Combinatorial Chemistry & High Throughput Screening, **2018**, *21*, **1-11**

RESEARCH ARTICLE

QSAR Analysis for Antioxidant Activity of Dipicolinic Acid Derivatives

Vesna Rastija[*a], Maja Molnar[b], Tena Silađi[b], and Vijay Hariram Masand[c]

[a]*Department of Chemistry, Faculty of Agriculture, J. J. Strossmayer University, Osijek 31 000, Croatia;* [b]*Department of Applied Chemistry and Ecology, Faculty of Food Technology, J.J. Strossmayer University, Osijek, Croatia;* [c]*Department of Chemistry, Vidya Bharati College, Camp, Amravati, Maharashtra, India*

ARTICLE HISTORY

Received: September 21, 2017
Revised: November 17, 2017
Accepted: February 3, 2018

DOI:
10.2174/1386207321666180213092352

Abstract: ***Aim and Objective*:** The aim of this study was to derive robust and reliable QSAR models for clarification and prediction of antioxidant activity of 43 heterocyclic and Schiff bases dipicolinic acid derivatives. According to the best obtained QSAR model, structures of new compounds with possible great activities should be proposed.

***Methods*:** Molecular descriptors were calculated by DRAGON and ADMEWORKS from optimized molecular structure and two algorithms were used for creating the training and test sets in both set of descriptors. Regression analysis and validation of models were performed using QSARINS.

***Results*:** The model with best internal validation result was obtained by DRAGON descriptors (*MATS4m*, *EEig03d*, *BELm4*, *Mor10p*), split by ranking method (R^2 = 0.805; R^2_{ext} = 0.833; F = 30.914). The model with best external validation result was obtained by ADMEWORKS descriptors (*NDB, MATS5p*, *MDEN33, TPSA*), split by random method (R^2 = 0.692; R^2_{ext} = 0.848; F = 16.818).

***Conclusion*:** Important structural requirements for great antioxidant activity are: low number of double bonds in molecules; absence of tertial nitrogen atoms; higher number of hydrogen bond donors; enhanced molecular polarity; and symmetrical moiety. Two new compounds with potentially great antioxidant activities were proposed.

Keywords: QSAR, antioxidant activity, dipicolinic acid, polar surface area, hydrogen bond donors, derivatives.

1. INTRODUCTION

Dipicolinic acid (DPA) (2,6-pyridinedicarboxylic acid) is naturally inherent in the spores of *Bacillus* [1] and *Clostridium genusus* [2] that exists in the core in the form of chelates with calcium ions [3]. DPA contributes to the spore resistance to UV radiation [4], wet heat [5], and protecting spore DNA from damage [6]. DPA forms stable chelates with metal ion, and these complex compounds show a variety of biological activities, such as antimicrobial, antifungal [7-9], anticancer [10] and antioxidant activities [11].

It was discovered that a series of substituted mono- and bis-dipicolinic derivatives possessed antimicrobial and antioxidant activities [12-15]. In our previous work, we have reported synthesis of Schiff bases [16] and heterocyclic compounds [17] derived from DPA. Many of the obtained compounds exhibited significant antifungal and antioxidant activities. The above-mentioned studies identified significant chemical features for the most active compounds. In a series of synthesized mono- and bis-dipicolinic acid heterocyclic derivatives – thiosemicarbazides, triazoles, oxadiazoles and thiazolidinones, and thiosemicarbazides showed predominant antioxidant activity. Influence of different substituents is obvious, since in almost all compounds phenyl substitution results in a better 2,2-diphenyl-1-picrylhydrazyl (DPPH) radical scavenging activity compared to the alkyl substitution [17]. In a series of Schiff bases derived from dipicolinic acid, compounds derived from di-hydrazide showed higher antioxidant activity than the ones derived from mono-hydrazide [16].

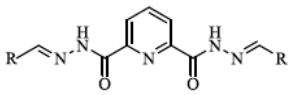
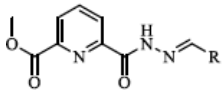
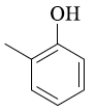
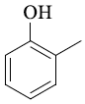
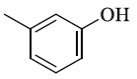
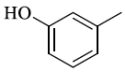
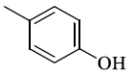
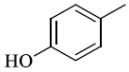
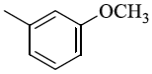
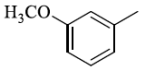
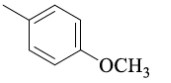
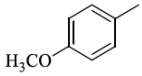
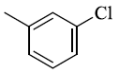
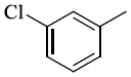
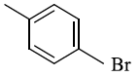
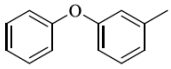
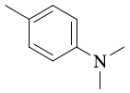
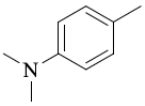
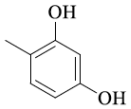
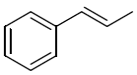
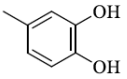
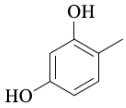
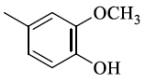
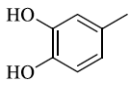
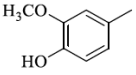
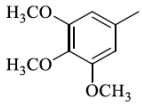
The structure-activity relationship (SAR) of the antibacterial and antiproliferative potential of some 1-pyridinecarbonyl-4-substituted thiosemicarbazide derivatives showed that substitution at the position 2 of the pyridine ring enhances biological activity.

In the present study, the main goal was to build robust and reliable QSAR models for the description and prediction of antioxidant activity of heterocyclic and Schiff bases dipicolinic acid derivatives. The relevance of the best QSAR model should also be to provide a chemical and structural explanation of antioxidant activities of the most active compounds. Also, the aim of this study was to compare the statistical performance of different algorithms for splitting data into training and test set, as well as models obtained by two sets of descriptors calculated by different computer programs.

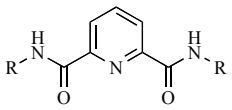
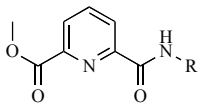
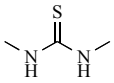
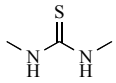
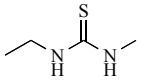
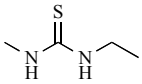
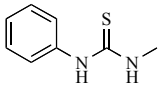
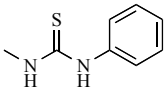
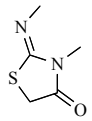
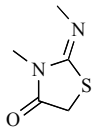
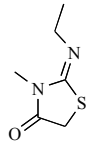
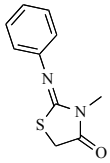
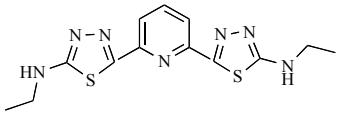
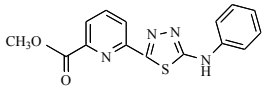
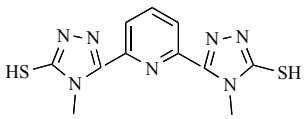
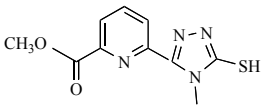
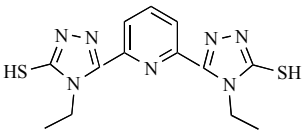
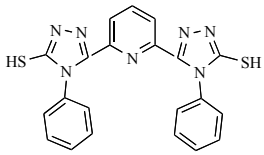
*Address correspondence to this author at the Department of Chemistry, Faculty of Agriculture, J. J. Strossmayer University, P.O. Box: 719, 31 000 Osijek, Croatia; Tel: +385-31-554-903; Fax: +385-31-554-853;
E-mail: vrastija@pfos.hr

1386-2073/18 $58.00+.00

Table 1. Structures along with experimentally determined antioxidant activities (expressed as log % DPPH) of 2,6-pyridinedicarboxylic acid derivatives [16, 17].

					
Cpd.	**R**	**log %DPPH**	**Cpd.**	**R**	**log %DPPH**
1		-0.31	2		0.894
3		0.782	4		0.582
5		0.701	6		0.845
7		0.834	8		0.786
9		0.671	10		0.68
11		0.748	12		0.705
13		0.768	14		0.656
15		0.608	16		-0.495
17		-0.854	18		0.955
19		0.874	21		0.856
20		-0.42	23		1.692
22		0.874	25		1.429
24		1.732			
26		1.621			
27		0.736			

(Table 1) Contd….

					
Cpd.	**R**	**log %DPPH**	**Cpd.**	**R**	**log %DPPH**
31		1.061	**28**		1.583
32		1.566	**29**		1.68
33		1.688	**30**		1.936
41		0.301	**40**		0.114
42		0.996			
43		0.362			
36		1.134	**34**		1.1
37		1.487	**35**		1.401
38		1.504	**39**		1.49

2. MATERIALS AND METHODS

2.1. Data Set

The synthesis and antioxidant evaluation of 43 DPA derivatives have been described in our previous studies [16, 17]. Antioxidant activities, expressed as % scavenging activity on DPPH (using ascorbic acid as standard), were converted in the form of the logarithm (log %DPPH). Higher value of log %DPPH means more powerful antioxidant activity. Structural details of all studied molecules along with experimental log %DPPH are shown in Table **1**.

2.2. Descriptor Calculation and Selection

The three-dimensional (3D) structures of 43 DPA derivatives were optimized using the molecular mechanics force field (MM+) [18] applying the Avogadro 1.2.0. (University of Pittsburgh, Pittsburgh, PA, USA). Subsequently, all structures were submitted to the geometry optimization using semiempirical AM1 method [19]. The molecular structures were optimized using Polak-Ribiere algorithm until the root mean square gradient (RMS) was 0.001 kcal/(Åmol).

The Polar Surface Area (PSA) surface was generated from an optimized structure by VEGA ZZ (Department of Pharmaceutical Sciences of the University of Milan, Milan, Italy) [20]. The Polar Surface Area (PSA) was calculated considering polar and apolar atom surfaces, as dotted shape, with probe radius 1.4 and density 60.

Two set of descriptors were calculated by two different softwares 1. DRAGON descriptors – were calculated using Parameter Client (Virtual Computational Chemistry Laboratory, an electronic remote version of the Dragon program [21]. Seventeen groups of two-dimensional (2D) and three-dimensional (3D) Dragon's descriptors were used to generate QSAR models: constitutional, topological, walk and path counts, connectivity, information, 2D auto-correlations, edge adjacency, BCUT (Burden eigenvalues), topological charge, eigenvalue-based, geometrical, RDF (Radial Distribution Function), 3D-MoRSE (3D-molecular representation of structures based on electron diffraction), WHIM (WeigHted Covariance Matrices), GETAWAY (Geometry, Topology, and Atom-Weights AssemblY) descriptors, functional group counts, and molecular properties [22]. In order to reduce huge number of calculated descriptors (about 1260), firstly, zero values descriptors were excluded from initial pool. Further exclusion was performed using QSARINS-Chem 2.2.1 [23]: constant and semi-constant descriptors, i.e. those having chemical compounds with a constant value for more than 80 %, and too inter-correlated descriptors (> 85%) were rejected.

2. ADMEWORKS descriptors – calculated by ADMEWORKS ModelBuilder, tool for building mathematical models (Version 7.9.1.0(build.187.4934.2102) Enterprise Edition Copyright (C) 2011 Fujitsu Kyushu Systems Limited). Numerous groups of 2D and 3D molecular descriptors have been calculated, such as: charged partial-surface-area (CPSA) descriptor, atom-specific CPSA descriptor (DATOM), carbon type (CTYPE), molecular distance edge descriptor (DEDGE), fragment descriptors generation routine (DMFRAG), hydrogen bond specific descriptors for pure (HBPURE) and mixed (HBMIX) compounds, measure of the conformational flexibility, Hückel molecular orbital calculation (HMO), MOPAC descriptor, molecular strain energy calculation (STRAIN), etc. [24]. Elimination of irrelevant descriptors was performed using Feature Selection command of ADMEWORKS ModelBuilder that includes following tests: a) missing values test - which excludes descriptors with missing values; b) zero test - which excludes descriptors with less than the specified percentage of non-zero values; c) automated correlations test - deletes all parameters that have single or multiple correlations to other parameters, with the R^2 value larger than the specified threshold (0.7).

2.3. Training and Test Set Compounds Selection

Two algorithms were used for creating the training and test sets in both set of descriptors:

1. Data sets were randomly divided into training (80 %, N_{train} = 35) and prediction (20 %, N_{test} = 8) set using QSARINS.

2. Data sets were split by ranking method: Compounds were ranked by the activities from the most active to the least active compound. Then, the activities were divided into the bins by grouping the values into the six class intervals. Finally, one or two compounds were selected randomly from each class for the test set (N_{train} = 35, N_{test} = 8) [25].

2.4. Regression Analysis and Validation of Models

The best QSAR models were obtained by using a Genetic Algorithm (GA) using QSARINS. In order to avoid the overfitting the smallest number of descriptors that can adequately model the activity of the compounds in the study should be used. According the Topliss-Costello rule [26], the number of variables should be higher or equal to 5. In this study, considering we worked with the small data set (35 compounds in the training set), the number of descriptors (k) in the multiple regression equation was limited to four. Additional descriptors, will be resulted in overfittied and not predictive models. The models have been assessed by: fitting criteria; internal cross-validation using leave-one out (LOO) method and Y-scrambling; and external validation. Fitting criteria included: the coefficient of determination (R^2), adjusted R^2 (R^2_{adj}), cross-validate R^2 using leave-one-out method (Q^2_{LOO}), global correlation among descriptors (Kxx), difference between global correlation between molecular descriptors and y the response variable, and global correlation among descriptors (ΔK), standard deviation of regression (s), and Fisher ratio (F) [27-29]. Collinearity among the descriptors has also validated by variance inflation factor (VIF). VIF is the reciprocal of tolerance: $1/(1-R^2_i)$ where R^2_i is the squared multiple correlation of the ith independent variable regressed on the other independent variables in the analysis [30].

Internal and external validations also included the following parameters: the coefficient of determination of test set (R^2_{ex}), root-mean-square error of the training set ($RMSE_{tr}$); root-mean-square error of the training set determined through cross validated LOO method ($RMSE_{cv}$), root-mean-square error of the external validation set ($RMSE_{ex}$), concordance correlation coefficient of the training set (CCC_{tr}), test set using LOO cross validation (CCC_{cv}), and of the external validation set (CCC_{ex}) [27], mean absolute error of the training set (MAE_{tr}), mean absolute error of the internal validation set (MAE_{cv}) and mean absolute error of the external validation set (MAE_{ex}) [27], predictive residual sum of squares determined through cross-validated LOO method ($PRESS_{cv}$) in the training set and in the external prediction set ($PRESS_{ex}$). The analysed external validation parameters also include. Robustness of QSAR models was tested by Y-randomisation test.

Investigation of the applicability domain of a prediction model was performed by leverage plot (plotting residuals vs. leverage of training compounds). The warning leverage h^* is defined as $3p'/n$, where n is the number of training compounds and p' is the number of model adjustable parameters [31]. Tools of regression diagnostic as residual plots and Williams plots were used to check the quality of the best models and define their applicability domain using QSARINS.

3. RESULTS AND DISCUSSION

Four models were obtained by two different sets of descriptors (calculated by Dragon and ADMEWORKS software) and two different methods (random and ranking) for creating training and prediction set.

Models obtained by DRAGON descriptors:

The random splitting model

log %DPPH = -1.073 – 0.492 *IVDE* – 0.492 *Mor08u* – 11.755 *Dp* + 1.736 *R5u* (1)

Compounds in the test set: **15**, **20**, **23**, **31**, **33**, **36, 37**, **42**

The ranking splitting model

log %DPPH = 2.954 – 1.427 *MATS4m* + 1.093 *EEig03d* – 4.2 *BELm4* + 0.974 *Mor10p* (2)

Compounds in the test set: **1**, **2**, **7**, **16, 29**, **33**, **34**, **35**

Models obtained by ADMEWORKS descriptors:

The random splitting model

log %DPPH = 0.545 – 0.4 *NDB* – 2.2 *MATS5p* – 0.182 *MDEN33* + 0.014 *TPSA* (3)

Compounds in the test set: **1**, **4**, **6**, **7**, **11**, **13**, **28, 29**

The ranking splitting model

log %DPPH = 3.361 – 0.281 *NDB* – 0.432 *GATS8v* – 1.235 *MATS5p* + 0.129 *NUMHBD* (4)

Compounds in the test set: **1**, **2**, **7**, **16, 29**, **33**, **34**, **35**

The statistical results for the obtained models are presented in Table **2**.

Description of descriptors included is given in Table **3**. In order to exclude collinearity of descriptors included in same model, correlation matrix was generated (Tables **4-7**). Descriptors included in models (1-4) are not mutually correlated (correlation coefficient, $R \leq 0.7$). *VIF* values of individual descriptors from each model were also presented in Tables **4-7**. Linear dependence within the correlation descriptors sets has been rejected since the all *VIF* < 5 [32]. Low collinearity is also verified by the low values of *Kxx* (Table **2**). The molecular descriptor values have been tabulated in Supplementary File 1 (SF 1). Experimental and calculated log %DPPH by model (1-4) are shown in Supplementary File 2 (SF 2).

Satisfaction of fitting criteria implies the following: the closer R^2 values are to unity, the more similar calculated values are to the experimental ones, that is, $R^2 \geq 0.60$. The minimum acceptable statistical parameters for internal and external predictivity include the following conditions: $R^2_{ext} \geq 0.60$; $CCC \geq 0.85$; *RMSE* and *MAE* close to zero; and $RMSE_{tr} < RMSE_{cv}$. Robust QSAR models should have R^2_yscr and Q^2_yscr < 0.2, as R^2_yscr > Q^2_yscr [33]. Also, larger *F* statistic and lower standard deviation means that the model is more significant. Analysis of Table **2** indicates that all four models satisfy threshold for most of the internal validation parameters. However, models (1) and (2), created by Dragon descriptors, have better fitting performances (higher R^2, R^2_{adj}, *F*, and CCC_{tr}, and lower *s*, $RMSE_{tr}$, MAE_{tr} than models (3) and (4), obtained by ADMEWORKS descriptors. Also, models (1) and (2) showed a better performance in the

Table 2. The statistical results for the QSAR models for antioxidant activity.

	Model (1)	Model (2)	Model (3)	Model (4)
N_{tr}	35	35	35	35
N_{ex}	8	8	8	8
Fittinig criteria				
R^2	0.821	0.805	0.692	0.646
R^2_{adj}	0.797	0.779	0.650	0.599
s	0.281	0.278	0.381	0.374
F	34.455	30.914	16.818	13.689
p	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$	$< 10^{-5}$
Kxx	0.357	0.325	0.251	0.192
ΔK	0.034	0.004	0.031	0.074
$RMSE_{tr}$	0.260	0.257	0.353	0.257
MAE_{tr}	0.212	0.19	0.290	0.257
CCC_{tr}	0.902	0.892	0.818	0.785
Internal validation criteria				
Q^2_{LOO}	0.753	0.725	0.576	0.522
$RMSE_{cv}$	0.306	0.305	0.414	0.403
MAE_{cv}	0.250	0.226	0.340	0.300
$PRESS_{cv}$	3.268	0.849	5.989	5.670
CCC_{cv}	0.866	0.849	0.750	0.715
R^2_Yscr	0.118	0.118	0.119	0.116
Q^2_Yscr	-0.224	-0.232	-0.208	-0.215
External validation criteria				
$RMSE_{ext}$	0.674	0.550	0.231	0.520
MAE_{ext}	0.611	0.365	0.197	0.381
$PRESS_{ext}$	3.637	2.424	0.426	2.160
R^2_{ext}	0.717	0.833	0.848	0.639
Q^2_{F1}	-0.009	0.55	0.842	0.562
Q^2_{F2}	-0.078	0.507	0.84	0.56
Q^2_{F3}	-0.203	0.106	0.868	0.203
CCC_{ext}	0.678	0.657	0.908	0.666
$\overline{r_m^2}$	0.261	0.083	0.765	0.26
Δr_m^2	0.366	0.651	0.12	0.4
Applicability domain				
N compounds outlier	2 (**16**,**17**)	3 (**1**,**16**,**27**)	0	2 (**16**,**17**)
N compounds out of app.dom.	3 (**39**, **42**, **20**)	1 (**17**)	0	0

internal validation (higher Q^2_{LOO}, CCC_{cv}, and lower $RMSE_{cv}$, MAE_{cv}, and $PRESS_{cv}$. The results of Y-scrambling demonstrated that all models were not obtained by chance correlation (Q^2_yscr < 0.2, and R^2_yscr< 0.2; R^2_yscr > Q^2_yscr). In spite of a good fitting performance and good internal validation, the real predictive power of model (1) and (2) is failed according to the external validation parameters and outliers and compounds outside of applicability domain. Moreover, negative values of Q^2_{F1}, Q^2_{F2} and Q^2_{F3} of model 1, as well as very low values (< 0.7) of model (2) and (4), including values of $\overline{r_m^2}$ < 0.6 indicate these models are useless for external predictivity.

Model (3) was generated by ADMEWORKS descriptors and random splitting methods. In spite of its powerless fitting and internal performances, model possess real predictivity for the chemicals in the validation set according to high values of parameters for external validations (R^2_{ext} = 0.848; CCC_{ext} = 0.908) and small difference between $RMSE_{tr}$ and $RMSE_{ex}$, and between MAE_{tr} and MAE_{ex}.). Also, Williams plot for same model reveals no outliers, and no compounds outside of applicability domain (Fig. **1**). A scatter plot of experimentally obtained antioxidant activity by model (3) is shown in Fig. (**2**). Model (4) has weakest parameters of external validation, as well as one outlier (**16**), and one compound out of the applicability domain (**17**).

Despite of difference in predictive potential of the proposed models, included molecular descriptors may relieve in elucidation of important physicochemical and structural requirements for the antioxidant activity of heterocyclic and Schiff bases dipicolinic acid derivatives. Negative sign of variable that represents a number of double bonds (*NDB*) in equations (3) and (4), means that this type of covalent bond is unfavourable for the antioxidant power (except double bonds in phenyl or heterocyclic ring). Thus, the most active compound **30** (log %DPPH = 1.936) has 3 double bonds, while the compounds with low antioxidant activity have 6 double bonds (**41,** log %DPPH = 0.301; **17**, log %DPPH = -0.854). Negative sign of descriptor *MDEN33* in model (3) indicates that higher distance between tertial nitrogen atoms negatively influences on the antioxidant activity. Compounds without tertial nitrogen atoms have a higher antioxidant potential, such as thiosemicarbazides (**28-33**). Since DPPH test is based on the capability of stable free radical 2,2-diphenyl-1-picrylhydrazyl to react with H-donors [34], positive coefficient of hydrogen bond donors (*NUMHBD*) in model (4) is expected. That implies that higher number hydrogens attached to the oxygen, nitrogen or sulphur atoms, positively influences on antioxidant activity, such as secondary nitrogen atoms in thiosemicarbazides (**28-33**). This supports our recent findings that enhanced value of hydrophilic factor, which is related with number of –OH, –NH_2, and >NH groups in molecule, is favourable for antioxidant activity of coumarinyl Schiff bases [35]. Also, it corresponds with previously statement about negative influence of enhanced number of tertial nitrogen atoms on antioxidant activity. However, five compounds (**22**, **24**, **31-33**) have more than five hydrogen bond donors, and according the Lipinski's rule of five [36], as potential drug, they could have poor absorption or permeation. Topological polar surface area (*TPSA*) in model (3), is descriptor that also characterized drug absorption, including intestinal absorption, bioavailability and blood–brain barrier penetration. Polar surface area (*PSA*) is sum of surface of polar atoms (oxygen, nitrogen, sulphur, etc.) [37]. Although model (3) implies that molecules with enhanced values of *TPSA* have higher antioxidant activity, these molecules could poorly penetrate through the cell membranes [15, 38]. According the data presented in Supplementary File 1, value of *TPSA* for the most active compound (**30**) is larger (124.44) than for the least active (**17**) (114.27). *TPSA* is a useful descriptor in QSAR models, which a sign its coefficient can indicate whether a more polar ligand is favoured or disfavoured for enhanced activity. However, it is based only on the contribution of tabulated polar fragments, not from the 3D conformations of these chemical groups. Also, *TPSA* does not include the influence of positional changes of functional groups [38]. In order to additionally clarify sense of PSA for antioxidant activity of Schiff bases dipicolinic acid derivatives, we were calculated and compared PSA of the most active (**30**) and the least active compound (**17**). Apolar atom surface takes into account C and H atoms bonded to C atoms. Polar atoms are O, S, N, P and H, not bonded to C. Because of higher number of polar atoms in molecule (**30**), their PSA (203.4 $Å^2$) is much higher than molecule (**17**) (165.6 $Å^2$). Moreover, apolar surface area of the most active compound (**30**) is significantly lower (590.3 $Å^2$) than at least active compound (**17**) (983.2 $Å^2$). Fig. (**3**) shows a mapped PSA of the most active molecule (**30**) and the least active molecule (**17**), for comparison. Apolar and polar surfaces are presented as gradient of two color codes: black (apolar surface) and grey (polar surface).

3D-MoRSE descriptors, *Mor08u* and *Mor10p*, are involved in models (1) and (2). These descriptors were generated from electron diffraction studies and reflect the three-dimensional arrangement of atoms in a molecule [39]. Descriptor *Mor08u* denotes unweighted descriptors with scattering parameter s = 7 $Å^{-1}$. Since it is unweighted, the descriptor has no discriminative ability and treats each atom equally. It has the possibility to distinguish the difference between the bond lengths of any kinds of atoms at least 0.12 Å. Consider that their final values are derived mostly from long distances atoms, it has great power to distinguish mono- from bis-substituted dipicolinic derivatives (Table **S1**). Negative coefficient of *Mor08u* in Eq. (1) implies that lower values of that descriptor are favourable for the exhibition of antioxidant activity. Descriptor *Mor10p* reflects the contribution of 3D distribution of atomic polarizability (p) at a scattering parameter s = 9 $Å^{-1}$. According to model (2), it is expected that increased values of *Mor10p* indicates higher antioxidant activity. As can be noticed from Table **S1**, this descriptor is extremely sensitive to the position of the sulphur atoms (atom with higher polarizability, p = 19.6) in molecules.

MATS5p, *MATS4m* and *GATS8v* belong to the 2D autocorrelation molecular descriptors that describe how a considered property is distributed along a topological molecular structure. *MATS5p* corresponds to the Moran autocorrelation –lag 5 / weighted by atomic polarizability, while *MATS4m* is the Moran autocorrelation descriptor –lag 5 / weighted by atomic masses.

Table 3. Summary table for descriptors included in models 1-4.

Symbol	Descriptor name	Groups of descriptors
IVDE	Mean information content on the vertex degree equality	Information
Mor08u	3D-MoRSE signal 08 / unweighted	3D-MoRSE
Mor10p	3D-MoRSE signal 10 / weighted by polarizability	3D-MoRSE
Dp	D total accessibility index / weighted by polarizability	WHIM
R5u	R autocorrelation of lag 5 / unweighted	GETAWAY
MATS4m	Moran autocorrelation of lag 4 weighted by mass	Autocorrelations
MATS5p	Moran autocorrelation of lag 4 weighted by atomic polarizabilities	Autocorrelations
GATS8v	Geary autocorrelation of lag 8 weighted by van der Waals volume	Autocorrelations
EEig03d	Eigenvalue 03 from edge adj. matrix weighted by dipole moments	Edge adjacency
BELm4	Lowest eigenvalue 4 of Burden matrix / weighted by atomic masses	BCUT
NDB	Number of double bonds	DMFRAG
MDEN33	Molecular distance edge between all tert tert N	EDGE
TPSA	Topological polar surface area	Mol. properties
NUMHBD	Number of hydrogen bond donors	H acceptor/donor

Table 4. Correlation matrix for the descriptors included in model 1 and variance inflation factor (VIF) for individual descriptor.

	VIF	***IVDE***	***Mor08u***	***Dp***	***R5u***
IVDE	1.45	1.00			
Mor08u	1.69	0.38	1.00		
Dp	1.23	0.00	0.01	1.00	
R5u	1.45	0.18	0.33	0.37	1.00

Table 5. Correlation matrix for the descriptors included in model 2 and variance inflation factor (VIF) for individual descriptor.

	VIF	***MATS4m***	***EEig03d***	***BELm4***	***Mor10p***
MATS4m	1.49	1.00			
EEig03d	1.49	0.43	1.00		
BELm4	1.11	-0.28	0.09	1.00	
Mor10p	1.37	-0.41	-0.34	0.17	1.00

Table 6. Correlation matrix for the descriptors included in model 3 and variance inflation factor (VIF) for individual descriptor.

	VIF	**NDB**	**MATS5p**	**MDEN33**	**TPSA**
NDB	1.96	1.00			
MATS5p	1.15	-0.20	1.00		
MDEN33	1.96	-0.49	-0.01	1.00	
TPSA	1.49	0.21	-0.01	0.34	1.00

Table 7. Correlation matrix for the descriptors included in model 4 and variance inflation factor (VIF) for individual descriptor.

	VIF	NDB	GATS8v	MATS5p	NUMHBD
NDB	1.49	1.00			
GATS8v	1.35	-0.13	1.00		
MATS5p	1.39	-0.20	0.01	1.00	
NUMHBD	1.85	0.43	-0.38	-0.02	1.00

GATS8v is Geary autocorrelation of lag 8 weighted by van der Waals volume [22]. These descriptors reflect contribution of pairs of atoms different polarizability/mass/ van der Waals volume at the defined topological distance or spatial lag. Therefore, *MATS5p*, *MATS4m* and *GATS8v* indicate dependence of one atom on value of polarizability /mass/ van der Waals volume of other atoms at the topological distance 5, 4, and 8, respectively. Their negative regression coefficients in models (3) and (4) suggest in unfavorable effect of increased autocorrelation contents of five-, four-, eight-member structural graphs weighted by atomic polarizability, mass, and van der Waals volume for the activity.

WHIM is geometrical descriptor calculated on the projections of the atoms along principal axes [24]. Descriptor *Dp* is total accessibility index weighted by atomic polarizability, which values are strongly influenced by kind and position of substituents on phenyl ring. Negative coefficient of *Dp* in model (1) implies that 3D distribution of substituents with the increased atomic polarizability (Cl, Br atoms) could negatively influenced to the antioxidant power. It explains weak antioxidant activity of compounds with Br (**16**) and Cl atoms (**14**, **15**). Descriptor *IVDE* is mean information content on the vertex degree equality [24]. Since it is a measure of the lack of structural homogeneity or the diversity of a molecule [40], its negative coefficient in model (1) suggests that molecules with symmetrical moiety are potentially more active.

Descriptor *BELm4* in model (2) belongs to the BCUT descriptors. BCUT descriptors are the eigenvalues of a modified connectivity matrix, the Burden matrix, which capture useful measurement of molecular diversity [41]. Highest value of *BELm4* have most inactive compounds, **17** and **20** (*BELm4* = 1.894 and 1.607, respectively, Table **S1**), which correspond with a negative coefficient of that variable in model (2). Descriptor *EEig03d* in model (2) is related to the molecular polarity, which mainly described the electronic effect of molecule and the hydrophobic properties. Same as *TPSA* in model (3), positive regression coefficient of *EEig03d* in model (2) suggests that enhanced molecular polarity positively influence on antioxidant activity of compounds.

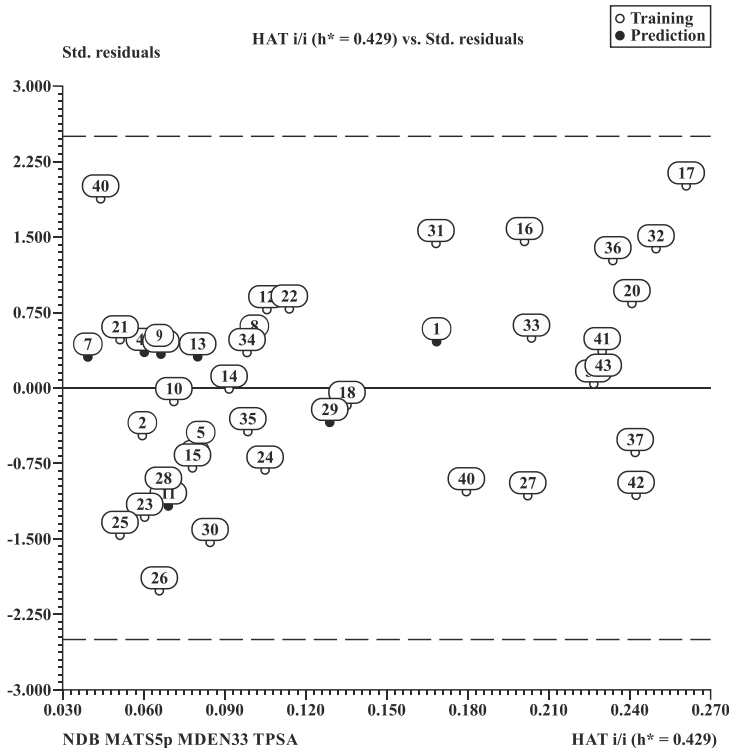


Fig. (1). Applicability domain of the QSAR model for antioxidant activity calculated by model (3).

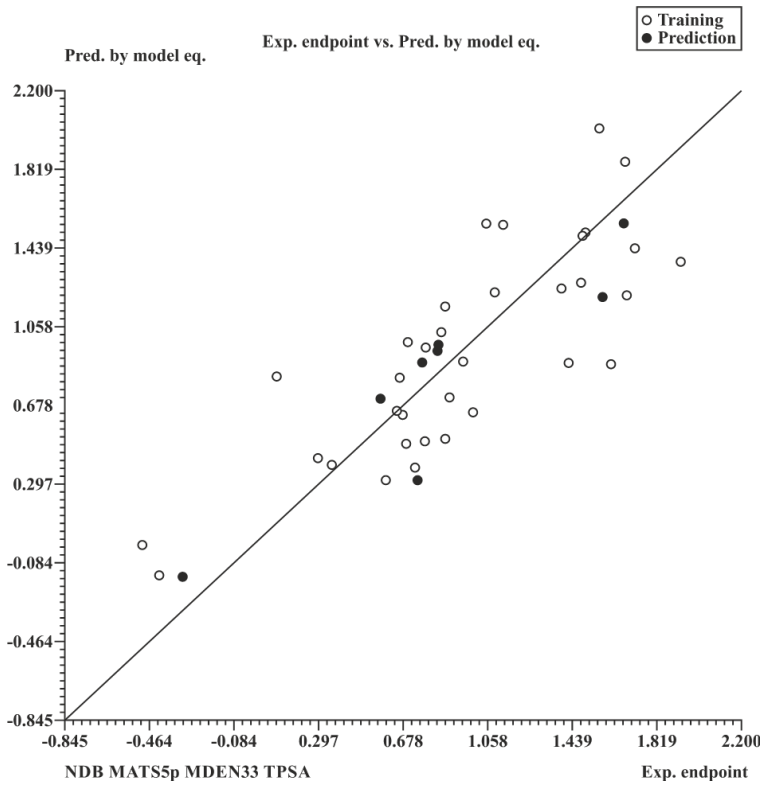


Fig. (2). Observed versus predicted log %DPPH for the 43 dipicolinic acid derivatives calculated by model (3).

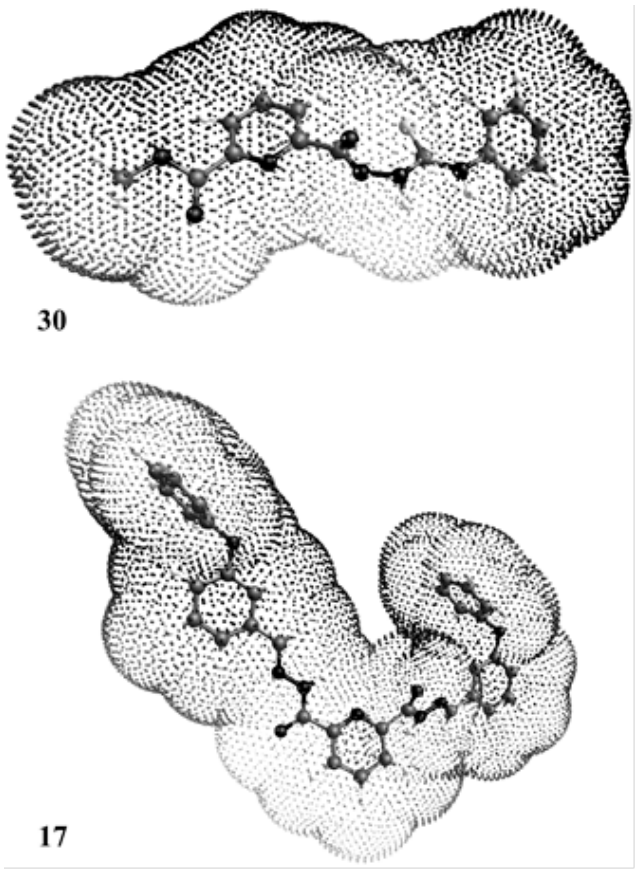
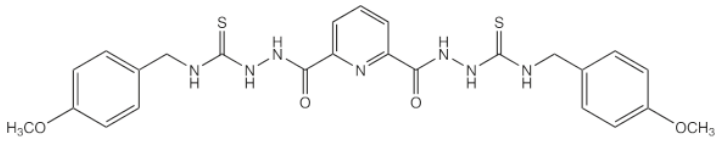
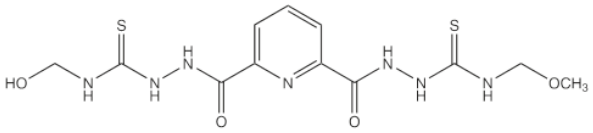


Fig. (3). Dotted polar surface area (PSA) for the compound with highest (30), and compound (17) with lowest antioxidant activity. PSA is presented as gradient of two colors: black (apolar surface) and grey (polar surface).



44 log %DPPH$_{pred.}$ = 1.953



45 log %DPPH$_{pred.}$ = 2.015

Fig. (4). Structures of the proposed molecules as promising antioxidant agents. Antioxidant activities have been predicted according the model (3).

Based on the conclusions given in the QSAR analysis, structures of two new compounds (**44**, **45**) with possible great activity are proposed (Fig. **4**). Antioxidant activities of the proposed compounds have been predicted by means of the model (3) (log %DPPH = 1.953 and 2.015, for **44** and **45**, respectively). Calculated descriptors are shown in Supplementary File 1. Applicability domain of the proposed new potentially active derivatives has been verified. Leverage (HAT) values of both compounds are inside the applicability domain of a model (0.267 for **44,** 0.321 for **45**; $h^* = 0.429$). The Williams plot of the regression is presented in Supplementary File 3 (SF 3).

CONCLUSION

The results of the QSAR analysis suggest that derivatives of dipicolinic acid with the following structural feature may exhibit great antioxidant activity: low number of double bonds in molecules; absence of tertial nitrogen atoms; higher number of hydrogen bond donors; enhanced molecular polarity; substituents without halogen atoms; and symmetrical moiety. Model with the best external validation result was obtained by ADMEWORKS descriptors, and the test set was generated by random method. Obtained models could help in suggesting design of novel molecules with improved activity profile. According to developed model, structures of two new compounds with possible great activities were proposed. Thus, the model provides a practical tool for the prediction of antioxidant activity of new and untested antioxidant.

CONSENT FOR PUBLICATION

Not applicable.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

This work was supported in by the two J. J. Strossmayer University Grants: INGI-2015-20 and „Synthesis and biological activity of dipicolinic acid derivatives“ (2013).

SUPPLEMENTARY MATERIAL

Supplementary materials are available on the publisher's web site along with the published article:

SF 1. Values of the most relevant descriptors in QSAR models (1)-(4).

SF 2. Values of the experimental and calculated log %DPPH by QSAR models (1)-(4).

SF 3. Applicability domain of the proposed new potentially active derivatives (44, 45) calculated by model (3).

REFERENCES

[1] Goodacre, R.; Shann, B.; Gilbert, R.J.; Timmins, Ė.M.; McGovern, A.C.; Alsberg, B.K.; Kell, D.B.; Logan, N.A. Detection of the dipicolinic acid biomarker in Bacillus spores using Curie-Point Pyrolysis Mass Spectrometry and Fourier Transform Infrared Spectroscopy. *Anal. Chem*., **2000**, *72*, 119-127.

[2] Tabor, M.W.; MacGee, J.; Holland, J.W. Rapid determination of dipicolinic acid in the spores of Clostridium species by gas-liquid chromatography. *App. Env. Microbiol*., **1976**, *31*, 25-28.

[3] Gerhardt, P.; Marquis, R.E. In: *Spore thermoresistance mechanisms*, In: *Regulation of prokaryotic development*, Smith, R.A.; Slepecky, R.A.; Setlow, P., Eds.; American Society for Microbiology, Washington, USA, **1989**, pp. 43-63.

[4] Slieman, T.A.; Nicholson, W.L. Role of dipicolinic acid in survival of *Bacillus subtilis* spores exposed to artificial and solar UV radiation. *Appl. Environ. Microbiol.* **2001**, *67*, 1274-1279.

[5] Kort, R.; O'Brien, A.C.O.; van Stokkum, I.H.M; Oomes, S.C.M.; Crielaard, W.; Hellingwert, K.J.; Brul, S. Assessment of heat resistance of bacterial spores from food product isolates by fluorescence monitoring of dipicolinic acid release. *Appl. Environ. Microbiol*., **2005**, *71,* 3556-3564.

[6] Setlow, B.; Atluri, S.; Kitchel, R.; Koziol-Dube, K.; Setlow, P. Role of dipicolinic acid in resistance and stability of spores of Bacillus subtilis with or without DNA-protective α/β-type small acid-soluble proteins. *J. Bacteriol.*, **2006**, *188*, 3740-3747.

[7] Colak, A.T.; Colak, F.; Yesilel, O.Z.; Orhan, B. Synthesis, characterization, crystal structure and biological activities of supramolecular compounds of Mn(II) and Zn(II) with dipicolinic acid and 8-hydroxyquinoline. *J. Coord. Chem*., **2009**, *62*, 1650-1660.

[8] Yenikava, C.; Poyraz, M.; Sari, M.; Demirci, F.; Ilkimen, H.; Orhan, B. Synthesis, characterization and biological evaluation of a novel Cu(II) complex with the mixed ligands 2,6-pyridinedicarboxylic acid and 2-aminopyridine. *Polyhedron*, **2009**, *28*, 3526-3532.

[9] Tamer, O.; Sariboga, B.; Ucar, I.; Orhan, B. Spectroscopic characterization, X-ray structure, antimicrobial activity and DFT

calculations of novel dipicolinate copper(II) complex with 2,6-pyridinedimethanol. *Spectrochim. Acta, Part A*, **2011**, *84*, 168-177.

[10] Uzun, N.; Colak, A.T.; Emen, F.M.; Kismali, G.; Meral, O.; Alpay, M.; Cigli, G.K.; Sahin, E. The syntheses, crystal structure, thermal analysis, and anticancer activities of novel dipicolinate complexes. *J. Coord. Chem*., **2015**, *68*, 949-967.

[11] Murakami, K.; Yoshino, M. Dipicolinic acid as an antioxidant: Protection of glutathione reductase from the inactivation by copper, *Biomed. Res.*, **1999**, *20*, 321-326.

[12] Ghozlan, S.A.S.; Mohamed, S.F.; Amr, A.E.-G.E.; Mustafa, E.-S.E.; El-Wahab, A.A.A. Synthesis and reactions of some new 2,6-bis-substituted pyridine derivatives as antimicrobial. *World J. Chem*., **2009**, *4*, 83-88.

[13] Ghozlan, S.A.S.; Al-Omar, M.A.; Amr, A.E.-G.E.; Ali, K.A.; El-Wahab, A.A.A. Synthesis and antimicrobial activity of some heterocyclic 2,6-bis(substituted)-1,3,4-thiadiazolo-, oxadiazolo-, and oxathiazolidino-pyridine derivatives from 2,6-pyridine dicarboxylic acid dihydrazide. *J. Heterocyclic Chem.,* **2011**, *48*, 1103-1110.

[14] Naeem, S.; Akhtar, S.; Asghar, N.; Sherwani, S.K.; Mushtaq, N.; Kamil, A.; Zafar, S.; Arif, M.; Saify, Z.S. Antimicrobial and antioxidant screening of N¢-substituted sulphonyl and benzoyl derivatives of 4-pyridine carboxylic acid hydrazide, *Pak. J. Pharm. Sci.,* **2015**, *28*, 2129-2134.

[15] Pitucha, M.; Woś, M.; Miazga-Karska, M.; Klimek, K.; Mirosław, B.; Pachuta-Stec, A.; Gładysz, A.; Ginalska, G. Synthesis, antibacterial and antiproliferative potential of some new 1-pyridinecarbonyl-4-substituted thiosemicarbazide derivatives, *Med. Chem. Res.*, **2016**, *25*, 1666-1677.

[16] Molnar, M.; Čačić, M.; Zec Zrinušić, S. Synthesis and antioxidant evaluation of Schiff bases derived from 2,6-pyridinedicarboxylic acid, *Lett. Org. Chem*., **2012**, *9*, 401-410.

[17] Molnar, M.; Pavić, V.; Šarkanj, B.; Čačić, M.; Vuković, D.; Klenkar, J. Mono- and bis-dipicolinic acid heterocyclic derivatives – thiosemicarbazides, triazoles,oxadiazoles and thiazolidinones as antifungal and antioxidant agents, *Heterocycl. Commun*., **2017**, *23*, 35-42.

[18] Hocquet A.; Langgård, M. An evaluation of the MM+ force field*, Mol. Model*., **1998**, *4*, 94-112.

[19] Dewar, M.J.S.; Zoebisch, E.G.; Healy, E.F.; Stewart, J.J.P. AM1: A new general purpose quantum mechanical model. *J. Am. Chem. Soc.,* **1985**, *107*, 3902-3909.

[20] Pedretti, A.; Villa, L.; Vistoli, G. Vega – an open platform to develop chemo-bio-informatics applications using plug-in architecture and script programming. *J. Comput. Aided. Mol. Des*., **2004**, *18*, 167-173.

[21] VCCLAB (Virtual Computational Chemistry Laboratory). http://146.107.217.178/lab/pclient/ (Accessed January 15, **2017**).

[22] Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*, Wiley-VCH Verlag GmbH., Weinheim, **2000**.

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

[24] Karelson, M. *Molecular Descriptors in QSAR/QSPR*, Wiley-Interscience, New York, **2000**.

[25] Golbraikh, A.; 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*, 358-369.

[26] Topliss, J.G.; Costello R.J. Chance correlations in structure–activity studies using multiple regression analysis. *J. Med. Chem*. **1972**, *15*, 1066-1068.

[27] Gramatica, P. Principles of QSAR models validation: internal and external. *QSAR Comb. Sci.*, **2007**, *26*, 694-701.

[28] Todeschini, R.; Consonni, V.; Maiocchi, A. The K correlation index: theory development and its application in chemometrics. *Chemom. Intell. Lab. Syst.*, **1999**, *46*, 13-29.

[29] Tropsha, A. Best practices for QSAR model development, validation, and exploitation. *Mol. Inform.* **2010**, *29*, 476-488.

[30] O'Brien, R.M. A caution regarding rules of thumb for variance inflation factors. *Qual*. *Quant*., **2007**, *41*, 673-690.

[31] Eriksson, L.; Jaworska, J.; Worth, A.P.: Cronin, M.T.D.; McDowell, R.M.; Gramatica, P. Methods for reliability and uncertainty assessment and for applicability evaluations of classification- and regression-based QSARs. *Environ. Health Perspect.*, **2003**, *111*, 1361-1375.

[32] Holder, A.J.; Ye, L.; Yourtee, D.M.; Agrawal, A.; Eick, J.D.; Chappelow, C.C. An application of the QM-QSAR method to predict and rationalize lipophilicity of simple monomers. *Dent. Mater.*, **2005**, *21*, 591-598.

[33] Masand, V.H.; Mahajan, D.T.; Nazeruddin, G.M.; Hadda, T.B.; Rastija, V.; Alfeedy, A.M. Effect of information leakage and method of splitting (rational and random) on external predictive ability and behavior of different statistical parameters of QSAR model. *Med. Chem. Res.*, **2015,** *24*, 1241-1264.

[34] Roginsky, V.; Lissi, E.A. Review of methods to determine chain-breaking antioxidant activity in food. *Food Chem.*, **2005**, *92*, 235-254.

[35] Molnar, M.; Komar, M.; Brahmbhatt, H.; Babić, J.; Jokić, S.; Rastija, V. Deep eutectic solvents as convenient media for synthesis of novel coumarinyl Schiff bases and their QSAR studies. *Molecules*, **2017**, *22,* 1482.

[36] Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug. Deliv. Rev*., **2001**, 23, 3-25.

[37] Clark, D.E. Rapid calculation of polar molecular surface area and its application to the prediction of transport phenomena. 2. Prediction of blood-brain barrier penetration. *J. Pharm. Sci*., **1999**, *88*, 815-821.

[38] Prasanna, S.; Doerksen, R.J. Topological polar surface area: A useful descriptor in 2D-QSAR. *Curr. Med. Chem.*, **2009**, *16*, 21-41.

[39] Devinyak, O.; Havrylyuk, D.; Lesyk, R. 3D-MoRSE descriptors explained. *J. Mol. Graph. Model.* **2014**, *54*, 194-203.

[40] Deshpande, S.; Solomon, V.R.; Katti, S.B.; Prabhakar, Y.S. Topological descriptors in modelling antimalarial activity: N1-(7-chloro-4-quinolyl)-1,4-bis(3-aminopropyl)piperazine as prototype. *J. Enzyme. Inhib. Med. Chem.*, **2009**, *24*, 94-104.

[41] Pearlman, R.S.; Smith, K.M. Metric validation and the receptor-relevant subspace concept. *J. Chem. Inf. Comput. Sci.*, **1999**, *39*, 28-35.