

## A quantitative structure-activity relationship (QSAR) study of some diaryl urea derivatives of B-RAF inhibitors

Sedighe Sadeghian-Rizi<sup>1</sup>, Amirhossein Sakhteman<sup>2</sup>, Farshid Hassanzadeh<sup>3,\*</sup>

<sup>1</sup>Department of Medicinal Chemistry, School of Pharmacy and Pharmaceutical Sciences, Isfahan University of Medical Sciences, Isfahan, I.R. Iran.

<sup>2</sup>Department of Medicinal Chemistry, School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, I.R. Iran.

<sup>3</sup>Department of Medicinal Chemistry and Novel Drug Delivery Systems Research Center, School of Pharmacy and Pharmaceutical Sciences, Isfahan University of Medical Sciences, Isfahan, I.R. Iran.

### Abstract

In the current study, both ligand-based molecular docking and receptor-based quantitative structure activity relationships (QSAR) modeling were performed on 35 diaryl urea derivative inhibitors of <sup>V600E</sup>B-RAF. In this QSAR study, a linear (multiple linear regressions) and a nonlinear (partial least squares least squares support vector machine (PLS-LS-SVM)) were used and compared. The predictive quality of the QSAR models was tested for an external set of 31 compounds, randomly chosen out of 35 compounds. The results revealed the more predictive ability of PLS-LS-SVM in analysis of compounds with urea structure. The selected descriptors indicated that size, degree of branching, aromaticity, and polarizability affected the inhibition activity of these inhibitors. Furthermore, molecular docking was carried out to study the binding mode of the compounds. Docking analysis indicated some essential H-bonding and orientations of the molecules in the active site.

**Keywords:** QSAR; B-RAF inhibitors; Diaryl Urea; Docking; Multiple linear regressions; PLS-LS-SVM

### INTRODUCTION

RAF is among tyrosine kinase type receptors with serine/threonine kinase activity (1). Its contribution is in mitogen activated protein kinase (MAPK) signaling pathway, which conducts signals from membrane-based receptors to the nucleus to mediate cell proliferation, differentiation, and survival (2). Numerous cancers are related to the constitutive activation of the above signaling pathway (3). B-RAF is one of the isoforms of the RAF kinase family that can regulate multiple downstream molecules and is also regulated by a variety of signaling molecules (4,5). In about 7% of human cancers, the mutation of B-RAF has been detected (6-10).

Some small molecule RAF kinase inhibitors by diverse scaffolds such as ureas, urea bioisosteres, imidazoles, benzamides, oxindoles, and aza-stilbenes have emerged in the recent past decades (11,12). But diaryl urea have been most extensively investigated because of

sorafenib success in clinical for renal and hepatocellular carcinoma (13-15).

It is of great importance to introduce computer-aided drug design (CADD) approach to accelerate the time-consuming process of conventional drug discovery (16). Quantitative structure activity relationships (QSAR) and molecular docking are two of the helpful methods of CADD for drug design and prediction of drug activity (17,18). In QSAR large number of compounds are usually evaluated resulting in models that can predict the potency or activity of new or even non-synthesized compounds (19).

When the three-dimensional structure of the target protein is available or can be modeled, molecular docking is often used for screening of compound libraries. Molecular docking predicts the conformation of a protein-ligand complex and calculates the binding affinity and investigates protein-ligand interactions (20,21).

\*Corresponding author: F. Hassanzadeh  
Tel: 0098 3137927096, Fax: 0098 3136680011  
Email: hassanzadeh@pharm.mui.ac.ir

#### Access this article online



Website: <http://rps.mui.ac.ir>

DOI: 10.4103/1735-5362.194869

In this study aimed to develop a robust and accurate model for the inhibitory activity of inhibitors in order to design potential B-RAF kinase inhibitor. We used different method to connect between structural parameters and B-RAF kinase inhibitory. These methods included multiple linear regressions (MLR) as linear method and partial least squares least squares support vector machine (PLS-LS-SVM) as a nonlinear approach. The latter method was used to carry out non-linear mappings on the physicochemical and biological descriptors of the molecules. In Support vector machines, nonlinear kernel based functions were used to solve both regression and classification problems. An advantage of this method is its reproducibility in data mapping (22). Our aim in this study was to develop more examples of modeling based on this approach. Finally docking study was performed to suggest a binding mode for the inhibitors on B-RAF target.

## MATERIALS AND METHODS

All calculations were made using an Intel Core i5 (CPU 2.6 GHZ) laptop running on windows 7 operating system. The m-files for MATLAB calculations were developed in our group.

### Dataset and descriptor generation

The dataset used in this study was taken from the work of Menard, *et al.* (23). Chemical structure of 35 studied compounds is provided in Table 1. This set contains diarylurea derivatives with inhibition potency against B-RAF kinase. The chemical structures of molecules were drawn and optimized by HyperChem 7.0 software (HyperCube Inc. USA). Energy minimizations for all compounds were performed by AM1 semi-empirical method with Polark-Ribiere algorithm until the root-mean-square gradient of 0.01 Kcal/mol was reached. The resulted geometries were transferred into Dragon program (developed by Milano Chemometrics and QSAR Group) to calculate descriptors. The physicochemical parameters were calculated utilizing HyperChem and Dragon softwares. Molecular descriptors computed using the Dragon software were constitutional,

functional, topological, and geometrical groups. Hyperchem was used to obtain chemical descriptors such as Log P, hydration energy, polarizability, molar refractivity, molecular volume, and molecular surface area. Gaussian 98 W package was employed to use HF method at 6-31G\* basis set for optimization and calculation of different quantum chemical descriptors including dipole moment, local charge on atoms, high-occupied molecular orbital (HOMO), and low-unoccupied molecular orbital (LUMO) energies. Indices of electronegativity, electrophilicity, hardness, and softness were calculated from the energies of HOMO and LUMO.

### Model building

The correlation of the calculated descriptors with each other was calculated and collinear descriptors (0.85) were specified. Those with higher correlation towards activity vector were retained and the others were eliminated.

Splitting of the matrix into calibration (train) and external (test) set was done using kenard-stone algorithm. Subsequently, stepwise multiple linear regressions and partial least square analysis were performed on the training set for MLR and support vector machine (SVM) methods. In case of PLS-LS-SVM, Gaussian RBF Kernel with two tuning parameters,  $\gamma$  (gama) and  $\sigma^2$  (sigma<sup>2</sup>) were used. Latent variables of partial least squares (PLS) were entered to the models using stepwise method. The predicted residual error sum of squares (PRESS) was used to select the best models for validity evaluations.

### Validation of QSAR models

To evaluate robustness of the models in the calibration subset, leave-one-outcross validation method was used. PRESS and  $Q^2_{LOO}$  values were thereafter calculated according to the equations 1 and 2.

$$Press = \sum_{i=1}^n [y(\text{exp}) - y(\text{pred})]^2 \quad (1)$$

$$Q^2_{LOO} = \text{press}/\text{ssy} \quad (2)$$

where, SSY is the sum of squares of the response values. The models obtained from calibration set were then used to predict the response values of the test set. PRESS and  $R^2$  of test set were calculated as described above.

**Table 1.** Chemical structure of B-RAF kinase inhibitor used in this study

ID	pIC50 (μM)	Ring B	Ring C	R <sub>1</sub> /R <sub>2</sub>	ID	pIC50 (μM)	Ring B	Ring C	R <sub>1</sub> /R <sub>2</sub>
1	0.023			H/H	19	0.006			H/H
2	0.001			H/H	20	0.012			H/H
3	0.025			H/H	21	0.007			H/H
4	0.106			H/H	22	0.037			H/H
5	0.03			H/H	23	0.007			H/H
6	10			H/H	24	0.22			H/H
7	0.022			H/H	25	0.079			H/H
8	6.5			H/H	26	0.014			H/H
9	0.021			H/H	27	0.638			H/H

Table 1. (Continued)

ID	pIC50 ( $\mu\text{M}$ )	Ring B	Ring C	R <sub>1</sub> /R <sub>2</sub>	ID	pIC50 ( $\mu\text{M}$ )	Ring B	Ring C	R <sub>1</sub> /R <sub>2</sub>
10	0.037			H/H	28	0.023			Me/H
11	0.090			H/H	29	0.156			H/Me
12	0.025			H/H	30	0.017			Me/H
13	0.023			H/H	31	0.167			H/Me
14	0.047			H/H	32	0.037			Me/H
15	0.056			H/H	33	0.43			H/Me
16	0.001			H/H	34	0.010			Me/H
17	0.016			H/H	35	0.024			H/Me
18	0.012			H/H					

Finally, chance correlation test was used by scrambling the response vectors for several times and recalculating error metrics (PRESS and  $R^2$ ). Applicability domain for the best model was also done using Williams plot. For this purpose standard residuals were plotted against leverage for the data set (24). Other parameters of validation including K and K' were also provided for both modeling approaches.

### Molecular docking

Docking was conducted using AutoDock 4.2 software package. For this purpose, the crystal structure of B-RAF was retrieved from Protein Data Bank (PDB code: 1UWH). Grid and docking parameter files was generated by AutoDockTools version 6.5. For the preparation of protein, the original ligand and water molecules were removed from the coordinates and Kollman charges together with polar hydrogens were added. For ligands, Gasteiger charges were assigned and non-polar hydrogens were merged. The grid map of  $60 \times 60 \times 60$  points in x, y, and z directions was centered on the binding site (Cys 531) with a spacing of 0.375 Å. Lamarckian genetic algorithm (LGA) was used for docking with the following settings: a maximum number of 25,000,000 energy evaluations, an initial population of 300, a maximum number of 27,000 generations, and 100 independent docking runs. Other parameters of .dpf file were remained as default. All parameters of docking were validated based on re-docking of the cognate ligand (BAY 43-9006) and measuring the root-mean-square deviation (RMSD) value between the best pose of the docked structure with its conformation at crystal state. The RMSD value of less than 2 Å can validate the protocol of docking. Results were ranked according to the binding free energy and mode of interactions. Ligplot program was used to visualize hydrophobic and hydrogen-bonding interactions between the ligand and receptor (25).

## RESULTS

Sixty two remained descriptors constructed an input file for stepwise selection based on MLR analysis. Comparison of statistical

parameters was employed to result the best model as seen in the models bellow (Table 2). The numbers of training set compounds in the first and second model were 31 and 28, respectively.

$$pIC_{50} = -34.171 (\pm 5.597) - 0.006 (\pm 0.002). \alpha_{zz} - 0.091 (\pm 0.009). G (N...N) + 3.260 (\pm 0.379). TI2 - 0.110 (\pm 0.016). DISPm + 22.523 (\pm 10.470). PW3 + 7.682 (\pm 1.542). BLI + 50.35 (\pm 17.976). PW4 + 2.055 (\pm 0.750). (PJI3). \quad (3)$$

where,  $pIC_{50}$  is  $-\log (IC_{50})$ ,  $N = 31$ ,  $F = 26.102$ ,  $R^2 = 0.905$ ,  $Se = 0.3131$ .

$$pIC_{50} = -9.576 (\pm 8.537) + 0.062 (\pm 0.12). G (N...N) + 1.981 (\pm 0.891). DISPm + 66.775 (\pm 12.261). BLI - 34.512 (\pm 10.247). (FDI). \quad (4)$$

where,  $N = 28$ ,  $F = 13.441$ ,  $R^2 = 0.837$ ,  $Se = 0.5144$ .

In this equation, the values in the parentheses show the standard deviation of the coefficients. N, R, Se, and F are numbers of components, coefficient, standard error of regression, and Fisher's F-ratio, respectively.

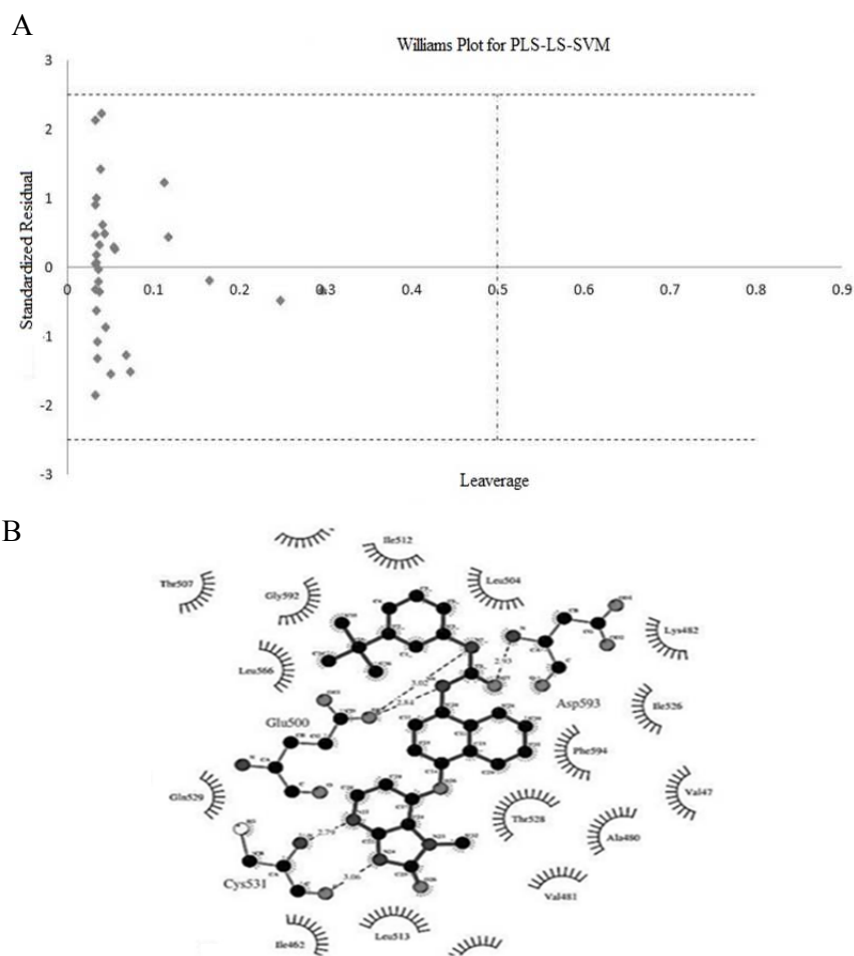
As a result for PLS-LS-SVM, the first three latent variables resulted in a model with reasonable  $R^2$  values (0.93 and 0.84).

Internal validation of the models was revealed by means of leave-one-out cross-validation. The results for cross-validation studies of both modeling approaches are listed in Table 2.

The more robustness of LS-SVM has been therefore shown based on both PRESS and  $R^2$  values. Furthermore, LS-SVM was more resistant to the number of compounds in training and test set with respect to MLR. Chance correlation test by means of Y-permutation ensured that the presented models were not achieved randomly.

As depicted in Fig. 1a, the molecules were not more than the cut-off h value of leverage. Applicability domain studies for the model based on PLS-LS-SVM have revealed that the model can be applied for structurally relevant compounds.

Docking analysis revealed important H-bonding interactions between urea moiety with Glu500 and Asp593 residues in allosteric pocket. H-bonding between tow nitrogen atoms of imidazopyridine moiety with Cys531 residue of hinge binding region was also significant.



**Fig. 1.** (a) applicability domain based on williams plot, (b) interaction model of docked inhibitor 32 on B-RAF kinase.

**Table 2.** Comparison of statistical parameters of the models

Model	R <sup>2</sup>	Q <sup>2</sup>	F	Number of descriptors	Number of compounds
Model 1	0.905	0.8145	26.102	8	31
Model 2	0.83	0.67	13.441	4	28

**Table 3.** The results for QSAR modeling using MLR and PLS-LS-SVM

Statistics	MLR	PLS-LS-SVM
Models	1	2
No. LVs	-	3
Function Type	-	Gaussian (RBF)
preprocessing	Auto scale	Auto scale
R <sup>2</sup> <sub>cal</sub>	0.905	0.9334
Q <sup>2</sup> <sub>looCV</sub>	0.8145	0.76
R <sup>2</sup> <sub>ex. Val.</sub>	0.87	0.87
R <sup>2</sup> <sub>y-perm. test</sub>	<0.33	<0.25
Press <sub>cal</sub>	2.1578	1.51
Press <sub>looCV</sub>	4.2631	5.22
Press <sub>ext. val</sub>	6.578	2.003
K	0.96961	0.93526
K <sup>2</sup>	0.79158 (Failed)	1.02743 (passed)
External test	7,19,24,26	7,9,10,11,24,27,33

**Table 4.** Binding energy resulted from Autodock software

Compound	Binding energy (kcal/mol)	Compound	Binding energy (kcal/mol)
1	-12.09	19	-14.10
2	-12.33	20	-12.08
3	-12.20	21	-11.45
4	-12.38	22	-14.03
5	-14.75	23	-11.88
6	-12.57	24	-13.51
7	-12.39	25	-13.58
8	-11.39	26	-15.69
9	-12.80	27	-14.75
10	-12.66	28	-15.11
11	-10.59	29	-14.19
12	-12.64	30	-14.49
13	-12.61	31	-13.62
14	-12.02	32	-16.19
15	-13.41	33	-15.05
16	-12.33	34	-14.53
17	-12.15	35	-14.85
18	-11.51		

## DISCUSSION

$R^2$ ,  $Q^2$ , and F ratio of two models were compared, model 1 revealed the best statistical parameters  $R^2 = 0.905$ ,  $Q^2 = 0.8145$ ,  $F = 26.102$  (Table 2). The more power of LS-SVM compared to MLR has been verified based on external validation including K and K' as displayed in Table 3. SVM model has passed the validation criteria and showed to be a more powerful modeling approach.

The best MLR model contained 8 descriptors including 4 topological parameters (TI2, BLT, PW3, PW4), 3 geometrical parameters (G(N..N), PJI3, DISPm), and one chemical parameter ( $\alpha_{zz}$ ). Based on descriptor coefficients, increasing  $\alpha_{zz}$ , G (N..N), and DISPm can decrease pIC<sub>50</sub> and increasing TI2, PJI3, BLI, PW3, and PW4 can increase the extent of pIC<sub>50</sub> of the structures.

TI2 is a topological descriptor belonging to the Mohar indices which can affect solubility. TIs can be sensitive to size, shape, symmetry, branching, and cyclicality.

G (N..N) is a 3D geometrical descriptor relating to the geometrical distance between nitrogen and nitrogen in molecules with more than one N-atom.

DISPm is related to the molecular geometry as well as molecular size and describes the conformational features of the molecules. Negative contribution of DISPm revealed that rigidity of molecules can increase inhibitory activity due to the absence of long and flexible substituents and the presence of unsaturated bonds. Kier benzene-likeness index (BLI) is

an aromaticity index calculated from molecular topology. Aromaticity might influence the transport or distribution of the chemical across cell membrane.

PW3 and PW4 as the path/walk count ratio are independent of molecular size. These descriptors can be considered as shape descriptors whose value increases with increased branching in the vertices.

PJI3 (Petitjean shape indices), a first Petitjean shape index, is a topological anisometry descriptor.  $\alpha_{zz}$  is the polarizability along zz axis of the molecule which represents how readily the molecular charge distribution is distorted by external electric fields. Almost 7 descriptors appeared in the SMLR equation belongs to the geometrical and topological groups of descriptors.

In Support vector machines, nonlinear kernel based functions are used to solve both regression and classification problems. To reduce the complexity of optimization, least squares support vector machines has emerged in the recent years. In the following regression equation,  $\omega$  and  $b$  are the slope and the intercept of the equation, respectively.

$$F(x) = \omega^T \varphi(x_i) + b_i = 1, 2, \dots \quad (5)$$

The function  $\varphi(x_i)$  is used to map the input space into a higher dimensional space.

To investigate the role of descriptors on the first three variables, correlation of all descriptors with the loading values of the three variables were calculated. It was observed that the factors 1, 2, and 3 have higher loadings for TE, Ms, and FDI respectively. This result was found in both modes with 28 and 31

compounds. Total energy (TE) of a molecular system is among quantum chemical descriptors and is the sum of the total electronic energy (E<sub>ee</sub>) and the energy of intern clear repulsion (E<sub>nr</sub>).

$$TE = E_{ee} + E_{nr} \quad (6)$$

The mean electro topological state (M<sub>s</sub>) is among constitutional descriptor which is calculated by dividing S<sub>s</sub> by the number of non-hydrogen atoms (nSK):

$$M_s = S_s / nSK \quad (7)$$

Folding degree index (FDI) is a geometrical descriptor. Negative contribution of TE revealed that decrease in total energy for molecular systems can increase their inhibitory activity. Whereas positive values of M<sub>s</sub> and FDI descriptors show that the indicated descriptor contribute positive effect on the values of pIC<sub>50</sub>. According to experimental QSAR results and docking binding energies performed in this study, the modification of the phenyl middle ring using hydrophobic group such as halogen, thiomethyl substitution (2,6 compounds) or replacement by naphthyl group (24-35, 22, 19, and 5 compounds) can enhance the inhibitory activity due to an increase in the potency of binding and this result was in accord with the previously reported SAR result (23).

The binding energy of all compounds is reported in Table 4 and the compound with naphthyl moiety have the best binding energy. In Fig. 1b, orientation of the interactions for compound 32 with the lowest binding energy is depicted.

## CONCLUSION

In this work, molecular docking and structure-based QSAR method were performed to explore structural features, orientations, and conformations of some inhibitors interacting with B-RAF kinase. A structure-based QSAR model using MLR and LS-SVM were developed. The result indicated that the performance of LS-SVM model was better than the MLR model.

The selected descriptors indicated that size, degree of branching, aromaticity, and polarizability affected the inhibition activity of

these inhibitors. Docking analysis indicated some essential H-bonding and orientations in active site. This information will guide us through design of novel B-RAF kinase inhibitors.

## ACKNOWLEDGEMENTS

The content of this paper is extracted from the PhD thesis (No. 394159) which was financially supported by the Isfahan University of Medical Sciences, Isfahan, Iran.

## REFERENCES

1. Leicht DT, Balan V, Kaplun A, Singh-Gupta V, Kaplun L, Dobson M, *et al.* RAF kinases: function, regulation and role in human cancer. *Biochim Biophys Acta.* 2007;1773(8):1196-1212.
2. Peyssonnaud C, Eychene A. The RAF/MEK/ERK pathway: new concepts of activation. *Biol Cell.* 2001;93(1-2):53-62.
3. Malumbres M, Barbacid M. RAS oncogenes: the first 30 years. *Nat Rev Cancer.* 2003;3(6):459-465.
4. Roskoski R Jr. RAF protein-serine/threonine kinases: Structure and regulation. *Biochem Biophys Res Commun.* 2010;399(3):313-317.
5. Hagemann C, Rapp UR. Isozyme-specific functions of RAF kinases. *Exp Cell Res.* 1999;253(1):34-46.
6. Huang T, Zhuge J, Zhang WW. Sensitive detection of BRAFV600E mutation by Amplification Refractory Mutation System (ARMS)-PCR. *Biomark Res.* 2013;1:3.
7. Lee B, Mukhi N, Liu D. Current management and novel agents for malignant melanoma. *J Hematol Oncol.* 2012;5:3.
8. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, *et al.* Mutations of the BRAF gene in human cancer. *Nature.* 2002;417(6892):949-954.
9. Domingo E, Laiho P, Ollikainen M, Pinto M, Wang L, French AJ, *et al.* BRAF screening as a low-cost effective strategy for simplifying HNPCC genetic testing. *J Med Genet.* 2004;41(9):664-668.
10. Rahman MA, Salajegheh A, Smith RA, Lam AK. B-RAF mutation: A key player in molecular biology of cancer. *Exp Mol Pathol.* 2013;95(3):336-342.
11. Li HF, Lu T, Zhu T, Jiang YJ, Rao SS, Hu LY, *et al.* Virtual screening for RAF-1 kinase inhibitors based on pharmacophore model of substituted ureas. *Eur J Med Chem.* 2009;44(3):1240-1249.
12. Wong KK. Recent developments in anti-cancer agents targeting the Ras/RAF/ MEK/ERK Pathway. *Recent Pat Anticancer Drug Discov.* 2009;4(1):28-35.
13. Pratilas CA, Solit DB. Targeting the mitogen-activated protein kinase pathway: physiological feedback and drug response. *Clin Cancer Res.* 2010;16(13):3329-3334.



14. Stenner F, Chastonay R, Liewen H, Haile SR, Cathomas R, Rothermundt C, *et al.* A pooled analysis of sequential therapies with soRAFenib and sunitinib in metastatic renal cell carcinoma. *Oncology*. 2012;82(6):333-340.
15. Wu J, Zhu AX. Targeting insulin-like growth factor axis in hepatocellular carcinoma. *J Hematol Oncol*. 2011;4:30.
16. Zhang S. Computer-Aided Drug Discovery and Development. *Methods Mol Biol*. 2011;716:23-38.
17. Hansch C, Hoekman D, Gao H. Comparative QSAR: toward a deeper understanding of chemicobiological interactions. *Chem Rev*. 1996;96(3):1045-1076.
18. Foroumadi A, Sakhteman A, Sharifzadeh Z, Mohammadhosseini N, Hemmateenejad B, Moshafi MH, *et al.* Synthesis, antituberculosis activity and QSAR study of some novel 2-(nitroaryl)-5-(nitrobenzylsulfinyl and sulfonyl)-1,3,4-thiadiazole derivatives. *Daru*. 2007;15(4):218-226.
19. Pirard B. Computational methods for the identification and optimisation of high quality leads. *Comb Chem High Throughput Screen*. 2004;7(4):271-280.
20. Cavasotto CN, Orry AJ. Ligand docking and structure-based virtual screening in drug discovery. *Curr Top Med Chem*. 2007;7(10):1006-1014.
21. Kitchen DB, Decornez H, Furr JR, Bajorath J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov*. 2004;3(11):935-949.
22. Khoshneviszadeh M, Sakhteman A. Exploring quantitative structure-activity relationship (QSAR) models for some biologically active catechol structures using PC-LS-SVM and PC-ANFIS. *J Appl Biol Chem*. 2016;59(3):433-441.
23. Menard D, Niculescu-Duvaz I, Dijkstra HP, Niculescu-Duvaz D, Suijkerbuijk BM, Zambon A, *et al.* Novel potent BRAF inhibitors: toward 1 nM compounds through optimization of the central phenyl ring. *J Med Chem*. 2009;52(13):3881-3891.
24. Tetko IV, Sushko I, Pandey AK, Zhu H, Tropsha A, Papa E, *et al.* Critical assessment of QSAR models of environmental toxicity against *Tetrahymena pyriformis*: focusing on applicability domain and overfitting by variable selection. *J Chem Inf Model*. 2008;48(9):1733-1746.
25. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng*. 1995;8:127-134.