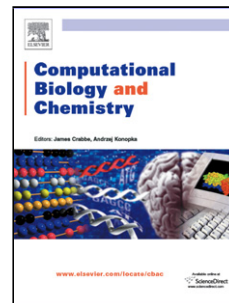


Accepted Manuscript

Title: A comparative QSAR analysis and molecular docking studies of phenyl piperidine derivatives as potent dual NK₁R antagonists/serotonin transporter (SERT) inhibitors

Author: Somayeh Zare Masood Fereidoonnezhad Davoud Afshar Zahra Ramezani



PII: S1476-9271(16)30431-5
DOI: <http://dx.doi.org/doi:10.1016/j.compbiolchem.2016.12.004>
Reference: CBAC 6618

To appear in: *Computational Biology and Chemistry*

Received date: 1-9-2016
Revised date: 1-11-2016
Accepted date: 15-12-2016

Please cite this article as: Zare, Somayeh, Fereidoonnezhad, Masood, Afshar, Davoud, Ramezani, Zahra, A comparative QSAR analysis and molecular docking studies of phenyl piperidine derivatives as potent dual NK₁R antagonists/serotonin transporter (SERT) inhibitors. *Computational Biology and Chemistry* <http://dx.doi.org/10.1016/j.compbiolchem.2016.12.004>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

**A comparative QSAR analysis and molecular docking studies of
phenyl piperidine derivatives as potent dual NK₁R
antagonists/serotonin transporter (SERT) inhibitors**

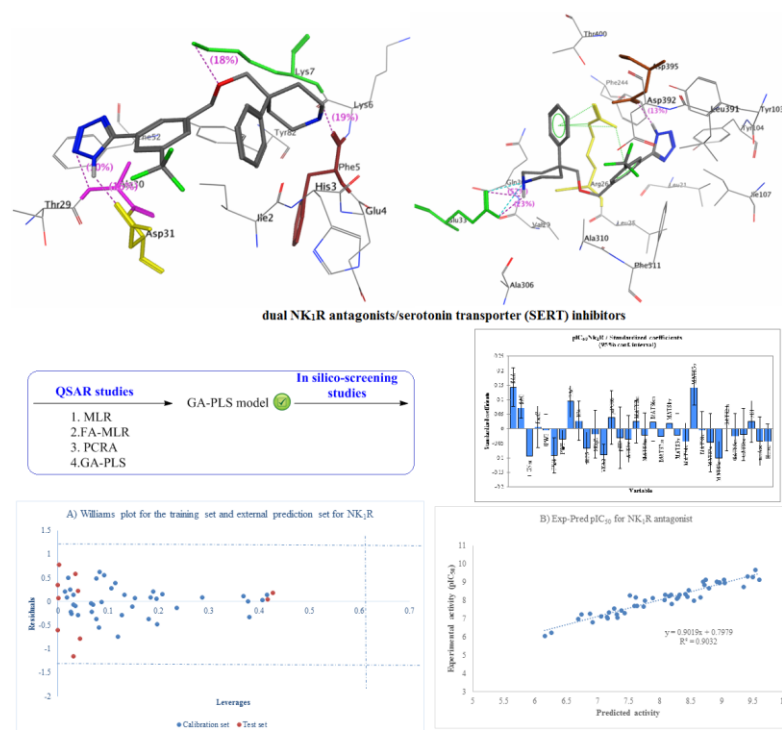
Somayeh Zare¹, Masood Fereidoonnezhad^{1,2*}, Davoud Afshar¹, Zahra Ramezani^{1,2}

¹ Department of Medicinal Chemistry, School of Pharmacy, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran.

² Cancer, Environmental and Petroleum Pollutants Research Center, Ahvaz Jundishapur University of Medical Sciences, Ahvaz, Iran.

*Corresponding author. Tel.: +98 61 32425305; Fax: +98 61 32424126; Email: Fereidoonnezhad-m@ajums.ac.ir

Graphical abstract



Highlights

- A collection of chemometrics methods such as MLR, FA-MLR, PCR, GA-PLS to make relations between structural characteristics and NK₁R antagonism/SERT inhibitory of novel phenyl piperidine derivatives
- an *in silico*-screening study to introduce new potent lead compounds based on new structural patterns
- Molecular docking studies of these compounds on both NK₁R /serotonin transporter (SERT) targets.
- validated docking protocols on both targets

Abstract

Depression is a critical mood disorder that affects millions of patients. Available therapeutic antidepressant agents are associated with several undesirable side effects. Recently, it has been shown that Neurokinin 1 receptor (NK₁R) antagonists can potentiate the antidepressant effects of serotonin-selective reuptake inhibitors (SSRIs). In this study, a series of phenyl piperidine derivatives as potent dual NK₁R antagonists/serotonin transporter (SERT) inhibitors were applied to quantitative structure–activity relationship (QSAR) analysis. A collection of chemometrics methods such as multiple linear regression (MLR), factor analysis–based multiple linear regression (FA-MLR), principal component regression (PCR), and partial least squared combined with genetic algorithm for variable selection (GA-PLS) were applied to make relations between structural characteristics and NK₁R antagonism/SERT inhibitory of these compounds. The best multiple linear regression equation was obtained from GA-PLS and MLR for NK₁R and SERT, respectively. Based on the resulted model, an *in silico*-screening study was also conducted and new potent lead compounds based on new structural patterns were designed for both targets. Molecular docking studies of these compounds on both targets were also conducted and encouraging results were acquired. There was a good correlation between QSAR and docking results. The results obtained from validated docking studies indicate that the important amino acids inside the active site of the cavity that are responsible for essential interactions are Glu33, Asp395 and Arg26 for SERT and Ala30, Lys7, Asp31, Phe5 and Tyr82 for NK₁R receptors.

Keywords: QSAR, Molecular Docking, Neurokinin 1 receptor antagonists, serotonin transporter (SERT) inhibitors, *in silico*-screening

Introduction

Major depressive disorder (MDD), is a mood disorder characterized by an extending and tenacious low mood that is accompanied by feeling of sadness and loss of interest. Over 90% of people who die by suicide have MDD or another diagnosable mental disorder. Lifetime prevalence varies widely, from 3% in Japan to 17% in the United States (1, 2). Older treatment modalities focus on using drugs that potentiate the activity of monoamines, particularly serotonin (5-hydroxytryptamine; 5-HT) and norepinephrine (3).

Serotonin transporter (SERT) is the target for several antidepressant drugs including the selective serotonin reuptake inhibitors (SSRIs). (4, 5). However, due to the fact that transporter defects are associated with a large number of psychiatric diseases, identification of novel and improved SERT modulators are urgently needed. The available SSRIs have multiple side-effects, such as anxiety, impotency, weight gain, sexual dysfunction and sleep disorders, as well as a delayed onset of action (6). Therefore, therapeutic agents with more effective and safer profile are needed.

The tachykinin receptor 1 (TACR₁) also known as neurokinin 1 receptor (NK₁R) or substance P receptor (SPR) is a G protein coupled receptor found in the central nervous system (CNS) and peripheral nervous system (PNS). Substance P is the endogenous ligand for this receptor that as a neurotransmitter released in response to acute stressors in stress-sensitive brain regions (7). NK₁ antagonists are a novel class of medications that possesses unique antidepressant properties. They alone may not be sufficient in treating depression in humans. The first NK₁ receptor antagonists were developed in the 1980s. Kramer et al. reported that NK₁R antagonists acts as antidepressant in clinical trials in 1998 (8).

Other forms of treatment that may be a useful target in the development of medications for the treatment of depression is a potential combination strategy that involve SSRIs and NK₁R antagonists. Degnan et al. showed that development of a dual NK₁R/SERT antagonists appear to have great antidepressant effects. They undertook that high NK₁ receptor occupancy reduce the potential SSRI side effects (9). So it is conceivable that development of the dual NK₁R/SERT is useful strategy to progress the medications for depressing patients (10).

In the present paper, two different drug design methodologies namely, QSAR and molecular docking simulations have been applied for a series of phenyl piperidine derivatives with the ability to antagonist NK₁R and inhibit serotonin transporter. In a comprehensive study of this

system we used a very large descriptor set including topological, geometrical, constitutional, functional group, 2D autocorrelation, atom-centered fragments, electrostatic, quantum and chemical to describe the physicochemical properties of the molecules. Different statistical methods were applied to model the relationship between the structural features and the serotonin transporter inhibitory and NK₁R antagonist activity of the studied compounds. These methods were: (i) multiple linear regression (MLR) (ii) factor analysis–based multiple linear regression (FA-MLR), (iii) principal component regression (PCR) and (iv) partial least squared combined with genetic algorithm for variable selection (GA-PLS). Validated molecular docking simulation technique was also performed on all compounds of dataset as well as the designed compounds to reach the detailed molecular binding models for these compounds interacting with the key active site amino acids of targets. This usually helps the medicinal chemist to further understand the structure-activity relationships of the studied molecules.

Materials and methods

Data set

A data set consisting of 49 phenyl piperidine derivatives as a series of potent dual NK₁R antagonists/ SERT inhibitors were selected for the current study (9). The structural features and biological activity details of these compounds are listed in Table 1. The biological data reported as IC₅₀ values and converted to the pIC₅₀ and finally used for the QSAR modeling studies.

Molecular descriptors

Two dimensional structures of the ligands were constructed using ChemBioDraw 12.0 software (11). The ligands were subjected to minimization procedures by means of an *in house* TCL script using Hyperchem (Version 8, Hypercube Inc., Gainesville, FL, USA). Each ligand was optimized with different minimization methods such as molecular mechanics (MM+) followed by quantum based semi-empirical method (AM1) using Hyperchem package. Z-matrices of the structures were provided by the software and transferred to the Gaussian 98 program (12). Large number of molecular descriptors was calculated using Hyperchem, Gaussian 98 and Dragon package (13). Some chemical parameters including molecular volume (V), molecular surface area (SA), hydrophobicity (LogP), hydration energy (HE) and molecular polarizability (MP) were calculated using Hyperchem Software. Highest occupied molecular orbital (HOMO) and

lowest unoccupied molecular orbital (LUMO) energies, the most positive and the most negative net atomic charges, the average absolute atomic charge and molecular dipole moment were calculated by Gaussian98 software. Quantum chemical indices of hardness ($\eta = 0.5$ (HOMO+LUMO)); softness ($S = 1/\eta$); electronegativity ($\chi = -0.5$ (HOMO-LUMO)); and electrophilicity ($\omega = \chi^2/2\eta$) were calculated according to the equations proposed by Thanikaivelan et al. (14). Dragon calculated different topological, geometrical, charge, empirical and constitutional descriptors for each molecule. 2D autocorrelations, aromaticity indices, atom-centered fragments and functional groups was also calculated by dragon. The brief description of some of them is listed in Table 2.

Variable importance in the projection (VIP)

Variable important in projection (VIP) was employed in order to survey the relative importance of the variable appeared in the final model obtained by GA-PLS method (15). The importance of terms in PLS model is reflected in VIP values. As it was explained by Erikson et al., X-variables (predictor variables) could be classified pursuant to their relevance in explaining y (predicted variable), so that $VIP > 1.0$ and $VIP < 0.8$ mean highly or less influential, respectively, and $0.8 < VIP < 1.0$ means moderately influential (16).

Docking procedure

The docking studies were carried out by means of an *in house* batch script (DOCKFACE) (17, 18) of AutoDock 4.2. For docking procedure, each ligand was optimized with MM⁺ then AM1 minimization method using HyperChem 8. Then the partial charges of atoms were calculated using Gasteiger-Marsili procedure implemented in the AutoDock Tools package (19). Non-polar hydrogens of compounds were merged and then rotatable bonds were assigned. The output structures were converted to PDBQT using MGLtools 1.5.6 (20).

The three dimensional crystal structure of SERT (PDB ID: 3GWW) and NK₁R (PDB ID: 1NK1) were retrieved from protein data bank (<http://www.rcsb.org/pdb/home/home.do>). All water molecules were removed, missing hydrogens were added and after determining the Kollman united atom charges non-polar hydrogens were merged into their corresponding carbons using

AutoDock Tools (21). As the final part of this process, desolvation parameters were assigned to each protein atom. Among the three different search algorithms performed by AutoDock 4.2 the commonly used Lamarckian Genetic Algorithm (LGA) was applied (22, 23). Subsequently, the enzymes were converted to PDBQT using MGLTOOLS 1.5.6.

For Lamarckian GA, a maximum number of 2,500,000 energy evaluations, 27000 maximum generations; 150 population size, a gene mutation rate of 0.02; and a crossover rate of 0.8 were applied. The grid maps of the receptors were calculated using AutoGrid tools of AutoDock 4.2. The size of grid was set in a way to include not only the active site but also considerable portions of the encircling surface. A grid box of 55×59×69 and 45×45×45 points in x, y, and z directions was built and centered on the center of the ligand in the complex with a spacing of 0.375 Å for 1NK₁R and 3GWW, respectively. Number of points for 1NK₁R in x, y and z was 20.055, 18.173 and 25.706, and for 3GWW was 25.388, 20.809 and 22.282, consequently. AutoDock Tools was employed to produce both grid and docking parameter files i.e. gpf and dpf.

Cluster analysis was performed on the docked results using a root mean square deviation (RMSD) tolerance of 2 Å. For the internal validation phase, co-crystal ligand (fluoxetine) inside the pdb file of SERT (3GWW) was extracted using a viewer and treated the same as other ligands. All the docking protocols were done on validated structures with RMSD values below 2 Å.

Ligand-receptor interactions were all detected on the basis of docking results using Autodock tools program (ADT, Version 1.5.6), VMD software (24) and PLIP (fully automated protein–ligand interaction profiler) (25). These software visualize hydrogen bonding, π - π stacking and π -cationic as well as hydrophobic interactions which are established through the docking procedure.

Model development

The calculated descriptors were collected in a data matrix whose number of rows and columns were the number of molecules and descriptors, respectively. Four different regression methods were applied for QSAR equations including: simple multiple linear regression with stepwise

variable selection (MLR), Factor analysis as the data preprocessing step for variable selection (FA-MLR), Principal component regression analysis (PCRA) and Genetic algorithm–partial least squares (GA-PLS). These known methods are well applied in the QSAR studies (26).

Here, stepwise selection and elimination of variables was applied for developing QSAR models using SPSS software (version 21; SPSS Inc., IBM, Chicago, IL, USA). The resulted models were validated by leave-one-out cross-validation procedure to check their predictability and robustness using MATLAB 2015 software (version 8.5; Math work Inc., Natick, MA, USA).

FA-MLR was also performed on the dataset. Factor analysis (FA) was used to reduce the number of variables and to detect structure in the relationships between them. This data-processing step is applied to identify the important predictor variables and to avoid co-linearity among them (27). Principle component regression analysis, PCRA, was also tried for the dataset along with FA-MLR. Co-linearities among X variables are not involved as a disturbing factor with PCRA and the number of variables included in the analysis may exceed the number of observations (28). In this method, factor scores, as obtained from FA, are used as the predictor variables (27). In PCRA, all descriptors are assumed to be important while the aim of factor analysis is to identify relevant descriptors.

The partial least square (PLS) regression method was applied to the NIPALS-based algorithm existed in the chemometrics toolbox of MATLAB software. Leave-one-out cross-validation procedure was used to obtain the optimum number of factors based on the Haaland and Thomas F-ratio criterion (29, 30). The MATLAB PLS toolbox developed by eigenvector company was used for PLS and GA modeling. All calculations were run on a core i7 personal computer (CPU at 6 MB) with Windows 7 operating system.

Model validation

Statistical parameters such as standard error of regression (SE), correlation coefficient (R^2), variance ratio (F) at specified degrees of freedom, leave-one-out cross-validation correlation coefficient (Q^2), root mean square error of cross-validation (RMScv) and double cross validation (Cvcv) were employed for validity of regression equation. In order to test the developed model performances, 20 % of the molecules were selected as test set molecules. The predictive value of

a QSAR model that has not been taken into account during the process of developing the model should be tested on an external set of data. QSAR model was developed in more than three data sets and the best equation were selected as the best model.

Applicability domain

One of the great uses of a QSAR model is based on its precise prediction ability for new compounds. It is important to emphasize that no matter how valid, significant and validated a QSAR may be, it cannot be expected to reliably predict the modeled property for the entire space of chemicals. Therefore, before a QSAR is put into use for screening chemicals, its domain of application must be defined and predictions for only those chemicals that fall in this domain may be considered reliable. The applicability domain is appraised by the leverage values for each compound. A Williams plot (the plot of standardized residuals versus leverage values (h)) can then be used for an immediate and simple graphical detection of both the response outliers (Y outliers) and structurally influential chemicals (X outliers) in our model. In this graph, the applicability domain is established inside a squared area within $\pm x$ (standard deviations) and a leverage threshold h^* .

The numerical value of leverage has certain characteristic: (a) the value is always greater than zero, (b) the lower the value; the higher is the confidence in the prediction. A value of 1 indicates very poor prediction. A value of 0 indicates perfect prediction and will not be achieved. Another factor for analysis of the results is warning leverage (h^*). The threshold h^* is generally fixed at $3(k + 1) / n$ (k is the number of model parameters and n is the number of training set compounds), whereas $x = 2$ or 3 . Prediction must be considered unreliable for compounds with a high leverage value ($h > h^*$). A leverage greater than warning leverage h^* means that the predicted response is the result of substantial extrapolation of the model and therefore may not be reliable. From the other point of view, when the leverage value of a compound is lower than the threshold value, the probability of agreement between observed and predicted values is as high as that for the training set compounds (31, 32).

Results and Discussion

Here, we developed a detailed QSAR study using a combination of chemical, electronic and substituent constant, to explore structural parameters affecting phenyl piperidine derivatives as potent dual NK₁R antagonists/SERT inhibitors. Among the different chemometrics tools available for modeling the relationship between the biological activity and molecular descriptors, four methods (i.e. stepwise MLR, FA-MLR, PCRA, and GA-PLS) were applied.

MLR modeling

In the first step, separate stepwise selection-based MLR analyses were performed using different types of descriptors, and then, an MLR equation was obtained utilizing the pool of all calculated descriptors. The results for SERT and NK₁R are summarized in Table 3 and Table S1 (available in supplementary file) respectively. Correlation coefficient (r^2) matrix for the descriptors used in different MLR equations is shown in Table 4. Collinear descriptors degrade the performance of MLR equations and such models have lowered prediction ability. The correlation coefficient (r^2) matrix for the descriptors used in MLR equation 1, shows that no significant correlation exists between pairs of descriptors (Table 4). The same results were obtained for NK₁R (Table S2).

The statistical parameters calculated for each target such as R^2 , correlation coefficient (R^2_p) of test set, SE, F at specified degrees of freedom, Q^2 , Cvcv and RMScv were used for validating the goodness of fit of the resulted QSAR equations are represented in both Table 3 and Table S1. Equation 1 (in Table 3) and 1a (in Table S1) was selected as the best equation in the MLR model in both tables. The selected variables in table 3 demonstrate that Chemical (SAA, SAG), topological (TIC2), 2D autocorrelations (MATS2e, MATS4V, MATS2V) descriptors and in Table S1, constitutional (RBN), functional (nCIC, nHDon, nNHR) and topological (x3sol, D/Dr10) descriptors affect the NK₁R antagonist and SERT inhibitory activity of the studied compounds, respectively.

FA-MLR and PCRA

Table 5 shows the five factor loadings of the variables (after VARIMAX rotation) for the compounds based on their SERT inhibitory (factor 1, 2, 3, 5 and 12). As it is observed in the table, about 72% of variances in the original data matrix could be explained by selected five factors. As it was shown in Table S3, factors 1, 2, 5, 6, 7 and 15 explain about 77% of variances in the original data matrix for NK₁R antagonist activity.

Table 4 revealed that, descriptors such as nHAcc, G(F..F), MATS4m, MATS6m, GATS2m, GATS2m and ATS7e are the highest loading values for factor 1. The highest loading values for factor 2 are associated with Ref, VAR, IVDM, PCWTe, VEA2, X2V, nNHR and DipY whereas ATS1V, ATS3V and ATS6V are the highly loaded descriptors of factor 5. Table 5 revealed that, factors 1 and 2 are moderately loaded with SERT inhibitory activity. Interestingly, the former possessed the highest loadings from functional (nHAcc), geometrical (G(F..F)) and 2D autocorrelations (MATS4m, MATS6m, GATS2m, GATS2m and ATS7e) whereas the latter is containing the information from chemical (Ref), topological (X2V, IVDM, VAR and VEA2), charge (PCWTe), functional (nNHR) and quantum (ASP) descriptors. The subsequent FA-MLR equation using highly loaded descriptors is shown in Table 2, Eq.2.

As it was shown in Table S3, the highest loading values for factor 2 are associated with SAA, SAG, MW, X2V, ISIZ, IVDM, TIC2, logp, H3D and piPC06 whereas X5Av, SEigZ, nHAcc, MATS3e, GATS4m, GATS2m, G(F..F) and ATS7e are the highly loaded descriptors of factor 1. Table 5 revealed that, factors 1 and 2 are moderately loaded with NK₁R antagonist activity. Interestingly, the former possessed the highest loadings from chemical (MW, logp, SAA and SAG), topological (X2V, ISIZ, IVDM, piPC06 and TIC2), geometrical (H3D) descriptors whereas the latter is containing the information from functional (nHAcc), geometrical (G(F..F)) and 2D autocorrelations (MATS3e, GATS4m, GATS2m and ATS7e), topological (X5Av, SEigZ) descriptors. The subsequent FA-MLR equation using highly loaded descriptors is shown in Table S1, Eq.2a.

PCRA

When factor scores were used as the predictor parameters in a multiple regression equation using forward selection method (PCRA), equation 3 was obtained. Factor scores, instead of selected descriptors, contain information for the different descriptors, so the chance of loss of information is decreased. By principle component method, five (Table 5) and six (Table S3) factors scores

were used as independent parameters for developing QSAR equations. These factor scores were used as independent parameters for developing QSAR equations. The variables used in Eq. 3 shows statistical quantities similar to those obtained by FA-MLR method. But, it indicates partly higher calibration and lower cross-validation statistics with respect to Eq.2.

In Table 5, Factor score 1 signifies the importance of nHAcc, G(F..F), MATS4m, MATS6m, GATS2m, GATS2m and ATS7e descriptors. Factor score 2 indicates the importance of Ref, VAR, IVDM, PCWTe, VEA2, X2V, nNHR and DipY descriptors. Factor score 3 indicates the importance of ATS6v, ATS3v and ATS1v. Factor score 5 indicates the importance of nCIC, PiID, D/Dr10 and Jhept descriptors. Factor score 12 signifies the importance of RNCG, PCWTe and qneg descriptors.

In Table S3, Factor score 1 indicates the importance of X5Av, SEigZ, nHAcc, MATS3e, GATS4m, GATS2m, G(F..F) and ATS7e. Factor score 2 signifies the importance of Ref, X2V and ISIZ descriptors. Factor score 7 indicates the importance of Jhetz, Jhetp and x1A descriptors. The factor score 6 reveals the importance of GATS7p and ATS8v descriptors. Factor score 5 indicates the importance of SAG and ASP Factor score 15 signifies the importance of, GATS5p, MATS5v, MATS5m and PJI2 descriptors.

GA-PLS

In PLS analysis, the descriptors data matrix is decomposed to orthogonal matrices with an inner relationship between the dependent and independent variables. Therefore, unlike MLR analysis, the multicollinearity problem in the descriptors is omitted by PLS analysis. Since a minimal number of latent variables are used for modeling in PLS; this modeling method coincides with noisy data better than MLR. So, many different GA-PLS runs were done using different initial set of populations. The statistical parameters calculated for this model are represented in Table 3 for SERT inhibitory and in Table S1 for NK₁R antagonist activity.

Table 3 shows that a combination of 2D autocorrelations (ATS4p, MATS6m and MATS4m), topological (STN), geometrical (J3D) and charge (RNCG) descriptors have been selected by GA-PLS to account for the SERT inhibitory activity. As it is shown in Table S1, Eq. 4a, a combination of 2D autocorrelations (MATS5v, MATS7m, MATS3m and MATS3e), geometrical (SEigZ), and topological (piPC06 and PiID) descriptors have been selected by GA-PLS to account for the NK₁R antagonist activity. In this table, Eq. 4a with high statistical quality

parameters was obtained from the pool of calculated descriptors (i.e., $R^2 = 0.89$ and $Q^2 = 0.84$) and, the predictive R^2 value for the test set was found to be 0.75. Table S1 shows that none of the suggested QSAR models were obtained by chance and the best set of calculated descriptors was selected by genetic algorithm because of its greatest statistical parameters. Therefore the best predictive results were observed.

The most convenient GA-PLS model that resulted in the best fitness contained 32 indices for NK₁R and 40 indices for SERT. The PLS estimate of coefficients for the descriptors of NK₁R are given in Figure 1. As it observed, a combination of chemical, topological, geometrical, quantum, 2D-autocorrelations and functional descriptors have been selected by GA-PLS to account the NK₁R antagonist activity of phenyl piperidine derivatives. To measure the significance of the 32 selected PLS descriptors in the NK₁R antagonist activity; VIP was calculated for each descriptor. The VIP analysis of PLS equation is shown in Figure 2. VIP shows that SAA, SAG, MATS5v, MATS7e, BAC, GNar, PW4, VEA2 and SIC5 are the most important indices in the QSAR equation derived by PLS analysis. In addition, $VIP < 0.8$ mean less influential parameters. The PLS estimate of coefficients and the VIP analysis for the descriptors of SERT are given in figure S1 and figure S2, respectively. VIP shows that nNHR, RBN, STN, piID, ATS5v, MATS4v, GATS4p, RPCG, DipY, MATS5v, MATS1v, FDI, SAA and RNCG are the most important indices in the QSAR equation derived by PLS analysis for SERT inhibitory. GATS6m and DECC have been found to be moderately influential parameters.

By considering the GA-PLS equation on these targets (Table 3 and Table S1), It is obvious that the dominant descriptors in both equations is 2D-autocorrelation descriptors. These descriptors depict the topological structure of the compounds, but are more complicated in nature with respect to the classical topological descriptors. The calculation of these descriptors includes the summations of different autocorrelation functions related to different structural lags and leads to different autocorrelation vectors corresponding to the lengths of the substructural fragments. As a result, these descriptors address the topology of the structure or parts thereof in association with a specific physicochemical property.

In silico screening

In silico screening studies can be used to replace expensive and time consuming *in vivo* experiments in the early stages of drug development. It is thought to have the potential to speed the rate of discovery, predicting and identifying new pharmaceutical compounds. Virtual screening is a forceful technique to identify potentially active compounds from molecular databases and it was applied by deletion, insertion and substitution of different substitutes on the parent molecules and the effects of the structural modifications on the biological activity were investigated.

The domain application of QSAR model was determined to use the model for screening new compounds. The applicability domain (AD) of QSAR model was applied to verify the prediction reliability, to recognize the troublesome compounds and to predict the compounds with acceptable activity that falls within this domain. We employed the important descriptors selected in both GA-PLS (SERT inhibitors) and MLR (NK₁R antagonists) model for designing new active compounds because of their greatest statistical parameters compared to the others. Analyzing the model AD in the Williams plot of the GA-PLS model of NK₁R (figure 3A) and MLR model of SERT (figure 3B) based on the whole data set, appeared that none of the compounds were identified as an obvious outlier for these activity. As it is cleared, none of the compounds have leverage (*h*) values greater than the threshold leverages (*h**). The warning leverage (*h**), was found to be 0.61 for NK₁R and 0.54 for SERT. The compounds that had a standardized residual more than three times of the standard deviation units were considered to be outliers. For both the training set and prediction set of targets, the presented model matches the high quality parameters with good fitting power and the capability of assessing external data. Moreover, almost all of the compounds were within the applicability domain of the proposed model and were evaluated accurately. While chemicals with a leverage value higher than *h** were considered to be influential or high leverage chemicals (29, 30).

According to the developed QSAR model, the *in silico* screening was used to design new compounds with improved potential NK₁R/SERT inhibitor activity. Then, the *in silico* screen was applied by substituting diverse groups in different places. The results of this investigation are summarized in Table S4. As it was shown in Table S4, 30 novel compounds were designed and their predicted activities for SERT inhibitory and NK₁R antagonist activity based on MLR and GA-PLS equations, respectively as well as their docking binding energies were obtained. Leverage values show that all of the designed compounds were within the applicability domain.

Among these molecules, the compound **1a**, **6a**, **8a**, **10a**, **1b**, **6b**, **9b**, **14b** and **17b** showed the best activity. These compounds have a good potential for becoming antidepressant agent.

Finally, this result confirms the reliability of the models and it shows that it is possible to identify new synthetic compounds for drug discovery, with the aim of the *in silico* screening QSAR studies.

To have a consideration on the cross-validated prediction results, the predicted activity data are plotted against the experimental activities in figure 4. As it was mentioned above, the least scattering of data was obtained from GA-PLS and MLR for NK₁R and SERT, respectively. High regression ratio ($R^2 = 0.89, 0.90$ for SERT and NK₁R respectively) in this plot shows the great agreement between cross-validated predicted values of activity and the experimental activity.

Docking Studies

In the present study, molecular docking simulations were performed on 49 compounds of dataset as well as 30 designed compounds, to elucidate the interactions between SERT and NK₁R targets and their inhibitors further and to gain some insight into their molecular binding mode. The results obtained from this part of study including the estimated free binding energy values (ΔG_{bind}) for the best position of the docked compounds, expressed in kcal/mol⁻¹, along with the corresponding favorable interactions with the key amino acid residues at the active site of enzyme are summarized in Table 1, S1 and figures 5-11.

In the validity evaluation step of docking process, redocking of fluoxetine, which is the natural substrate of SERT with a high affinity for this enzyme, indicated that the X-ray crystallography conformer was extremely identical to the docked conformer. RMSD of docking for Fluoxetine in comparison with its coordination in the crystal structure was 0.043. Figure 5A displays that the best docked pose of the fluoxetine (green) within the enzyme cavity is in accordance with that in the crystal structure (yellow) which points toward the reliability of the molecular docking procedure. Figure 5B shows that amine side chain group of fluoxetine is involved in hydrogen bonds with Asp395. There also exist donor hydrogen bonds between hydrogen of piperidine ring with Asp31. The two phenyl groups are also involved in arene-cation interaction with Arg26.

As it was shown in Table 1, the studied compounds can be classified to five different classes based on their structures. The structural changes in these compounds have affected the docking results. The

ΔG_{bind} values of the best docked poses of class 1 (compounds **1-6**) are within the range of -9.07 to -9.93 kcal.mol⁻¹ for SERT and -5.87 to -6.74 kcal.mol⁻¹ for NK₁R. The best docking binding energies of class 2 (compounds **7-13**) are within the range of -8.81 to -10.55 kcal.mol⁻¹ for SERT and -6.29 to -7.71 kcal.mol⁻¹ for NK₁R. Class 3 (compounds **14-36**) docking binding energies are within the range of -9.16 to -10.93 kcal.mol⁻¹ for SERT and -6.12 to -8.84 kcal.mol⁻¹ for NK₁R. The best docked poses of class 4 (compounds **37-44**) are within the range of -9.67 to -10.52 kcal.mol⁻¹ for SERT and -7.36 to -8.96 kcal.mol⁻¹ for NK₁R. ΔG_{bind} values at the range of -7.72 to -10.01 kcal.mol⁻¹ for SERT and -7.72 to -8.08 kcal.mol⁻¹ is seen for class 5 (compounds **45-49**). Based on these results, it can be concluded that class 4 has the best docking binding energies especially on NK₁R. It should be mentioned that all of the studied compounds has the higher docking binding energy in compared to fluoxetine (the co-crystal ligand of SERT) with ΔG_{bind} values of -6.73 kcal.mol⁻¹.

As indicated in figure 6, the hydrogen atom attached to nitrogen of piperidine and tetrazole ring, is involved in hydrogen bond interactions with residues Glu33 and Asp395, respectively. The phenyl ring attached to piperidine and tetrazole ring, is involved in arene-cation interactions with residues Arg26. The side chains consisting of amino acid residues Ile107, Phe244, Phe311, Val29, Ala310 and Asp392 make direct van der Waals contacts with compound **4**.

In NK₁R binding mode (figure 7), Compound **4** interacts via acceptor hydrogen bonds through nitrogen of tetrazole with Ala30 and through oxygen group with Lys7. There also exist donor hydrogen bonds between hydrogen attached to the nitrogen of tetrazole with Asp31 and hydrogen of piperidine ring with Phe5. The amino acid residues consist of Ile2, His3, Thr29, Phe52 and Tyr82 are also involved in the hydrophobic site of active site.

Docking studies on compound **11** with SERT receptor (Figure 8) revealed that, three arene-cation interactions are existed between thiazole ring, phenyl ring attached to piperidine and phenyl ring bearing CF₃ group with Arg26. The hydrogen atom attached to nitrogen of piperidine is involved in hydrogen bond interactions with residues Glu33. The other amino acid residues in active site such as Ala310, Val29, Leu21, Phe244, Leu25, Ile107, Phe311 and Asp395 are involved in hydrophobic interactions.

Binding mode of compound **11** with NK₁R (figure 9) shows that, oxygen group is involved in two acceptor hydrogen bonds with Tyr82 and Lys7. There also exist donor hydrogen bonds

between hydrogen of piperidine ring with Asp31. The phenyl attached to piperidine ring is involved in arene-cation interaction with Lys7.

As it was shown in figure 10, the key amino acids in binding mode of compound **23** and **41** are two arene-cation interactions between two phenyl ring attached to piperidine and phenyl ring bearing CF₃ group with Arg26. There is also existed hydrogen bond interaction between hydrogen attached to nitrogen of piperidine with residues Glu33.

Compound **23** interacts via hydrogen bond through its oxygen with Tyr82. As it was depicted in figure 11B, this interaction was also seen in compound **41**. Both of these compounds interact between through hydrogen of their piperidine ring with Asp31. There is evidence that arene-cation interaction is existed between phenyl attached to piperidine ring of compound **23** with Lys7. The phenyl bearing methoxy group is involved in this interaction with Lys7 in compound **41**.

The results obtained from this docking study indicate that the important amino acids inside the active site cavity that are in charge of essential interactions are Glu33, Asp395 and Arg26 for SERT and Ala30, Lys7, Asp31, Phe5 and Tyr82. The majority of the studied compounds took part in hydrogen bond formation especially through the hydrogen attached to nitrogen of piperidine with Glu33 and Asp31 for SERT and NK₁R inhibitory activity, respectively. It should be noted that compounds of high activity form one or more hydrogen bonds with the active site residues. Therefore, one of the essential requirements for optimum NK₁R antagonist/SERT inhibition is hydrogen bond formation by ligands. The docking results also suggest that apart from hydrogen bonding formation, binding of different compounds with the active site is stabilized by van der Waals and hydrophobic interactions with the nonpolar amino acids. The arene-cation interaction is also existed between ligands and their targets through their phenyl groups.

Molecular docking were applied on the designed compounds to gain some insight into their molecular binding mode on both targets. Binding mode of **1a** and **17b** in the active site of SERT and NK₁R are available in supplementary file.

As it was shown in figure 12, the results of molecular docking simulations were highly in accordance with the experimental data of SERT inhibition and activity, respectively.

Conclusions

Quantitative relationships between molecular structure and SERT inhibitory as well as NK₁R antagonist activity of a series phenyl piperidine derivatives were discovered by a collection of chemometrics methods including MLR, GA-PLS, FA-MLR and PCRA. The reliability, accuracy and predictability of the proposed models were evaluated by various criteria, including cross-validation, the root mean square error of prediction (RMSEP), root mean square error of cross-validation (RMSECV), validation through and Y-randomization. In this series a significant role of topological, geometrical, 2D-autocorrelations, and charge descriptors on the inhibitory activity was observed. A comparison between the different statistical methods employed indicated that GA-PLS and MLR represented superior results for NK₁R and SERT, respectively. According to the developed QSAR model, *in silico* screening was applied and new compounds such as **1a**, **6a**, **8a**, **10a**, **1b**, **6b**, **9b**, **14b** and **17b** with potential inhibitory activity on both targets were suggested for synthesis. The results of molecular docking simulations on both receptors were highly in accordance with the QSAR and experimental data.

Acknowledgments

The authors would like to thank research deputy Ahvaz Jundishapur University of medical sciences who support this work. Collaboration of medicinal chemistry, faculty of pharmacy, Ahvaz Jundishapur University of Medical Sciences, in providing the required facilities for this work is greatly acknowledged.

References

1. Inaba A, Thoits PA, Ueno K, Gove WR, Evenson RJ, Sloan M. Depression in the United States and Japan: gender, marital status, and SES patterns. *Social science & medicine*. 2005;61(11):2280-92.
2. Kessler RC, Bromet EJ. The epidemiology of depression across cultures. *Annual review of public health*. 2013;34:119.
3. Quesseveur G, Nguyen HT, Gardier AM, Guiard BP. 5-HT₂ ligands in the treatment of anxiety and depression. *Expert opinion on investigational drugs*. 2012;21(11):1701-25.

4. Hyttel J. Pharmacological characterization of selective serotonin reuptake inhibitors (SSRIs). *International Clinical Psychopharmacology*. 1994.
5. Schloss P, Williams DC. The serotonin transporter: a primary target for antidepressant drugs. *Journal of Psychopharmacology*. 1998;12(2):115-21.
6. Ferguson JM. SSRI antidepressant medications: adverse effects and tolerability. Primary care companion to the *Journal of clinical psychiatry*. 2001;3(1):22.
7. Goldsmith LE, Kwatra MM. NK1 (substance P) receptor. *UCSD Molecule Pages*. 2012;1(1).
8. Kramer MS, Cutler N, Feighner J, Shrivastava R, Carman J, Sramek JJ, et al. Distinct mechanism for antidepressant activity by blockade of central substance P receptors. *Science*. 1998;281(5383):1640-5.
9. Degnan AP, Tora GO, Han Y, Rajamani R, Bertekap R, Krause R, et al. Biaryls as potent, tunable dual neurokinin 1 receptor antagonists and serotonin transporter inhibitors. *Bioorganic & medicinal chemistry letters*. 2015;25(15):3039-43.
10. Wu Y-J, He H, Bertekap R, Westphal R, Lelas S, Newton A, et al. Discovery of disubstituted piperidines and homopiperidines as potent dual NK 1 receptor antagonists—serotonin reuptake transporter inhibitors for the treatment of depression. *Bioorganic & medicinal chemistry*. 2013;21(8):2217-28.
11. Ultra C. 12.0, Cambridge Soft, Cambridge, MA 2010. Received: May. 2013;10.
12. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al. *Gaussian 09*. Wallingford, CT, USA: Gaussian, Inc.; 2009.
13. A. Mauri VC, M. Pavan, R. Todeschini: D., RAGON Software: An Easy Approach to Molecular Descriptor Calculations. *MATCH Commun Math Comput Chem*. 2006;56: 237-48.
14. Thanikaivelan P, Subramanian V, Raghava Rao J, Unni Nair B. Application of quantum chemical descriptor in quantitative structure activity and structure property relationship. *Chem Phys Lett*. 2000;323(1–2):59-70.
15. Olah M, Bologa C, Oprea TI. An automated PLS search for biologically relevant QSAR descriptors. *Journal of computer-aided molecular design*. 2004;18(7-9):437-49.
16. Bishop JM. The molecular genetics of cancer. *Science*. 1987;235(4786):305-11.
17. Fereidoonhezahad M, Faghih Z, Mojaddami A, Tabaei S, Rezaei Z. Novel Approach Synthesis, Molecular Docking and Cytotoxic Activity Evaluation of N-phenyl-2, 2-

- dichloroacetamide Derivatives as Anticancer Agents. *Journal of Sciences*, Islamic Republic of Iran. 2016;27(1):39-49.
18. Fereidoon nezhad M, Faghih Z, Mojaddami A, Sakhteman AH, Rezaei Z. A Comparative Docking Studies of Dichloroacetate Analogues on Four Isozymes of Pyruvate Dehydrogenase Kinase in Humans. *Indian J Pharm Edu Res*. 2016; 50(2s):S32-S38.
 19. Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron*. 1980;36(22):3219-28.
 20. Morris GM, Huey R, Olson AJ. Using AutoDock for Ligand-Receptor Docking. *Current Protocols in Bioinformatics*: John Wiley & Sons, Inc.; 2002.
 21. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, et al. A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society*. 1984;106(3):765-84.
 22. Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of computational chemistry*. 1998;19(14):1639-62.
 23. Hamed A, Khoshnoud MJ, Tanideh N, Abbasi F, Fereidoon nezhad M, Mehrabani D. Reproductive toxicity of *Cassia absus* seeds in female rats: possible progesteronic properties of chaksine and b-sitosterol. *Pharmaceutical Chemistry Journal*. 2015;49(4):268-74.
 24. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *Journal of molecular graphics*. 1996;14(1):33-8, 27-8.
 25. Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder M. PLIP: fully automated protein–ligand interaction profiler. *Nucleic acids research*. 2015:gkv315.
 26. Fereidoon nezhad M, Khoshneviszadeh M, Faghih Z, Jokar E, Mojaddami A, Rezaei Z. QSAR, Molecular Docking and protein ligand interaction fingerprint studies of N-phenyl dichloroacetamide derivatives as anticancer agents. *Trends in Pharmaceutical Sciences*. 2016;2(2).
 27. Franke R, Gruska A, van de Waterbeemd H. Chemometrics Methods in molecular design. *Methods and Principles in Medicinal Chemistry*. 1995;2:113-9.
 28. Bhattacharya P, Roy K. QSAR of adenosine A₃ receptor antagonist 1, 2, 4-triazolo [4, 3-a] quinoxalin-1-one derivatives using chemometric tools. *Bioorganic & medicinal chemistry letters*. 2005;15(16):3737-43.

29. Asadollahi T, Dadfarnia S, Shabani AMH, Ghasemi JB, Sarkhosh M. QSAR Models for CXCR2 Receptor Antagonists Based on the Genetic Algorithm for Data Preprocessing Prior to Application of the PLS Linear Regression Method and Design of the New Compounds Using In Silico Virtual Screening. *Molecules*. 2011;16(3):1928-55.
30. Khoshneviszadeh M, Edraki N, Miri R, Foroumadi A, Hemmateenejad B. QSAR Study of 4-Aryl-4H-Chromenes as a New Series of Apoptosis Inducers Using Different Chemometric Tools. *Chem Biol Drug Des* 2012; 79:442-458.
31. Roy K, Kar S, Das RN. Chapter 7 - Validation of QSAR Models. *Understanding the Basics of QSAR for Applications in Pharmaceutical Sciences and Risk Assessment*. Boston: Academic Press; 2015. p. 231-89.
32. Weaver S, Gleeson MP. The importance of the domain of applicability in QSAR modeling. *J Mol Graphics Modell*. 2008;26(8):1315-26.

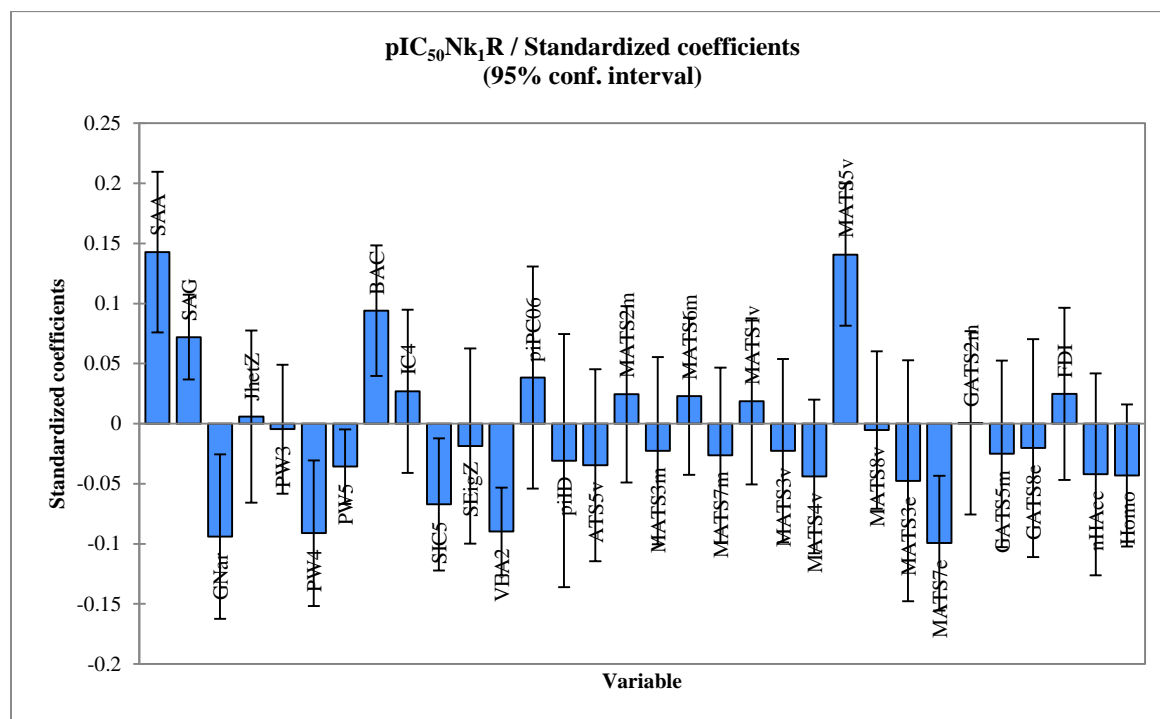


Figure 1. PLS regression coefficients for the variables used in GA-PLS model for NK_1R

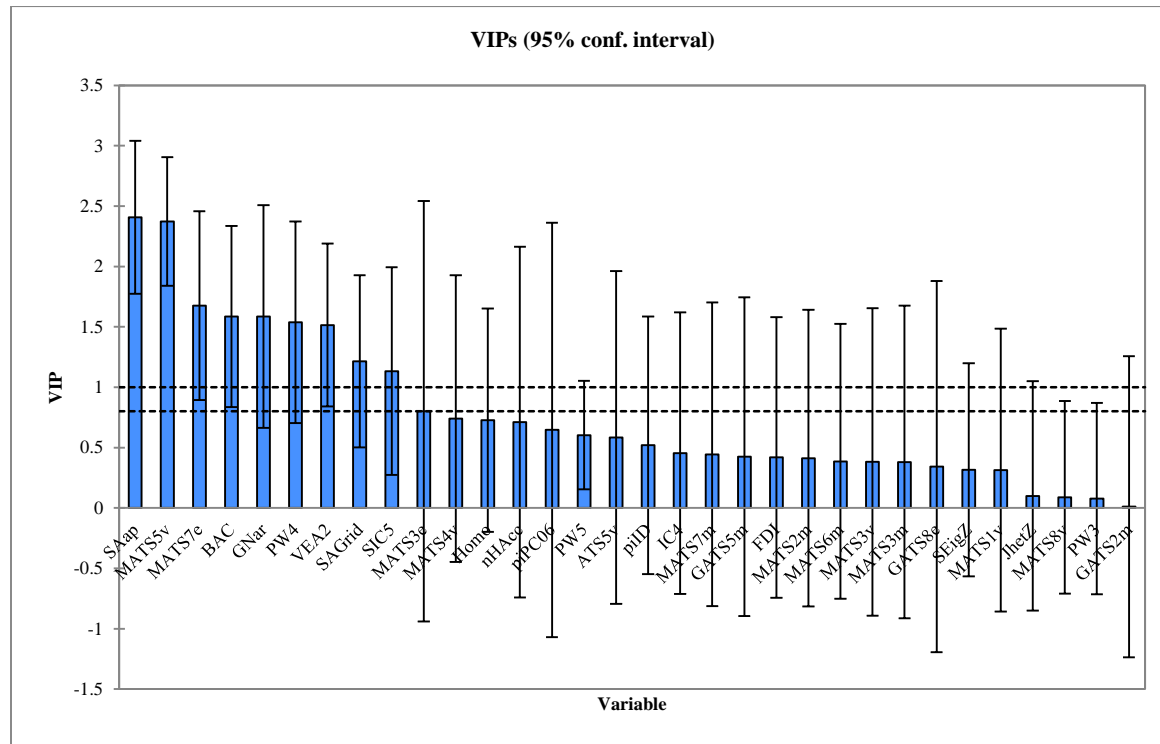


Figure 2. Variable importance in the projection (VIP) for the variables used in GA-PLS model for NK₁R.

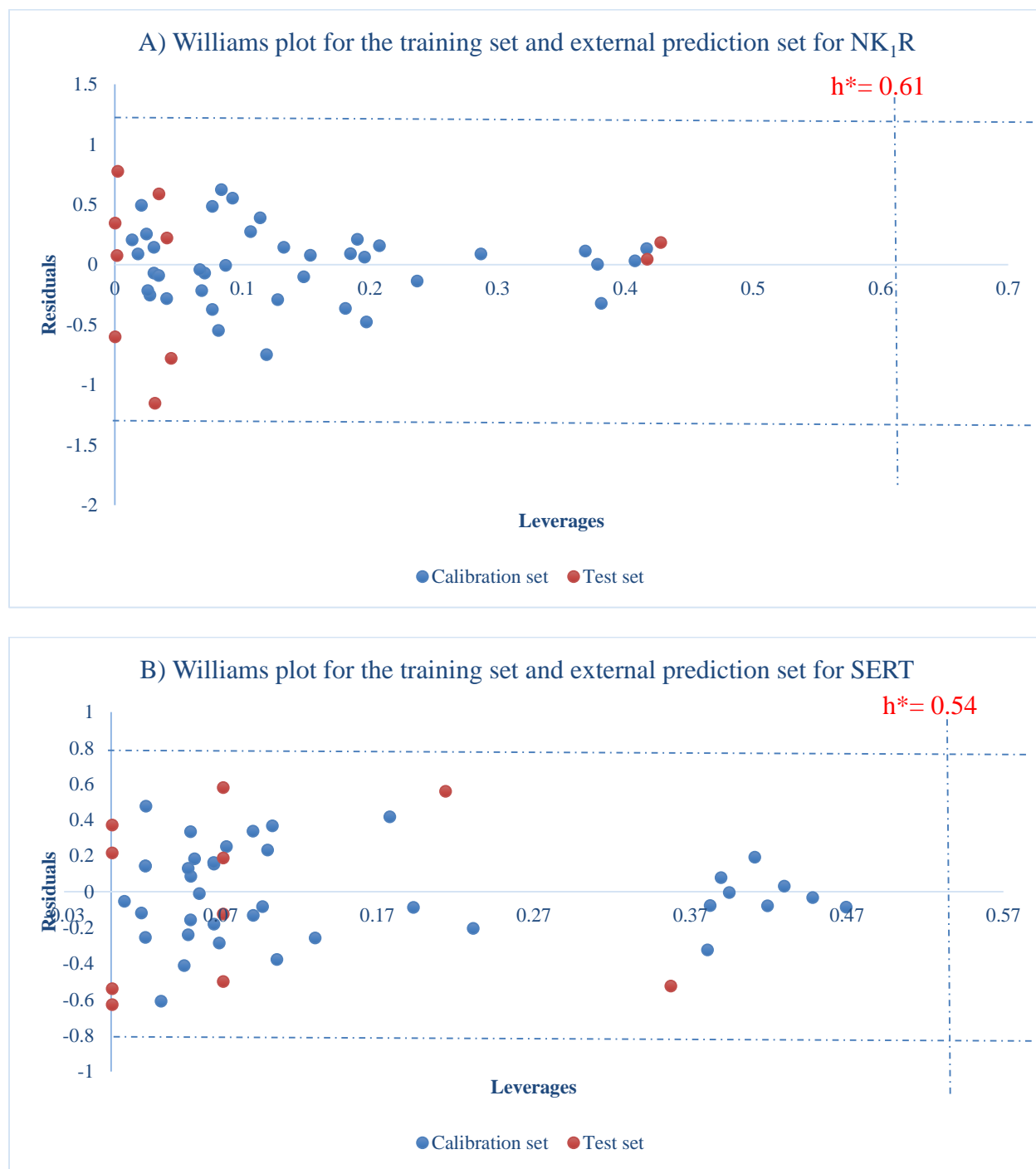


Figure 3. Williams plot for the calibration set and external prediction set for A) NK₁R antagonist activity B) SERT inhibition of studied compounds.

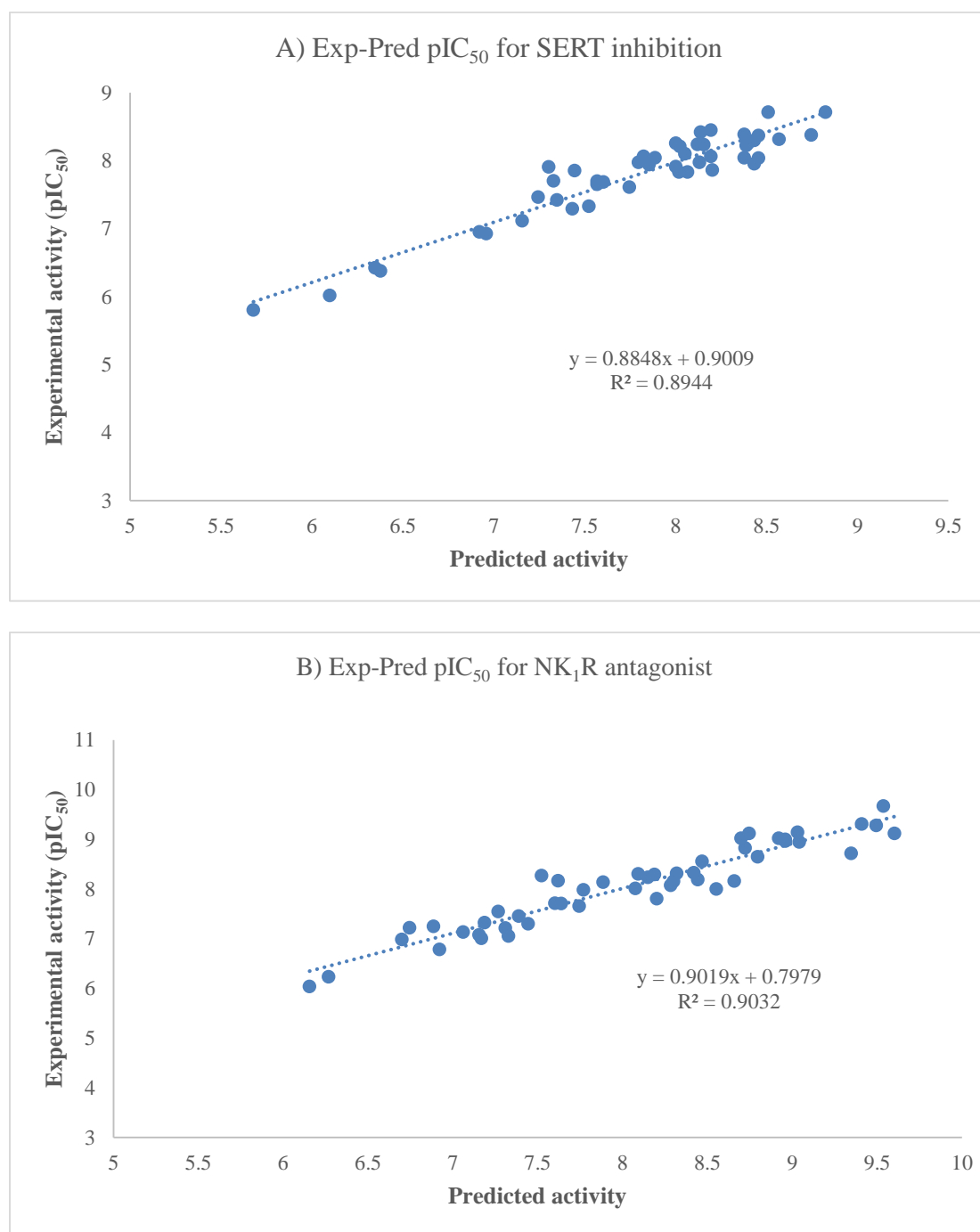


Figure 4. A) Plots of cross-validated predicted values of activity by MLR against the experimental values (SERT inhibition). B) Plots of cross-validated predicted values of activity by GA-PLS against the experimental values (NK_1R antagonist).

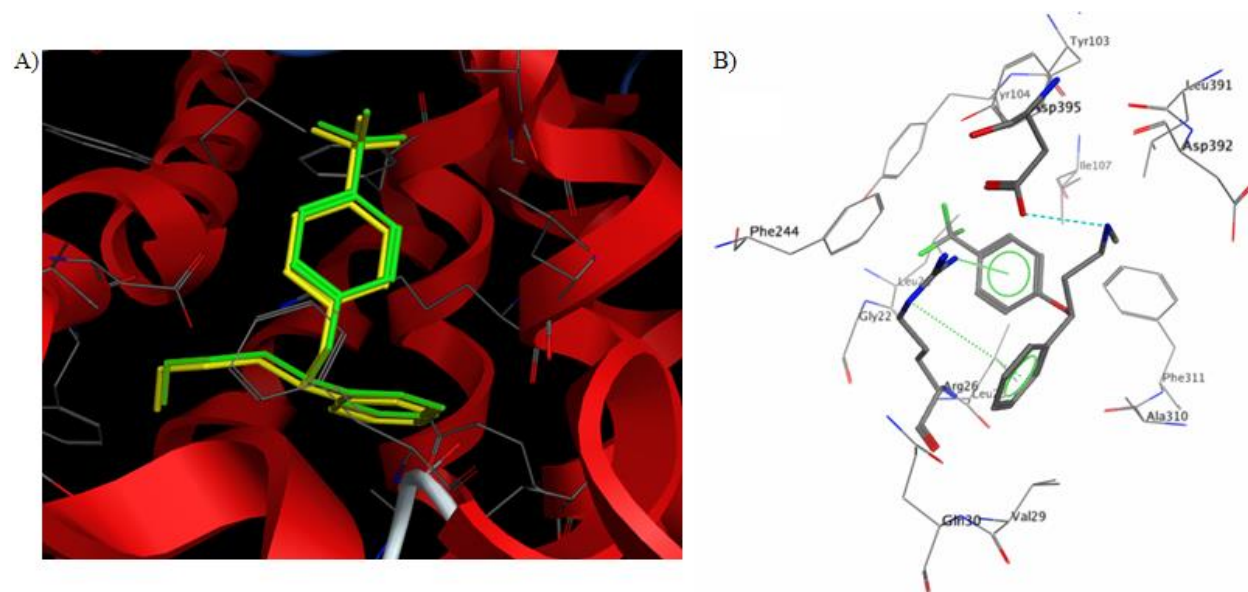


Figure 5. A) Comparison of two bound conformations of fluoxetine in the SERT active site: the yellow model shows the crystal orientation and the redocked result is shown as a green model. B) The structure of fluoxetine surrounded by the key residues in the active site of SERT.

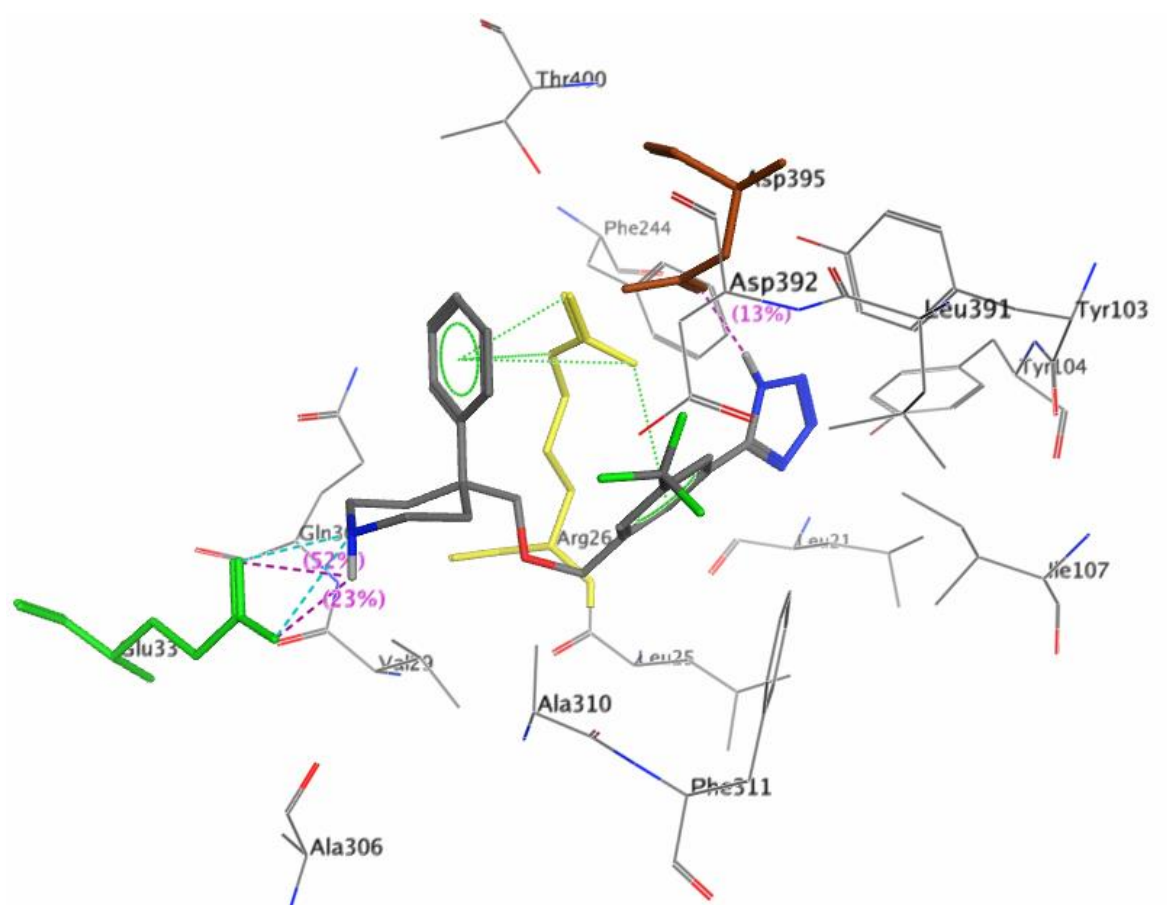


Figure 6. The structure of **4** surrounded by the key residues in the active site of SERT.

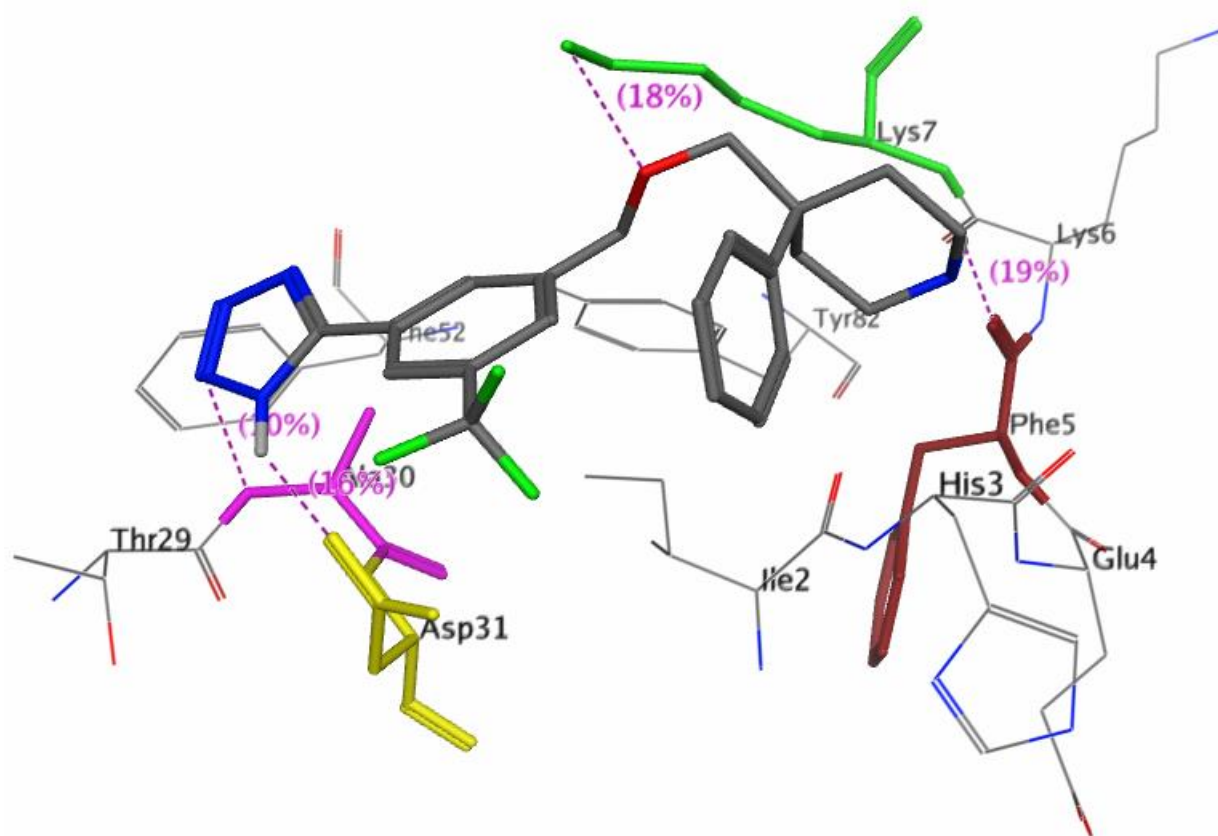


Figure 7. The structure of **4** surrounded by the key residues in the active site of NK₁R.

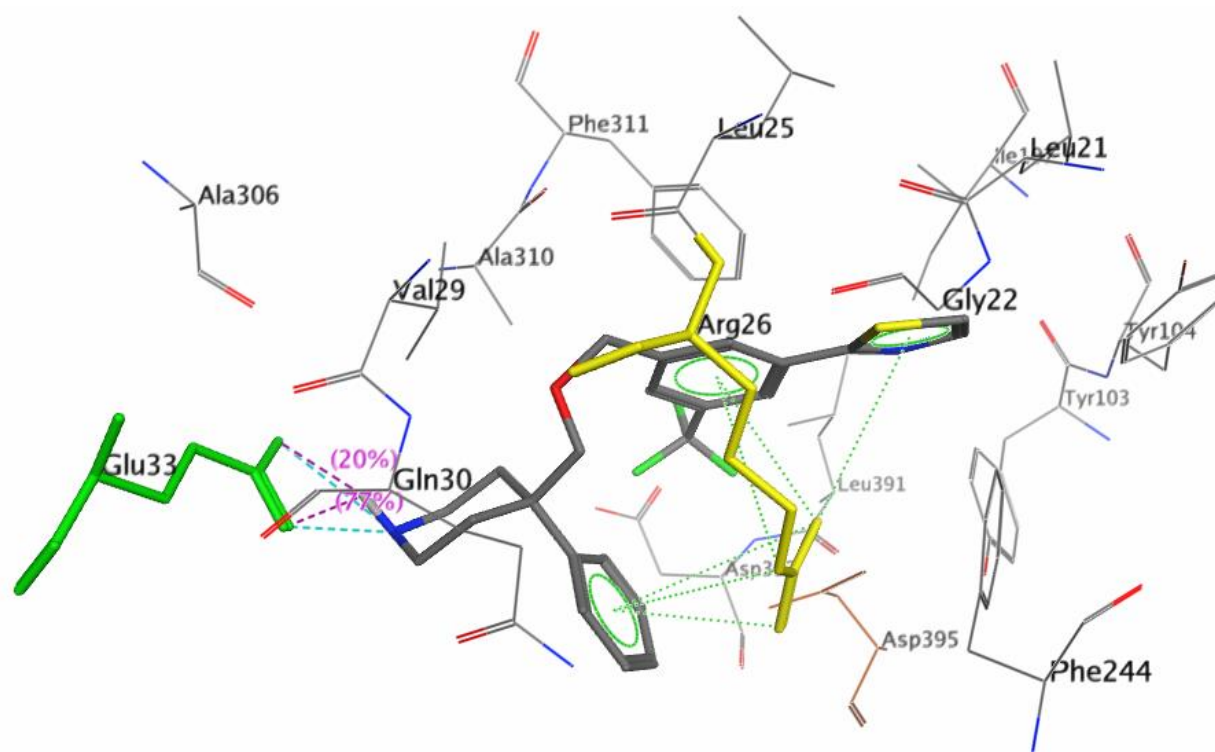


Figure 8. The structure of **11** surrounded by the key residues in the active site of SERT.

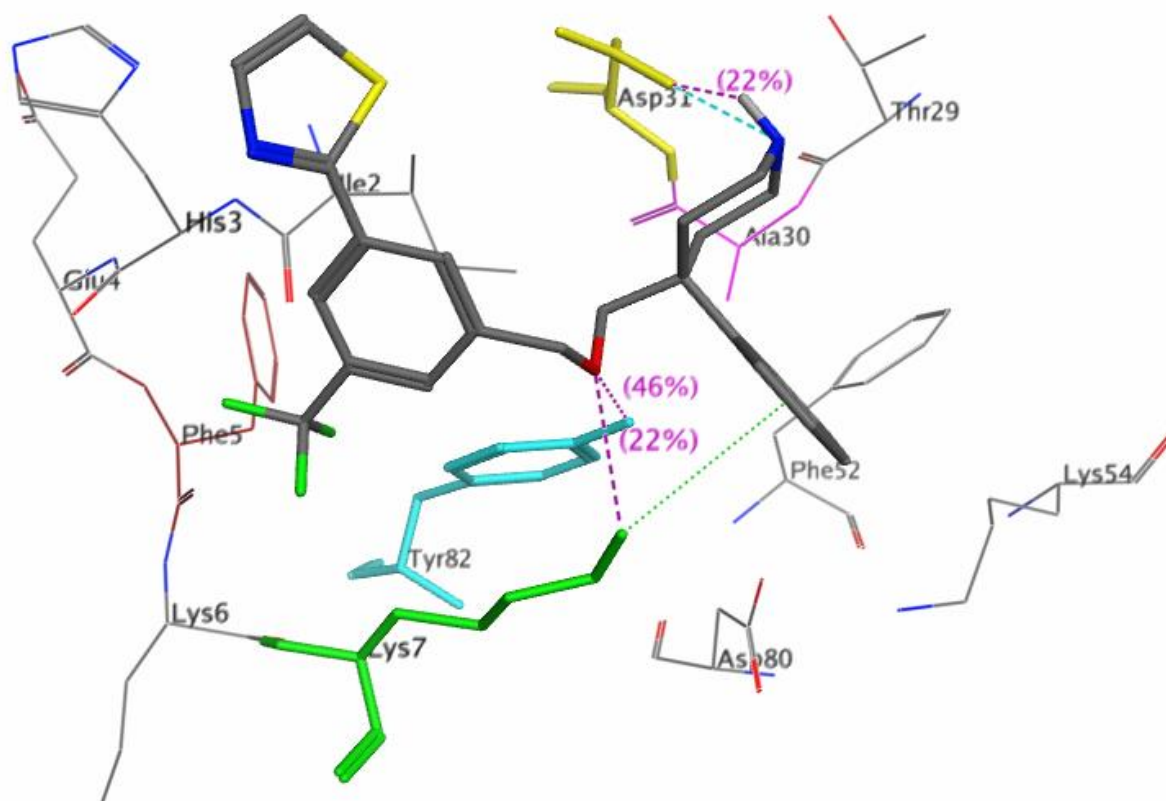


Figure 9. The structure of **11** surrounded by the key residues in the active site of NK₁R.

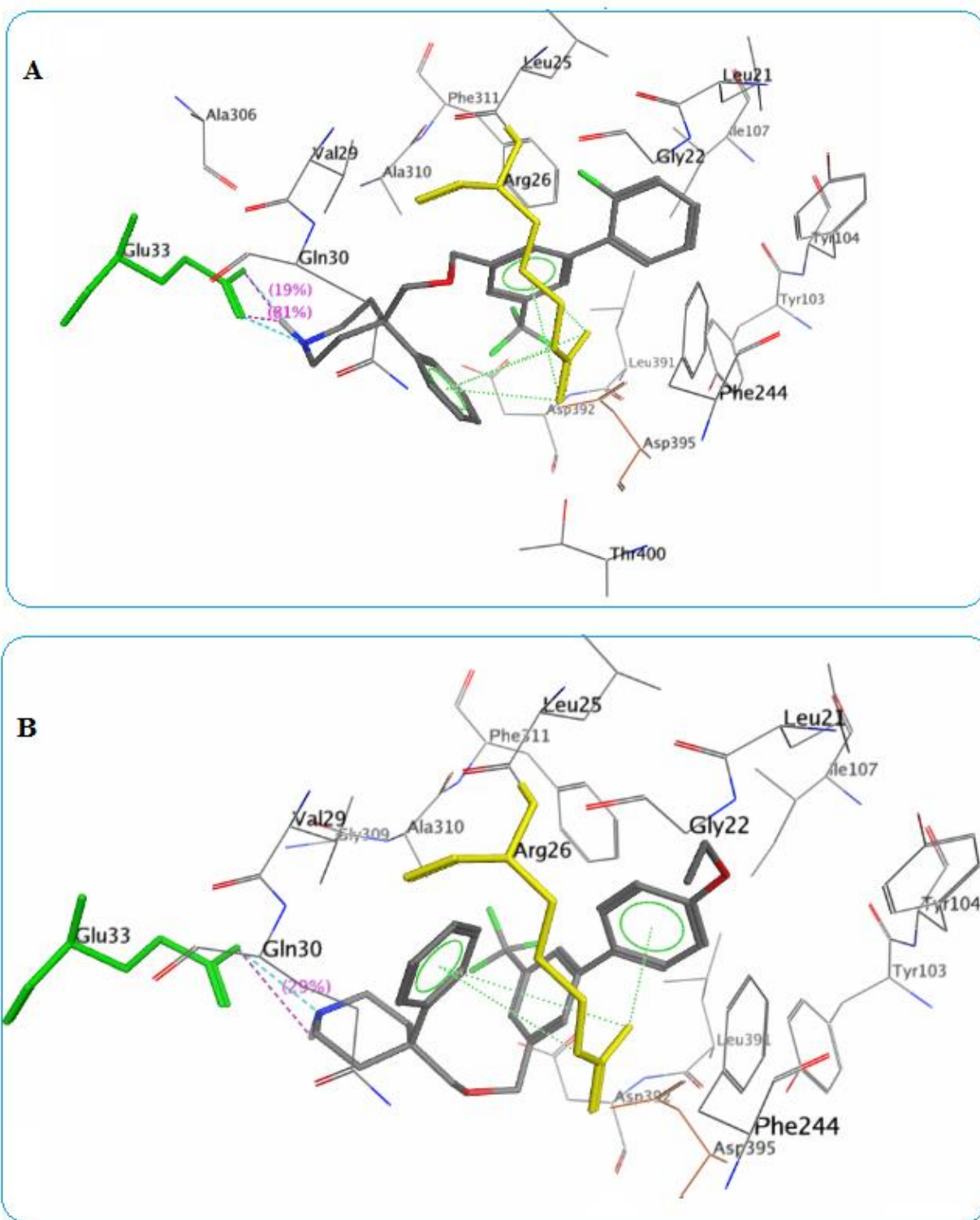


Figure 10. A) The structure of **23** surrounded by the key residues in the active site of SERT. B) The structure of **41** surrounded by the key residues in the active site of SERT

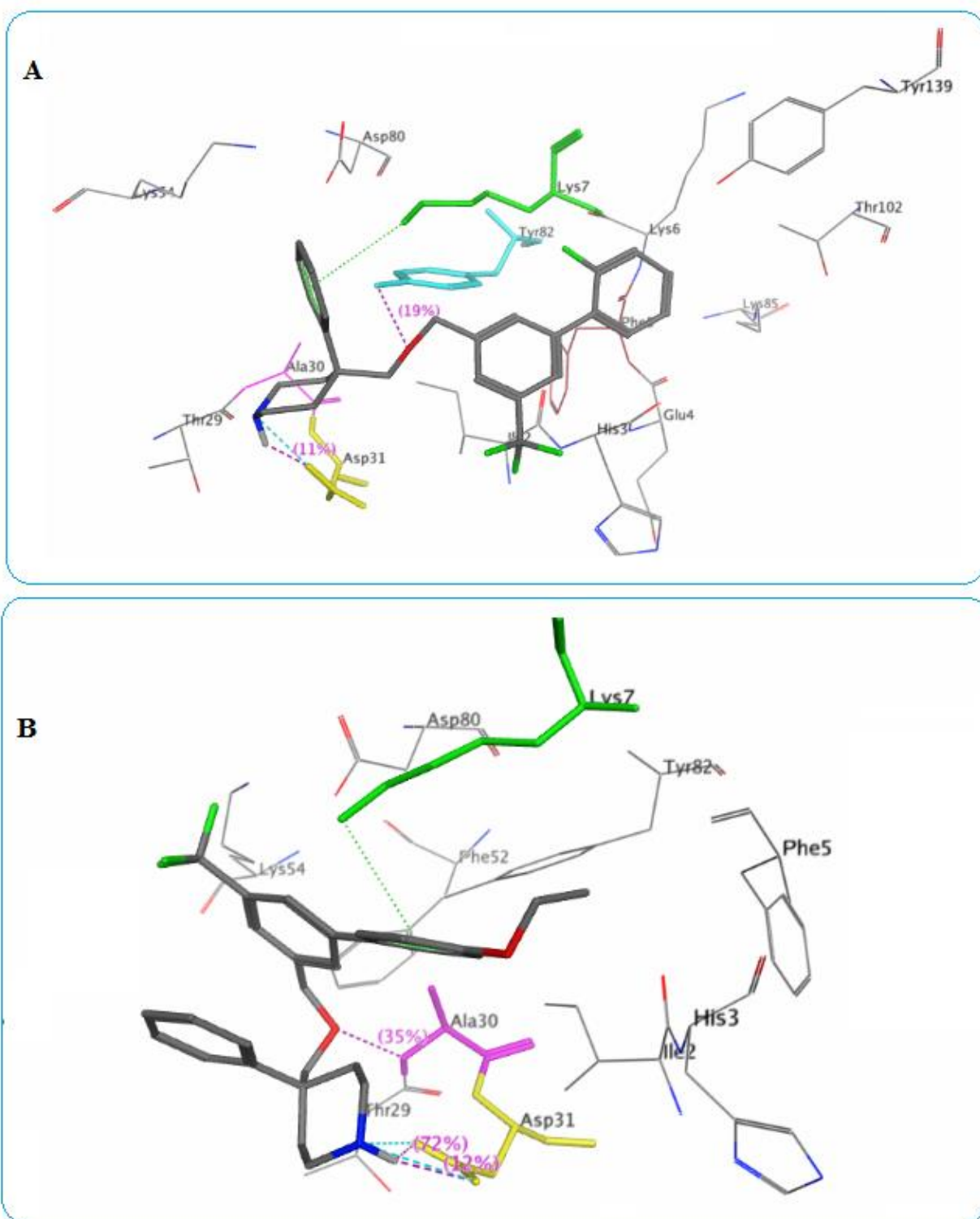


Figure 11. A) The structure of **23** surrounded by the key residues in the active site of NK₁R. B) The structure of **41** surrounded by the key residues in the active site of NK₁R.

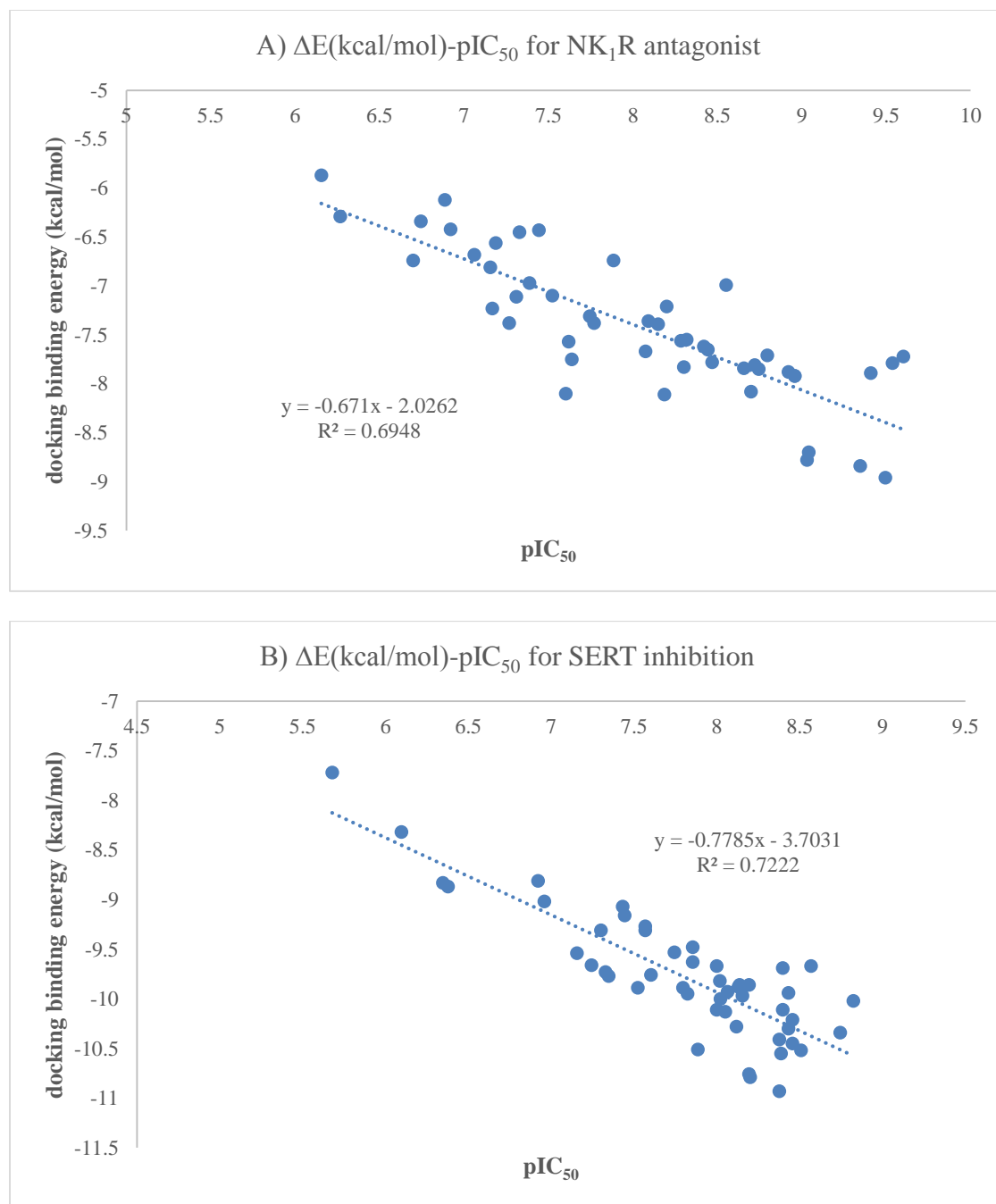
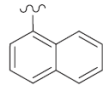
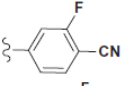
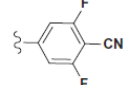
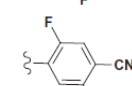
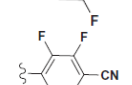
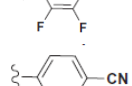
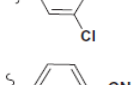
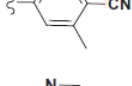
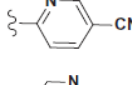


Figure 12. Plots of experimental pIC_{50} values versus docking binding energy, A) for NK_1R antagonist B) for SERT inhibition.

Table 1. Chemical structure of the compounds used in this study and their docking binding energy, experimental and cross-validated predicted activity on both targets (SERT, NK₁R).

		SERT			NK ₁ R		
Name	R	Exp.pIC ₅₀	Pred.pIC ₅₀	ΔE ¹ (kcal/mol)	Exp.pIC ₅₀	Pred.pIC ₅₀	ΔE (kcal/mol)
1-6							
7-13							
14-36							
37-44							
45-49							
1		7.43180	7.29358	-9.07	7.06048	7.13456	-6.68
2		7.15490	7.11435	-9.54	6.74473	7.22218	-6.34
3		7.24413	7.46327	-9.66	6.15490	6.04202	-5.87
4		8.06550	7.83271	-9.93	6.69897	6.98940	-6.74
5		7.30103	7.90908	-9.31	7.18709	7.32457	-6.56
6		8.01773	7.83389	-9.82	7.44370	7.30049	-6.43
7		8.43180	8.29871	-10.3	8.20066	7.81052	-7.21
8		7.79588	7.97621	-9.89	8.79588	8.65321	-7.71
9		8.13077	7.97621	-9.88	7.38722	7.45843	-6.97
10		7.74473	7.61231	-9.53	7.32790	7.05373	-6.45
11		8.38722	8.22359	-10.55	7.15490	7.07818	-6.81
12		6.95861	6.92776	-9.02	6.92082	6.78692	-6.42

13		6.92082	6.95331	-8.81	6.26761	6.23751	-6.29
14	2-Cl	8.19382	8.45031	-10.76	7.16749	7.01135	-7.23
15	2-F	8.11919	8.24168	-10.28	7.63827	7.70819	-7.75
16	2-Me	7.88606	8.04170	-10.51	6.88606	7.25101	-6.12
17	2-NO ₂	8.45593	8.36999	-10.21	8.18709	8.29347	-8.11
18	2-OMe	8.37675	8.04170	-10.41	7.26761	7.55062	-7.38
19	3-CH ₃	7.85387	7.97222	-9.63	7.61979	8.16836	-7.57
20	3-CN	8.00000	7.91234	-10.11	7.30980	7.21673	-7.11
21	3-F	8.20066	7.86313	-10.79	8.30103	8.15751	-7.83
22	3-NH ₂	7.52288	7.33074	-9.89	7.60206	7.71348	-8.1
23	3-NO ₂	8.37675	8.38663	-10.93	8.28400	8.07797	-7.56
24	3-OH	7.56864	7.65148	-9.31	7.88606	8.13910	-6.74
25	3,4-OCH ₂ O	7.34679	7.42436	-9.77	8.14874	8.23925	-7.39
26	4-CF ₃	7.44370	7.85456	-9.16	7.52288	8.27244	-7.1
27	4-Cl	8.74473	8.37799	-10.34	8.55284	8.00116	-6.99
28	4-CO ₂ Me	7.60206	7.68725	-9.76	7.76955	7.98603	-7.38
29	4-F	8.82391	8.71349	-10.02	8.65758	8.16384	-7.84
30	4-Me	8.19382	8.06294	-9.86	8.07572	8.01458	-7.67
31	4-NMe ₂	8.45593	8.03829	-10.45	7.74473	7.65713	-7.31
32	4-NO ₂	8.13668	8.42165	-9.86	8.42022	8.33056	-7.62

33	4-OEt	7.85387	7.94014	-9.48	8.46852	8.56214	-7.78
34	4-OH	7.56864	7.69970	-9.27	8.44370	8.18958	-7.65
35	4-OMe	7.82391	8.06294	-9.95	9.03152	9.14724	-8.78
36	4-CN	8.39794	8.25462	-10.11	9.34679	8.72278	-8.84
37		8.05061	8.10309	-10.13	8.95861	9.00057	-7.92
38		8.15490	8.23396	-9.97	9.49485	9.28372	-8.96
39		8.43180	7.95500	-9.94	9.04096	8.95015	-8.7
40		7.32790	7.70475	-9.73	8.31876	8.31605	-7.55
41		8.50864	8.71208	-10.52	8.95861	8.96431	-7.92
42		8.56864	8.31721	-9.67	8.92082	9.02304	-7.88
43		8.39794	8.25462	-9.69	8.09151	8.30863	-7.36
44		8.00000	8.25462	-9.67	8.72125	8.82914	-7.81
45	Me	8.02228	8.21128	-10.01	9.53760	9.67454	-7.79
46	Et	6.37675	6.38013	-8.87	9.60206	9.11887	-7.72
47	i-Pr	6.34679	6.42263	-8.83	9.40894	9.31278	-7.89
48	c-Pr	5.67778	5.80321	-7.72	8.74473	9.11887	-7.85
49	Bn	6.09691	6.01769	-8.32	8.69897	9.02233	-8.08

¹ docking binding energy

Table 2. Definitions of molecular descriptors present in the models

Descriptors	Brief description
SAA	Molecular surface area apparent
SAG	surface area grid
MATS1e	ran autocorrelation - lag 1 / weighted by atomic Sanderson electronegativities
MATS2e	ran autocorrelation - lag 2 / weighted by atomic Sanderson electronegativities
MATS3e	ran autocorrelation - lag 3 / weighted by atomic Sanderson electronegativities
MATS2V	Moran autocorrelation - lag 2 / weighted by atomic van der Waals volumes
MATS4V	Moran autocorrelation - lag 4 / weighted by atomic van der Waals volumes
MATS5V	Moran autocorrelation - lag 5 / weighted by atomic van der Waals volumes
MATS3m	Moran autocorrelation- lag 3 / weighted by atomic masses
MATS4m	Moran autocorrelation - lag 4 / weighted by atomic masses
MATS5m	Moran autocorrelation - lag 5 / weighted by atomic masses
MATS6m	Moran autocorrelation - lag 6 / weighted by atomic masses
MATS7m	Moran autocorrelation - lag 7 / weighted by atomic masses
GATS4p	Geary autocorrelation - lag 4 / weighted by atomic polarizabilities
GATS6m	Geary autocorrelation - lag 6 / weighted by atomic masses
piPC06	molecular multiple path count of order 06
ATS4p	Broto-Moreau autocorrelation of a topological structure - lag 4 / weighted by atomic polarizabilities
ATS5v	Broto-Moreau autocorrelation of a topological structure - lag 5 / weighted by atomic van der Waals volumes
PiID	conventional bond order ID number
nNHR	number of secondary amines (aliphatic)
nCIC	number of rings
nHDon	Number of donor atoms for H-bonds (N and O)
x3sol	solvation connectivity index chi-3
D/Dr10	distance/detour ring index of order 10
RBN	total number of rotatable bonds
Jhetp	Balaban-type index from polarizability weighted distance matrix
LP1	Lovasz-Pelikan index
DMY (DipY)	Molecular dipole moment at Y-direction
SPAM	average span R
RBF	rotatable bond fraction
STN	spanning tree number (log)
J3D	3D-Balaban index
RNCG	relative negative charge
TIC2	total information content index (neighborhood symmetry of 2-order)
SEigZ	the eigenvalue sum from Z weighted distance matrix
X5Av	average valence connectivity index chi-5 (linear and non-linear)

X2V	valence connectivity index chi-2
ISIZ	information index on molecular size
IVDM	mean information content vertex degree magnitude
H3D	3D-Harary index
PCWTe	partial-charge weighted topological electronic
VEA2	average eigenvector coefficient sum from adjacency matrix
GNar	Narumi geometric topological index
RPCG	relative positive charge
DECC	eccentric
FDI	folding degree index

Table 3. The results of different QSAR models with different type of dependent variables for SERT.

Model	Eq. no.	MLR Equation	n ¹	R ² _c	Q ²	Rm _{scv}	Cv _{cv}	F	SE	R ² _p
MLR	1	pIC ₅₀ = 2.165 nNHR(±0.298) – 1.15 nCIC (±0.291) -0.321 nHDon(±0.092) +0.574 x3sol (±0.0119) -0.192RBN (±0.069)-0.04D/Dr10 (±0.02) + 7.55 (±1.18)	39	0.85	0.81	0.27	3.57	31.1	0.12	0.76
FA-MLR	2	pIC ₅₀ = 2.039 nNHR(±0.26)+4.934 Jhetp(±1.232)-25.079 Lp1(±4.694)-0.359 nHDon (±0.114)+66.98 MATS5m(±16.359)+0.158 DipY(±0.065)-5.844 SPAM (±2.551) -3.353(±16.855)	39	0.81	0.64	0.41	5.29	19.1	0.22	0.86
PCRA	3	pIC ₅₀ = 0.126FAC1(±0.061) -0.380 FAC2 (±0.059) + 0.251 FAC3 (±0.066) + 0.218 FAC5 (±0.071) -0.147 FAC12 (±0.055) +7.806 (±0.064)	39	0.58	0.49	0.48	6.27	15.6	0.20	0.58
GA-PLS	4	pIC ₅₀ = -1.589 STN (±0.168) 44.28ATS4p (±7.00) -85.48 MATS6m (±15.22) +60.527 MATS4m (±13.853) + 9.394 MATS1e(±2.023) -2.246 J3D(±0. 491) - 15.628 RNCG (±7.399) +31.9033(±16.15)	39	0.80	0.73	0.33	4.26	17.8	0.07	0.74

¹Number of molecules of training set used to derive the QSAR modelsR²_c: Regression Coefficient for Calibration setQ²: Regression Coefficient for Leave One Out Cross ValidationRMS_{cv}: Root Mean Square Error of cross validationCv_{cv}: cross validation of cross validation

S.E: Standard Error

R²_p: Regression Coefficient for prediction set

Table 4. Correlation coefficient (R^2) matrix for descriptors represented in multiple linear regression eqn 1.

	D/Dr10	nHDon	X3sol	nCIC	RBN	nNHR
D/Dr10	1	-0.008	0.302	0.192	-0.210	0.070
nHDon		1	-0.340	0.146	-0.335	0.122
X3sol			1	0.288	0.232	-0.223
nCIC				1	-0.064	-0.246
RBN					1	-0.142
nNHR						1

Table 5. Factor loadings of some significant descriptors after VARIMAX rotation (SERT).

Descriptor	F1	F2	F3	F5	F12	Communalities
pIC ₅₀	0.612	0.593	0.212	-0.206	0.187	0.912
Ref	-0.181	0.928	0.153	-0.165	0.073	0.939
ATS6V	0.212	0.350	0.741	-0.246	-0.015	0.993
nCIC	-0.046	0.311	0.132	-0.905	0.046	0.988
VAR	0.017	0.918	0.091	0.050	-0.048	0.984
IVDM	0.143	0.849	0.330	-0.104	0.044	0.956
piID	0.032	0.366	0.323	-0.764	0.091	0.932
PCWTe	-0.021	0.340	-0.056	-0.069	0.877	0.956
Jhetp	0.137	-0.030	0.476	0.694	-0.062	0.970
VEA2	-0.291	-0.898	0.015	0.026	-0.061	0.979
RNCG	-0.057	-0.188	0.095	0.086	-0.929	0.965
qneg	-0.103	-0.019	0.001	0.084	-0.681	0.983
nHDon	0.012	-0.500	0.075	-0.016	-0.081	0.981
nHAcc	0.819	-0.110	-0.411	0.048	-0.036	0.940
RBN	-0.006	0.051	-0.036	0.046	0.000	0.894
D/Dr10	-0.059	0.058	0.257	-0.768	0.075	0.871
Lp1	0.039	0.028	-0.007	-0.012	0.006	0.939
G(F..F)	0.820	0.001	-0.018	0.001	0.005	0.919
MATS5m	0.463	0.489	0.125	0.053	-0.007	0.952
DipY	-0.219	-0.639	0.337	0.136	0.187	0.967
ATS4p	-0.268	0.151	0.706	-0.223	-0.081	0.929
MATS4m	0.689	-0.060	0.241	0.029	0.048	0.955
MATS1e	0.035	-0.079	-0.336	0.037	-0.067	0.943
J3D	-0.180	0.142	-0.149	0.624	0.046	0.910
MATS6m	0.574	0.117	-0.026	0.258	0.143	0.987
GATS1m	-0.390	0.118	-0.105	-0.016	-0.023	0.979
GATS2m	-0.897	0.040	-0.117	0.011	-0.029	0.963
GATS4m	-0.895	0.117	-0.012	0.076	-0.093	0.954
ATS7e	0.919	-0.094	0.037	0.138	-0.027	0.981
X2v	-0.110	0.888	0.227	-0.155	0.070	0.918
ATS3v	0.282	0.343	0.793	-0.158	0.059	0.976
ATS1v	0.172	0.264	0.839	0.084	-0.017	0.945
nNHR	0.181	-0.797	0.182	-0.104	-0.160	0.983