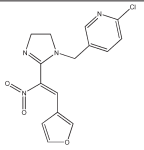
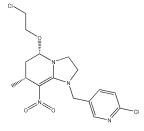
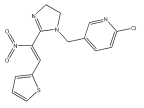
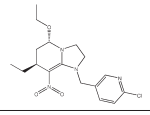
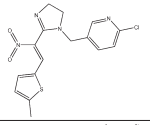
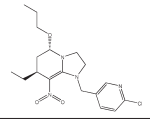
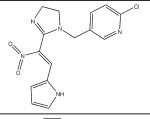
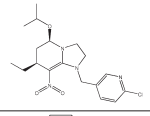
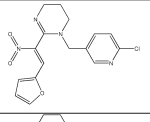
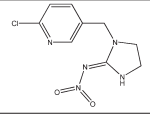
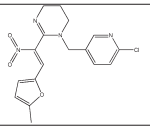
[1] Natural Science Laboratory, Toyo University, 5-28-20 Hakusan, Bunkyo-ku, Tokyo 112-8606, JAPAN
[2] Institute of Chemistry Timisoara of the Romanian Academy, 24 Mihai Viteazul Blvd., RO-300223 Timişoara, ROMANIA

targeting nAChRs (Matsuda et al., 2009). Mode-of-action studies have been conducted to show that both nitromethylene heterocycles and imidacloprid chemotypes act on insect nAChRs. Imidacloprid, the first commercially successful neonicotinoid, exhibits much higher photostability than related nitromethylene analogs (Matsuda et al., 2005). The selective neonicotinoid toxicity to insects over vertebrates has been shown to be related to their selectivity to nAChRs at least partially. Neonicotinoids have a common pharmacophore containing nitroimine, nitromethylene, or cyanoimine groups that determine insecticidal potency and selectivity. Basic residues, which are selectively present in insect nAChRs, are probably involved in differential neonicotinoid-nAChR interactions. The most remarkable result on neonicotinoid actions is the identification of basic residues in loop D of insect nAChRs as those likely to be involved in selective interactions with neonicotinoids.

In the present study, a series of neonicotinoid insecticides, active against the cowpea aphids (*Aphis craccivora*) was modeled using molecular and quantum mechanics approaches. The insecticidal activity was correlated to the structural descriptors derived from the minimum energy structures using the multiple linear regression approach. The predictive models thus obtained would be useful to design new insecticides with improved activity.

Table 1. The neonicotinoid structures, their experimental insecticidal (pLC_{50}) and predicted ($pLC_{50\ pred}$) activity values obtained using the MLR1 model

No	Structure	pLC_{50exp}	pLC_{50pred}	No	Structure	pLC_{50exp}	pLC_{50pred}
1		5.43	5.15	15*		4.22	4.55
2		5.20	5.34	16		4.69	4.70
3		5.74	5.45	17		4.61	4.48
4		5.33	5.29	18**		3.63	

5		4.98	5.03	19		5.46	5.21
6		5.12	5.51	20*		3.75	3.77
7*		5.14	5.14	21		4.04	3.99
8		4.96	4.92	22		3.58	3.69
9		5.35	5.07	23		3.72	3.62
10*		5.37	5.21	24		3.88	3.78
11		5.51	5.58	25		3.64	3.90
12*		4.95	5.13	26		3.63	3.58
13		4.12	4.16	27		4.46	4.98
14**		3.16					

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

2. Methods

Definition of target property and molecular structures

27 neonicotinoid analogues bearing five-membered heterocycles and cis-nitromethylene neonicotinoids with known insecticidal activity were investigated (Shao et al., 2008; Shao et al., 2009). The insecticidal activity data against cowpea aphids (*Aphis craccivora*), expressed as pLC_{50} values (where LC_{50} represents the median lethal concentration of the chemical in air that kills 50% of the test animals during the observation period) was used 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. Then, the structures were energy minimized using the semiempirical PM7 Hamiltonian (Stewart, 2013) implemented in MOPAC 2016 program (MOPAC2016, James J. P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, HTTP://OpenMOPAC.net (2016).). A gradient norm limit of 0.01kcal/Å was set in the geometry optimization. The quantum chemical descriptors are presented in Table 2. 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.

Table 2. Quantum chemical descriptors (Schüürmann, 1990a; Schüürmann, 1990b) of the neonicotinoid derivatives*

No	ACOSMO ($Å^2$)	VCOSMO ($Å^3$)	$-\alpha$ (eV)	η (eV)	E_{HOMO} (eV)	E_{LUMO} (eV)	ΔH_f (kcal/mol)
1	318.12	365.17	5.61	3.72	-9.33	-1.89	59.35
2	309.09	387.91	5.45	3.87	-9.31	-1.58	44.80
3	322.72	410.00	5.44	3.87	-9.31	-1.57	39.15
4	330.63	408.10	5.39	3.86	-9.25	-1.52	36.67
5	309.97	392.87	5.60	3.86	-9.45	-1.74	6.23
6	308.55	388.63	5.63	3.77	-9.40	-1.86	52.58
7	332.72	403.03	5.70	3.72	-9.42	-1.99	64.92
8	334.76	401.16	6.02	3.51	-9.53	-2.51	61.89
9	318.21	368.32	5.45	3.94	-9.38	-1.51	57.11
10	328.92	383.26	5.57	3.74	-9.31	-1.83	93.14
11	309.18	400.97	5.41	3.81	-9.22	-1.60	74.73
12	324.98	367.48	5.45	3.90	-9.35	-1.56	85.37
13	328.87	386.87	5.58	3.88	-9.45	-1.70	48.26
14	314.76	406.77	5.24	3.88	-9.12	-1.35	30.50
15	313.84	404.73	5.47	3.90	-9.36	-1.57	39.23

16	309.45	403.22	5.40	3.81	-9.21	-1.59	77.51
17	330.66	423.26	5.31	3.97	-9.28	-1.34	68.88
18	322.11	374.30	5.54	3.74	-9.28	-1.80	47.71
19	314.56	364.08	5.62	3.79	-9.41	-1.83	71.10
20	310.14	380.04	5.06	4.10	-9.15	-0.96	-29.22
21	327.85	420.35	4.84	4.01	-8.85	-0.83	-29.99
22	342.94	425.46	4.85	4.16	-9.01	-0.69	-31.17
23	340.41	425.99	4.92	4.05	-8.97	-0.88	-34.08
24	344.23	425.20	4.58	4.25	-8.83	-0.34	-22.07
25	329.20	443.67	4.91	3.99	-8.90	-0.93	-36.92
26	368.79	450.25	4.77	4.20	-8.97	-0.57	-35.75
27	255.47	274.84	5.31	4.31	-9.61	-1.00	59.30

* ACOSMO-COSMO area; VCOSMO-COSMO VOLUME; - α - Mulliken electronegativity; η-Parr & Pople absolute hardness; EHOMO-HOMO energie; ELUMO-LUMO energie; ΔH_f-final heat of formation

The Multiple Linear Regression (MLR) method

The multiple linear regression approach was employed to correlate the experimental insecticidal activity to the structural calculated descriptors. The Genetic Algorithm (GA) was used as variable selection method (Depczynski et al., 2000), inside the QSARINS v. 2.2 program (Gramatica et al, 2013). 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.

Model validity

The dataset was splitted into a training set of 81.5% and a random test set of 18.5% of the total number of compounds (no. 7, 10, 12, 13, 22). Several model predictivity parameters were employed: 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 (with a lowest threshold value of 0.5 to be accepted).

For internal model validation, the model robustness and overfitting were checked using the Y-randomization test. 2000 randomizations were employed to develop MLR models with minimal r^2 and q^2 values (Eriksson et al., 2001). In addition, 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 coefficient were calculated, too.

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).

The Multi-Criteria Decision Making (MCDM) (Keller et al., 1991) was used as a single number (score) between 0 and 1 to summarize the performances of a certain number of simultaneous criteria. To each validation criteria a desirability function was associated, which values range from 0 to 1 (where 0 represents the worst validation criteria value and 1 the best). ‚MCDM all' scores (obtained by geometric average of all values obtained from the desirability functions) were calculated based on the fitting, cross validated and external criteria and were used to choose the best MLR models.

3. Results and discussion

The data was normalized using the auto-scaling method:

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

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.

Several MLR models were obtained using the genetic algorithm for variable selection. The internal and external validation criteria for these models are presented in Table 3-5.

Table 3. Internal validation parameters 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	0.913	0.880	0.867	0.897	0.211	0.160	0.955	0.157	-0.336	0.235	56.144
MLR2	0.940	0.910	0.890	0.929	0.175	0.136	0.969	0.159	-0.360	0.196	83.337
MLR3	0.932	0.904	0.878	0.920	0.186	0.142	0.965	0.159	-0.349	0.208	73.303
MLR4	0.891	0.841	0.819	0.871	0.236	0.187	0.943	0.157	-0.340	0.263	43.750

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

Table 4. External validation parameters of the MLR models (test set)*

Model	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	$RMSE_{ext}$	MAE_{ext}	CCC_{ext}
MLR1	0.907	0.907	0.933	0.186	0.141	0.949
MLR2	0.783	0.783	0.843	0.284	0.243	0.924
MLR3	0.804	0.804	0.857	0.270	0.246	0.887
MLR4	0.854	0.854	0.894	0.233	0.184	0.914

*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 5. The r_m^2 predictivity parameter, MCDM values and final selected descriptors*

Model	r_m^2	MCDM	Descriptors included in the MLR model*
MLR1	0.854	0.901	GATS3e, Mor03e, H-053
MLR2	0.745	0.882	ARR, BLI, H-053
MLR3	0.796	0.877	nR06, Mor03e, ASAHydrophobic
MLR4	0.788	0.867	nR06, Mor03v, ΔH_f

*GATS3e- Geary autocorrelation - lag 3 / weighted by atomic Sanderson electronegativities, Mor03e- 3D-MoRSE - signal 03 / weighted by atomic Sanderson electronegativities, H-053 - H attached to C0(sp3) with 2X attached to next C, ARR- aromatic ratio, BLI- Kier benzene-likeliness index, nR06- number of 6-membered rings, ASAHydrophobic-the water accessible molecular surface area of all hydrophobic atoms, Mor03v- 3D-MoRSE - signal 03 / weighted by atomic van der Waals volumes, ΔH_f-heat of formation.

Experimental versus predicted pLC_{50} values, Williams and Y-scramble plots are included for the MLR1 best model in Figure 1, 2 and 3, respectively.

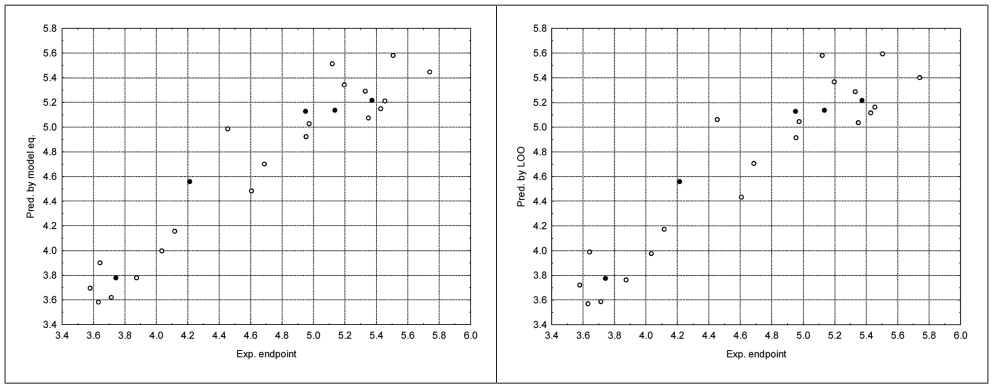


Figure. 1. Experimental versus predicted pLC50 values for the MLR1 model predicted by the model (left) and by the leave-one-out (right) crosvalidation approach (white circles-training compounds, black circles-test compounds).

Two compounds were found as outliers (no. 14 and 18) according to the studentized residuals and were not included in the final MLR models.

The Williams plot for the training set establishes the applicability domain of the models within $\pm 2.5\sigma$ and a leverage threshold h* of 0.600 ($h_i > h^*$; h_i =leverage of a given chemical; h^*= the warning leverage). All compounds in the dataset are within the applicability domain of the models, as presented in Figure 2 for the MLR1 model.

MLR1 model overfit was checked using the y-scrambling test. Significant low scrambled r^2 (r_{scr}^2) and cross-validated q^2 (q_{scr}^2) values obtained for MLR1 model, after 2000 trials, indicate the model robustness and no chance correlation. Figure 3 suggest that in

case of all the randomized models, the values of r^2_{scr} and q^2_{scr} were < 0.6 (r^2_{scr}/q^2_{scr} of 0.157/-0.336).

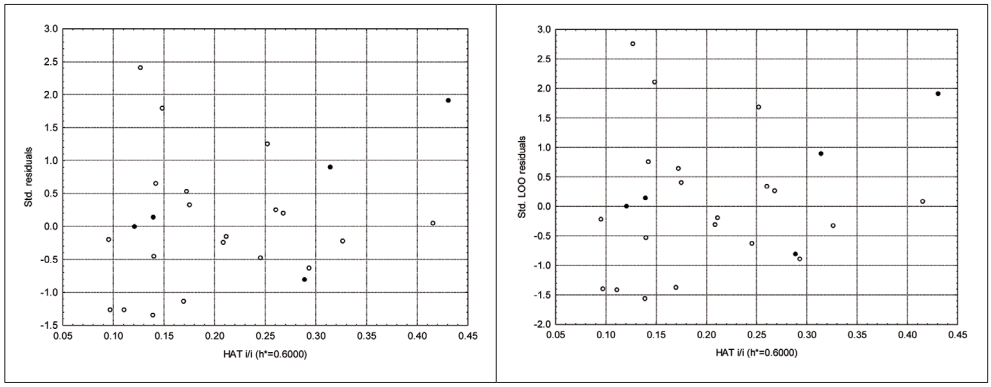


Figure. 2. Williams plot presented by the MLR1 model predicted (left) and by the leave-one-out (right) crosvalidation approach (white circles-training compounds, black circles-test compounds).

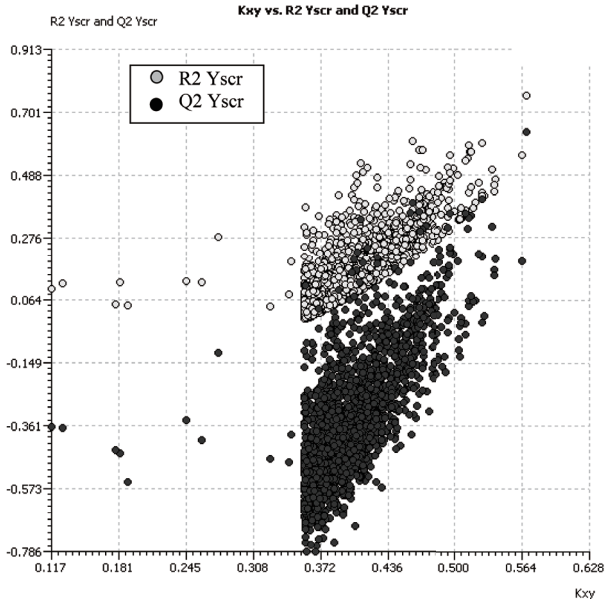


Figure. 3. Y-scramble plot for the MLR1 model.

The best MLR1 model has three descriptors: GATS3e, Mor03e and H-053. Higher values of GATS3e and Mor03e descriptors are beneficial for the insecticidal activity. The presence of H atom attached to a CO(sp3) group with 2X (where X is any electronegative (O, N, S, P, Se, halogens) atom) attached to next C atom decreases the insecticidal activity.

2D autocorrelations are molecular descriptors calculated from molecular graph in 2D space, which describe how a considered property is distributed along a topological molecular structure. GATS autocorrelations are calculated by the Geary coefficient (Geary, 1954).

3D-MoRSE descriptors are sensitive to molecular geometry (Todeschini and Consonni, 2009), taking into account several weighting scheme (physico-chemical properties such as atomic mass, partial atomic charges, and atomic polarizability).

In our best MLR1 model both 2D autocorrelation and 3D-MoRSE descriptors are weighted by the atomic Sanderson electronegativities.

High values of the number of 6-membered rings are not favorable for the insecticidal activity. Atoms responsible for hydrophobic molecular area and higher values of molecular aromatic ratio increase the insecticidal activity.

An intercorrelation analysis of the selected molecular descriptors from the MLR1 model is presented in Table 6. The three selected descriptors are not intercorrelated.

Table 6. Correlation matrix of the selected descriptors included in the MLR1 model

	GATS3e	Mor03e	H-053
GATS3e	1.0000		
Mor03e	-0.344	1.0000	
H-053	0.283	-0.606	1.0000

The statistical results and intercorrelation coefficients presented above confirm that the MLR method associated with a proper variable selection procedure generates an efficient QSAR model for predicting the insecticide activity of neonicotinoid insecticides.

4. Conclusion

Quantitative structure activity relationships of a series of 27 neonicotinoid insecticides against the cowpea aphids (*Aphis craccivora*) was investigated using several computational techniques. Structure optimization modeling using the semiempirical quantum chemical PM7 approach was used. Structural descriptors derived from the minimum energy structures were related to the insecticide activity. MLR models with good statistical results and predictive power were obtained. Higher 2D autocorrelation, 3D-MoRSE descriptor values, significant hydrophobic molecular moieties and aromaticity increase the insecticidal potency. The presence of 6-membered rings in the neonicotinoid structure does not favor the insecticidal activity.

Acknowledgements

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.

References

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**, 320-2335(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)

Depczynski, U., Frost, V.J., and Molt, K.: Genetic algorithms applied to the selection of factors in principal component regression. *Anal. Chim. Acta*, **420**, 217-227(2000).

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

Geary, R.C.; The contiguity ratio and statistical mapping, *The Incorporated Statistician*, **5**(3), 115-146 (1954)

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).

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.*, **34**, 2121-2132(2013).

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

Matsuda, K., Shimomura, M., Ihara, M., Akamatsu. M., Sattelle, D.B.: Neonicotinoids show selective and diverse actions on their nicotinic receptor targets: electrophysiology, molecular biology, and receptor modeling studies. *Biosci Biotechnol Biochem.*, **69**, 1442-1452(2005).

Matsuda, K., Kanaoka, S., Akamatsu, M., Sattelle, D. B.: Diverse actions and target-site selectivity of neonicotinoids: Structural insights, *Mol Pharmacol.*, **76**, 1-10(2009).

Nauen, R., Denholm, I.: Resistance of insect pests to neonicotinoid insecticides: Current satus and future prospects: *Archives of Insect Biochemistry and Physiology*, **58**, 200-

215 (2005)

Ren, L., Lou, Y., Chen, N., Xia, S., Shao, Xu, X., Li,Z.: Synthesis and Insecticidal Activities of Tetrahydroimidazo [1,2-a] pyridinones: Further Exploration on *cis*-Neonicotinoids, *Synthetic Commun.*, **44**, 858-867 (2014).

Roy, K. and 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).

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.*, **48**, 2140-2145(2008).

Schüürmann,G.: QSAR Analysis of the acute fish toxicity of organic phosphorothionates using theoretically derived molecular descriptors, *Environmental Toxicology and Chernislry*, **9**, 417-428 (1990a).

Schüürmann,G.: Quantitative structure-property relationships for the polarizability, solvatochromic parameters and lipophilicity, *Quant. Struct.-Act. Relat.*, **9**, 326-333 (1990b).

Shao, X., Zhang, W., Peng, Y., Li, Z., Tian, Z., Qian, X.: cis-Nitromethylene neonicotinoids as new nicotinic family: Synthesis, structural diversity, and insecticidal evaluation of hexahydroimidazo[1,2-α]pyridine, *Bioorganic & Medicinal Chemistry Letters*, **18**, 6513-6516(2008).

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).

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).

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).

Todeschini, R., Consonni V.: Molecular Descriptors for Chemoinformatics, Volumes I & II, Wiley-VCH, Weinheim, pp. 551, 2009.

和文要旨

ネオニコチノイド系農薬分子のササゲアブラムシ（*Aphis craccivora*）に対する殺虫活性の構造活性相関

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。27種類のヘテロ5員環構造を持つシス－ニトロメチレン系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）を検討した。その結果、予測性のある統計的に有意な重回帰分析モデルを構築することができ、農薬分子側の構造として、3D-MoRSE と2D-自己相関記述子の値が高いほど、また分子中の疎水性の部分と芳香族性が殺虫活性に寄与することを見出した。