Current Topics in Biochemical Research
Vol. 17, 2016

Original Communication

Modeling of fungicide activity of trifluoromethyl-substituted 1,2,4-triazoles using PLS, ANN and SVM

Simona Funar-Timofei[1], Daniela Ionescu[2] and Takahiro Suzuki[3,*]

[1]Institute of Chemistry Timisoara of the Romanian Academy, 24 Mihai Viteazul Blvd., RO 300223 Timişoara, Romania; [2]University of Medicine and Pharmacy, Faculty of Pharmacy, P-ta Eftimie Murgu 2-4, 300034 Timisoara, Romania. [3]Natural Science Laboratory, Toyo University, 5-28-20 Hakusan, Bunkyo-ku, Tokyo 112-8606, Japan.

ABSTRACT

1,2,4-triazole derivatives have a wide spectrum of biological activities, including anti-inflammatory, antiviral, analgesic, antimicrobial, anticonvulsant, anticancer, antioxidant and antidepressant. In this study, quantitative structure-activity relationships (QSAR) of trifluoromethyl-substituted 1,2,4-triazole derivatives with fungicidal activity against *Fusarium oxysporum* f. sp. *cucumerinum* were investigated by employing modern modeling approaches. In a previous work two types of *E* and *Z* isomers were found and several molecular descriptors were calculated from their minimum energy conformations. In this study, these descriptors were correlated with the experimental fungicide relative inhibition rate using artificial neural networks (ANNs) and support vector machine (SVM) methods. Partial least squares (PLS) calculations were also performed for the same sets. The predictive ability of these models was validated by several criteria using an external set of five out of eighteen compounds. Best results were obtained by means of the SVM model for *Z* isomers, which could be used for the prediction of new fungicides with higher activity.

KEYWORDS: QSAR, fungicide, PLS, SVM, ANN, 1,2,4-triazoles

∗Corresponding author: suzuki@toyonet.toyo.ac.jp

Dedicated to the 150th anniversary of the Romanian Academy.

INTRODUCTION

1,2,4-triazole and its derivatives represent a flexible class of biologically active compounds, possessing a wide spectrum of activities, including anticonvulsant, antidepressant, antioxidant, anti-inflammatory, analgesic, antinociceptive, antibacterial, antimycobacterial, antifungal, antiviral, anticancer, anti-parasitic, and anti-urease [1, 2].

1,2,4-triazole fragment is used in several therapeutically interesting drugs [3]. Some of the new drugs containing fused heterocycles with a triazole moiety that are worth mentioning are alprazolam, triazolam, estazolam (hypnotic, sedative, tranquilizer), trazodone (antidepressant, anxiolytic), trapidil (hypotensive), terconazole (antifungal), hexaconazole (antifungal), etizolam (amnesic, anxiolytic, anticonvulsant, hypnotic, sedative and skeletal muscle relaxant), rilmazafon (hypnotic, anxiolytic) and rizatriptan (antimigrane agent). Mannich bases with good antibacterial activity mainly contain morpholine, 4-benzylpiperazine, *N*-methylpiperidine and trifluoromethylphenylpiperazine in the aminomethyl group [4, 5].

In 1996, Williams reported that around 25% of all agrochemical compounds used commercially are chiral compounds, accounting for 26% of the total agrochemical market value [6, 7]. Triazole chirality (asymmetrical carbons existing at the position(s) immediate and/or vicinal to the triazole rings) is expected to play an important role in the bioactivities of triazole fungicides [8, 9]. Several chiral

N-substituted azoles are used as wide range fungicides, and as plant growth regulators [10].

Fluoroaryl derivatives of 1,2,4-triazole such as fluconazole, fosfluconazole, voriconazole, and itraconazole are inhibitors of the fungal cytochrome P450 enzyme 14 α-demethylase [11]. Some derivatives of 3-trifluoromethyl-1,2,4-triazole-5-thione have been shown to exhibit a high fungicide activity [12].

Triazoles are used as fungicides in fruits, vegetables, legumes and grain crops, both as pre- and postharvest applications [13]. The mechanism of their antifungal effect is based on the inhibition of ergosterol biosynthesis (being involved in the fungal cell-wall formation) and of sterol 14 α-demethylase. It was found that 3-amino-1,2,4-triazole is an inhibitor of mitochondrial and chloroplast function, and it is used as a herbicide and cotton defoliant. The antimicrobial, anticonvulsant and antidepressant characteristics of 1,2,4-triazoles were reported by Siwek *et al*. [14].

This paper presents a quantitative structure-fungicidal activity relationship study for a series of eighteen 1-[(4-substituted-benzylpiperazin-1-yl) methyl]-4-(substituted)benzylideneamino-3-trifluoromethyl-1*H*-1,2,4-triazole-5(4*H*)-thiones, which were tested against the *Fusarium oxysporum* f. sp. *cucumerinum* fungi test [12]. Three types of modern QSAR methods, namely partial least squares (PLS), artificial neural networks (ANNs) and support vector machines (SVMs) were applied to gain information on designing new fungicides with higher activity.

MATERIALS AND METHODS

Definition of target property and molecular structures

A series of 18 Mannich bases having trifluoromethyl-substituted 1,2,4-triazole containing substituted benzylpiperazine ring (Table 1) was used [12]. Their fungicidal relative inhibition rate (RIR) against *Fusarium oxysporum* f. sp. *cucumerinum* was employed as a dependent variable.

These fungicides were previously [15] modeled using a conformer ensemble for the herbicide structures, which was generated by means of the MMFF94s force field included in the Omega (version 2.5.1.4, OpenEye Scientific Software, Santa Fe, NM. http://www.eyesopen.com) software [16, 17]. Starting from the two types of *E* and *Z* isomers generated with respect to the C=N bond by Omega, two series of compounds were built. Structural 0D, 1D, 2D and 3D descriptors were calculated for the lowest energy structures using the DRAGON (Dragon Professional 5.5 (2007), Talete S. R. L., Milano, Italy), InstantJchem (which was used for structure database management, search and prediction) (InstantJchem 15.7.27, 2015, ChemAxon (http://www.chemaxon.com) and ChemProp (UFZ Department of Ecological Chemistry 2014. ChemProp 6.2; http://www.ufz.de/index.php?en=34593 softwares.

The partial least squares (PLS) method

The partial least squares (PLS) regression is a statistical technique in which independent responses are related to projections of factors [18]. In PLS a block of response variables are linked to a block of explanatory (even correlated) variables and thus stable and highly predictive models are accomplished [19].

In the PLS model of F dimension, significant principal components (t_{if} columns, as presented in equation (1), i = 1, ..., N) of N training set compounds are calculated from the **X** matrix of chemical descriptors.

$$x_{ik} = \bar{x}_k + \sum_{f=1}^{F} p_{fk} \cdot t_{if} + e_{ik} \quad (1)$$

where $\bar{x}_k$ represents the mean of variable k, p_{fk} the loading of variable k in dimension (factor) f, and e_{ik} the residuals [20]. A maximal covariance between the consecutive orthogonal latent variables (t_{if}) and the dependent variables (y) is obtained. The linear PLS inner relation is described by equation (2):

$$y_i = \bar{y} + \sum_{f=1}^{F} b_f \cdot t_{if} + e_i \quad (2)$$

where $\bar{y}$ represents the average of the y-variable and b_f the regression coefficients. They can be used to transform the biological activity y as a function of the original x_k descriptors.

PLS calculations were performed using the SIMCA (P+ 12.0.0.0 May 20 2008, Umetrics, Sweeden, http://www.umetrics.com/) package.

Table 1. Trifluoromethyl-substituted 1,2,4-triazoles structures, experimental *Fusarium oxysporum* f. sp. *cucumerinum* (RIR_{exp}) and predicted relative inhibition rate by the PLS, ANN and SVM models calculated for isomer *E* and *Z* series.

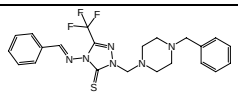
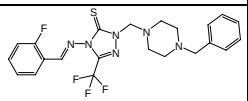
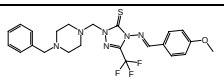
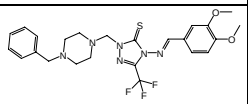
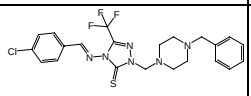
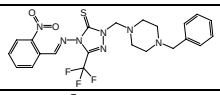
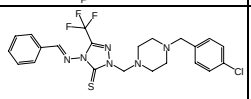
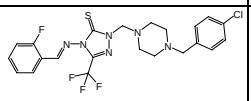
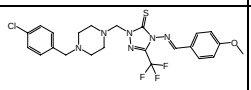
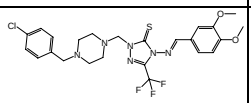
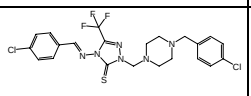
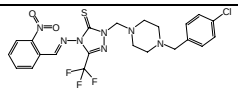
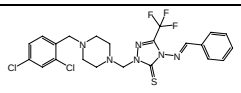
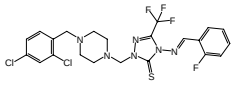
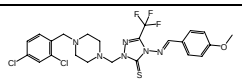
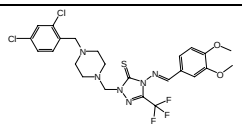
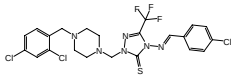
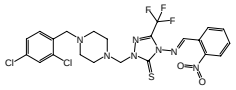
No.	Structure	RIR_{exp}	RIR_{pred} PLS_E	RIR_{pred} PLS_Z	RIR_{pred} ANN_E	RIR_{pred} ANN_Z	RIR_{pred} SVM_E	RIR_{pred} SVM_Z
1		0.101	0.190	0.505	0.178	0.107	0.083	0.083
2		0.804	0.453	0.571	0.229	0.121	0.237	0.083
3		0.187	0.187	0.403	0.143	0.100	0.168	0.083
4		0	0.122	0.020	0.113	0.085	0.151	0.083
5		0	0.359	0.083	0.252	0.378	0.301	0.319
6		0.402	0.348	0.376	0.301	0.312	0.298	0.386
7		0.509	0.396	0.638	0.566	0.732	0.486	0.676
8		0.719	0.631	0.698	0.443	0.369	0.368	0.386
9		0.604	0.337	0.542	0.297	0.312	0.294	0.386
10		0.401	0.258	0.454	0.319	0.348	0.275	0.386
11		0.303	0.431	0.165	0523	0.525	0.462	0.543
12		0.502	0.526	0.491	0.621	0.711	0.458	0.623
13		0.708	0.627	0.756	0.935	0.768	0.614	0.743

Table 1 continued..

14		0.826	0.861	0.817	0.647	0.622	0.525	0.623
15		0.504	0.663	0.700	0.629	0.658	0.500	0.623
16		0.705	0.614	0.644	0.508	0.653	0.453	0.623
17		0.607	0.690	0.811	0.692	0.654	0.546	0.691
18		0.608	0.738	0.596	0.178	0.107	0.83	0.083

SVM approach

The support vector machines (SVMs) are a set of supervised learning methods used for classification, regression and detection of outliers [21]. SVM has several advantages over other methods such as MLR and ANN [22]. SVMs have been applied to solve regression problems for data sets by choosing a suitable cost function (ε-insensitive loss function).

The regression function (SVR), in the LIBSVM software [23] was employed for this modeling. In order to accurately evaluate the predictive performance of SVM models, two parameters have to be optimized: *gamma* (g) and *cost* (c) in LIBSVM.

ANN method

Artificial neural networks (ANNs) have an inherent ability to provide nonlinear and cross product terms for QSAR modeling. ANNs are especially useful when a rigid theoretical basis and/or mathematical relationship describing a phenomenon to be modeled are not available beforehand as in the case of SVMs. However, it has significant disadvantages, including local minima, over-fitting, over-training, and long processing time.

Among the many ANN approaches, differing both in architecture and in learning algorithms, the three-layer ANNs with the back–propagation of errors [24] were employed in this study. The hidden layer contained variable nodes, and the input and hidden variables each had a bias neuron.

Input and output data were normalized between 0.1 and 0.9, and models were evaluated on the basis of correlation coefficient (r) and root-mean-square error (RMSE). ANN calculations were carried out using an in-house program. A commonly used log sigmoid function and the delta rule for the error correction formula were used in the networks.

Model validation

The data over-fitting and model applicability were controlled by comparing the root-mean-square errors (RMSE) and the mean absolute error (MAE) [25] of the training and validation sets.

For internal validation results several measures of robustness were employed: Y-scrambling [26] and q^2 leave-one-out and leave-seven-out cross-validation function coefficients.

The concordance correlation coefficient (CCC) [27] was used to test the robustness and predictive power of the model. A threshold value higher than 0.85 (which has been rigorously determined by a simulation study [28]) was employed.

To test the predictive power of the model, other parameters were calculated: the predictive r^2

$\left(r^2_{pred}\right)$ [29] and r^2_m [30] parameters. It is considered that for a predictive QSAR model, the value of r^2_{pred} and r^2_m should be higher than 0.5. In addition, several types of variances explained in external prediction like Q^2_{F1} [31], Q^2_{F2} [32], and Q^2_{F3} [33] were considered too (models with values higher than 0.7 were considered as acceptable [28]).

Among other statistical measures to check the model predictivity, the following parameters were used [34]: (i) squared correlation coefficient (r^2) between the predicted and observed activities as well as squared correlation coefficient by cross-validation (Q^2); (ii) coefficient of determination for linear regressions with intercepts set to zero, i.e. R_0^2 (predicted versus observed activities), and $R_0^{`2}$ (observed versus predicted activities); (iii) slopes k and k' of the above-mentioned two regression lines. The following conditions should be satisfied for a model with acceptable predictive ability:

$$q^2 > 0.5 \quad \text{(a)}$$

$$r^2 > 0.6 \quad \text{(b)}$$

$$(R^2 - R_0^2)/ R^2 < 0.1 \text{ and } 0.85 \le k \le 1.15 \text{ or} \quad \text{(c)}$$

$$(R^2 - R_0^{`2})/ R^2 < 0.1 \text{ and } 0.85 \le k` \le 1.15 \quad \text{(d)}$$

$$\left| R^2 - R_0^{`2} \right| < 0.3 \quad \text{(e)}$$

RESULTS AND DISCUSSION

PLS modeling results

Four clusters were found on the basis of the partition against medoids (PAM) algorithm [35] using the R software (*R Development Core Team*, ISBN 3-900051-07-0 (2010); retrieved from www.r-project.org), based on the silhouette values. The silhouette index measures how well an object has been classified by comparing its dissimilarity within its cluster and with its nearest neighbor. Following compounds were chosen as a test set to validate the final models: 1, 7, 10, 12, and 17 (Table 1) in both series, containing the *E* and *Z* isomers.

PCA models for both *E* and *Z* series of isomers were built for the entire **X** matrix (including $N = 18$ compounds and $X = 1464$ descriptors). The first three (from the total number of six and eight for *E* and *Z* isomers, respectively) significant principal components explained 72.4% and 76.7% of the information content of the descriptor matrix for the *E* and *Z* isomers, respectively.

PLS calculations were performed to correlate the RIR experimental values with all the calculated descriptors. The PLS model was constructed using the training set, and one principal component PLS model was obtained for both datasets: $R^2X(cum) = 0.726$, $R^2Y(cum) = 0.576$, $Q^2(cum) = 0.533$ for *E* isomers, and $R^2X(cum) = 0.756$, $R^2Y(cum) = 0.823$, $Q^2(cum) = 0.815$ for *Z* isomers, where $R^2Y(cum)$ is the cumulative sum of squares of the entire Y's explained by all extracted principal components and $Q^2(cum)$ is the fraction of the total variation of the Y's that can be predicted for all the extracted principal components.

Table 2 presents the type [36] of significant descriptors included in the final PLS models.

The importance of a given x variable for the **Y** matrix is proportional to its distance from the origin in the loading space. These lengths correspond to

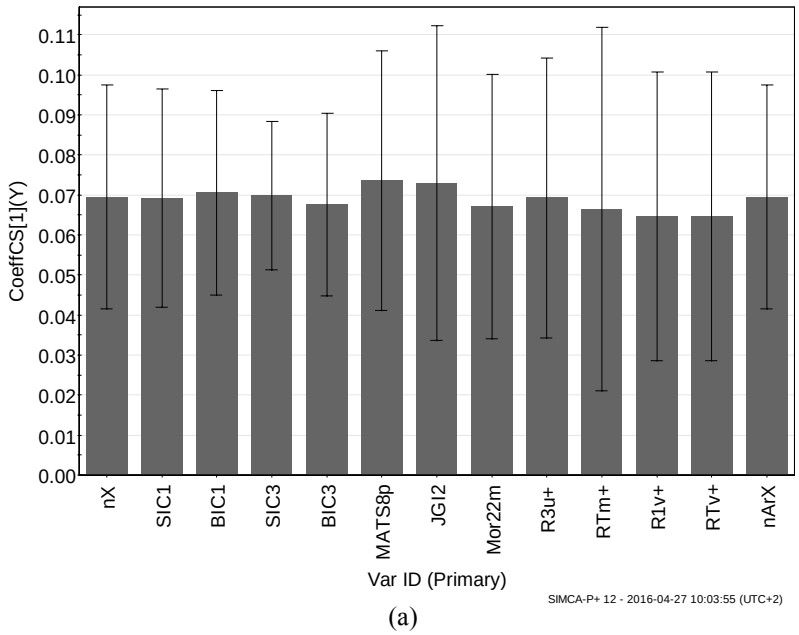
Table 2. Descriptors [36] included in the final PLS models.

Constitutional descriptors:	nX*
Topological descriptors:	BIC1*, SIC3*, SIC1*, BIC3*
2D autocorrelations:	MATS8p*
Galvez topological charge indices:	JGI2*
Functional group count:	nArX*
WHIM descriptors:	E1v**, E1p**, E1e**, E1s**, L3m**, L3s**
GETAWAY descriptors:	H3e**, R3u+*, H3u**, RTm+*, R1v+*,**, RTv+*,**
3D-MoRSE descriptors:	Mor28v**, Mor28p**, Mor28e**, Mor22m*, Mor28u**

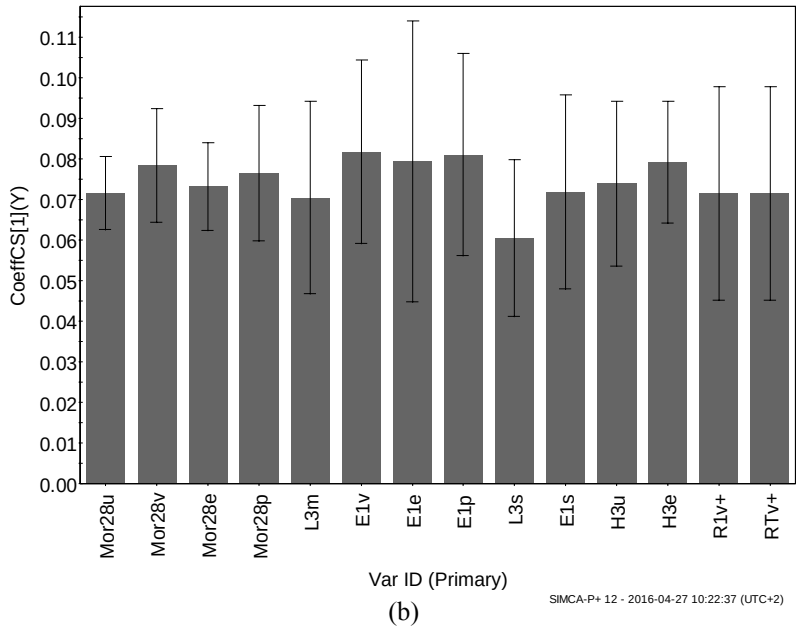
*for *E* isomer set
**for *Z* isomer set

the PLS regression coefficients derived from the first component. The importance of descriptors was evaluated by the VIP (variable influence on projection) values [37], which summarizes the importance of the x variables for both **Y** and **X** matrices in the model. This is a weighted sum of squares of the PLS-weights, with the weights calculated from the amount of Y variance of each PLS component. The noise variables (variable coefficient values close to 0) were excluded to reduce the model over-fit. The PLS coefficients and VIP plots are presented in figures 1 and 2.

Y-randomization test and leave-seven-out crossvalidation runs were performed to check the

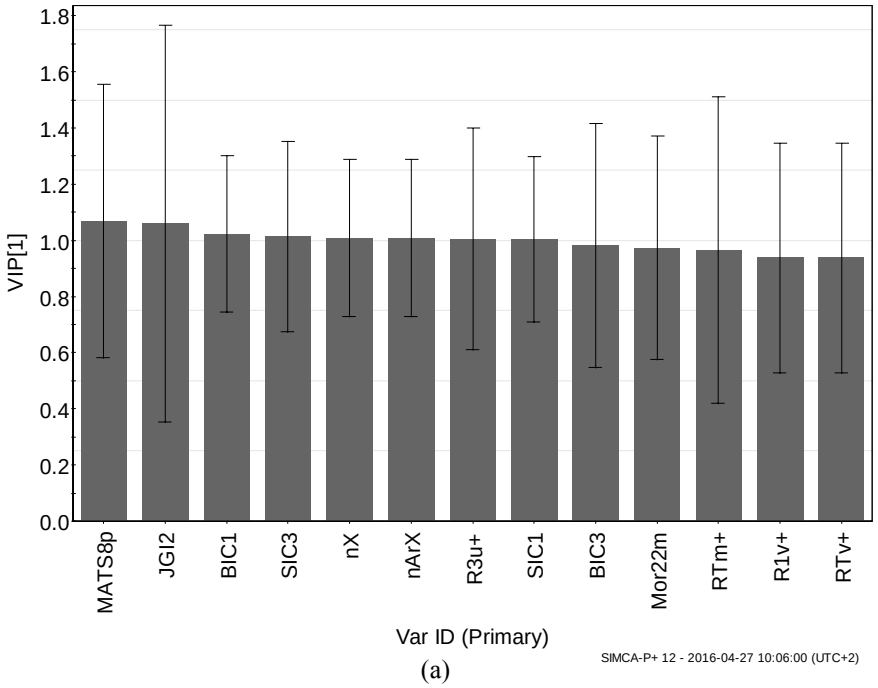


(a)

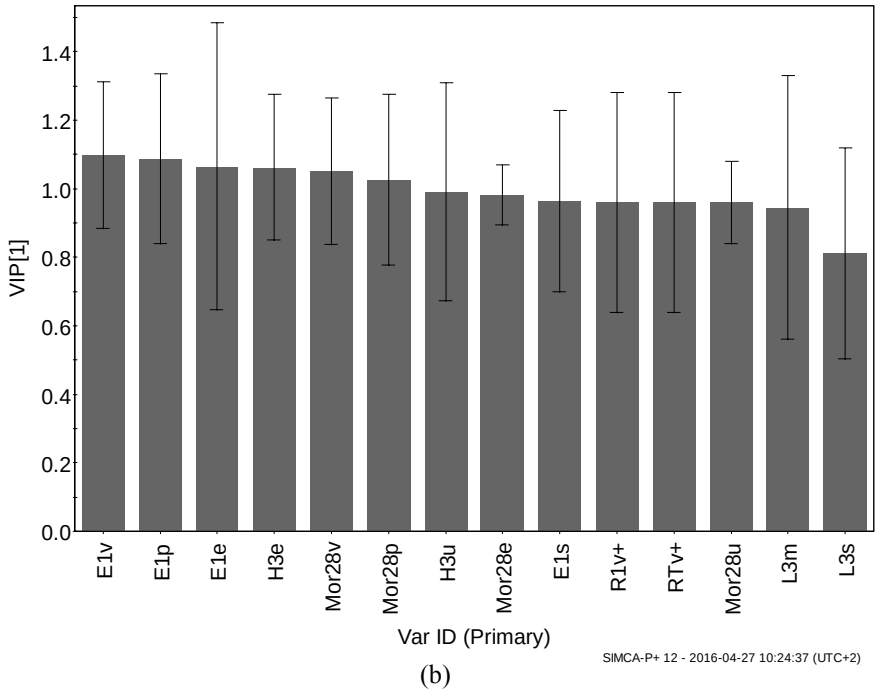


(b)

Figure 1. PLS regression coefficients' plot of the one-component PLS model, for *E* (a) and *Z* (b) isomer sets. The bars indicate 95% confidence intervals based on jack-knifing.



(a)



(b)

Figure 2. VIP plots of the x-variables of the one-component PLS models for *E* (a) and *Z* (b) isomer sets.

robustness and internal predictive ability of the PLS models. The risk of chance correlation was verified by the Y-scrambling procedure, which was repeated 999 times. The extremely low values of scrambled R^2 (0.111) and Q^2 (-0.172) for the *E* isomer series, and R^2 (0.0316) and Q^2 (-0.244) for the *Z* isomer series indicate no chance correlation for the chosen model.

The fitting and predictivity criteria for the PLS models are presented in tables 3-5.

Table 3. Internal validation parameters of the PLS, ANN and SVM models (training, *tr*, and cross-validation, *cv*, sets)*.

Model	$r^2_{training}$	q^2_{LOO}	q^2_{L7O}	$RMSE_{tr}$	MAE_{tr}	CCC_{tr}	$RMSE_{cv}$	MAE_{cv}	CCC_{cv}
PLS_E	0.576		0.533	0.181	0.144	0.731	0.190	0.150	0.690
PLS_Z	0.823		0.815	0.117	0.086	0.903	0.119	0.090	0.896
ANN_E	0.821	0.632		0.115	0.106	0.901	0.147	0.124	0.792
ANN_Z	0.887	0.855		0.092	0.070	0.938	0.222	0.197	0.855
SVM_E	0.632	0.576		0.165	0.143	0.734	0.124	0.104	0.731
SVM_Z	0.855	0.823		0.104	0.096	0.920	0.122	0.113	0.891

*$r^2_{training}$ - correlation coefficient; q^2_{LOO} - leave-one-out correlation coefficient; q^2_{L7O} - leave-seven-out correlation coefficient; $RMSE_{tr}$ and $RMSE_{cv}$ – training and cross-validation root-mean-square errors; MAE_{tr} and MAE_{cv} – training and cross-validation mean absolute errors; CCC_{tr} and CCC_{cv} – training and cross-validation concordance correlation coefficients.

Table 4. External validation parameters of the PLS, ANN and SVM models (test set)*.

Model	Q^2_{F1}	Q^2_{F2}	Q^2_{F3}	$RMSE_{ext}$	MAE_{ext}	CCC_{ext}	r^2_m	R^2_{pred}
PLS_E	0.720	0.679	0.874	0.099	0.090	0.846	0.575	0.720
PLS_Z	-0.292	-0.478	0.417	0.099	0.160	0.376	0.335	-0.292
ANN_E	0.560	0.537	0.811	0.119	0.102	0.730	0.618	0.560
ANN_Z	0.690	0.674	0.867	0.099	0.072	0.633	0.548	0.690
SVM_E	0.489	0.462	0.781	0.128	0.106	0.873	0.618	0.560
SVM_Z	0.889	0.883	0.952	0.060	0.043	0.945	0.775	0.889

*Q^2_{F1} [31], Q^2_{F2} [32], Q^2_{F3} [33] - external validation parameters; $RMSE_{ext}$ - root-mean-square errors; MAE_{ext} - mean absolute error; CCC_{ext} - external concordance correlation coefficient.

Table 5. Tropsha *et al*. [34] criteria calculated for external validation of the PLS, ANN and SVM models (test set)*.

Model	r^2_{test}	$(r^2 - r_0^2)/r^2$	$(r^2 - r_0'^2)/r^2$	k	k’	$\lvert r_0^2 - r_0'^2 \rvert$
PLS_E	0.720	0.056	0.009	0.995	0.959	0.034
PLS_Z	0.348	0.004	3.665	0.995	1.234	1.276
ANN_E	0.669	0.009	0.111	1.160	0.820	0.069
ANN_Z	0.776	0.111	0.003	1.051	0.909	0.083
SVM_E	0.849	0.018	0.001	1.306	0.747	0.015
SVM_Z	0.908	0.024	0.003	1.024	0.961	0.019

*r^2_{test} - squared correlation coefficient between the predicted and observed activities; r_0^2 - coefficient of determination for linear regressions with intercepts set to zero, i.e. (predicted versus observed activities); $r_0'^2$ - coefficient of determination for linear regressions with intercepts set to zero (observed versus predicted activities); k and k’- slopes of the above-mentioned two regression lines.

Best statistical results for fitting, but poor predictive model power was observed in the case of PLS model computed for *Z* stereoisomers. The statistical results of the PLS model for *E* isomers were worse from the fitting point of view, compared to results obtained for *Z* isomers, but better in terms of predictive power. In the case of *E* isomers, the types of most relevant variables which influence the fungicidal activity were 2D autocorrelations, Galvez topological charge indices, topological descriptors and number of halogen atoms and those in the case of *Z* isomer series were WHIM, GETAWAY and 3D-MoRSE descriptors.

ANN and SVM results

In a preliminary MLR study [15] performed for the same series of compounds to find significant descriptors for the nonlinear modeling, it was found that geometrical size descriptors significantly contribute to the fungicidal activity. The descriptor Mor19m (which represents the 3D-MoRSE - signal 19 / weighted by atomic masses) was found to have the highest correlation with RIR values for the *E* stereoisomers and the strongest basic pKa whereas the T(N..F) (the sum of topological distances between N..F) descriptor gave the best statistical results for the *Z* isomers' dataset. These descriptors were not included in the final PLS models because the variable coefficient values were close to 0 and VIP values did not exceed the value of 1.

Starting from these descriptors nonlinear ANN and SVM models for both *E* and *Z* datasets were developed. The architectures of the ANN models optimized were (1 input + 1 bias) : (2 hidden-layer nodes + 1 bias) : (1 output) for ANN_E and (2 inputs +1 bias) : (2 hidden-layer nodes + 1 bias) : (1 output) for ANN_Z, respectively.

Comparing the results of PLS, SVM and ANN, the best model was obtained using the SVM approach for the *Z* isomers. The optimized SVM model obtained for the *Z* isomers' set has the best fitting and predictive power. For this model the data over-fitting and model applicability were controlled by comparing the root-mean-square errors (RMSE) and the mean absolute error (MAE) calculated for the training, cross-validation and validation sets. The leave-one-out crossvalidation results show that the model is stable. The small difference of 5.4% between $r^2_{training}$ and q^2_{LOO} and the calculated RMSE and MAE values indicate an internally predictive model.

The calculated concordance correlation coefficient values for the training (CCC_{tr} = 0.920), crossvalidation (CCC_{cv} = 0.891) and test (CCC_{ext} = 0.945) sets indicate a robust model with good predictive power, which was confirmed by the r^2_{pred} value of 0.889 and by all the other predictivity parameters (Q^2_{F1} = 0.889, Q^2_{F2} = 0.883, Q^2_{F3} = 0.952, r^2_m = 0.775 and all parameters of Tropsha *et al*., see table 5).

The fitting and predictivity results for all the other PLS, SVM and ANN models were worse. The SVM for *Z* isomers model had comparable, but better results compared to the MLR ones obtained for *Z* isomers and same test set (for which following results were obtained: $r^2_{training}$ = 0.820, q^2_{LOO} = 0.705, $RMSE_{tr}$ = 0.12, $RMSE_{cv}$ = 0.15, $RMSE_{ext}$ = 0.05, MAE_{tr} = 0.11, MAE_{cv} = 0.14, MAE_{ext} = 0.04, CCC_{tr} = 0.901, CCC_{cv} = 0.840, CCC_{ext} = 0.949, Q^2_{F1} = 0.907, Q^2_{F2} =0.903, Q^2_{F3} = 0.960, r^2_m = 0.893, and r^2_{pred} = 0.907).

CONCLUSION

The fungicidal activity of the trifluoromethyl-substituted 1,2,4-triazoles against the *F. oxysporum* f. sp. *cucumerinum*, in terms of relative inhibition rate, was correlated with structural descriptors using the partial least squares (PLS), artificial neural networks (ANN) and support vector machine (SVM) methods. The compound stereoselectivity, with respect to the C=N bond, can influence the fungicidal activity. Better results were obtained for *Z* isomers compared to *E* isomers. Several criteria of internal and external validation to check the model robustness and predictivity were used to compare the results obtained by these approaches. We conclude that new trifluoromethyl-substituted 1,2,4-triazole derivatives with improved fungicidal activity against the *F. oxysporum* f. sp. *cucumerinum* could be best predicted by the SVM approach.

ACKNOWLEDGEMENTS

This project was financially supported by the Project No. 1.1 of the Institute of Chemistry Timisoara of the Romanian Academy (Timisoara, Romania).

CONFLICT OF INTEREST STATEMENT

The authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest, or non-financial interest in the subject matter or materials discussed in this manuscript.

REFERENCES

1. Koparir, M., Orek, C., Koparir, P. and Sarac, K. 2013, Spectrochim. Acta A, 105, 522.
2. Güniz Küçükgüzel, Ş. and Çıkla-Süzgün, P. 2015, Eur. J. Med. Chem., 97, 830.
3. Sahu, J. K., Ganguly, S. and Kaushik, A. 2013, Chin. J. Nat. Med., 11(5), 0456.
4. Roman, G. 2015, Eur. J. Med. Chem., 89, 743.
5. Koparir, M., Orek, C., Parlak, A. E., Söylemez, A., Koparir, P., Karatepe, M. and Dastan, S. D. 2013, Eur. J. Med. Chem., 63, 340.
6. Garrison, A. 2011, Chiral Pesticides: Stereoselectivity and Its Consequences, ACS Symposium Series Vol. 1085; Chapter 1, American Chemical Society, Washington, DC, 1.
7. Williams, A. 1996, Pestic. Sci., 46, 3.
8. Wu, Y. S., Lee, H. K. and Li, S. F. Y. 2001, J. Chromatogr. A, 912, 171.
9. Sekhon, B. S. 2009, J. Pestic. Sci., 34(1), 1.
10. Kurihara, N., Miyamoto, J., Paulson, G. D., Zeeh, B., Skidmore, M. W., Hollingworth, R. M. and Kuiper, H. A. 1997, Pure Appl. Chem., 69, 1335.
11. Ostrovskii, V. A. and Trifonov, R. E. 2014, Fluorine in Heterocyclic Chemistry, V. Nenajdenko (Ed.), Volume 1, Ed., Springer, Cham, Heidelberg, New York, Dordrecht, London, 495.
12. Wang, B.–L., Liu, X.–H., Zhang, X.–L., Zhang, J.–F., Song, H.–B. and Li, Z.–M. 2011, Chem. Biol. Drug. Des., 78, 42.
13. Kumar, S. S. and Kavitha, H. P. 2013, Mini-Rev. Org. Chem., 10, 40.
14. Siwek, A., Wujec, M., Dobosz, M., Jagiełło-Wójtowicz, E., Chodkowska, A., Kleinrok, A. and Paneth, P. 2008, Cent. Eur. J. Chem., 6(1), 47.
15. Suzuki, T., Funar-Timofei, S., Borota, A., Bora, A. and Avram, S. 2016, Journal of Toyo University, Natural Science, 60, 113.
16. Hawkins, P. C. D., Skillman, G. A., Warren, G. L., Ellingson, B. A. and Stahl, M. T. 2010, J. Chem. Inf. Model., 50, 572.
17. Hawkins, P. C. D. and Nicholls, A. 2012, J. Chem. Inf. Model., 52, 2919.
18. Wold, H. 1985, Encyclopedia of Statistical Sciences, S. Kotz and N. L. Johnson (Eds.), (Vol. 6), Wiley, New York, 581.
19. Höskuldsson, A. 1988, J. Chemometrics, 2(3), 211.
20. Hellberg, S., Wold, S., Dunn III, W. J., Gasteiger, J. and Hutchings, M. G. 1985, Quant. Struct.-Act. Relat., 4, 1.
21. Vapnik, V. 1995, The Nature of Statistical Learning Theory, Springer-Verlag, New York.
22. Yao, X. J., Panaye, A., Doucet, J. P., Zhang, R. S., Chen, H. F., Liu, M. C., Hu, Z. D. and Fan, B. T. 2004, J. Chem. Inf. Model., 44, 1257.
23. Chang, C. C. and Lin, C. J. 2001, LIBSVM: a library for support vector machine. Software. https://www.csie.ntu.edu.tw/~cjlin/libsvm/,
24. Zupan, J. and Gasteiger, J. 1995, Neural Networks for Chemistry and Drug Design, 2nd Ed., Wiley VCH, Weinheim.
25. Gramatica, P. 2013, Computational Toxicology, Volume II, Methods in Molecular Biology, B. Reisfeld and A. N. Mayeno (Eds.),Vol. 930, Springer, 499.
26. Todeschini, R., Consonni, V. and Maiocchi, A. 1999, Chemometr. Intell. Lab., 46, 13.
27. Chirico, N. and Gramatica, P. 2011, J. Chem. Inf. Model, 51, 2320.
28. Chirico, N. and Gramatica, P. 2012, J. Chem. Inf. Model., 52, 2044.
29. Roy, P. P., Paul, S., Mitra, I. and Roy, K. 2009, Molecule, 14, 1660.
30. Roy, K. and Mitra, I. 2012, Mini-Rev. Med. Chem., 12, 491.
31. Shi, L. M., Fang, H., Tong, W., Wu, J., Perkins, R., Blair, R. M., Branham, W. S. S., Dial, L., Moland, C. L. and Sheehan, D. M. 2001, J. Chem. Inf. Model., 41, 186.
32. Schüürmann, G., Ebert, R. U., Chen, J., Wang, B. and Kühne, R. 2008, J. Chem. Inf. Model., 48, 2140.

33. Consonni, V., Ballabio, D. and Todeschini, R. 2009, J. Chem. Inf. Model., 49, 1669.
34. Alexander, D. L. J., Tropsha, A. and Winkler, D. A. 2015, J. Chem. Inf. Model., 55, 1316.
35. Kaufman, L. and Rousseeuw, P. J. 1990, Finding Groups in Data: An Introduction to Cluster Analysis, New York, Wiley.
36. Todeschini, R. and Consonni, V. 2009, Molecular Descriptors for Chemoinformatics, Vol. I and II, Wiley-VCH, Weinheim.
37. Eriksson, L., Johansson, E., Kettaneh-Wold, N. and Wold, S. 2001, Multi- and Megavariate Data Analysis. Principles and Applications, Umetrics AB, Umeå, Sweden, 94.