Buletinul USAMV-CN, nr. 62/2006 (35-40)
ISSN 1454-2382

MOLECULAR DESCRIPTORS FAMILY ON STRUCTURE-ACTIVITY RELATIONSHIPS: MODELING HERBICIDAL ACTIVITY OF SUBSTITUTED TRIAZINES CLASS

Bolboacă SD[1], L Jäntschi [2]

[1]"Iuliu Hatieganu" University of Medicine and Pharmacy, 13 Emil Isac, 400023 Cluj-Napoca, Romania, sbolboaca@umfcluj.ro
[2]Technical University of Cluj-Napoca, 15 Constantin Daicoviciu, 400020 Cluj-Napoca, Romania, lori@academicdirect.org

Key words
Molecular Descriptor Family on Structure-Activity Relationships (MDF-SAR); Herbicidal Activity; Triazines; Multiple Linear Regressions (MLR)

Summary
Herbicidal activity of a set of thirty 1,3,5-substituted-triazines were studied using an original structure-activity relationships approach. The cross-validation leave-one-out correlation score, the training vs. test analysis, and the model stability sustained the prediction ability of the best performing multi-varied model with four variables. The comparison with the previous reported model was performed by the use of correlated correlation analysis. The obtained multi-varied MDF-SAR model with four-descriptors shows that the herbicidal activity of 1,3,5-substituted-triazines is of geometrical and topological nature and is strongly depended on partial charges and number of directly bonded hydrogen's.

INTRODUCTION

The Quantitative Structure-Activity Relationships is use today for finding the link between the activity and structure of chemical compounds in order to obtain new compounds with better properties, lowest expenses, and without time-consuming experiments [Peijnenburg, 1994].

The herbicidal activity of some 1,3,5-substituted-triazines, heterocyclic ring structures analogous to the six-members benzene ring with three carbon from positions 1, 3 and 5 replaced by nitrogen, were previously studied using orthogonalized molecular connectivity indices [Soskic at all, 1996] and topological substituent descriptors [Diudea at all, 2002]. The models and their statistical characteristics reported by Diudea & all [Diudea at all, 2002] were:

$$\text{Est } pI_{50} = 9.614 - 0.153 \cdot X_5 - 58.888 \cdot 1/V_5 - 2.430 \cdot 1/N_3$$

$$n = 30;\ r^2 = 0.9694;\ F = 274.3;\ r_{loo} = 0.9778$$

where X_5 = topological descriptor for substituent number 5, V_5 = fragmental volumes of the substituent in the position 5 (cm^3/mol); N_3 = total number of hydrogen's in the substituent 3; r^2 = squared correlation coefficient; F = Fisher parameter; and r_{loo} = squared correlation coefficient obtained by leave-one-out analysis.

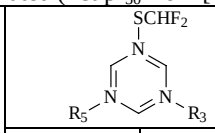
According with the concepts of quantitatively correlating structure of compounds with their biological activities [Kumar, 2001], starting from the successful results obtained by an original molecular descriptors family on structure-activity relationships (MDF-SAR) [Jäntschi, 2004; Bolboacă & Jäntschi, 2005; Bolboacă at all, 2006; Jäntschi at all, 2005] herbicidal activity of a set of thirty 1,3,5-substituted-triazines was modeled by the use of

MDF-SAR methodology, and estimation and prediction abilities of the multi-varied models were analyzed.

MATERIAL AND METHOD

The inhibition activity of thirty 1,3,5-substituted-triazines on Chorella, express as pI_{50} (the negative logarithm of concentration required for 50% inhibition of Hill reaction), was taken from a previous study [Morita at all, 1987]. The abbreviation of the compounds, the subsistent in the positions 3 and 5 (R_3 and R_5 according with the below presented generic structure), the measured activity (pI_{50}), and previous estimated activity are in table 1.

Table 1. The substituent R_3 and R_5 of 1,3,5-triazines, measured (pI_{50}) and previous estimated (Est pI_{50} from [Diudea at all, 2002]) activity

Abb.			pI_{50}	$EstpI_{50}$
	R_3	R_5		
t_01	NH_2	NH_2	3.82	3.88
t_02	$NHCH_3$	NH_2	5.20	5.09
t_03	NHC_2H_5	NH_2	5.34	5.50
t_04	NH-i-C_3H_7	NH_2	5.83	5.70
t_05	$NHCH_3$	$NHCH_3$	6.01	6.10
t_06	NHC_2H_5	$NHCH_3$	6.39	6.51
t_07	NHC_3H_7	$NHCH_3$	6.75	6.71
t_08	NH-i-C_3H_7	$NHCH_3$	6.76	6.71
t_09	NHC_4H_9	$NHCH_3$	6.74	6.83
t_10	NH-s-C_4H_9	$NHCH_3$	6.76	6.83
t_11	NH-t-C_4H_9	$NHCH_3$	6.78	6.83
t_12	NHC_5H_{11}	$NHCH_3$	7.12	6.91
t_13	NHC_3H_7	NHC_2H_5	6.82	6.59
t_14	NHC_3H_7	NHC_2H_5	6.74	6.79
t_15	NH-i-C_3H_7	NHC_2H_5	6.89	6.79
t_16	NHC_3H_7	NHC_0H_5	6.95	6.91
t_17	NH-i-C_4H_9	NHC_2H_5	7.01	6.91
t_18	NH-s-C_4H_9	NHC_2H_5	6.87	6.91
t_19	NH-t-C_4H_9	NHC_2H_5	6.97	6.91
t_20	NHC_5H_{11}	NHC_2H_5	6.94	6.99
t_21	NHC_6H_{13}	NHC_2H_5	7.21	7.05
t_22	NHC_7H_{15}	NHC_2H_5	7.01	7.09
t_23	NHC_8H_{17}	NHC_2H_5	6.81	7.13
t_24	NHC_3H_7	NHC_3H_7	6.45	6.52
t_25	NHC_3H_7	NH-i-C_3H_7	6.75	6.65
t_26	NH-i-C_3H_7	NH-i-C_3H_7	6.75	6.65
t_27	NHC_4H_9	NH-i-C_3H_7	6.71	6.77
t_28	NH-s-C_4H_9	NH-i-C_3H_7	6.88	6.77
t_29	NH-t-C_4H_9	NH-i-C_3H_9	6.70	6.77
t_30	NHC_5H_{11}	NH-i-C_3H_7	6.69	6.85

The steps applied in MDF-SAR modeling [Jäntschi, 2005] were: (1) Sketch of the thirty 1,3,5-substituted-triazines compounds; (2) Creation of *triazines.txt* file (contain measured herbicidal activity for 1,3,5-substituted-triazines); (3) Generation of molecular descriptors family for studied 1,3,5-substituted-triazines; (4) Identification of MDF-SAR models; (5) Validation of MDF-SAR models; and (6) Analysis of the best performing MDF-SAR model in terms of estimation and prediction and comparison MDF-SAR model with previous

reported QSAR.

The MDF SAR methodology was applied for modeling of herbicidal activity of the substituted triazines class in order to find a relationship between information obtained from the compounds structure and their herbicidal activity.

The validation of the best performance multi-varied MDF-SAR model(s) was analyzed in training versus test sets by the use of ***Training vs. Test*** application [***, Training vs. Test Experiment, 2005]. The abilities of MDF-SAR models were analyzed by the use of a correlated correlation approach [Steiger, 1980] in which correlation coefficients obtained by MDF-SAR models were compared with correlation coefficients obtained by previous reported QSAR [Diudea at all, 2002].

RESULTS AND DISCUSSIONS

In MDF-SAR modeling of herbicidal activity of 1,3,5-substituted-triazines, from the total possible number of molecular descriptors (787968), 298462 proved to had real and distinct values. A number of 74467 MDF members were significantly different molecular descriptors. By the use of a series of MLR procedures, a pair of two descriptors, three descriptors and two pairs of two descriptors were correlated with measured herbicidal activity obtaining the best performing multi-varied models with two, three and respectively four descriptors. It was defined as best performing MDF-SAR model the one which obtained greater value for the squared correlation coefficient and for leave-one-out squared correlation coefficient. The best identified MDF-SAR models were:

- MDF-SAR model with two-variables:

$\hat{Y}_{2v} = 5.52 - 8112.2 \cdot iSMMWHg + 194.35 \cdot iSMmEQt$

- MDF-SAR model with three-variables:

$\hat{Y}_{3v} = 1.74 - 9261 \cdot iSMMWHg + 10.34 \cdot iAMdEHg + 3.89 \cdot INDRLQg$

- MDF-SAR model with four-variables:

$\hat{Y}_{4v} = 5.75 + 199 \cdot iSMmEQt - 9010 \cdot iSMMWHg - 0.071 \cdot LADmkQt + 2.86 \cdot INPRJQg$

where $\hat{Y}_i$ (i = 1,…,4) is the estimator of the herbicidal activity and *iSDRFHg*, *iSMmEQt*, *iSMMWHg iAMdEHg*,*LADmkQt*, and *INPRJQg* are MDF members.

The statistics associated with the MDF-SAR models are in table 2.

Table 2. Statistical characteristics of the best performing MDF SAR models

Characteristic (notation)	Value		
Number of descriptors used in the model (n)	4	3	2
Correlation coefficient (r)	0.994	0.991	0.987
Squared correlation coefficient (r^2)	0.988	0.983	0.975
Adjusted squared correlation coefficient (r^2_{adj})	0.987	0.981	0.973
Standard error (s_{est})	0.081	0.096	0.114
Fisher parameter (F_{est})	537**	501**	533**
Cross-validation leave-one-out correlation score ($r^2_{cv(loo)}$)	0.985	0.977	0.971
Fisher parameter for leave-one-out analysis (F_{pred})	409**	361**	449**
Standard error - leave-one-out analysis (s_{loo})	0.092	0.113	0.124
Model stability ($r^2 - r^2_{cv(loo)}$)	0.003	0.006	0.004

** $p < 0.001$

The squared correlation coefficient between each descriptor and measured herbicidal activity (r^2(d, pI_{50}), and the statistics of the MDF-SAR model with four-variables (express as coefficients of regression and associated 95% confidence interval (95%CI), standard error (StErr), t test parameter (t Stat) and its significance) are in table 3.

The best performing MDF-SAR model with four-variable combines the geometrical shape (*iSMMWH**g***, *INPRJQ**g***) as well as the topological shape of the molecules (*iSMmEQ**t***,

*LADmkQ**t***), and as atomic property the partial charges of the molecule (*iSMmE**Q**t*, *LADmk**Q**t*, *INPRJ**Q**g*) and the number of directly bounded hydrogen's (*iSMMW**H**g*). According with the best performing multi-varied MDF-SAR, model with four-variables, the herbicidal activity of studied compounds it is like to be of geometrical and topological nature and depend on the partial charges as well as by the directly bonded hydrogen's of the molecules.

Table 3. The MDF-SAR model with four-variables: results of MLR

	$r^2(d, pI_{50})$	Coefficient [95%CI]	StErr	t Stat
Intercept	-	5.75 [5.28, 6.23]	0.23	24.88*
iSMmEQt	0.6282	199 [161.27, 236.26]	18.21	10.92*
iSMMWHg	0.9138	-9006 [-9712, -8300]	342.6	-26.28*
LADmkQt	0.3224	-0.071 [-0.11, -0.03]	0.02	-4.05*
INPRJQg	0.1599	2.86 [1.69, 4.03]	0.57	5.04*

* $p < 0.05$

Ninety-nine percent of variation in herbicidal activity it is explainable by its linear relation with *iSMmEQt*, *iSMMWHg*, *LADmkQt*, and *INPRJQg* descriptors. Almost ninety-three percent of variation in herbicidal activity can be explainable by its linear relation with *iSMMWHg* descriptor and eighty-two percent by its linear relation with *iSMmEQt* descriptor. The values of squared correlation coefficient ($r^2 = 0.9885$) demonstrate the goodness of fit of the multi-varied MDFSAR model with four descriptors.

The power of the MDF-SAR model with four descriptors in prediction of the herbicidal activity of 1,3,5-substituted-triazines is demonstrate by the cross-validation leave-one-out correlation score ($r^2_{cv(loo)} = 0.9849$), procedure which did not take into consideration one molecule from the whole set. The stability of the best performing MDF-SAR model is give by the difference between the squared correlation coefficient and the cross-validation leave-one-out correlation score ($r^2 - r^2_{cv(loo)} = 0.0035$).

In order to evaluate the prediction ability of the MDF-SAR model with four-variables, the compounds were randomly split into two sets, training and test. A random routine pick out a specified number of compounds (*n*) from whole molecules (30), include them in the training set, and rebuild the MDF-SAR model. The prediction ability of the MDF-SAR model with four-descriptors was validated on 14 test sets (30 - *n*). The number of molecules in training sets varied from 10 to 23 (in test sets from 20 to 7) and the results are in table 4.

Table 4. The results of the training vs. test sets analysis using MDF-SAR model with four-descriptors

Coefficients					Training set			Test set			r_{tr}vs.r_{ts}
a_0	a_1	a_2	a_3	a_4	No_{tr}	r^2	F_{tr}	No_{ts}	r^2	F_{ts}	$F_{Z\text{-test}}$
5.32	284.8	-12604	-0.12	4.009	10	0.897	11*	20	0.994	18**	3.22**
5.63	214.6	-9261	-0.084	2.562	11	0.989	138**	19	0.986	245**	0.28†
5.83	184.2	-8488	-0.054	1.678	12	0.995	353**	18	0.968	90**	2.22*
6.28	158.8	-8404	-0.096	2.154	13	0.991	230**	17	0.971	79**	1.43†
5.85	201.4	-9220	-0.084	1.763	14	0.987	177**	16	0.985	143**	0.18†
5.74	188.2	-8563	-0.043	3.16	15	0.989	230**	15	0.987	126**	0.21†
5.91	185.1	-8690	-0.075	2.096	16	0.99	280**	14	0.953	38**	1.91*
5.91	187.7	-8759	-0.074	2.033	17	0.977	125**	13	0.989	168**	0.90†
5.94	184.9	-8717	-0.086	2.359	18	0.995	612**	12	0.972	44**	2.06*
5.59	210.9	-9184	-0.069	3.137	19	0.993	516**	11	0.97	43**	1.69*
5.92	175.7	-8284	-0.066	2.813	20	0.908	37**	10	0.995	108**	3.29**
5.63	199.6	-8739	-0.041	2.538	21	0.988	330**	9	0.99	55**	0.19†
6.07	172.8	-8713	-0.075	3.329	22	0.981	220**	8	0.992	69*	0.87†
5.84	196.9	-9169	-0.09	3.462	23	0.989	411**	7	0.987	21*	0.15†

** p value < 0.001; * 0.05 < p-value < 0.001; † p > 0.05

In table 4, was used the following generic equation: $a_0 + a_1 \cdot iSMmEQt + a_2 \cdot iSMMWHg + a_3 \cdot LADmkQt + a_4 \cdot INPRJQg$, and the results are express as squared correlation coefficients

(r^2_{tr} - for training set and r^2_{ts} - for test set), Fisher parameter and associated significance (less than 0.0001 if the value has one star (*) and between 0.0001 and 0.05 if the value has associated two stars (**)) for the MDF-SAR models, and the Fisher Z- test ($_{FZ\text{-}test}$) which test the null hypothesis that there were not significant differences between correlation coefficient obtained in training set and the correlation coefficient obtained in the associated test set.

The prediction ability of the best performing MDF-SAR model with four descriptors is sustained by the results obtained in training vs. test analysis. The difference between leave-one-out procedure and training vs. test procedure is represented by the omission of more than one compound in training versus test analysis. It was not used an independent set for validation of the MDF-SAR model with four variables because the whole sample of thirty 1,3,5-substituted-triazines was used just for generating the list of descriptors. The algorithm of descriptors list generation is strictly based on the structure of the compounds. Any time the algorithm is used, for the same compound, the list of descriptors is the same; thus, splitting the compounds into training and test sets is not useful in descriptors list generation. The average of the squared correlation coefficients obtained for test sets (0.9814) it is not statistically grater comparing with the average of the squared correlation coefficients obtained for training sets (0.9765) and sustained the prediction ability of the model.

Assessment of the MDF-SAR models were performed by the use of a correlated correlation analysis (Steiger's Z test), which took into consideration MDF SAR models with two-, three- and respectively four-variables and compared them, in terms of correlation coefficients, with previous reported QSAR [Diudea at all, 2002]. The results of comparison are in table 5.

Table 5. The results of comparison between MDF-SAR models and previous reported QSAR

Characteristic	Value		
Number of MDF descriptors	4	3	2
$r(pI_{50}, \hat{Y}_{MDF\text{-}SAR})$	0.994	0.991	0.987
$r(pI_{50}, \hat{Y}_{QSAR})$	0.985	0.985	0.985
$r(\hat{Y}_{MDF\text{-}SAR}, \hat{Y}_{!SAR})$	0.988	0.981	0.986
Steiger's Z parameter	2.828^*	$1.563^\dagger$	$0.651^\dagger$

* $p < 0.05$; † $p > 0.05$

In the best performing MDF-SAR model with four descriptors, even if the 3rd and 4th descriptors seems to be insignificant, the MLR model which took into consideration just 1st and 2nd descriptors did not obtained a squared correlation coefficient significant different comparing with previous reported model (Steiger's Z test parameter = 0.651, $p > 0.05$).

Comparing with the previous reported model [Diudea at all, 2002] which use topological descriptors, the correlation coefficient obtained with the multi-varied MDF-SAR model with four descriptors is significantly greater (see table 5), sustaining its ability in prediction of herbicidal activity of 1,3,5-substituted-triazine compounds.

A software which allows to used the accumulated knowledge through learning of behavior on MDF-SAR models was developed in order to be apply to the new 1,3,5-subtituted-triazine compounds. The software is free to be use at:

http://vl.academicdirect.org/molecular_topology/mdf_findings/sar/

and is able to predict the activity of interest of a compound based on a choused class of compounds, choused model and on a *.hin file of compound of interest.

Thus, by using of the software from above address, the herbicidal activity of new 1,3,5-substituted-triazine can be calculated without any experiments. Unfortunately, in the stage in that the research was performed based on the obtained models, was proved that the model is useful just to be use to obtain new compounds. In the future research we intended to explore the physicochemical nature of each descriptor, but, as it results from the manual of the

program these are very complex.

The future MDF-SAR study of herbicidal activity of a substituted triazines class will straighten on physicochemical properties of each descriptor and on mechanism of drug-descriptor interaction, in order to found the usefulness of MDF in exploring drug-action.

CONCLUSIONS

Based on the values of squared correlation coefficients, the values of leave-one-out squared correlation coefficients and according with the correlated correlation analysis, it can be consider that the best performing MDF-SAR is the model with four-variable.

According with the best performing multi-varied MDF-SAR, model with four-variables, the herbicidal activity of studied compounds it is like to be of geometrical and topological nature and depend on the partial charges as well as by the directly bonded hydrogen's of the molecules.

Even if it is a time-consuming method, the MDF-SAR methodology gives a solution in predicting the herbicidal activity of 1,3,5-substituted-triazines providing a stable and performing multi-varied MDF-SAR model with four descriptors.

ACKNOWLEDGEMENT

The research was in part supported by UEFISCSU Romania through project ET36/2005.

REFERENCES

- Bolboacă S., L. Jäntschi, 2005, Molecular Descriptors Family on Structure Activity Relationships 3. Antituberculotic Activity of some Polyhydroxyxanthones, Leonardo Journal of Sciences, 7, 58-64.
- Bolboacă S., Șt. Țigan, L. Jäntschi, 2006, Molecular Descriptors Family on Structure-Activity Relationships on anti-HIV-1 Potencies of HEPTA and TIBO Derivatives, Assa Reichert, George Mihalaş, Lăcrămioara Stoicu-Tivadar, Ştefan Schulz, Rolf Engelbrech (Eds.), Proceedings of the European Federation for Medical Informatics Special Topic Conference, April 6-8, Timişoara, Romania, p. 222-226.
- Diudea V.M., L. Jäntschi, L. Pejov, 2002, Topological Substituent Descriptors. Leonardo Electronic Journal of Practices and Technologies, 1, 1-18.
- Jäntschi L., 2004, Delphi Client - Server Implementation of Multiple Linear Regression Findings: a QSAR/QSPR Application, Applied Medical Informatics, 15, 48-55.
- Jäntschi L., 2005, Molecular Descriptors Family on Structure Activity Relationships 1. The review of Methodology, Leonardo Electronic Journal of Practices and Technologies, 6, 76-98.
- Jäntschi L., M.L. Ungureşan, S.D. Bolboacă., 2005, Integration of Complex Structural Information in Modeling of Inhibition Activity on Carbonic Anhydrase II of Substituted Disulfonamides, Applied Medical Informatics, 17, 12-21.
- Kumar A.D., 2001, Quantitative Structure-Activity Relationship (QSAR) Paradigm - Hansch Era to New Millennium, Mini. Rev. Med. Chem., 1(2), 187-195.
- Morita K., T. Nagare, Y. Hayashi, 1987, Quantitative structure-activity relationships for herbicidal 2-Difluoromethylthio-4,6-bis(monoalkylamino)-1,3,5-triazines, Agric. Biol. Chem., 51, 1955-1957.
- Peijnenburg W.J.G.M., 1994, Structure-Activity Relationships for Biodegradation: A Critical Review, Pure &Appl. Chem., 66(9), 1931-1941.
- Soskic M., D. Plavsic, N. Trinajstic, 1996, 2-Difluoromethylthio-4, 6-bis(monoalkylamino)-1, 3, 5-triazines as Inhibitors of Hill Reaction: A QSAR Study with Orthogonalized Descriptors, J Chem. Inf. Comput. Sci., 36, 146-150.
- Steiger J.H., 1980, Tests for comparing elements of a correlation matrix, Psychol. Bull., 87, 245-251.
- ***, Training vs. Test Experiment. ©2005, Virtual Library of Free Software [cited 2006 March]. Available from: URL: http://vl.academicdirect.org/molecular_topology/qsar_qspr_s/.