# Modeling of Mannich Bases Fungicidal Activity by the MLR Approach

**Simona Funar-Timofei, Ana Borota, Alina Bora, Ramona Curpan, Sorin Avram**

*Computational Chemistry Department, Institute of Chemistry Timisoara of the Romanian Academy, 24Mihai Viteazul Bvd., 300223, Timisoara, Romania*

## ABSTRACT

In the present paper, we have carried out quantitative structure-fungicidal activity relationships analysis on a novel series of Mannich bases with trifluoromethyl-1,2,4-triazole and substituted benzylpiperazine moieties reported to have improved fungicidal activity against *Fusarium oxysporum f.sp. cucumerinum*. The chemical structures were energy minimized based on semiempirical quantum chemical method RM1. The molecular descriptors were calculated using the DRAGON, InstantJchem and ChemProp software. Several models for the prediction of fungicidal activity have been drawn up by using the multiple regression technique (MLR). The genetic algorithm approach was employed for variable selection method to search for the best ranking models. The predictive ability of the MLR models was validated using an external test set of 5 out of 18 molecules. The best MLR model was chosen by observing acceptable $r^2$, $r^2_{adj}$ and $q^2_{LOO}$ values, low residual errors and high Multi-Criteria Decision Making (MCDM) scores. The MLR equation suggests the positive impact of GETAWAY and edge adjacency matrix descriptors on the fungicidal activity. The high acidic character of the molecule increase the fungicidal activity.

## INTRODUCTION

Triazoles are often used in pharmacology, medicine and agriculture, having a broad spectrum of biological activities such as antimicrobial, cytotoxic, antihistaminic, anticonvulsant, analgesic, anti-inflammatory, insecticidal, antimycotic, antimycobacterial, anticancer, antiprotozoal, antimalarial and anti-ulcer activity [1].

Molecules containing thiazole ring systems are important because of their low toxicity and excellent biological activity [2].

Triazoles undergo different types of reactions to yield other heterocyclic compounds, e.g., mannich bases, thioureas, thioethers, schiff bases, triazolothiadiazoles, triazolothiazines, triazolothiazepines and triazolothiadiazines. They are not only transition compounds but they are also very effective organic compounds [3].

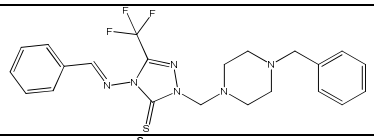
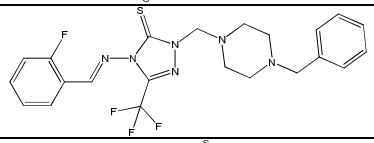
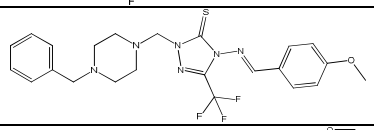
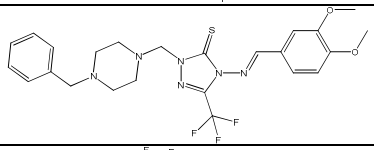
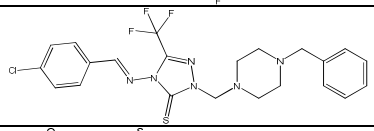
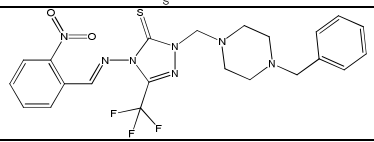
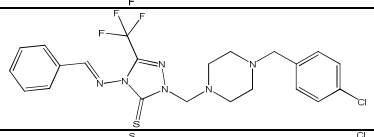
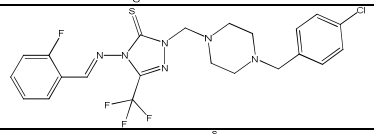
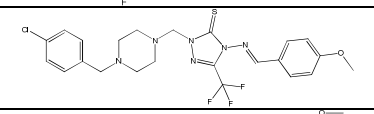
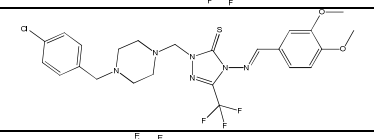
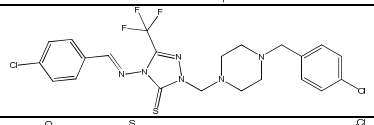
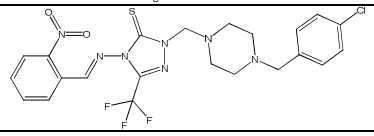
Triazole compounds have shown a great efficacy against antifungal infections. The mechanism of inhibition of fungal growth is well established. Thus, the azoles antifungal action is performed in two steps: (i) inhibition of ergosterol synthesis, a major component of fungal membrane and (ii) the blocking of P450-dependent enzyme *i.e.*, lanodterol 14-α-demethylase (CYP 51) [4]. Triazole fungicides are widely used broad-spectrum fungicides that inhibit the sterol 14-α-demethylase, an enzyme involved in the biosynthesis of ergosterol [5].

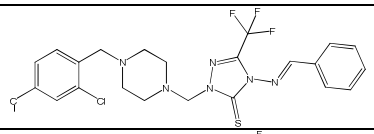
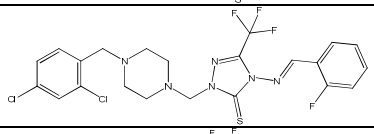
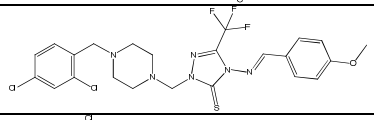
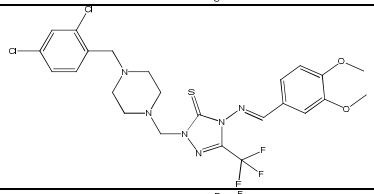
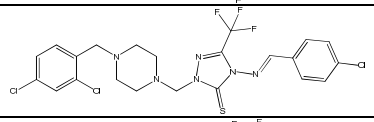
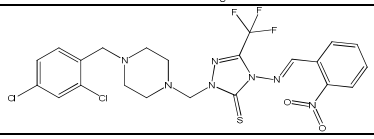
A series of novel 18 trifluoromethyl-substituted 1,2,4-triazole Mannich bases containing substituted benzylpiperazine ring have been synthesized and investigated for their herbicidal, fungicidal and plant growth regulators activity [6] (Table 1).

The current paper presents a quantitative structure-activity relationships study for this series of 1-[(4-substituted-benzylpiperazin-1-yl)methyl]-4-(substituted)benzylideneamino-3-trifluoromethyl-1H-1,2,4-triazole-5(4H)-thiones using multiple linear regression (MLR). These compounds were optimized using the RM1 semiempirical molecular orbital method

[7]. Descriptors calculated for the RM1 geometries were related to the mycelial growth inhibition activity against the *Fusarium oxysporum f. sp. cucumerinum* fungi test [6].

**Table 1**. The chemical structure of trifluoromethyl-substituted 1,2,4-triazole Mannich bases and their *Fusarium oxysporum f. sp. Cucumerinum* experimental relative inhibition rates (RIR)*

| No | Structure | RIR | HATS8u | R2u | EEig11r | Strongest basic pKa | $\Delta H_f$ kcal/mol |
|---|---|---|---|---|---|---|---|
| 1 |  | 0.101 | 0.402 | 2.004 | 2 | 7.74 | -3.69 |
| 2 |  | 0.804 | 0.271 | 2.103 | 2.167 | 7.74 | -47.91 |
| 3 |  | 0.187 | 0.426 | 1.972 | 2 | 7.74 | -81.18 |
| 4 |  | 0 | 0.428 | 1.966 | 2.167 | 7.74 | -86.31 |
| 5 |  | 0 | 0.398 | 2.032 | 2.167 | 7.74 | -73.76 |
| 6 |  | 0.402 | 0.383 | 2.015 | 2.332 | 7.74 | -40.62 |
| 7 |  | 0.509 | 0.341 | 2.082 | 2.167 | 6.76 | -36.52 |
| 8 |  | 0.719 | 0.271 | 2.129 | 2.333 | 6.76 | -16.05 |
| 9 |  | 0.604 | 0.405 | 1.995 | 2.167 | 6.76 | -49.72 |
| 10 |  | 0.401 | 0.357 | 2.061 | 2 | 6.76 | -83.93 |
| 11 |  | 0.303 | 0.398 | 2.03 | 2.167 | 6.76 | -17.43 |
| 12 |  | 0.502 | 0.401 | 2.009 | 2.332 | 6.76 | -11.65 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 13 |  | 0.708 | 0.355 | 2.109 | 2 | 6.01 | -19.18 |
| 14 |  | 0.826 | 0.286 | 2.116 | 2.167 | 6.01 | -64.77 |
| 15 |  | 0.504 | 0.388 | 2.014 | 2 | 6.01 | -58.49 |
| 16 |  | 0.705 | 0.416 | 2.041 | 2.167 | 6.01 | -93.71 |
| 17 |  | 0.607 | 0.389 | 2.043 | 2 | 6.01 | -45.43 |
| 18 |  | 0.608 | 0.386 | 2.034 | 2.167 | 6.01 | -48.14 |

* HATS8u represents everage-weighted autocorrelation of lag 8 / unweighted (GETAWAY descriptor); R2u - R autocorrelation of lag 2 / unweighted (GETAWAY descriptor); EEig11r - eigenvalue 11 from edge adj. matrix weighted by resonance integrals (Edge adjacency index); heat of formation ($\Delta H_f$) of the energy optimized structure.

## MATERIALS and METHODS

### Definition of target property and molecular structures

A series of 18 trifluoromethyl-substituted 1,2,4-triazole Mannich bases containing substituted benzylpiperazine ring (Table 1) was used, having the fungicidal *Fusarium oxysporum f. sp. Cucumerinum* relative inhibition rate (RIR, expressed in %) as dependent variable.

All geometries of the title fungicides were minimized with the semiempirical RM1 quantum chemical approach [7] using the semiempirical NDDO module of Schrödinger software (Schrödinger, LLC, New York, NY, 2015). The following quantum chemical descriptors were derived for the RM1 geometries: electronegativity, hardness, chemical potential, electrophilicity, HOMO and LUMO molecular orbital energies, heat of formation, dipole moment, molecular surface area, softness, maximum average local ionization energy on the molecular surface, minimum average local ionization energy on the molecular surface, mean average local ionization energy on the molecular surface, maximum electrostatic potential on the molecular surface, minimum electrostatic potential on the molecular, mean electrostatic potential on the molecular surface, electrophilic superdelocalizability, nucleophilic superdelocalizability, radical superdelocalizability, atom self polarizability. The outlines of the calculated quantum chemical parameters provide additional information about the activity of the studied compounds.

Structural 0D, 1D, 2D and 3D descriptors were calculated for the lowest energy compounds 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=6738) software.
The variables were normalized using the following equation (1):

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

where for each variable *m*, $XT_{mj}$ and $X_{mj}$ are the values *j* for the variable *m* after and before scaling respectively, $\overline{X}_m$ is the mean and $S_m$ the standard deviation of the variable.
Structural descriptors were correlated with the fungicide relative inhibition rate by multiple linear regression (MLR). MLR calculations were combined with a genetic algorithm for variable selection included in the QSARINS v.2.2 program [8]. The RQK fitness function, with leave-one-out cross-validation correlation coefficient was used as constrained function to be optimized. The dataset was divided into training set and a randomly selected (30% of the total number of compounds) test set. Compounds 7, 9, 10, 13, 18 (Table 1) were included in the test set. Validation is a crucial aspect of any quantitative structure–activity relationship (QSAR) analysis [9, 10]. In this light, the developed MLR models were validated using internal and external validation.

**Model validation**

All the statistical tests were performed at a significance level of 5 %. In MLR models, outliers were detected by a value of residual greater than 2.5 times, the value of standard error in calculation.
For internal validation results several measures of robustness were employed: leave-one-out cross-validation ($Q^2_{LOO}$), Y-scrambling and $Q^2_{LMO}$ leave-more-out (LMO) cross-validation (carried out for 30% of data out of training, each run).
Y-scrambling testing was repeated 2000 times. It is used for checking the robustness of a QSAR model and the statistical significance of the estimated predicted power. Satisfactory leave-one-out cross-validation values are stable and predictive if validated by the leave-more-out (LMO) procedure.
The data over fitting and model applicability was controlled by comparing the root-mean-square errors of training ($RMSE_{tr}$) and validation ($RMSE_{ext}$) sets. To test the predictive power of the model, several parameters were calculated: $Q^2_{F1}$ [11], $Q^2_{F2}$ [12], $Q^2_{F3}$ [13], $RMSE_{ext}$, $MAE_{ext}$ (mean absolute error for test set) and the predictive $r^2$ ($r^2_{pred}$) test [14]. It is considered that for a predictive QSAR model, the value of $r^2_{pred}$ should be higher than 0.5.
The Multi-Criteria Decision Making (MCDM) [15] is a technique that summarizes the performances of a certain number of criteria simultaneously, as a single number (score) between 0 and 1. A desirability function, takes values ranging from 0 to 1 (where 0 represents the worst validation criteria value and 1 the best) and is associated to every validation criteria. The geometric average of all the values obtained from the desirability functions gives the MCDM value. The ‚MCDM all' scores were calculated using all the criteria: fitting, cross-validated and external and were used to choose the best MLR models.

**RESULTS AND DISCUSSION**

A training set of 12 compounds and five test compounds (no.: 7, 9, 10, 13, 18) were used to build the models and to measure their performances. Compound 2 was found as outlier and was excluded from the final MLR models. Starting from all calculated descriptors several one and two descriptor models were generated (Table 2). Structural parameters derived from the

InstantJChem, Dragon and ChemProp programs and quantum chemical descriptors obtained from the RM1 geometries were employed in the MLR calculations. Variable selection was carried out by the genetic algorithm, using the leave-one-out fit criterion as constrained function to be optimized. Several fitting and predictability criteria were employed for model validation (see Tables 2 and 3). Satisfactory MLR models were obtained. Good fitting results were obtained for all MLR models. The predictive ability of models 3 and 4 is acceptable (except the $Q^2_{F2}$ value), the "MCDM all" scores indicating as satisfactory models 3 and 4, too.

**Table 2**. Internal validation parameters of the MLR models (training set)

| Model | Variables | $r^2_{training}$ | $r^2_{adj}$ | $q^2_{LOO}$ | $RMSE_{tr}$ | $MAE_{tr}$ | $r^2_{scr}$ | $q^2_{scr}$ | $q^2_{LMO}$ | MCDM all | F |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Strongest basic pKa HATS8u | 0.839 | 0.803 | 0.735 | 0.110 | 0.095 | 0.188 | -0.494 | 0.699 | 0 | 23.38 |
| 2 | Strongest basic pKa R2u | 0.823 | 0.783 | 0.715 | 0.116 | 0.095 | 0.180 | -0.499 | 0.667 | 0 | 20.86 |
| 3 | Strongest basic pKa EEig11r | 0.818 | 0.777 | 0.683 | 0.117 | 0.105 | 0.183 | -0.468 | 0.636 | 0.610 | 20.16 |
| 4 | Strongest basic pKa | 0.705 | 0.675 | 0.583 | 0.149 | 0.127 | 0.092 | -0.312 | 0.572 | 0.663 | 23.87 |

* $r^2_{training}$ -correlation coefficient; $r^2_{adj}$ -adjusted correlation coefficient; $q^2_{LOO}$ - leave-one-out cross-validation correlation coefficient; $RMSE_{tr}$-root-mean-square errors; $MAE_{tr}$-mean absolute error; $r^2_{scr}$ - correlation coefficient of the randomized responses; $q^2_{scr}$ - cross-validation correlation coefficient of the randomized responses; $q^2_{LMO}$ -leave-more-out cross-validation correlation coefficient; MCDM all-Multi-Criteria Decision Making scores using all the fitting, cross-validated and external criteria; F-Fischer test.

**Table 3**. External validation parameters of the MLR models (test set)

| Model | $Q^2_{F1}$ | $Q^2_{F2}$ | $Q^2_{F3}$ | $RMSE_{ext}$ | $MAE_{ext}$ | $r^2_{pred}$ |
|---|---|---|---|---|---|---|
| 1 | 0.699 | -0.030 | 0.853 | 0.105 | 0.074 | 0.699 |
| 2 | 0.537 | -0.583 | 0.774 | 0.131 | 0.102 | 0.537 |
| 3 | 0.731 | 0.081 | 0.869 | 0.100 | 0.092 | 0.731 |
| 4 | 0.811 | 0.352 | 0.908 | 0.084 | 0.072 | 0.810 |

* $Q^2_{F1}$, $Q^2_{F2}$, $Q^2_{F3}$ -external validation parameters; $RMSE_{ext}$-root-mean-square errors; $MAE_{ext}$-mean absolute error; $r^2_{pred}$ -predictive $r^2$

The best MLR model was chosen by observing the acceptable $r^2_{training}$, $r^2_{adj}$, $q^2_{LOO}$ and $r^2_{pred}$, values, high 'MCDM all' scores and low residual errors. Based on these criteria, the best MLR model could be considered equation 3 (Table 2):

$$RIR = 0.583(\pm 0.07) - 0.553(\pm 0.09)\text{Strongest basic pKa} + 0.2445(\pm 0.10)\text{EEig11r}$$

$$N_{training} = 12 \quad N_{test} = 5 \quad r^2_{training} = 0.818 \quad SEE = 0.135 \quad r^2_{adj} = 0.777 \quad q^2_{LOO} = 0.683 \quad q^2_{LMO} = 0.636$$

where: SEE represents the standard error of estimates, F – the Fischer test

The differences between $r^2_{training}$ and $r^2_{adj}$ of 0.0406, between $r^2_{training}$ and $q^2_{LOO}$ of 0.1345, and between $q^2_{LOO}$ and $q^2_{LMO}$ of 0.0474, indicate that model 3 is robust and has low over fitting effects. The low differences between the root-mean-square errors and between the mean absolute errors of the training and validation sets point to good fitting results and a robust model ($RMSE_{tr}$–$RMSE_{ext}$ = 0.017; $MAE_{tr}$–$MAE_{ext}$ = 0.013).

In order to check the reliability of the proposed equation, the observed versus predicted activities RIR values according to the QSAR equation using molecular descriptors, the Williams and the Y-scramble plots predicted by the MLR 3 model are outlined in Figures 1, 2 and 3, respectively.

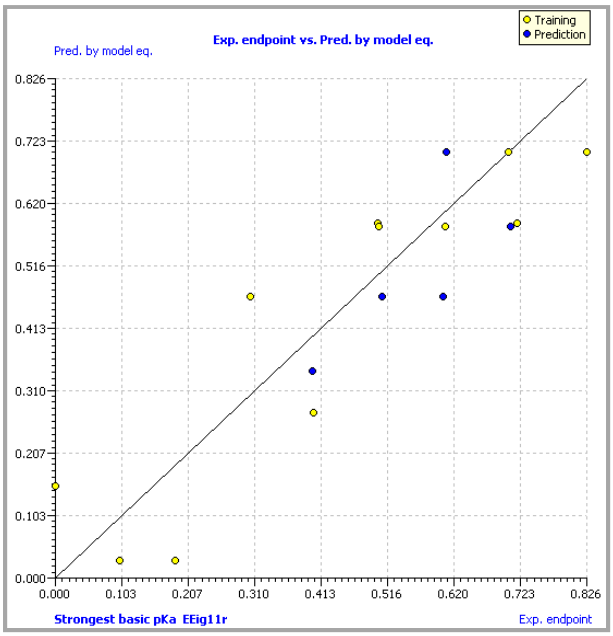



**Fig. 1**. Experimental *versus* predicted RIR values for the MLR3 model (Table 2).

Generally, the Williams plot is used to identify compounds with the greatest structural influence ($h_i > h^*$; $h_i$ =leverage of a given chemical; $h^*$= the warning leverage) in developing the model. The Williams plot for the training set presented in Figure 2, establishes applicability domain of the model within ±2.5σ and a leverage threshold h* = 0.750. It is obvious from Figure 2 that all the compounds in the dataset are within the applicability domain of the model.

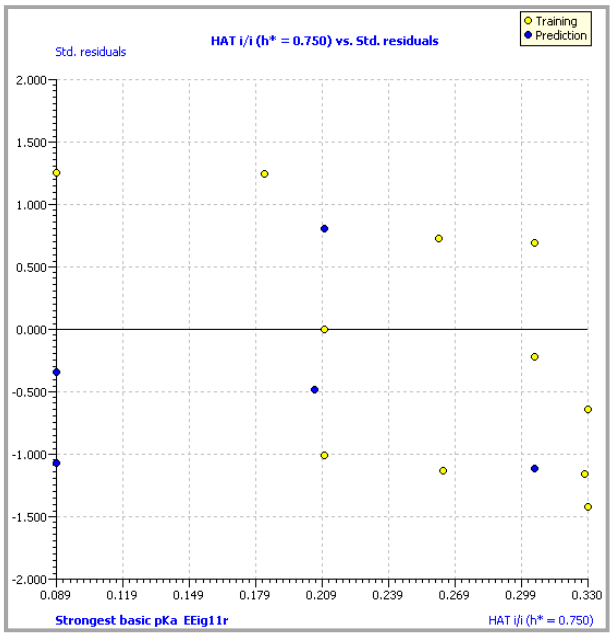



**Fig. 2**. Williams plot predicted by the MLR3 model (Table 2).

Y-scramble test was verified if the developed QSAR model is robust and not derived due to chance. The models are expected to have significant low scrambled $r^2$ ($r^2_{scr}$) and cross-validated $q^2$ ($q^2_{scr}$) values for several trials, which confirm the robustness of the developed models. From Figure 3 one can observe that in case of all the randomized models, the values of $r^2_{scr}$ and $q^2_{scr}$ were < 0.5. The low calculated $r^2_{scr}$ and $q^2_{scr}$ values (Table 2, Figure 3) indicate no chance correlation for the chosen model.

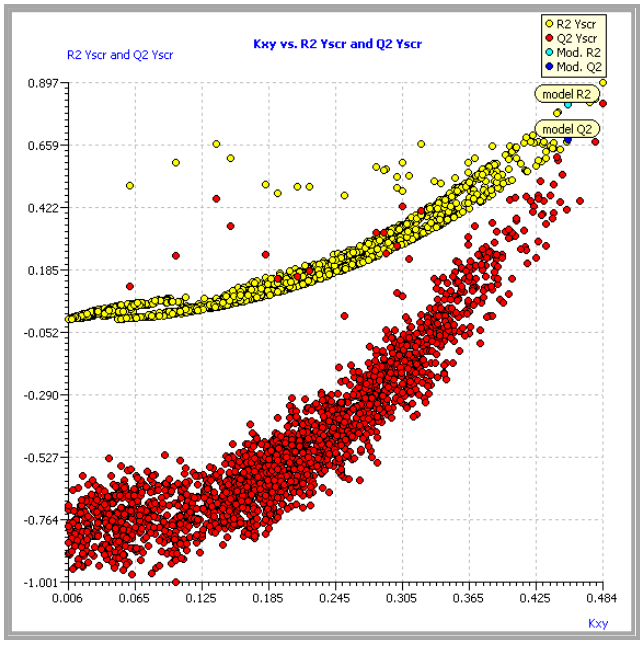



**Fig. 3**. Y-scramble plots for the MLR 3 model.

The predictive ability of the MLR models 3 and 4 is acceptable, according to the $Q^2_{F1}$, $Q^2_{F3}$ and $r^2_{pred}$ values, model 4 having lower fitting results compared to model 3.

## CONCLUSIONS

In this study we developed MLR models for a series of trifluoromethyl-1,2,4-triazole derivatives with fungicide activity against *Fusarium oxysporum f.sp. cucumerinum*. Cross-validation (LOO and LMO), 'MCDM all' scores, *y*-scrambling test and applicability domain analysis validate the internal and external predictabilities of the models developed using the training and test sets. The *y*-randomization test outcomes ensure that the developed MLR model is robust and not derived merely due to chance. Moreover, the applicability domain evaluation confirms that the developed model is reliable to make predictions, which were checked by several external validation criteria.

The chosen regression equation 3 indicates that low values of the 'strongest basic pKa' descriptor (more acidic fungicides) and high values of the EEIG11r descriptor increase the RIR values, respectively the fungicide activity.

We conclude that GETAWAY and edge adjacency matrix descriptors provide the highest contribution to the fungicidal activity for the data set studied herein, the acidic ability influencing the fungicide inhibition rate.

## ACKNOWLEDGEMENT

This project was financially supported by the Project No. 1.1 of the Institute of Chemistry of Romanian Academy, Timisoara. The authors are indebted to the Chemaxon Ltd., Prof. Paola Gramatica from The University of Insubria (Varese, Italy) and Prof. Gerrit Schüürmann from Helmholtz Centre for Environmental Research (UFZ, Leipzig, Germany) for giving access to their software.

## REFERENCES

[1]. Kumar S.S; Kavitha H.P. (2013). Synthesis and Biological Applications of Triazole Derivatives - A Review, *Mini-Reviews in Organic Chemistry*, 10(1), p. 40-65.

[2]. Qin X., Yu H. B., Dai H., Qin Z. F., Zhang X., Bing G. F., Wang T. T., Fang J. X. (2010). Synthesis and plant-growth regulatory activities of novel imine derivatives containing 1H-1,2,4-triazole and thiazole rings. *Chinese Chem. Lett.*, 21, p. 283–286.

[3]. Güniz Küçükgüzel S., Çıkla-Süzgün P. (2015). Recent advances bioactive 1,2,4-triazole-3-thiones. *Eur. J. Med. Chem*. 97, p. 830-870.

[4]. Khana I. A., Ahmada M., Aslamb S., Saifc M. J., Zahoora A. F., Raza Naqvia S. A., Mansha A. (2015). Recent advances in the synthesis of triazole derivatives, *AFINIDAD LXXII, Enero - Marzo*, 569, p.64-77.

[5]. Buerge J.I., Poiger T., Buser H.R. (2006). Influence of pH on the stereoselective degradation of the fungicides epoxiconazole and cyproconazole in soils. *Environ. Sci.Technol*., 40, p. 5443–5450.

[6]. Wang B.–L., Liu X.–H., Zhang X.–L., Zhang J.–F., Song H.–B., Li Z.–M. (2011). Synthesis, structure and biological activity of novel 1,2,4-triazole Mannich bases containing a substituted benzylpiperazine moiety. *Chem. Biol. Drug. Des*., 78, p. 42–49.

[7]. Rocha G.B., Freire R.O., Simas A.M., Stewart J.J.P. (2006). RM1: a Reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I. *J. Comput. Chem*., 27(10), p. 1101-1111.

[8]. Gramatica P., Chirico N., Papa E., Cassani S., Kovarich S. (2013). QSARINS: A new software for the development, analysis, and validation of QSAR MLR models. *J. Comput. Chem*., 34, p. 2121–2132.

[9]. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models. (2007), OECD Environment Health and Safety Publications Series on Testing and Assessment No. 69. OECD: Paris, (http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=env/jm/mono%282007%292)

[10]. Topliss J. G., Edwards R. P. (1979). Chance factors in studies of quantitative structure-activity relationships. *J. Med. Chem*., 22, p.1238–1244.
[11]. 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. (2001). QSAR models using a large diverse set of estrogens. *J Chem. Inf. Model*. 41, p.186–195.
[12]. Schüürmann G., Ebert R.U., Chen J., Wang B., Kuhne R. (2008) External validation and prediction employing the predictive squared correlation coefficient test set activity mean vs training set activity mean. *J. Chem. Inf. Model*, 48, p. 2140–2145.
[13]. Consonni V., Ballabio D., Todeschini R. (2009). Comments on the definition of the Q2 parameter for QSAR validation. *J. Chem. Inf. Model*, 49, p. 1669–1678.
[14]. Roy P.P., Paul S., Mitra I., Roy K. (2009). On two novel parameters for validation of predictive QSAR models. *Molecules, 14*, p. 1660-1701.
[15]. Keller H.R., Massart D.L., Brans J.P. (1991). Multicriteria decision making: a case study. *Chemom. Int.Lab. Syst*., 11**,** p.175-189.