SHORT NOTES

INSIGHT INTO THE STRUCTURAL REQUIREMENTS OF SULFONYL DERIVATIVES BASED ON TWO AND THREE-DIMENSIONAL DESCRIPTORS: QSAR STUDIES

ABSTRACT

The present work provides the rationale to the changes in the structure to have more potent analogs sulfonyl derivatives were reported to possess potent activity for the angiotensin AT_1 receptor. We report here 2D QSAR and *k-nearest neighbor* molecular field analysis based model for sulfonylureas compounds as AT_1 receptor. The here *k-nearest neighbor* contour plots provided further understanding of the relationship between structural features of substituted sulfonyl derivatives and their activities which should be applicable to design newer potential AT_1 receptor .

Keywords: QSAR; Stepwise variable; PLS; AT_1 receptor

INTRODUCTION

The renin-angiotensin system (RAS) plays a key role in regulating cardiovascular homeostasis and electrolyte fluid balance in normotensive and hypertensive subjects[1]. Activation of the renin-angiotensin cascade begins with renin secretion from the juxtaglomerular apparatus of the kidney and culminates in the formation of the octapeptide angiotensin II, which then interacts with specific receptors present in different tissues[2]. Two basic types of receptors: the AT_1 receptor, responsible for the majority of effects attributed to this peptide, and the AT_2 receptor, with a functional role yet uncertain[3]. Angiotensins II are the regulation of blood pressure through vasoconstriction, the regulation of volemia through the stimulated release of vasopressin and aldosterone, which induces saline retention, and the regulation of the adrenocorticotropic hormone[4]. Quantitative structure activity relationship (QSAR) which has become an accepted tool for establishing quantitative relationship between biological activity and descriptors representing physicochemical properties of the compounds in a series using statistical methods and it helps to predict the biological activities of newly designed analogues contributing to the drug discovery processes[5]. Many different approaches to QSAR have been developed over the years [6,7] has led to the development of 3D structural descriptors and associated 3D QSAR methods. The optimum selection of variables can be achieved by combining statistical and stochastic search methods in conjunction with partial least square (PLS) analysis for model development[8,9]. We report here 2D QSAR and kNN-MFA-based stepwise variable selection QSAR models developed for sulfonylureas derivatives as angiotensin AT_1 antagonists. Sphere exclusion method of data selection and stepwise forward–backward method of variable selection[10] was found more suitable to generate these models. The best model can be used for predicting the biological activity of newly designed analogs prior to their synthesis.

MATERIALS AND METHODS

The molecular structure of the training set and test were sketched using V-life MDS (Molecular Design Suite) ™ 3.5 software supplied by V-life Sciences Technologies Pvt. Ltd., Pune, India[11]. QSAR studies were performed on HP computer having genuine Intel Pentium Dual Core Processor and Windows XP operating system.

Dataset and descriptor generation

In the present study a dataset of 31 molecules of sulfonylureas derivatives angiotensin AT_1 reported was taken for the study[12]. The biological activity values [IC_{50} (nM)] reported in literature were converted to their molar units and then further to negative logarithmic scale and subsequently used as the dependent variable for the QSAR analysis. Table I shows a summary of IC_{50} along with the structure of the 31 compounds in the series.

Generation of 2D/3D- Molecular Descriptors

Energy minimization and batch optimization using Merck Molecular Force Field[13], fixing Root Mean Square Gradients to 0.01 kcal/mol Å. Statistical 2D-QSAR modelling was carried out as describe in the VLifeMDS software. Total of 218 2D descriptors were calculated which encoded different aspects of molecular structure and consists of electronic, thermodynamic, spatial, and structural descriptors, e.g., retention index (chi), atomic valence connectivity index (chiV), path count, chain path count, dipole moment, different topological descriptors, element count, estate number, semi-empirical, molecular

weight, molecular refractivity, logP, and topological index.

Optimized molecules were aligned (Fig. 1) by template-based method[14] using the most active molecule as reference molecule 6 as a template. For calculation of 3D field descriptor values, using Tripos force field, electrostatic, steric and hydrophobic field types, with cut-offs of 10.0 and 30.0 kcal/mol, respectively, were selected and charge type was selected as by Gasteiger and Marsilli. This resulted in calculation of 5100 field descriptors (1700 for each electrostatic, steric and hydrophobic) for all the compounds in separate columns.

In Stepwise forward-backward variable selection method (SW)[15,16], the search procedure begins with developing a trial model step by step with a single independent variable and to each step, independent variables are added one at a time, examining the fit of the model by using the PLS cross-validation procedure. Thus, the model is repeatedly altered from the previous one by adding or removing a predictor variable in accordance with the 'stepping criteria' (in this case, F =4 for inclusion and F = 3.99 for exclusion for the forward– backward selection method).

Internal validation was carried out using leave-one-out (q^2, LOO) method[17]. The q^2 was calculated using the equation which describes the internal stability of a model.

$$q^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{mean})^2}$$

Where and are the actual and predicted activity of the i^{th} molecule in the training set, respectively, and y_{mean} is the average activity of all molecules in the training set. The predictive ability of the selected model was also confirmed by external validation of test set compounds which is also denoted with pred_r^2. The pred_r^2 value is calculated as follows.

$$\text{pred\_}r^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - y_{mean})^2}$$

Where and are the actual and predicted activity of the i^{th} molecule in the training set, respectively, and y_{mean} is the average activity of all molecules in the training set.

RESULTS AND DISCUSSION

In this work, in order to analyze the structure activity relationship of these inhibitors and investigate the structural requirements for AT_1 receptor. The two-dimensional models were developed using SW-PLS method, which show both significant statistical quality and predictive ability.

pIC_{50}= +0.9136(±0.2013) T_2_N_1-0.5439(±0.1952) SulfursCount-0.2179(±0.0639) T_C_O_6 + 0.2157 (±0.0907) T_T_O_3+ 0.0910(±0.0352) T_2_N_6

$N_{training}$ = 24, N_{test}= 7, r^2= 0.7632, q^2= 0.7061, F test =46.653, r^2_se = 0.3044, q^2_ se =0.5858, pred_r^2 = 0.7110.

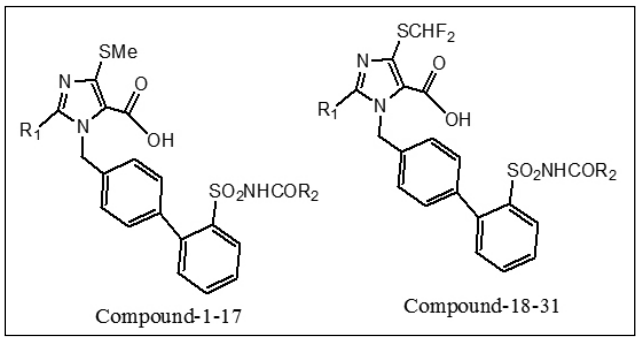
Model 1 with PLS method coefficient of determination (r^2) = r^2=0.7632, is capable of explaining 76 % of variance in the observed activity values. In accordance with model 1, shows a positive correlation with T_2_N_1, T_T_O_3, T_2_N_6 and negative contribution SulfursCount, T_C_O_6. The graph for model 1 showing observed versus predicted activity is presented in Fig.2 (a) and contribution charts of the descriptors for the model 2(b). As a positive contributing T_2_N_1, T_2_N_6 (count of number of double bounded atoms (i.e. any double bonded atom, T_2) separated from nitrogen atom by 1 and 6 bonds respectively plays an important role in determining activity. The model reveals that the descriptors T_T_O_3 double bounded atoms (i.e. any double bonded atom, T_T) separated from oxygen atom by 3 bonds in a molecule plays an important role in determining activity. Descriptor like sulfurs count number of sulphur atoms in a compound, T_C_O_6 (the count of number of carbon atoms separated from oxygen atom by 6 bonds in the molecule) are inversely proportional to the biological activity indicated that good values leads to activity. The correlation matrix is shown in Table III which shows good correlation of selected parameters with biological activity. The observed and predicted pIC_{50} along with values are shown in Table II.

pIC_{50}=+1.3261(±0.2822) T_2_N_1 -0.3682(±0.0603) XlogP -0.3408(±0.0786) T_2_N_3 + 0.0452(±0.0065) T_2_C_6

$N_{training}$ = 24, N_{test}= 7, r^2= 0.7704, q^2= 0.6941, F test =29.069, r^2_se = 0.3090, q^2_ se =0.3862, pred_ r^2=0.6795

Model 2 explains 77.4 % (r^2= 0.7704) of the total variance in the training set as well as it has internal (q^2) and external (pred_r^2) predictive ability of 70% and 68%

Table I: Structure and activity of Imidazolyl biphenyl sulfonylureas



Sr. No	R^1	R^2	IC_{50} [a]	pIC_{50} [b]	Traning/Test Set
1	Bu	NHMe	1.9	8.721	Traning Set
2	Bu	NHEt	0.2	9.698	Traning Set
3	Bu	NHPr	0.5	8.920	Traning Set
4	Bu	NHBu	0.2	8.657	Test Set
5	Bu	$NHCH_2$cHexyl	0.1	10	Traning Set
6	Bu	$NHCH_2Ph$	0.1	10	Traning Set
7	Bu	$NHCHPh_2$	0.4	9.397	Traning Set
8	Bu	$NHPh_2$	1.2	8.920	Test Set
9	Pr	NHcHexyl	0.5	9.301	Traning Set
10	Pr	$NHCH_2$cHexyl	0.1	10	Traning Set
11	Pr	$NHCH_2Ph$	0.2	9.698	Test Set
12	Pr	NHPh	2.3	8.638	Traning Set
13	Pr	NHPh-4-Me	1	9	Traning Set
14	Pr	$NHCH_2$-2-Thienyl	0.1	10	Traning Set
15	Pr	$NHCH_2$cPentyl	0.1	10	Traning Set
16	Et	$NHCH_2Ph$	0.1	10	Traning Set
17	Et	$NHCH_2$cHex	1.3	8.886	Test Set
18	Bu	Pr	0.5	9.301	Traning Set
19	Bu	CH_2Ph	0.15	9.823	Traning Set
20	Bu	CH_2cHexyl	0.2	9.698	Traning Set
21	Pr	CH_2cHexyl	0.07	10.154	Test Set
22	Pr	CH_2Ph	0.2	9.698	Traning Set
23	Pr	CHMePh	0.3	9.522	Traning Set
24	Pr	CH_2-Ph-2Cl	0.3	9.522	Traning Set
25	Pr	CH_2-Ph-4F	0.1	10	Test Set
26	Pr	Piperonyl	0.3	9.522	Traning Set
27	Pr	CH_2-2-Naphthyl	0.1	10	Traning Set
28	Pr	CH_2CH2Ph	0.1	10	Traning Set
29	Pr	$CH_2CH_2CH_2Ph$	0.7	9.154	Test Set
30	Pr	CH_2-2-Thienyl	0.07	10.154	Traning Set
31	Pr	CH_2CH_2-2-Thienyl	0.03	10.522	Traning Set

[a] *IC_{50} of specific binding of [^{125}I- Ang $IIAT_1$ receptor rat liver*

[b] *-log IC_{50} to generate equation.*

Table II: Comparative observed and predicted activities Imidazolyl biphenyl sulfonylureas derivatives by best QSAR models

pIC_{50}	2D QSAR model 1		2D QSAR model 2		3D QSAR model 3	
	Pred.	Res.	Pred.	Res.	Pred.	Res.
8.721	8.546	0.175	8.658	0.063	8.605	0.116
9.698	9.512	0.186	9.712	-0.014	9.666	0.032
8.92	8.911	0.009	8.936	-0.016	8.889	0.031
8.657	8.621	0.036	8.683	-0.026	8.569	0.088
10	9.992	0.008	9.983	0.017	9.962	0.038
10	9.981	0.019	10.12	-0.12	9.966	0.034
9.397	9.332	0.065	9.223	0.174	9.416	-0.019
8.92	8.836	0.084	8.856	0.064	8.967	-0.047
9.301	9.258	0.043	9.354	-0.053	9.342	-0.041
10	9.956	0.044	10.34	-0.34	10.68	-0.68
9.698	9.512	0.186	9.736	-0.038	9.802	-0.104
8.638	8.614	0.024	8.564	0.074	8.436	0.202
9	9.263	-0.263	9.286	-0.286	8.991	0.009
10	9.993	0.007	9.906	0.094	9.036	0.964
10	10.076	-0.076	9.893	0.107	10.08	-0.08
10	10.132	-0.132	10.361	-0.361	9.988	0.012
8.886	8.796	0.09	8.903	-0.017	8.966	-0.08
9.301	9.463	-0.162	9.186	0.115	9.044	0.257
9.823	8.796	1.027	9.783	0.04	9.696	0.127
9.698	9.589	0.109	9.555	0.143	9.777	-0.079
10.154	10.176	-0.022	10.091	0.063	10.099	0.055
9.698	9.763	-0.065	9.736	-0.038	9.586	0.112
9.522	9.496	0.026	9.621	-0.099	9.599	-0.077
9.522	9.587	-0.065	9.488	0.034	9.468	0.054
10	9.936	0.064	9.886	0.114	9.983	0.017
9.522	9.412	0.11	9.373	0.149	9.636	-0.114
10	10.22	-0.22	9.949	0.051	10.089	-0.089
10	10.08	-0.08	9.976	0.024	9.961	0.039
9.154	9.325	-0.171	9.268	-0.114	9.242	-0.088
10.154	10.084	0.07	10.232	-0.078	9.913	0.241
10.522	10.547	-0.025	10.465	0.057	10.632	-0.11

Table 3. Correlation matrix between descriptors present in the 2D QSAR model 1.

Parameter	T_2_N_1	T_T_O_3	T_2_N_6	SulfursCount	T_C_O_6
T_2_N_1	1.0000				
T_T_O_3	0.3638	1.0000			
T_2_N_6	0.2971	0.5936	1.0000		
SulfursCount	-0.1981	-0.4219	-0.6264	1.0000	
T_C_O_6	0.1876	0.5411	0.6487	0.7764	1.000

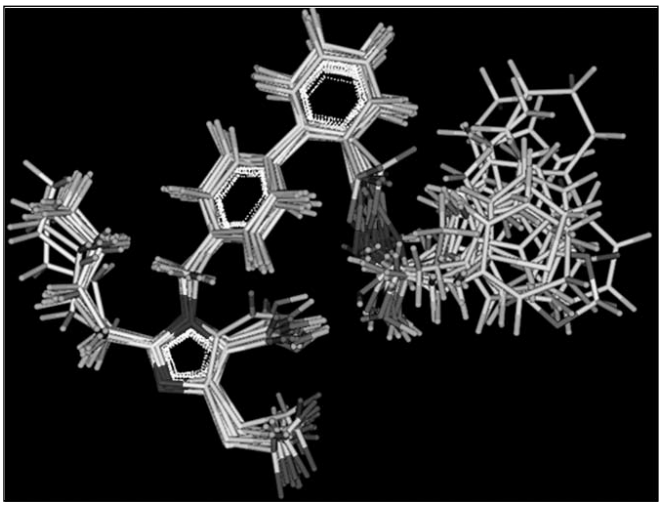


Fig. 1: Template based alignment of molecules

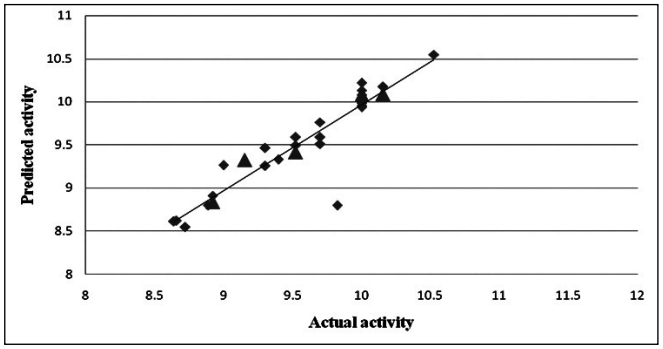


Fig. 2 (a): Graphical representation of observed vs. predicted activity Model-1

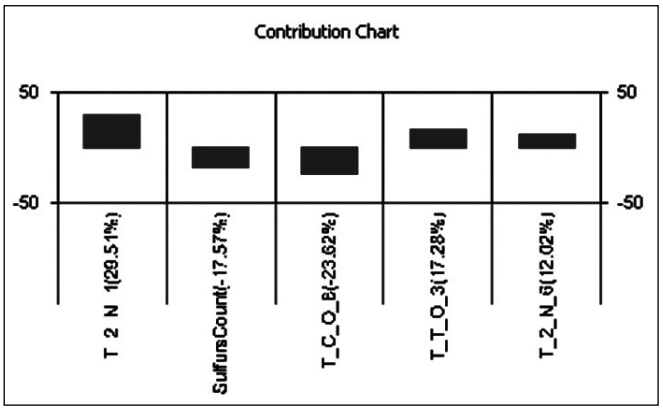


Fig. 2(b): Contribution chart of various descriptors in biological activity for Model-1

respectively. In the QSAR model, the positive coefficient value of T_2_N_1 and negative contribution descriptor T_2_N_3 count of number of double bounded atoms (i.e. any double bonded atom, T_2) separated from Nitrogen atom by 1 and 3 bonds respectively on the biological

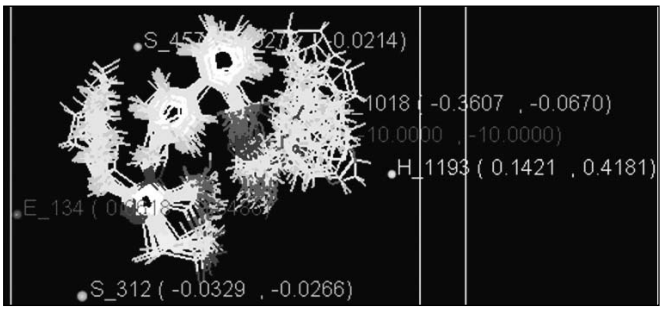


Fig 3(a): Contribution plot for electrostatic, steric and hydrophobic interactions model 3

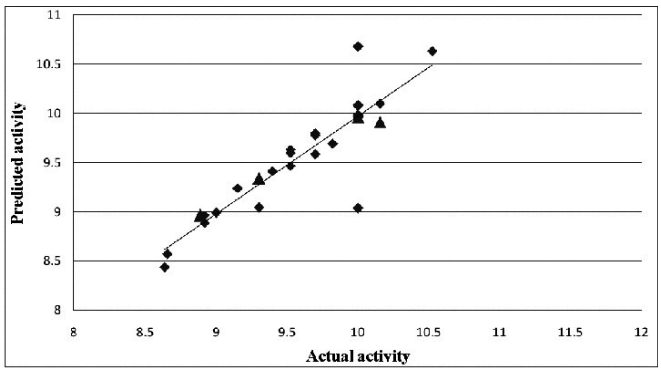


Fig 3 (b): Graphical representation of observed vs. predicted activity 3D-Model-3

activity indicated that leads to decreases activity. Model indicates that the descriptors XlogP mainly contribute to the activity; indicating that octanol water partition coefficient of the enhance activity. T_2_C_6 number of double bounded atoms (i.e. any double bonded atom, T_2) separated from carbon atom by six bonds indicates that, the increases in their value leads to activity.

For 3D-QSAR, a kNN-MFA with stepwise backward variable selection method was used resulted in several statistically significant models, of which the corresponding best model is reported herein. In this equation, the electrostatic (E), steric (S) and hydrophobic (H) descriptors specify the regions, where variations in the structural features (steric or electrostatic) of different compounds in the training set lead to increase or decrease in activities. Model 3 3D-QSAR studies helped to find out the importance of electrostatic, steric and hydrophobic groups for biological activity. Fig. 3a showed E_820 negative value in electrostatic field descriptors indicates that negative electronic potential is required to increase activity and more electronegative groups are preferred in that position, E_134 positive range indicates that group that imparting positive electrostatic potential is favorable for activity so less electronegative group is preferred in that region. H_1193 has positive range indicates that positive

hydrophobic is favorable for activity. Steric descriptor S_1018, S_312 and S_457 with negative range in steric descriptors indicates that negative steric potential is favorable for activity, and less bulky substituents group is preferred in that region. Values of q^2 (0.7954), pred_r^2 (0.6856) prove that QSAR equation so obtained is statistically significant and shows the predictive power of the model is 76% (internal validation). The contribution plot representations of the three-dimensional QSAR results for angiotensin AT_1 activity are presented in Figure 3(b). The observed and predicted pIC_{50} along with values are shown in Table I.

CONCLUSIONS

Two dimensional and three quantitative structure activity relationships study partial least square with stepwise forward-backward variable selection method was used for developed the best models was performed on a series of Sulfonylureas derivatives as AT_1 receptor. The QSAR model showed that the descriptors T_2_N_1, SulfursCount, T_C_O_6, T_T_O_3 and T_2_N_6 contributing to biological activity. 3D-QSAR studies helped to find out the importance of electropositive, electronegative, steric and hydrophobic groups for biological activity. Finally, it is concluded that the work presented here will play an important role in understanding the relationship of physiochemical parameters with structure and biological activity.

ACKNOWLEDGMENTS

The author thanks VLife Science Technologies Pvt. Ltd for providing the trial version software for the study.

Conflict of interest

The author declares no conflict of interest.

REFERENCES

1. Ferrario C.M.: The renin-angiotensin system: importance in physiology and pathology. J. Cardiovasc. Pharma.1990, 15(3),S1-S5.
2. Vallotton M. B.: The renin angiotensin system. **Trend. Pharmacol. Sci.**1987, 8, 69-74
3. Nahmias C. and Strosberg A.D.: The angiotensin AT_2 receptor: searching for signal-transduction pathways and physiological function. **Trend. Pharmacol. Sci.** 1995, 16,223–225.
4. Berecek K. H., King S.J. and Wu J.N.: Angiotensin-Converting Enzyme and Converting Enzyme Inhibitors. Cellular and Molecular Biology of the Renin-Angiotensin System; CRC Press: Boca Raton, FL, 1993: 183-220.
5. Ferreira M.M.C.: Multivariate QSAR. **J. Braz. Chem. Soc.** 2002,13,742.
6. Martin Y.C.: Overview of concepts and methods in computer-assisted rational drug design. **Method. Enzymol**. 1991,203,587-613.
7. Tropsha A.: History of quantitative structure activity relationships. In: Abraham DJ (ed) Burger's medicinal chemistry and drug discovery. Wiley Interscience, Hoboken,2003
8. Sutter J. M., Dixon S. L. and Jurs P. C.: Automated Descriptor Selection for Quantitative Structure-Activity Relationships Using Generalized Simulated Annealing. **J. Chem. Inf. Comput. Sci.**1995, 35,77-84.
9. Rogers D. and Hopfinger, A.J.: Application of Genetic Function Approximation to Quantitative Structure-Activity Relationships and Quantitative Structure-Property Relationships **J. Chem. Inf. Comput. Sci.**1994, 34, 854–866.
10. Golbraikh A. and Tropsha A.: Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection. **J. Comput. Aided. Mol. Des.** 2002,16,357-369.
11. VLife MDS 3.5. Molecular design suite. Vlife Sciences Technologies Pvt. Ltd. Pune, India, 2008.
12. Deprez P., Bertrand H., Alain C.,Jean-Paul V., Michel F. and Jacques G.: Balanced AT_1 and AT_2 angiotensin II antagonists. I. New orally active 5-carboxyl imidazolyl biphenyl sulfonylureas.**Bioorg. Med. Chem.Lett**.1995, 5(22), 2605-2610.
13. Halgren T.A.: Merck molecular force field. III. Molecular geometries and vibrational frequencies for MMFF94. **J. Comput. Chem.** 1996, 17,553-586.
14. Ajmani S., Jhadav K. and Kulkarni S.A.: Three-dimensional QSAR using the *k-nearest neighbor* method and its interpretation. **J. Chem. Inf. Mod.** 2006,46,24-31.
15. Darlington R.B.: Regression and Linear Models, McGraw-Hill, NewYork, 1990.
16. Zheng W. and Tropsha A.: Novel variable selection quantitative structure–property relationship approach based on the *k-nearest neighbor* principle. **J. Chem. Inf. Comput. Sci.** 2000;40:185–194.
17. Cramer R.D., Patterson D.E.and Bunce J.D.: Comparative Molecular Field Analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. **J. Am. Chem. Soc.** 1988,110, 5959-5967.

[a] School of Pharmacy, Devi Ahilya University, Takshashila Campus, Indore - 452 001, Madhya Pradesh, India

Sharma M. C.[a]*

***For Correspondence: E-mail: drmukeshcsharma@gmail.com**

(Received 08 July 2017) (Accepted 14 December 2018)

https://doi.org/10.53879/id.56.01.11088