



Exploring quantitative structure–activity relationship (QSAR) models for some biologically active catechol structures using PC-LS-SVM and PC-ANFIS

Mehdi Khoshneviszadeh^{1,2} · Amirhossein Sakhteman^{1,2,3}

Received: 16 December 2015 / Accepted: 16 February 2016 / Published online: 29 February 2016
© The Korean Society for Applied Biological Chemistry 2016

Abstract Exploring predictive QSAR models for dopamine catechol structures could be used in designing more potent ligands. In this study, efforts were taken to find out the most important molecular features responsible for the biological activity of catechol structures. All 2D descriptors of Dragon including constitutional, topological, molecular walk counts, BCUT descriptors, Galvez topological, 2D autocorrelations, functional groups, atom-centred fragments, empirical descriptors and properties were calculated for the structures. Two non-linear modelling methods (PC-LS-SVM and PC-ANFIS) were used and compared in this QSAR study. The results revealed the more predictive ability of PC-LS-SVM in the QSAR analysis of the compounds with catechol substructure. The roles of topological properties and number of hydrogen bond donors group as molecular features responsible for the activity of the compounds were discussed. The obtained QSAR models can be used in future studies of drug development for human dopamine D2 receptor.

Keywords Catechol structures · Human dopamine D2 receptor · Quantitative structure–activity relationship

Introduction

Dopamine neurotransmission plays an important role in several sensory processing functions such as novelty detection, attention, memory formation and coding of rewarding stimuli (Spuhler and Hauri 2013; Narendran et al. 2014). Dopaminergic pathways are also involved in the manifestation of CNS pathologies including Parkinson's disease, schizophrenia and substance abuse (Tinsley et al. 2009).

The cell-borne receptors of dopamine can be classified into two pharmacologic families (D₁ and D₂-like) which are encoded by at least five genes (Levey et al. 1993; Flisikowski et al. 2009). The D₁-like receptors stimulate the activity of adenylyl cyclase and their pharmacological functions from known ligands which are more or less identical. The D₂-like receptors, i.e. D₂, D₃ and D₄ receptors have long third intracellular loops with short C-terminal tails and inhibit adenylyl cyclase (Mizuta et al. 2013; Plouffe and Tiberi 2013).

From a pharmacological point of view, the D₂-like family, including D₂, D₃ and D₄, appears to be involved in schizophrenia (Dzirasa et al. 2006). The antagonists of D₂-like dopamine receptor are therefore being used in the treatment of schizophrenia (Dzirasa et al. 2006).

It is therefore of great importance to achieve a more comprehensive view of the structural features which are responsible for the activity of these types of compounds (Yeagle and Albert 2007; Tinsley et al. 2009). The results of modelling research in this area could provide a rational pattern for the interaction of these compounds with D₂ receptor. The obtained results are useful in design and synthesis of more potent pharmaceutical agents in future studies. Quantitative structure–activity relationship (QSAR) can extend our

✉ Amirhossein Sakhteman
asakhteman@sums.ac.ir; nava20ir@gmail.com

¹ Department of Medicinal Chemistry, School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran

² Medicinal and Natural Products Chemistry Research Center, Shiraz University of Medical Sciences, Shiraz, Iran

³ Department of Medicinal Chemistry, Faculty of Pharmacy, Shahid Sadoughi University of Medical Sciences, Yazd, Iran

knowledge about the role of molecular features responsible for the activity of dopamine antagonists (Yeagle and Albert 2007). The common approaches used in QSAR studies are normally based on linear methods such as Multiple Linear Regression (MLR) and Principal Component Regression or non-linear methods like support vector machines, Neural networks and Fuzzy mappings (Buyukbingol et al. 2007; Afuni-Zadeh and Azimi 2010). A QSAR study based on MLR (Multiple Linear Regression) and Artificial Neural Network methods has been recently developed for a set of 58 benzamide antagonists (Fatemi and Dorostkar 2010). In some other studies the role of electronic descriptors on the activity of dopamine antagonists were explained (Pontiki et al. 2005). QSAR modelling based on adaptive network-based fuzzy inference systems (ANFIS) and least squares support vector machines (LS-SVM) can present an interactive method and overcome some limitations of linear traditional QSAR methods. This fact is a result of their ability to carry out non-linear mappings on the physicochemical and biological descriptors of the molecules (Dastorani et al. 2010; Balabin and Lomakina 2011; Zhang et al. 2011). Using two authoritative and trust worthy non-linear techniques of QSAR (ANFIS and SVM) and their subsequent evaluations in a sample data set of catechol structures, it is possible to obtain a rational model of dopaminergic activity for the synthesized compounds.

In continuation to our work on developing QSAR models, in the present study we have explored ANFIS and SVM models on the eigenvectors of some catechol derivatives as antagonists of human D2 receptor. The results for PC-ANFIS were compared with those of PC-SVM. Finally, the accuracy of the developed models were illustrated using leave-one-out (LOO), leave-many-out (LMO), Y-randomization and external validation techniques. The results can be used to estimate the activity of similar compounds in future studies.

Materials and methods

All calculations were done using an intel core i5 (CPU 2.6 GHZ) laptop running on windows 8. The calculations including Kennard–Stone, cross-validation, chance correlation and external validation were based on literature (Kiralj and Ferreira 2009; Goodarzi and Freitas 2010; O’Boyle et al. 2011; QSAR_Tools 2011; Sakhteman et al. 2011; Chirico and Gramatica 2012; Gaulton et al. 2012; Bento et al. 2014). The m-files for calculations were developed in our group.

Preparation of the dataset

A series of compounds and their affinities as K_i values towards human dopamine D2 receptor were retrieved from

CHEMBL database as SMILES strings (Gaulton et al. 2012; Bento et al. 2014). The compounds with catechol ethylamine substructure were extracted using substructure search as implemented in Open babel 2.3.2 (O’Boyle et al. 2011). Another exclusion criterion for removing the compounds from data set was diversity of bioassay techniques. For this purpose, the compounds with different assay methods were excluded from the data. Finally, 43 compounds were extracted based on the used filters for the next step of QSAR analysis. Iterative runs of open babel using a batch shell script provided a primary 3D generation of the compounds as mol2 coordinates (Sakhteman et al. 2011). The 3D structures were thereafter minimized using Hyperchem (Hypercube Inc. Gainesville, FL 32601, United States). MM + force field as implemented in hyperchem and Polak–Ribiere minimization algorithm was used for optimization procedure. Tcl scripting as implemented in Hyperchem was used for batch minimization of the structures. The final minimized compounds were subjected to Dragon 2.1 for calculation of 1051 two-dimensional (2D) descriptors. The matrix of the descriptors for the compounds and their corresponding affinity vector as p-function ($-\log K_i$) were entered into Matlab software for further analysis.

QSAR modelling

A primary data reduction was used in order to remove autocorrelation from the dataset. During this procedure the Pearson correlation coefficients between each pair of descriptors were calculated. Among the two descriptors with a correlation value of more than 0.85, the one with less correlation towards the affinity response vector was excluded from the data set. The matrix of the remaining descriptors was divided into calibration and test set using Kennard–Stone method (Li et al. 2011). A subset of the compounds in the dataset was selected for external test set. The resulted matrices were subjected to principal component analysis and the eigenvectors were generated for both subsets. Subsequently, LS-SVM and ANFIS were used as two modelling methods in order to train the calibration subset. For least squares support vector machines, gaussian RBF kernel with two tuning parameters, γ (gam) and σ^2 (sig2) were used. In case of ANFIS, sugeno-type fuzzy models were built using different membership functions and the number of epochs for training was set to 50 (Goodarzi and Freitas 2010). In order to obtain the optimum number of principal components for both modelling methods, the eigenvectors of principal component analysis (PCA) were entered to the models using stepwise method. The predicted residual error sum of squares (PRESS) was used as an error metric to select the best models for further validity evaluations.

Validation of QSAR models

Different validation methods were used to evaluate the predictive ability and robustness of the models. In order to assess the robustness of the models in the calibration subset, LOO (leave-one-out) and LMO 20 % (leave-many-out) cross-validation methods were used. In case of LOO, each compound was once excluded from the data set and its response value was predicted using the obtained model. To perform LMO, the data set was divided into *N* blocks with 20 % of the compounds in each one. The blocks were consecutively excluded from the model and predicted as described for LOO. Press, Q_{LOO}^2 and Q_{LMO}^2 values were thereafter calculated according to the Eqs. 1 and 2 (Chirico and Gramatica 2012).

$$\text{PRESS} = \sum [y_{(\text{exp})} - y_{(\text{pred})}]^2, \quad (1)$$

$$Q_{\text{LOO}}^2 = 1 - \left(\frac{\sum [y_{(\text{exp})} - y_{(\text{pred})}]^2}{\sum [y_{(\text{exp})} - y_{(\text{mean})}]^2} \right). \quad (2)$$

To evaluate the predictive ability of the models for external data, external validation analysis was performed. The models obtained from calibration sets were therefore used to predict the response values of the test set based on the optimum principal component eigenvectors. PRESS and R^2 of external validation were used to represent the predictive ability of the models against external data sets. Finally, response permutation test was used to ensure that the obtained values of correlation were not obtained by chance for training set and LOO cross-validation. During this procedure the randomly scrambled response vectors (*Y*) were used several times to train the data of calibration set. The high error values of the random models revealed that the presented model was not obtained by chance. Since two machine learning methods were used in this study, it was of great importance to check out that the models were not overfitted. To ensure that the obtained models were not over trained, Golbraikh and Tropsha acceptable model criteria was also calculated for both models. For this purpose, ExternalValidation1.0 java application was used for the calculation of r^2_m , K and K' . CCC as another metric for correlation coefficient together with MAE (mean average error) and RMSE (root mean square error) were also calculated for both models. (Kiralj and Ferreira 2009; QSAR_Tools 2011; Chirico and Gramatica 2012).

The Pearson correlation coefficient of each descriptor with the eigenvectors of principle component analysis was calculated to estimate the contribution of each descriptor in the used PCs. QSAR applicability domain was obtained by means of Williams plot. For this purpose leverage values were calculated and plotted against standard residuals (Tetko et al. 2008).

Results and discussion

The compounds used in QSAR modelling are displayed in Table 1. As it can be seen, all compounds are similar in terms of bearing a common catechol ethylamine substructure. Different amino groups including primary, secondary, tertiary and quaternary can be found among the compounds of the data set. Other substituents and fragments attached to catechol substructure were however diverse in the compounds of dataset. The most potent structure was compound 31 bearing dibenzo [de,g] quinoline ring moiety ($K_i = 3.8$ nM). The inactive structure 22 revealed a K_i value equivalent to 810,000 nM. All possible 2D descriptors of Dragon including constitutional, topological, molecular walk counts, BCUT descriptors, Galvez topological, 2D autocorrelations, functional groups, atom-centred fragments, empirical descriptors and properties were calculated for the structures. Based on pretreatment of primary data, 134 molecular descriptors were selected for developing a QSAR model with respect to the response affinity vector. Since too many descriptors were still present in the data matrix before non-linear modelling, principal component analysis was used as a dimension reduction strategy on data matrix. As described in the methods section, Kennard–Stone was used for splitting the dataset into calibration and test sets. Kennard–Stone algorithm is able to select a set of representative data which are uniformly distributed in the space. This technique is based on selecting the first two objects with the maximum distance from each other. The third object is thereafter selected in such a way to have the farthest distance from the primarily selected data. The (*M* + 1) th sample is therefore selected based on the criterion Max (Min distance (d_1C, d_2C, \dots, d_mC)). Where C is denouncing the candidate sample in the dataset. The compounds selected as test set are determined with T superscript in Table 1. As described earlier, two non-linear modelling methods were used to develop a reasonable QSAR model in this study. Autoscaling preprocessing method was tested for both modelling approaches and showed to be effective only in case of PC-ANFIS.

The ANFIS is used for solving problems related to parameter identification. A hybrid learning rule combining the back-propagation gradient descent and a least squares method is normally used in the ANFIS models. In ANFIS, a hybrid learning algorithm is used to identify the membership function using the parameters of single output for Sugeno-type FIS (fuzzy inference systems) (Sugeno and Taniguchi 2004). The ANFIS structure is consisting of 5 layers, the first of which is the fuzzification layer composed of $n \times p$ nodes. Where *n* is representing the number of input variables and *p* is representative of membership

Table 1 Compounds with catechol ethylamine substructure and their K_i values for human Dopamine D_2 receptor (Superscript T is used for compounds in test set)

Ki (nM)	Structure	Ki (nM)	Structure	Ki (nM)	Structure
350		1300		217 ^T	
510		12000		810000	
340		1900		1900	
8000		400		5400	
1700		2500		3900	
810		4400 ^T		3400	
3600		9100		1900 ^T	
2200		520 ^T		11400	
1400		340 ^T		1550	
710		3700 ^T		4670	
3.8		70.3		80	
88		2607 ^T		200	
29		328		115	
210		19.1		39.1 ^T	
21.7 ^T					

Table 2 The results for QSAR modelling of catechol structures using PC-LS-SVM and PC-ANFIS

	LS-SVM	ANFIS
Used PCs	1–6	1 AND 4
Function type	Gaussian RBF	Gaussian (input = 2, epochs = 50)
Preprocessing	–	Auto scaling
$R^2_{\text{Calibration}}$	0.950	0.84
$Q^2_{\text{LOO Cross-validation}}$	0.82	0.63
$Q^2_{\text{LMO(20 \% Cross-validation)}}$	0.80	0.58
$R^2_{\text{External validation}}$	0.83	0.66
$R^2_{\text{Y-permutation test}}$	<0.2	<0.24
$R^2_{\text{Y-permutation test LOO}}$	<0.26	<0.3
Press _{calibration}	1.11	3.48
Press _{LOO Cross-validation}	3.90	8.014
Press _{LMO Cross-validation}	3.95	8.6
Press _{External validation}	1.38	7.32
$Rm^2_{\text{External Validation}}$	0.63 (passed)	0.4 (failed)
CCC	0.84	0.6
RMSE	0.41	0.8
MAE	0.12	0.19
K	0.984 (passed)	0.95 (failed)
K'	0.996 (passed)	0.93 (failed)

functions. The membership functions of the input variables are present in this layer. In most studies bell-shaped functions are being used for fuzzification. Meanwhile, back propagation is being taken to calculate the adaptive parameters. The nodes in the second layer are used to multiply the input values and produce the products based on the below equation.

$$W_i = \mu_{A_i}(x) \times \mu_{B_i}(y). \quad (3)$$

In the above equation x has the linguistic value A_i and y has the linguistic value B_i .

The number of rules in the second layer is defined by the P^n nodes. If two rules are present in the system, four possible rules will emerge in the ANFIS structure. Each node in the third layer computes its firing strength to the sum of all the rules firing strength (Sugeno and Taniguchi 2004). The result of this layer is a normalized firing strength.

In layer 4, the nodes compute a parameter function on the third layer output. Parameters in this layer are called *consequent parameters*. The overall output is computed as the summation of all incoming signals in layer 5 according to the below equation.

$$\text{Overall output} = \sum \hat{W}_i F_i / \sum W_i. \quad (4)$$

In this study, a combination of least squares and back-propagation gradient descent method was used for training FIS membership function parameters in order to model a

given set of input/output data. In the first experiment all PCs were entered to the ANFIS structure based on their eigen values. Since no acceptable model was obtained using eigenvalue ranking approach, the PCs were entered into the ANFIS structure using stepwise method. The two metrics, PRESS and R^2 were used to evaluate the validity of the obtained models. As displayed in Table 2, the model obtained by the two PCs, 1 and 4 led to a reasonable R^2 for the data in the calibration set (0.84).

The structure of ANFIS model is also depicted in Fig. 1. In this study the combination of two gaussian membership functions was found to be optimum for building the best possible ANFIS model. As seen in Fig. 1, 4 rules were used in the ANFIS structure. The second modelling approach was using least squares support vector machines. Support vector machines are capable of solving both regression and classification problems via non-linear kernel-based functions. SVM is based on translating the input space to a higher dimensional feature space in order to cope with non-linearities. Least squares support vector machines has been thereafter emerged to reduce the complexity of optimization (Suykens et al. 2001). A set of linear equations was used instead of the quadratic programming as implemented in the original version of the support vector machines.

Keeping in the mind the below regression equation, ω and b are the slope and the intercept of the equation, respectively.

Fig. 1 The ANFIS structure for mapping the activity of compounds in data set

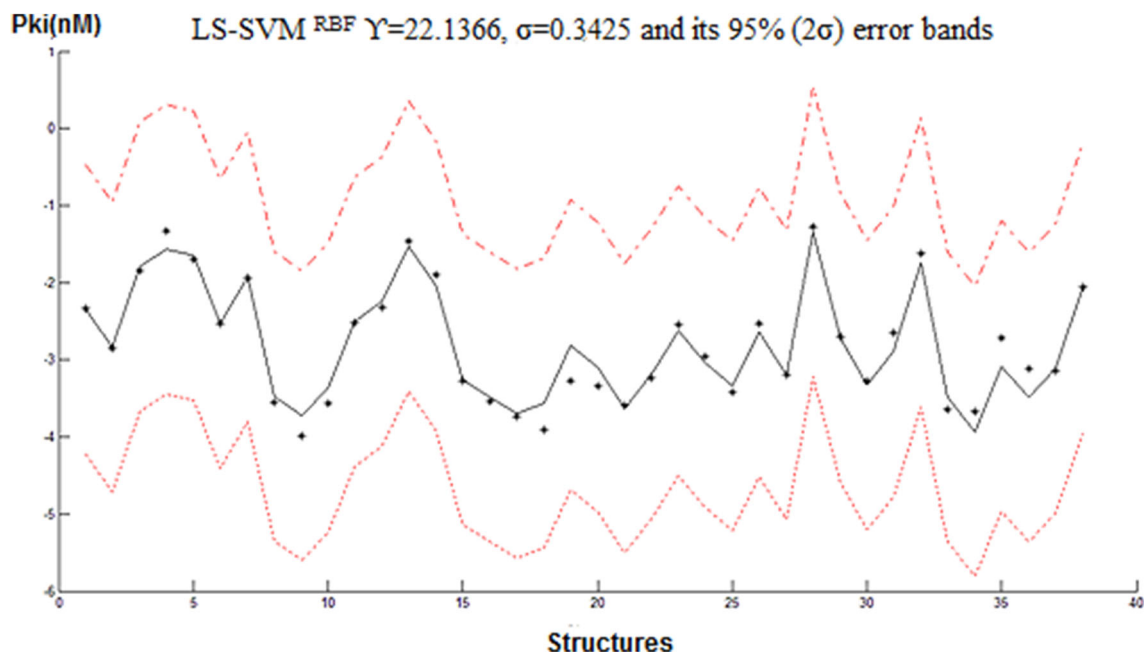
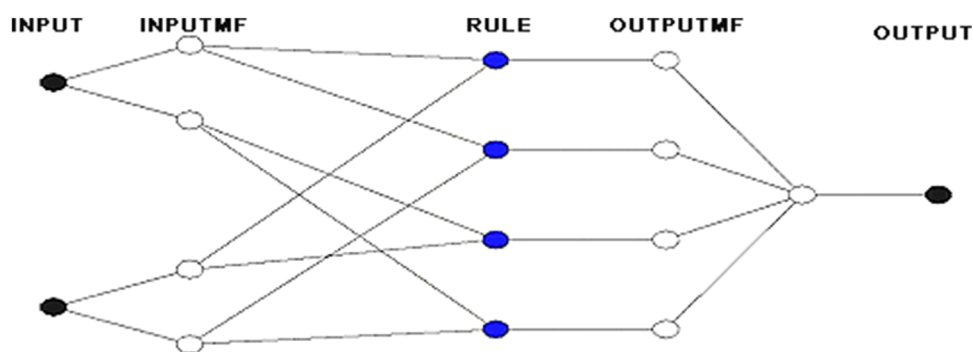


Fig. 2 Training step using LS-SVM

$$f(x) = \omega^T \varphi(xi) + b \quad i = 1, 2, \dots \quad (5)$$

$\varphi(xi)$ is the function for mapping the input space into higher dimensions. The following equation is used to minimize the function J .

$$\min J(\omega, b) = \frac{1}{2} \left(\omega^T \omega + C \frac{1}{2} \sum \varepsilon_i \right). \quad (6)$$

In the above equation, C is the regularization parameter to trade off the minimization of training error against that of model complexity.

In this study, the combination of the first six PCs based on eigenvalue ranking resulted in a model with the best R^2 value (0.95). The plot of training procedure with the used parameters for calibration set is displayed in Fig. 2. γ is the regularization parameter, determining the trade-off between the training error minimization and smoothness of the model. σ^2 is the squared bandwidth of gaussian function.

As seen, all structures were reasonably fitted within the model. Being satisfied with the training step, it was necessary to evaluate the quality of both models (ANFIS network and support vector machines) using different validation techniques. For this purpose cross-validation was performed using both LOO and LMO methods to evaluate the robustness of the obtained models. The result of cross-validation studies for both models are displayed in Table 2. It was observed that the robustness of LS-SVM is more than ANFIS model in terms of all metrics (PRESS and $Q_{LOO, LMO}^2$ values). Since the difference range between 0.1–0.2 is normally accepted for R^2 and Q^2 values, the robustness of PC-LS-SVM model was verified. Although R^2 m values were calculated for both models, it was revealed that only PC-LS-SVM is showing an acceptable value for K and K' meaning that this model was not overfitted. CCC, MAE and RMSE values were also more reasonable for PC-LS-SM compared to PC-ANIFS. The

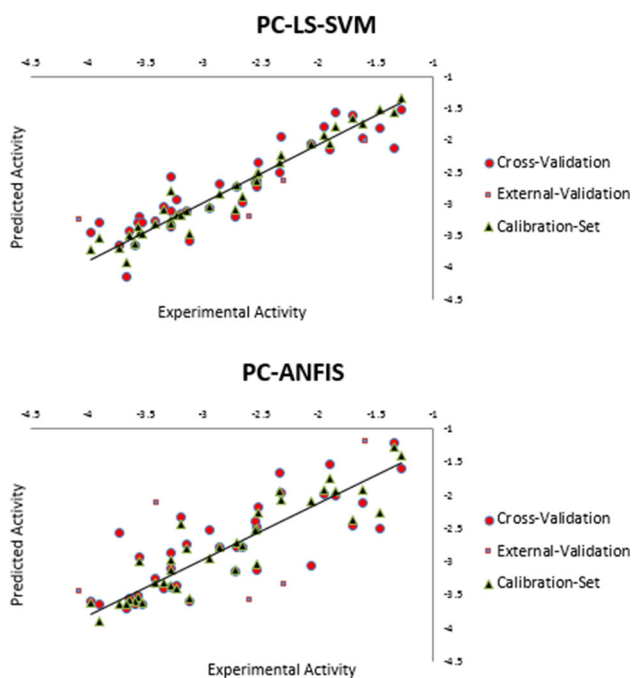


Fig. 3 Comparison of PC-LS-SVM and PC-ANFIS for prediction of activities in calibration and validation sets

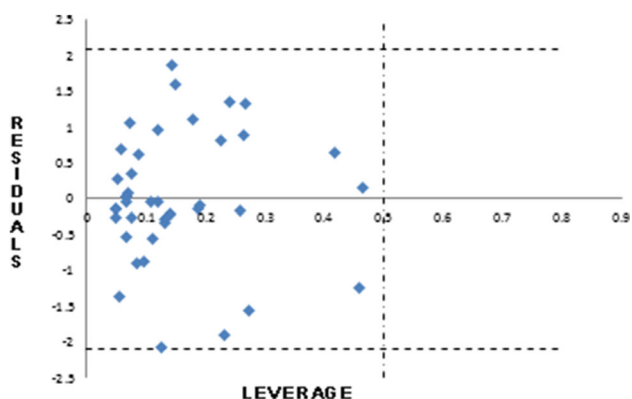


Fig. 4 Williams Plot for PC-LS-SVM model with six PCs

plot of experimental data versus calculated values for both models are also depicted in Fig. 3.

This result was in accord with a classification study for the more ability of SVM models in comparison with ANFIS (Ankishan and Yilmaz 2013).

In order to evaluate the predictive ability of QSAR models against external data, both models were subjected to test set. The results of external validation were also showing the more predictive ability of PC-LS-SVM against external data. Other external validation metrics including Golbraikh–Tropsha criteria was passed for LS-SVM but failed in the case of PC-ANFIS. Based on these validation studies, the model obtained by LS-SVM was proposed as a reasonable method for QSAR studies of compounds with catechol ethylamine substructure. The learning approach

was not overfitted and could be used in further studies of drug design. Y-responses permutation test was also performed on LS-SVM model in order to ensure that the proposed model was not obtained by chance. As seen in Table 2, the errors obtained by modelling random scrambled vectors with the matrix of principal components are significantly higher than those obtained for the real data set (24–54 vs. 1.11–3.90). This verifies that the primary model of support vector machines was not achieved by chance. QSAR studies must be also evaluated for applicability domain to ensure how the predicted data are reliable for structurally similar compounds. For this purpose, Williams graph was generated using standardized residuals and leverage data. As depicted in Fig. 4, no molecule was considered to be in the zone more than the cut-off h value for leverage. It is therefore showing that no predicted value is calculated by extrapolation in this data set.

In order to investigate the correlation of PCs with structural features of the compounds, Pearson correlation of each descriptor with the loading values of the first 6 PCs was calculated.

The results are displayed in Table 3. As it is concluded from Table 3, in case of each PC, the loading values for one or many descriptors are much higher than those of the others. The correlation values in Table 3 are indicative of descriptor weights inside each PC.

It should be noted that each PC is bearing information from all descriptors but the contribution of descriptors in different factors are not equal. As displayed in Table 3, HNar (Narumi harmonic topological index) is highly correlated to PC6. This descriptor is the number of non-hydrogen atoms divided by the reciprocal vertex degree. This descriptor can explain the more potency of those structures with extended scaffolds such as compound 31 and 43 in Table 1. For example, factors 1 and 2 have higher loadings for TIE and D/Dr_{10} , respectively. TIE is denouncing E-state which is a geometrical descriptor. This parameter is related to the characteristics of different atoms in the molecule. The information about D/Dr_{12} is highly incorporated in factor 4. D/Dr corresponds to distance/detour ring index and is among topological descriptors for the ring systems. D/Dr descriptors are reflecting the effects of different ring systems on the activity of the study structures. On the other hand, the loadings of factor 5 implied higher contribution for nHDon (number of donor atoms such as N and H for H-bonds). nHDon is a functional group descriptor where its presence in our QSAR mode represents the role of H-bonds. As seen in Table 3, H-050 (H attached to heteroatom) is another descriptor related to the ability of H-bond formation in this series of compounds. These two descriptors are in accord with other molecular docking studies of dopamine D2 receptor. It was reported that residues such as Ser193 and 194 as well as ASP14 in the

Table 3 Pearson correlation of the loadings for each PC with descriptors

	PC1	PC2	PC3	PC4	PC5	PC6
‘HNar’	−0.1303	0.4341	−0.1689	0.1204	−0.0979	−0.8353
‘D/Dr10’	0.0062	0.7517	−0.7756	0.0430	−0.1441	−0.0589
‘D/Dr12’	−0.1593	0.5060	−0.3829	−0.8206	−0.0775	−0.0208
‘TIE’	0.9997	−0.6681	−0.4047	0.0074	−0.0992	−0.0409
‘nHDon’	−0.2192	0.3230	0.1748	0.2760	−0.7352	0.2896
‘H-050’	−0.1423	0.0898	0.1997	0.1399	−0.8156	0.2391

receptor are able to interact with dopaminergic ligands. This might be actually explaining the more activity of compound 31 with an extra OH at para position of another phenyl ring with respect to compound 32. To propose new structures based on the QSAR model the compound 31 was selected as the most potent scaffold. Compounds with two OH groups were designed and resulted in an equipotent compound compared to 31 with K_i value of 4.2 nM. Based on the obtained results, the proposed model of QSAR can be therefore a leading strategy in future studies of design for this target.

Acknowledgments This manuscript was a report of research project (1579) performed at Shahid Sadoughi university of Medical Sciences, Yazd, Iran. The authors would like to acknowledge council of research at Shiraz University of Medical Sciences for support. The authors would like to thank Nazanin Bagherzadeh for her kind contribution in language editing of the manuscript.

References

- Afiuni-Zadeh S, Azimi G (2010) A QSAR study for modeling of 8-azaadenine analogues proposed as A1 adenosine receptor antagonists using genetic algorithm coupling adaptive neuro-fuzzy inference system (ANFIS) Analytical sciences: the international journal of the Japan Society for. *Anal Chem* 26:897–902
- Ankishan H, Yilmaz D (2013) Comparison of SVM and ANFIS for snore related sounds classification by using the largest Lyapunov exponent and entropy. *Comput Math Methods Med* 2013:238937
- Balabin RM, Lomakina EI (2011) Support vector machine regression (LS-SVM)—an alternative to artificial neural networks (ANNs) for the analysis of quantum chemistry data? *Phys Chem Chem Phys* 13:11710–11718
- Bento AP, Gaulton A, Hersey A, Bellis LJ, Chambers J, Davies M, Kruger FA, Light Y, Mak L, McGlinchey S, Nowotka M, Papadatos G, Santos R, Overington JP (2014) The ChEMBL bioactivity database: an update. *Nucleic Acids Res* 42:D1083–D1090
- Buyukbingol E, Sisman A, Akyildiz M, Alparslan FN, Adejare A (2007) Adaptive neuro-fuzzy inference system (ANFIS): a new approach to predictive modeling in QSAR applications: a study of neuro-fuzzy modeling of PCP-based NMDA receptor antagonists. *Bioorg Med Chem* 15:4265–4282
- Chirico N, Gramatica P (2012) Real external predictivity of QSAR models. Part 2. New intercomparable thresholds for different validation criteria and the need for scatter plot inspection. *J Chem Inf Model* 52:2044–2058
- Dastorani MT, Moghadamnia A, Piri J, Rico-Ramirez M (2010) Application of ANN and ANFIS models for reconstructing missing flow data. *Environ Monit Assess* 166:421–434
- Dzirasa K, Ribeiro S, Costa R, Santos LM, Lin SC, Grosmark A, Sotnikova TD, Gainetdinov RR, Caron MG, Nicolelis MA (2006) Dopaminergic control of sleep-wake states. *J Neurosci* 26:10577–10589
- Fatemi MH, Dorostkar F (2010) QSAR prediction of D2 receptor antagonistic activity of 6-methoxy benzamides. *Eur J Med Chem* 45:4856–4862
- Flisikowski K, Schwarzenbacher H, Wysocki M, Weigend S, Preisinger R, Kjaer JB, Fries R (2009) Variation in neighbouring genes of the dopaminergic and serotonergic systems affects feather pecking behaviour of laying hens. *Anim Genet* 40:192–199
- Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, Light Y, McGlinchey S, Michalovich D, Al-Lazikani B, Overington JP (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res* 40:D1100–D1107
- Goodarzi M, Freitas MP (2010) MIA–QSAR coupled to PCA–ANFIS for the modeling of the anti-HIV reverse transcriptase activities of TIBO derivatives. *Eur J Med Chem* 45:1352–1358
- Kiralj R, Ferreira M (2009) Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. *J Braz Chem Soc* 20:770–787
- Levey AI, Hersch SM, Rye DB, Sunahara RK, Niznik HB, Kitt CA, Price DL, Maggio R, Brann MR, Ciliax BJ (1993) Localization of D1 and D2 dopamine receptors in brain with subtype-specific antibodies. *Proc Natl Acad Sci USA* 90:8861–8865
- Li H, Wang JX, Xing ZN, Shen G (2011) Influence of improved Kennard/Stone algorithm on the calibration transfer in near-infrared spectroscopy. *Guang pu xue yu guang pu fen xi=Guang pu* 31:362–365
- Mizuta K, Zhang Y, Xu D, Mizuta F, D’Ovidio F, Masaki E, Emala CW (2013) The dopamine D1 receptor is expressed and facilitates relaxation in airway smooth muscle. *Respir Res* 14:89
- Narendran R, Mason NS, Paris J, Himes ML, Douaihy AB, Frankle WG (2014) Decreased prefrontal cortical dopamine transmission in alcoholism. *Am J Psychiatry* 171:881–888
- O’Boyle NM, Banck M, James CA, Morley C, Vandermeersch T, Hutchison GR (2011) Open babel: an open chemical toolbox. *J Cheminform* 3:33
- Plouffe B, Tiberi M (2013) Functional analysis of human D1 and D5 dopaminergic G protein-coupled receptors: lessons from mutagenesis of a conserved serine residue in the cytosolic end of transmembrane region 6. *Methods Mol Biol* 964:141–180
- Pontiki EA, Hadjipavlou-Litina DJ, Demertzis AM, Hadjidakis I, Kovala-Demertzi D (2005) Quantitative structure activity relationships (QSAR) of substituted (S)-phenylpiperidines as preferential dopamine autoreceptor antagonists. *J Enzyme Inhib Med Chem* 20:5–12
- QSAR_Tools (2011) http://teqip.jdvu.ac.in/QSAR_Tools/

- Sakhteman A, Lahtela-Kakkonen M, Poso A (2011) Studying the catechol binding cavity in comparative models of human dopamine D2 receptor. *J Mol Graph Model* 29:685–692
- Spuhler IA, Hauri A (2013) Decoding the dopamine signal in macaque prefrontal cortex: a simulation study using the Cx3Dp simulator. *PLoS One* 8:e71615
- Sugeno M, Taniguchi T (2004) On improvement of stability conditions for continuous Mamdani-like fuzzy systems. *IEEE Trans Syst Man Cybernet B* 34:120–131
- Suykens JA, Vandewalle J, De Moor B (2001) Optimal control by least squares support vector machines. *Neural Netw* 14:23–35
- Tetko IV, Sushko I, Pandey AK, Zhu H, Tropsha A, Papa E, Oberg T, Todeschini R, Fourches D, Varnek A (2008) Critical assessment of QSAR models of environmental toxicity against *Tetrahymena pyriformis*: focusing on applicability domain and overfitting by variable selection. *J Chem Inf Model* 48:1733–1746
- Tinsley RB, Bye CR, Parish CL, Tziotis-Vais A, George S, Culvenor JG, Li QX, Masters CL, Finkelstein DI, Horne MK (2009) Dopamine D2 receptor knockout mice develop features of Parkinson disease. *Ann Neurol* 66:472–484
- Yeagle PL, Albert AD (2007) G-protein coupled receptor structure. *Biochim Biophys Acta* 1768:808–824
- Zhang Y, Chai T, Wang H (2011) A nonlinear control method based on ANFIS and multiple models for a class of SISO nonlinear systems and its application. *IEEE Trans Neural Netwo* 22:1783–1795