SAR and QSAR in Environmental Research, 21(3-4), 369-388

A QSAR study on serotonin 5-HT$_6$ receptor ligands: indolyl and piperidinyl sulfonamides

B. K. Sharma[a,*], P. Singh[a], K. Sarbhai[a], and Y. S. Prabhakar*[b]

[a]Department of Chemistry, S. K. Government College, Sikar-332 001, India; [b]Medicinal and

Process Chemistry Division, Central Drug Research Institute, CSIR, Lucknow-226 001, India

Abstract

The serotonin 5-HT$_6$ binding affinity of indolyl- and piperidinyl-sulfonamide derivatives has been analyzed with the topological and molecular features from Dragon software. Analysis of the structural features in conjunction with the biological endpoints in Combinatorial Protocol in Multiple Linear Regression (CP-MLR) led to the identification of 25 descriptors for modeling the activity. The study clearly suggested the role of average Randic-type eigenvector-based index from adjacency matrix, VRA2, number of secondary aliphatic amines, nNHR, sum of topological distance between N and O, T(N..O), ring tertiary carbon atoms, nCrHR, and CH2RX type fragment, C-006, in a molecular structure to optimize the 5-HT$_6$ binding affinities of titled compounds. The PLS analysis has confirmed the dominance of information content of CP-MLR identified descriptors for modeling the activity when compared to those of leftover ones.

Keywords: Indolyl and piperidinyl sulfonamides; Serotonin 5-HT$_6$ receptor; Dragon descriptors; Combinatorial protocol in multiple linear regression (CP-MLR); Partial least square (PLS) analysis.

---

*Corresponding author; email: bksharma_sikar@rediffmail.com; yenpra@yahoo.co

1. Introduction

5-Hydroxytryptamine (serotonin) plays a vital role in various CNS disorders [1]. At present, 14 distinct 5-HT receptor subclasses have been reported in mammalian CNS [2]. On the basis of molecular cloning, amino acid sequence, pharmacology and signal transduction the reported serotonin receptors are grouped into seven subfamilies (5-HT_{1-7}) [3]. In 5-HT receptors, except the 5-HT_3 receptor, all other are G protein-coupled receptors. The 5-HT_6 receptor is positively coupled to adenylyl cyclase [4-6]. It is mainly localized in olfactory tubercles, striatum, nucleus accumbens, and hippocampus. Lower levels have been found in amygdale, hypothalamus, substantia nigra, cerebellum, or cerebral cortex. Studies, such as microdialysis [7-11], in situ hybridization [12], and electrophysiology [13] revealed the appearance of 5-HT_6 receptor in regulating several neurotransmitter systems including dopamine, noradrenaline, glutamate, aspartate and acetylcholine. The unique distribution in the brain and high affinity for therapeutic antipsychotics [14] and antidepressants [15] suggest a possible role of 5-HT_6 in CNS disorders. Studies have demonstrated that 5-HT_6 receptor variants are also associated with schizophrenia [16-20], bipolar affective disorders [21], Parkinson's disease [22] and Alzheimer's disease [23-25]. Potent and selective ligands with defined functionality are the important tools to elucidate the functional role of the 5-HT_6 receptor. These novel ligands include 5-methoxytryptamine, bromocriptine, octoclothepin, and the neuroleptics (such as clozapine, olanzapine, loxapine, chloropromazine and fluphenazine). The reported selective antagonists are the benzenesulfonamides (Ro 04-6790 and Ro 63-0563 [26], benzo[*b*]thiophenesulfonamide (SB-271046) [27] and the sulfonylindoles (ALX1161, ALX1175 [28] and MS-245 [29]). Recently, a novel series of indolyl sulfonamides, which is based on the medicinal chemistry guided hypothetical pharmacophore model, have been designed and synthesized as selective and high-affinity serotonin 5-HT_6 receptor ligands [30]. Furthermore, as a result of a screening of Esteve's discovery library against the cloned h5-HT_6 receptor, a series of benzoxazinepiperidinyl sulfonamide derivatives was identified as selective 5-HT_6 agents [31]. These compounds may allow for the detailed study of the role of the 5-HT_6 receptor in relevant animal models of disorders such as cognition deficits, depression, anxiety and obesity. In recent modeling studies, the graph theoretical descriptors from the molecular structures are increasingly put to use for exploring the QSAR rationales [32-40]. They suggest the important structural fragments for the phenomena under investigation and thereby the insight of the receptor topography via the structural parts of the probe molecules. In view of this, a QSAR study of 5-HT_6 receptor binding affinity of indolyl/piperidinyl sulfonamide derivatives is attempted using the 0D- to 2D-descriptors from DRAGON software [41].

2. Materials and method

2.1 *Data set*

The QSAR study has used reported series of compounds along with their human 5-HT6 receptor binding affinities (K_i) [30]. The structural variations in these compounds and their K_i values are shown in Table 1. The reported K_i values (Table 1) were determined in stably transfected HEK-293 cells using the radioligand [^{3}H]-LSD [30, 42, 43]. The general structures of the compounds are given in Figure 1. Compounds **1-28** (Figure 1A) are the representative of N-Aminoalkylindolyl sulfonamides whereas, compounds **29-52** (Figure 1B) contain variations at 3-position of indole. Compounds **53-57** (Figure 1C) belong to N-(3-Glyoxamidoindolyl) sulfonamides and **58-63** (Figure 1D) are the derivatives of 3-piperidinylindolyl sulfonamides. The benzoxazinepiperidinyl sulfonamide derivatives are presented through compounds **64-73** (Figure 1E). In present QSAR study the reported binding affinities were transformed into logarithm of reciprocal K_i, on molar basis, and expressed as pK_i. Since biological activity data are generally found to be skewed, the logarithmic transformation moves it to nearly normal distribution. Thus in QSAR studies, the concentration [C] values (e.g., K_i, IC_{50}, EC_{50}, etc.) are often expressed as log[C] or log 1/[C]. Also, log[C] or log 1/[C] are used in Hansch type of QSAR equations to emulate the linear free energy relationship with the molecular parameters.

2.2 *Computational procedure*

DRAGON software [41] has been used for the parameterization of the aforesaid compounds. The structures of the compounds under study have been drawn in 2D ChemDraw [44] and then converted into 3D modules using the default conversion procedure implemented in the CS Chem 3D Ultra. The generated 3D-structures of the compounds were subsequently subjected to energy minimization in the MOPAC module, using the AM1 procedure for closed shell systems. This was done to ensure a well defined conformer relationship across the compounds of the study. All these energy minimized structures were exported to DRAGON software for the computation of descriptors of the compounds in Table 1. This software offers a large number of descriptors corresponding to 0D-, 1D- and 2D- characteristics of the molecules. The descriptor classes considered in the study along with their definition and scope in addressing the structural features have been given in Table 2. For each of these classes, the computed descriptors are characteristic of the molecule under multi-descriptor class environment. The combinatorial protocol in multiple linear regression (CP-MLR) [45] and partial least-squares (PLS) [46-48] procedures have been used in the present work for developing QSAR models. Before the application of CP-MLR procedure, all those descriptors which are intercorrelated beyond 0.90 and showing a correlation of less than 0.1 with the biological endpoints (descriptor *vs* activity, $r < 0.1$) were

excluded. This procedure has reduced the total descriptors from 497 to 148 as relevant ones to explain the biological actions of titled compounds.

2.3 *Model Development*

The CP-MLR is a filter based variable selection procedure for model development in QSAR studies. Its procedural aspects and implementation are discussed in some of the recent publications [49-53]. The thrust of this procedure is in its embedded 'filters'. They are briefly as follows: filter-1 seeds the variables by way of limiting inter-parameter correlations to predefined level (upper limit ≤ 0.79); filter-2 controls the variables entry to a regression equation through *t*-values of coefficients (threshold value ≥ 2.0); filter-3 provides comparability of equations with different number of variable in terms of square-root of adjusted multiple correlation coefficient of regression equation, r-bar; filter-4 estimates the consistency of the equation in terms of cross-validated R^2 or Q^2 with leave-one-out (LOO) cross-validation as default option (threshold value $0.3 \leq Q^2 \leq 1.0$). In order to collect the descriptors with higher information content and explanatory power, the threshold of filter-3 was successively incremented with increasing number of descriptors (per equation) by considering the r-bar value of the preceding optimum model as the new threshold for next generation.

2.4 *Y-randomization*

Furthermore, in order to find out any chance correlations associated with the models recognized in CP-MLR, each cross validated model has been put to a randomization test [54,55] by repeated randomization of the activity to discover the chance correlations, if any, associated with them. For this every model has been subjected to 100 simulation runs with scrambled activity. The scrambled activity models with regression statistics better than or equal to that of the original activity model have been counted to express the percent chance correlation of the model under scrutiny.

2.5 *Model validation*

Validation of the derived model is necessary to test its prediction and generalization within the study domain. The data set of the present study is randomly divided into training set for model development and test set for external prediction. Out of the total number of 65 active compounds, nearly one fourth compounds (15) have been included in the test set for the validation of QSAR models obtained from training set. The compounds of test set were randomly picked using an in-house written randomization program in such a way that at lest one representative compound from each structure type (Fig. 1A to 1E) will be present in the test set.

For each model, a number of statistical parameters were obtained to access its overall statistical significance. These are multiple regression coefficient (r), the standard deviation (s), the F-ratio between the variances of calculated and observed activities (F), the cross validated Q^2_{LOO} (leave-one-out) and Q^2_{L5O} (leave-five-out). Additional statistical parameters such as the Akaike's information criterion, AIC [56,57], the Kubinyi function, FIT [58,59], and the Friedman's lack of fit, LOF [60], have also been calculated to further validate the derived models. The AIC takes into account the statistical goodness of fit and the number of parameters that have to be estimated to achieve that degree of fit. The FIT, closely related to the F-value (Fischer ratio), proved to be a useful parameter for evaluating the quality of models. The main drawback of the F-value is the change in its sensitivity with the value of k. It shows higher sensitivity for smaller k values and the sensitivity decreases with increasing k value. The FIT criterion has a low sensitivity towards changes in k-values, as long as the values are small and a gradually increasing sensitivity for large k-values. The model with the minimum value of AIC and highest value of FIT is considered to be the most robust model. The LOF is concerned with the number of terms used in the equation and is not biased, as are other indicators, towards large number of parameters. The model having minimum value of LOF is statistically sound.

The LOO method creates a number of modified data sets by taking away one compound from the parent data set in such a way that each observation is removed once only. This reduced data set is used to develop a model and the response values of removed compounds are predicted from these models leading to the Q^2_{LOO} index. In leave-five-out procedure a group of five compounds is randomly kept outside the analysis each time in such a way that all compounds, for once, become the part of the predictive groups. In case of internal validation, cross validated Q^2_{LOO} (leave-one-out) and Q^2_{L5O} (leave-five-out), is used as a criterion of both robustness and predictive ability of the model [61]. A value greater than 0.5 of Q^2-index suggests a statistically significant model. The predictive power of derived model is based on test set compounds. The model obtained from training set has a reliable predictive power if the value of the r^2_{Test} (the squared correlation coefficient between the observed and predicted values of compounds from test set) is greater than 0.5.

2.6 *Applicability domain*

The usefulness of a model is based on its accurate prediction ability for new congeners. A model is valid only within its training domain and new compounds must be assessed as belonging to the domain before the model is applied. The applicability domain is evaluated by the leverage values for each compound [62]. A Williams plot (the plot of standardized residuals versus leverage values (h)) is constructed which can be used for a simple graphical detection of both the response outliers(Y outliers) and structurally influential chemicals (X outliers) in the model. In this plot, the applicability domain is established inside a squared

area within ± x standard deviations and a leverage threshold h* which is generally fixed at 3(k + 1)/n (n is the number of training set compounds and k is the number of model parameters) whereas x = 2 or 3. If the compounds have a high leverage value (h > h*) then the prediction is not trustworthy. On the other hand, when the leverage value of a compound is lower than the threshold value, the probability of accordance between predicted and observed values is as high as that for the training set compounds.

3. Results and discussion

3.1 *QSAR results*

In multi-descriptor class environment, exploring for best model equation(s) along the descriptor class provides an opportunity to unravel the phenomenon under investigation. In other words, the concepts embedded in the descriptor classes relate the biological actions revealed by the compounds. A total number of 148 significant descriptors from 0D-, 1D- and 2D-classes, identified in the initial stage, have been subjected to CP-MLR analysis with default 'filters' set in it. Statistical models in one, two, three and four descriptor(s) have been derived successively to achieve the best relationship correlating the pK_i. Though each individual descriptor class is enriched with information corresponding to the activity, different descriptors classes together have led to the models with optimum explained variance. These models (with collective descriptors) were identified in CP-MLR by successively incrementing the filter-3 with increasing number of descriptors (per equation). For this the optimum r-bar value of the preceding level model has been used as the new threshold of filter-3 for the next generation. All the calculations in CP-MLR have been performed at the upper limit for filter-1. The highest significant regression equations, emerged in CP-MLR, in three and four parameters are given below

$pK_i = 9.556 – 0.061(0.012)ATS5m – 3.160(0.959)GATS2e + 0.638(0.091)nCaR$
$n = 50, r = 0.750, s = 0.563, \quad F = 19.669, Q^2_{LOO} = 0.478, Q^2_{L5O} = 0.496,$
$r^2_{randY}(sd) = 0.233(0.084), FIT = 1.000, LOF = 0.377, AIC = 0.372 \qquad (6)$

$pK_i = 8.443 + 0.046(0.010)VRA2 – 0.019(0.005)T(N..O) + 0.727(0.131)nNHR$
$\quad – 0.389(0.068)C\text{-}006$
$n = 50, r = 0.807, s = 0.508, F = 21.055, Q^2_{LOO} = 0.582, Q^2_{L5O} = 0.581,$
$r^2_{randY}(sd) = 0.270(0.090), FIT = 1.276, LOF = 0.329, AIC = 0.315 \qquad (7)$

In above and all follow up regression equations, the values given in the parentheses are the standard errors of the regression coefficients. The signs of the regression coefficients suggest the direction of influence of explanatory variables in the models. The $r^2_{randY}(sd)$ is the mean random squared multiple correlation coefficient of the regressions in the activity (Y) randomization study with its standard deviation from 100 simulations. In the randomization study (100 simulations per model), none of the identified models has shown any chance correlation.

In above equations, the 2D-AUTO class descriptors, ATS5m (Broto-Moreau autocorrelation of a topological structure of lag 5 weighted by atomic masses) and GATS2e (Geary autocorrelation of lag 2 weighted by atomic Sanderson electronegativities), contribute negatively to the activity. Thus molecules leading to lower value of such topological lags (paths) weighted by relevant atomic property would be favorable to the binding activity. The positive regression coefficients of FUNC class descriptors, nCaR (number of substituted aromatic carbon) and nNHR (number of secondary aliphatic amines), advocate such functionality in a molecule for better activity. A higher value of average Randic-type eigenvector-based index from adjacency matrix (TOPO class descriptor, VRA2) and a lower value of sum of topological distances between N and O (TOPO class descriptor, T(N..O)) would supplement the activity. The ACF class descriptor C-006, representing a CH2RX fragment in a molecular structure, has shown negative influence on the activity suggesting absence of such type of fragment for improved activity.

The above model equations (6) and (7), in three and four descriptors, could explain 56.25 and 65.12 percent variance in the observed activity values, respectively. Considering the number of observation in the dataset, models with up to five descriptors were explored. It has resulted in 38 five-parameter models with test set $r^2 > 0.50$. These models have shared 25 descriptors among them (Table 3). The physical meanings of these descriptors to the molecular structure are elaborated in Table 3. In order to examine the subgroup relations in the structure classes (Fig. 1A to 1E) of the compounds (Table 1) in terms of the 25 identified descriptors, a principal component analysis (PCA) has been carried out on all the compounds involving the identified descriptors. Figure 2 shows the plot of compounds (Table 1) with respect to first two components. The PC scores of the compounds are provided in the supplementary information. Figure 2 indicates that structure class A (compounds 1-28; Fig.1A) and structure class B (compounds 29-52; Fig.1B) are closely spaced with each other, whereas structure class D (compounds 58-63; Fig.1D) is embedded in the midst of structure classes A and B. This confirms the chemical space proximity of structure classes A, B and D. Interestingly, structure class E (compounds 64-73; Fig.1E) is at the periphery of structure class 'A' and 'B' and structure class C (compounds 53-57; Fig.1C) has occupied exclusive space far away from rest of the structure classes. Thus the identified descriptors have reasonably well identified the patterns in the compounds. The following is the highest significant model identified in five-descriptors for correlating the activity.

$pK_i = 8.558 + 0.033(0.010)VRA2 – 0.018(0.005)T(N..O) – 0.440(0.154)nCrHR$

$$+ 0.801(0.124)\text{nNHR} - 0.351(0.065)\text{C-006}$$
$$n = 50,\quad r = 0.840,\ s = 0.471,\ F = 21.170,\ Q^2_{LOO} = 0.645,\ Q^2_{L5O} = 0.652,$$
$$r^2_{randY}(sd) = 0.304(0.095),\ FIT = 1.411,\ LOF = 0.306,\ AIC = 0.283,\ r^2_{Test} = 0.655 \qquad (8)$$

The derived F-value for above equation remained significant at 99% level. The model is able to explain nearly 71 percent variance in observed activity values. In order to improve the significance of above equation, the compounds showing high activity residuals (observed − calculated) were considered as outliers. An outlier to a QSAR is normally identified by means of the residuals of the dependent variable. The outliers generally reflect large residuals in dependent variable and can restrict the applicability of QSAR models [63]. In this study, the compounds **4** and **42** have shown large residuals and were considered as "outliers". Although the compound **4** is the compound with best affinity in set A, no plausible explanation could be assigned for its outlier behavior. There are many reasons for the occurrence of outlier(s) in QSAR studies; for example, chemicals might be acting by a mechanism different from that of the majority of the data points. It is also likely that outlier might be a result of a random experimental error that could be significant when analyzing a large data set. The model obtained, on removal of these two compounds from the training set, is shown in correlation Eq. (9).

$$pK_i = 8.555 + 0.031(0.008)\text{VRA2} - 0.020(0.004)\text{T(N..O)} - 0.484(0.125)\text{nCrHR}$$
$$+ 0.918(0.104)\text{nNHR} - 0.319(0.053)\text{C-006}$$
$$n = 48,\ r = 0.898,\ s = 0.381,\ F = 34.813,\ Q^2_{LOO} = 0.744,\ Q^2_{L5O} = 0.754,$$
$$r^2_{randY}(sd) = 0.330(0.094),\ FIT = 2.384,\ LOF = 0.203,\ AIC = 0.187,\ r^2_{Test} = 0.680 \qquad (9)$$

The statistical parameters of Eq. (9) have shown considerable improvement over those of Eq. (8). The Eq. (9) has explained 80.64 percent of variance in observed activity values, and Q^2-index has accounted comparatively for a better robust model. The decreased values of parameters AIC and LOF and increased value of FIT have further shown the superiority of this model over that of the model given by Eq. (8). Equation (9) was also subjected to randomization process on observed activities, where 100 simulations were carried out but none of the identified models has shown any chance correlation. The calculated activity values, using Eq. (9), for the training and test set compounds, are in close agreement to the observed activity values. These values are provided in supplementary material. The goodness of fit between the experimental (observed) and calculated activity (through Eq. 9) for all the reported active compounds is shown in Figure 3. The predictions of the fifteen test-set compounds based on external validation are found to be satisfactory as reflected in the r^2_{Test} value. The obtained value, 0.680, of r^2_{Test} shows that the model has reliable predictive power. The residuals of test-set compounds from Eq. (9) are given in Table 4.

It is apparent from Equation (9) that a higher value of average Randic-type eigenvector-based index from adjacency matrix and more number of secondary aliphatic amines in a molecular structure are helpful to augment the activity. Additionally, a lower value of sum of topological distance between N and O, less number of ring tertiary carbon atoms (FUNC class descriptor, nCrHR) and absence of CH2RX type fragment (des in a molecular structure would be advantageous to enhance the activity. Thus the descriptors identified for rationalizing the activity give avenues to modulate the structure to a desirable biological end point. The Equation (9) has further been used to explore some new sulfonamide derivatives as 5-HT_6 receptor ligands. The potential structural variations and predicted activity values for these compounds are given in Table 5.

A PLS (partial least squares) analysis has been carried out on these 25 CP-MLR identified descriptors (Table 3) to facilitate the development of a 'single window' structure-activity model. For the purpose of PLS, the descriptors have been autoscaled (zero mean and unit s.d.) to give each one of them equal weight in the analysis. In the PLS cross-validation, three components are found to be the optimum for these 25 descriptors and they explained 79.4 % variance in the activity ($r^2 = 0.794$, $Q^2_{LOO} = 0.750$, $s = 0.383$, $F = 56.596$, $r^2_{Test} = 0.690$). The MLR-like PLS coefficients of these 25 descriptors are given in Table 3. The calculated activity of training and test set compounds based on this analysis are given in supplementary information. For the sake of comparison, the plot showing goodness of fit between observed and calculated activities (through PLS analysis) for the training and test set compounds is given in Figure 4. Figure 5 shows a plot of the fraction contribution of normalized regression coefficients of these descriptors to the activity. This plot provides an opportunity to make a comparison of relative significance among the descriptors. The derived fraction contributions from the normalized regression coefficients of descriptors allow this comparison within modeled activity.

The PLS analysis has also suggested nNHR (a FUNC class descriptor) as the most determining descriptor for modeling the activity of the compounds (descriptor S. No. 20 in Table 3; Figure 5). The other nine significant descriptors in decreasing order of significance are C-006, GATS2e, nCaR, MATS7e, H-052, ATS5m, BEHv8, nROR and Ram (descriptors S. No. 23,17, 19, 16, 24, 13, 11, 21 and 4 in Table 3; Figure 5). Except, MATS7e, H-052, BEHv8, nROR and Ram all these descriptors are part of equations 6-9 and convey same inference in the PLS model as well. The descriptor MATS7e belongs to 2D-AUTO class. The positive regression coefficients of MATS7e (Moran autocorrelation of lag 7 weighted by atomic Sanderson electronegativities) and H-052 (H attached to C0(sp3) with one X attached to next C; ACF class descriptor) advocate molecular structures leading to a higher positive value of these descriptors are helpful to augment the activity. In other words structures resulting larger topological path 7 (weighted by atomic Sanderson electronegativities) and carrying CH-CX- type fragments are favorable for the activity. Similarly, a lower value of ramification index (TOPO class descriptor, Ram) and highest eigenvalue n.8 of Burden matrix weighted by atomic van der Waals volumes (BCUT class descriptor, BEHv8) are also beneficial to improve the activity. The topological descriptor Ram addresses the branching in the molecule. Its regression coefficient in the model suggest in favor of less branched molecular structures for the activity. The negative regression coefficient of BEHv8 for the activity suggests in

favor of limiting the number of eight-length topological paths in the structures. Additionally, nROR (a FUNC class descriptor) suggested a lower number of aliphatic ether functionality in a molecule for improved activity. In comparison to these ten descriptors, the remaining ones appear in lower order of significance to influence the activity of the compounds (Table 3; Figure 5). The structural inferences of all these descriptors to the activity are explained in Table 3. It is also observed that PLS model from the data set devoid of 25 descriptors is inferior in explaining the activity of the analogues.
To compare the models, the root mean square error (RMSE) and the predictive residual sum of squares (PRESS) for both, the training and the test set, has also been computed. These values, for the training set and the test set, are given in Table 6. The model with the smallest RMSE and PRESS is judged as the best predictor. Thus the model (Eq. 9) derived from CPMLR analysis has shown the best predictive power.

3.2 *Applicability domain*

On analyzing the model applicability domain (AD) in the Williams plot (Figure 6) of the model based on the whole data set (n=65, r= 0.838, s= 0.497, F=27.751, Q^2_{LOO} = 0.644), it has appeared that compounds **4** and **42** were identified as an obvious outlier for the 5-HT_6 binding affinity if the limit of normal values for the Y outliers (response outliers) was set as 2(standard deviation) units. None of the compounds was found to have leverage (h) value greater than the threshold leverage (h^* = 0.360). For both the training set and test set, the suggested model matches the high quality parameters with good fitting power and the capability of assessing external data. Furthermore, almost all of the compounds was within the applicability domain of the proposed model and were evaluated correctly.

4. Conclusions

The derived QSAR models have provided a rational approach for the development of new sulfonamide derivatives, based on medicinal chemistry guided hypothetical pharmacophore model, as serotonin 5-HT_6 receptor ligands. The descriptors identified in CP-MLR analysis have highlighted the role of molecular features accounting average Randic-type eigenvector-based index from adjacency matrix, VRA2, number of secondary aliphatic amines, nNHR, sum of topological distance between N and O, T(N..O), ring tertiary carbon atoms, nCrHR, and CH2RX type fragment, C-006, in a molecular structure to optimize the 5-HT_6 binding affinities of titled compounds. The PLS analysis has confirmed the dominance of information content of CP-MLR identified descriptors for modeling the activity when compared to those of leftover ones. The study has led to the identification of structural variations for the design of potential 5-HT_6 receptor ligands.

Acknowledgements

Authors are thankful to their Institutions for providing necessary facilities to complete this study. C.D.R.I. Communication No.

References

[1] A. Slassi, M. Isaac, and A. O'Brien, *Recent progress in 5-HT_6 receptor antagonists for the treatment of CNS diseases*, Expert Opin. Ther. Pat. 12 (2002), pp. 513-527.
[2] D. Hoyer, and G.R. Martin, *5-HT receptor classification and nomenclature: towards a harmonization with the human genome,* Neuropharmacology 36 (1997), pp. 419-428.
[3] D. Hoyer, D.E. Clarke, J.R. Fozard, P.R. Hartig, G.R. Martin, E. Mylecharane, P.R. Saxena, and P.P.A. Humphrey, *International union of pharmacology classification of receptors for 5-hydroxytryptamine (serotonin)*, Pharmacol. Rev. 46 (1994), pp. 157-204.
[4] R. Kohen, M.A. Metcalf, N. Khan, T. Druck, K. Huebner, J.E. Lachowicz, H.Y. Meltzer, D.R. Sibley, B.L. Roth, and M.W. Hamblin, *Cloning, characterization, and chromosomal localization of a human 5-HT_6 serotonin receptor*, J. Neurochem. 66 (1996), pp. 47-56.
[5] M. Ruat, E. Traiffort, J.-M. Arrang, J. Tardivel-Lacombe, J. Diaz, R. Leurs, and J.-C. Swartz, *A novel serotonin (5-HT_6) receptor: molecular cloning, localisation and stimulation of cAMP accumulation.* Biochem. Biophys. Res. Commun. 193 (1993), pp. 269-276.
[6] M. Sebben, H. Ansanay, J. Bockaert, and A. Dumuis, *5-HT_6 receptors positively coupled to adenylyl cyclase in striatal neurones in culture*, Neuroreport 5 (1994), pp. 2553-2557.
[7] L.A. Dawson, H.Q. Nguyen, and P. Li, *In vivo effects of the 5-HT_6 antagonist SB-271046 on striatal and frontal cortex extracellular concentrations of noradrenaline, dopamine, 5-HT, glutamate and aspartate*. Br. J. Pharmacol. 130 (2000), pp. 23-26.
[8] L.A. Dawson, H.Q. Nguyen, and P. Li, *The 5-HT_6 receptor antagonist SB-271046 selectively enhances excitatory neurotransmission in the rat frontal cortex and hippocampus*, Neuropsychopharmacology 25 (2001), pp. 662-668.
[9] L.A. Dawson, H.Q. Nguyen, and P. Li, *Potentiation of amphetamine-induced changes in dopamine and 5-HT by a 5-HT_6 receptor antagonist*. Brain Res. Bull. 59 (2003), pp. 513-521.

[10] L.P. Lacroix, L.A. Dawson, J.J. Hagan, and C.A. Heidbreder, *5-HT_6 receptor antagonist SB-271046 enhances extracellular levels of monoamines in the rat medial prefrontal cortex*, Synapse 51 (2004), pp. 158-164.
[11] C. Riemer, E. Borroni, B. Levet-Trafit, J.R. Martin, S. Poli, R.H. Porter, and M. Bös, *Influence of the 5-HT_6 receptor on acetylcholine release in the cortex: pharmacological characterization of 4-(2-bromo-6-pyrrolidin-1-ylpyridine-4-sulfonyl)phenylamine, a potent and selective 5-HT_6 receptor antagonist*, J. Med. Chem. 46 (2003), pp. 1273-1276.
[12] D.J. Healy, and J.H. Meador-Woodruff, *Ionotropic glutamate receptor modulation of 5- HT6 and 5-HT7 mRNA expression in rat brain*, Neuropsychopharmacology 21 (1999), pp. 341-351.
[13] Y. Minabe, Y. Shirayama, K. Hashimoto, C. Routledge, J.J. Hagan, and C.R. Ashby Jr., *Effect of the acute and chronic administration of the selective 5-HT6 receptor antagonist SB-271046 on the activity of midbrain dopamine neurons in rats: an in vivo electrophysiological study*, Synapse 52 (2004), pp. 20-28.
[14] B.L. Roth, S.C. Craigo, M.S. Choudhary, A. Uluer, F.J. Monsma Jr., Y. Shen, H.Y. Meltzer, and D.R. Sibley, *Binding of typical and atypical antipsychotic agents to 5- hydroxytryptamine-6 and 5-hydroxytryptamine-7 receptors*, J. Pharmacol. Exp. Ther. 268 (1994), pp. 1403-1410.
[15] F.J. Monsma, Y. Shen, R.P. Ward, M.W. Hamblin, and D.R. Sibley, *Cloning and expression of a novel serotonin receptor with high affinity for tricyclic psychotropic drugs*. Mol. Pharmacol. 43 (1993), pp. 320-327.
[16] T. Shinkai, O. Ohmori, H. Kojima, T. Terao, T. Suzuki, and K. Abe, *Association study of the 5-HT6 receptor gene in schizophrenia*, Am. J. Med. Genet. 88 (1999), pp. 120-122.
[17] S.J. Tsai, H.J. Chiu, Y.C. Wang, and C.J. Hong, *Association study of serotonin-6 receptor variant (C267T) with schizophrenia and aggressive behavior*, Neurosci. Lett. 271 (1999), pp. 135-137.
[18] I.R. Vogt, D. Shimron-Abarbanell, H. Neidt, J. Erdmann, S. Cichon, T.G. Schulze, D.J. Müller, W. Maier, M. Albus, M. Borrmann-Hassenbach, M. Knapp, M. Rietschel, P. Propping, and M.M. Nothen, *Investigation of the human serotonin 6 [5-HT_6] receptor gene in bipolar affective disorder and schizophrenia*, Am. J. Med. Genet. 96 (2000), pp. 217-221.
[19] O. Ohmori, T. Shinkai, H. Hori, and J. Nakamura, *Novel polymorphism in the 5¢- upstream region of the human 5-HT_6 receptor gene and schizophrenia*. Neurosci. Lett. 310 (2001), pp. 17-20.
[20] S.Z. East, P.W. Burnet, R.A. Leslie, J.C. Roberts, and P.J. Harrison, *5-HT_6 receptor binding sites in schizophrenia and following antipsychotic drug administration: autoradiographic studies with [^{125}I]SB-258585*, Synapse 45 (2002), pp. 191-199.
[21] W.H. Wu, S.J. Huo, C.Y. Cheng, C.J. Hong, and S.J. Tsai, *Association study of the 5- HT_6 receptor polymorphism (C267T) and symptomatology and antidepressant response in major depressive disorders*, Neuropsychobiology 44 (2001), pp. 172-175.
[22] D. Messina, G. Annesi, P. Serra, G. Nicoletti, A. Pasqua, F. Annesi, C. Tomaino, I.C. Ciro-Candiano, S. Carrideo, M. Caracciolo, P. Spadafora, M. Zappia, G. Savettieri, and A. Quattrone, *Association of the 5-HT_6 receptor gene polymorphism C267T with Parkinson's disease*, Neurology 58 (2002), pp. 828-829.
[23] J. Thome, W. Retz, M. Baader, B. Pesold, M. Hu, M. Cowen, N. Durany, G. Adler, F.A. Henn, and M. Rosler, *Association analysis of HTR6 and HTR2A polymorphisms in Sporadic Alzheimer's disease*, J. Neural Transm. 108 (2001), pp. 1175-1180.
[24] H.C. Liu, C.J. Hong, C.Y. Liu, K.N. Lin, S.J. Tsai, T.Y. Liu, C.W. Chi, and P.N. Wang, *Association analysis of the 5-HT_6 receptor polymorphism C267T with depression in patients with Alzheimer's disease*, Psychiatry Clin. Neurosci. 55 (2001), pp. 427-429.
[25] M. Garcia-Alloza, W.D. Hirst, C.P. Chen, B. Laheras, P.T. Francis, and M.J. Ramirez, *Differential involvement of 5-HT(1B/1D) and 5-HT6 receptors in cognitive and noncognitive symptoms in Alzheimer's disease*, Neuropsychopharmacology 29 (2004), pp. 410-416.
[26] A.J. Sleight, F.G. Böss, M. Bös, B. Levet-Trafit, C. Riemer, and A. Bourson, *Characterization of Ro 04-6790 and Ro 63-0563: potent and selective antagonists at human and rat 5-HT_6 receptors*, Br. J. Pharmacol. 124 (1998), pp. 556-562.
[27] S.M. Bromidge, A.M. Brown, S.E. Clarke, K. Dodgson, T. Gager, H.L. Grassam, P.M. Jeffrey, G.F. Joiner, F.D. King, D.N. Middlemiss, S.F. Moss, H. Newman, G. Riley, C. Routledge, and P. Wyman, *5-Chloro-N-(4-methoxy-3-piperazin-1-ylphenyl)-3-methyl-2- benzothiophenesulfonamide (SB-271046): A potent, selective, and orally bioavailable 5-HT_6 receptor antagonist*, J. Med. Chem. 42 (1999), pp. 202-205.
[28] R.A. Glennon, M. Lee, J.B. Rangisetty, M. Dukat, B.L. Roth, J.E. Savage, A. McBride, L. Rauser, L. Hufeisen, and D.K.H. Lee, *2-Substituted tryptamines: agents with selectivity for 5-HT_6 serotonin receptors*, J. Med. Chem. 43 (2000), pp. 1011-1018.
[29] M.G.N. Russell, R.J. Baker, L. Barden, M.S. Beer, L. Bristow, H.B. Broughton, M. Knowles, G. McAllister, S. Patel, and J.L. Castro, *N-Arylsulfonylindole derivatives as serotonin 5-HT_6 receptor ligands*. J. Med. Chem. 44 (2001), pp. 3881-3895.
[30] J. Holenz, R. Mercè, J.L. Díaz, X. Guitart, X. Codony, A. Dordal, G. Romero, A. Torrens, J. Mas, B. Andaluz, S. Hernández, X. Monroy, E. Sánchez, E. Hernández, R. Pérez, R. Cubí, O. Sanfeliu, and H. Buschmann, *Medicinal chemistry driven approaches toward novel and selective serotonin 5-HT_6 receptor ligands*, J. Med. Chem. 48 (2005), pp. 1781-1795.

[31] A. Torrens, J. Mas, A. Dordal, and M.A. Fisas, Compuestos sulfonamı´dicos derivados de benzoxazinona, supreparacio´n y uso como medicamentos. Spanish Patent Application ES 2003-01812, July 30, 2003.

[32] R. Todeschini, M. Lasagni, and E. Marengo, *New molecular descriptors for 2D and 3D structures theory: Part 1*, J. Chemometrics 8 (1994), pp. 263–272.

[33] J. Galvez, R. Garcia-Domenech, J.V. de Julian-Ortiz, and R. Soler, *Topological approach to drug design*, J. Chem. Inf. Comput. Sci. 35 (1995), pp. 272-284.

[34] J. Devillers, and A.T. Balaban, *Topological indices and related descriptors in QSAR and QSPR*, Gordon and Breach Science Publishers. 1999, 296.

[35] R. Gozalbes, J.P. Doucet, and F. Derouin, *Application of topological descriptors on QSAR and drug design: history and new trends*, Curr. Drug Targets Infect. Disord. 2 (2002), pp. 93-102.

[36] M.P. González, C. Terán, M. Teijeira, P. Besada, and M.J. González-Moa, *BCUT descriptors to predicting affinity toward A3 adenosine receptors*, Bioorg. Med. Chem. Lett. 15 (2005), pp. 3491-3495.

[37] M.C. Bagchi, D. Mills, and S.C. Basak, *Quantitative structure-activity relationship (QSAR) studies of quinolone antibacterials against M. fortuitum and M. megmatis using theoretical molecular descriptors*, J. Mol. Model. 13 (2007), pp. 111–120.

[38] L. Saíz-Urra, M.P. González, and M. Teijeira, *2D-autocorrelation descriptors for predicting cytotoxicity of naphthoquinone ester derivatives against oral human epidermoid carcinoma*, Bioorg. Med. Chem. 15 (2007), pp. 3565-3571.

[39] A.M. Helguera, R.D. Combes, M.P. González, and M.N. Cordeiro, *Applications of 2D descriptors in drug design: a DRAGON tale*, Curr. Top. Med. Chem. 8 (2008), pp. 1628- 1655.

[40] R. Kühne, R.-U. Ebert, and G. Schüürmann, *Chemical domain of QSAR models from atom-centered fragments*, J. Chem. Inf. Model. 49 (2009), pp. 2660–2669.

[41] Dragon Software (version 1.11-2001) by R. Todeschini, and V. Consonni, Milano, Italy.

[42] B.L. Roth, S.C. Craigo, M.S. Choudhary, A. Uluer, F.J. Jr. Monsma, Y. Shen, H.Y. Meltzer and D.R. Sibley, *Binding of typical and atypical antipsychotic agents to 5- hydroxytryptamine-6 and 5-hydroxytryptamine-7 receptors*. J. Pharmacol. Exp. Ther. 268 (1994), pp. 1403-1410.

[43] P.J. Munson, and D. Robbard, *LIGAND: A versatile, computerized approach for characterization of ligand-binding systems*. Anal. Biochem. 107 (1980), pp. 220-239.

[44] Chemdraw Ultra 6.0 and Chem3D Ultra, Cambridge Soft Corporation, Cambridge, USA.

[45] Y.S. Prabhakar, *A combinatorial approach to the variable selection in multiple linear regression: analysis of Selwood et al. data set– a case study*. QSAR Comb. Sci. 22 (2003), pp. 583-595.

[46] S. Wold, *Cross-validatory estimation of the number of components in factor and principal components models*. Technometrics 20 (1978), pp. 397-405.

[47] N. Kettaneh, A. Berglund, and S. Wold, *PCA and PLS with very large data sets*, Comput. Stat. Data Anal. 48 (2005), pp. 69-85.

[48] L. Stahle, and S. Wold, *Multivariate data analysis and experimental design*, In: G.P. Ellis, and W.B. West, eds. Biomedical Research, Progress in Medicinal Chemistry, Elsevier Science Publishers, BV, vol. 25 (1998), pp. 291-338.

[49] S. Sharma, Y.S. Prabhakar, P. Singh, and B.K. Sharma, *QSAR study about ATP-sensitive potassium channel activation of cromakalim analogues using CP-MLR approach*. Eur. J. Med. Chem. 43 (2008), pp. 2354-2360.

[50] S. Sharma, B.K. Sharma, S.K. Sharma, P. Singh, and Y.S. Prabhakar, *Topological descriptors in modeling the agonistic activity of human A_3 adenosine receptor ligands: the derivatives of 2-Chloro-N^6-substituted-4´-thioadenosine-5´-uronamide*. Eur. J. Med. Chem. 44 (2009), pp. 1377-1382.

[51] P. Singh, R. Kumar, B.K. Sharma, and Y.S. Prabhakar, *Topological descriptors in modeling malonyl coenzyme A decarboxylase inhibitory activity: N-Alkyl-N- (1,1,1,3,3,3-Hexafluoro-2-hydroxypropylphenyl)amide derivatives*. J. Enzyme Inhib. Med. Chem. 24 (2009), pp. 77-85.

[52] B.K. Sharma, S.K. Sharma, P. Singh, S. Sharma, and Y.S. Prabhakar, *Modeling of vascular endothelial growth factor receptor 2 (VEGFR2) kinase inhibitory activity of 2-Anilino-5-aryloxazoles using chemometric tools*. J. Enzyme Inhib. Med. Chem. 24 (2009), pp. 86-93.

[53] B.K. Sharma, P. Pilania, K. Sarbhai, P. Singh, and Y.S. Prabhakar, *Chemometric descriptors in modeling the carbonic anhydrase inhibition activity of sulfonamide and sulfamate derivatives*, Mol Divers (2009), in press, doi: 10.1007/s11030-009-9181-5

[54] S.-S. So, and M. Karplus, *Three-dimensional quantitative structure-activity relationship from molecular similarity matrices and genetic neural networks. 1. Method and validation*. J. Med. Chem. 40 (1997), pp. 4347-4359.

[55] Y.S. Prabhakar, V.R. Solomon, R.K. Rawal, M.K. Gupta, and S.B. Katti, *CP-MLR/PLS directed structure-activity modeling of the HIV-1 RT inhibitory activity of 2,3-Diaryl- 1,3-thiazolidin-4-ones*. QSAR Comb. Sci. 23 (2004), pp. 234-244.

[56] H. Akaike, *Information theory and an extension of the minimum likelihood principle*, In: B.N. Petrov, and F. Csaki, eds., Second International Symposium On Information Theory. Budapest, Akademiai Kiado, 1973, pp. 267-281.

[57] H. Akaike, *A new look at the statistical identification model*. IEEE Trans. Autom. Control AC-19 (1974), pp. 716-723.

[58] H. Kubinyi, *Variable selection in QSAR studies. I. An evolutionary algorithm*. Quant. Struct.-Act. Relat. 13 (1994), pp. 285-294.
[59] H. Kubinyi, *Variable selection in QSAR studies. II. A highly efficient combination of systematic search and evolution*. Quant. Struct.-Act. Relat. 13 (1994), pp. 393-401.
[60] J. Friedman, In: Technical Report No. 102. Laboratory for Computational Statistics. Stanford University, Stanford. 1990.
[61] D.M. Hawkins, S.C. Basak, and D. Mills, *Assessing model fit by cross-validation*, J. Chem. Inf. Comput. Sci. 43 (2003), pp. 579-586.
[62] P. Gramatica, *Principles of QSAR models validation: internal and external*, QSAR Comb. Sci. 26 (2007), pp. 694-701.
[63] R.L. Lipnick, *Outliers: their origin and use in the classification of molecular mechanisms of toxicity*, Sci. Total Environ. 109 (1991), pp. 131–153.

Table 1 Structural variations (Figure 1 for general structure) and experimental 5-HT_6 binding affinities of sulfonamide derivatives

Structure 1(A)[a]					
Cpd	n	NR_1R_2	R_3	R_4	K_i(nM)[b]
1	2	NMe_2	H	5-Chloro-3-methyl-benzo[b]thiophen-2-yl	57.7
2[c]	2	NMe_2	H	1-Naphthyl	64.7
3[c]	2	NMe_2	H	2-Naphthyl	14.9
4	2	NMe_2	H	6-Chloro-imidazo[2,1-b]thiazol-5-yl	1.9
5	2	NMe_2	H	1-Naphthyl	94.2
6	2	NMe_2	H	2-Naphthyl	57.7
7	2	NMe_2	H	Benzyl	-[d]
8	2	NMe_2	Me	6-Chloro-imidazo[2,1-b]thiazol-5-yl	74.6
9[c]	2	NMe_2	Me	1-Naphthyl	123.6
10[c]	2	NMe_2	Me	2-Naphthyl	85.8
11[c]	2	NEt_2	H	6-Chloro-imidazo[2,1-b]thiazol-5-yl	97.3
12	2	NEt_2	H	1-Naphthyl	116.2
13	2	NEt_2	H	2-Naphthyl	140.5
14	2	NEt_2	H	p-Biphenyl	-[d]
15	2	1-Pyrrolidinyl	H	6-Chloro-imidazo[2,1-b]thiazol-5-yl	9.8
16	2	1-Piperidinyl	H	6-Chloro-imidazo[2,1-b]thiazol-5-yl	31.4
17	3	1-Piperidinyl	H	6-Chloro-imidazo[2,1-b]thiazol-5-yl	-[d]
18	2	NMe_2	H	6-Chloro-imidazo[2,1-b]thiazol-5-yl	55.2
19	2	NMe_2	H	5-Chloro-3-methyl-benzo[b]thiophen-2-yl	90.2
20[c]	2	NMe_2	H	1-Naphthyl	19.8
21	2	NMe_2	H	2-Naphthyl	41.2
22	2	NMe_2	H	1-Ethyl-naphthalene	74.5
23	2	NMe_2	Me	5-Chloro-3-methyl-benzo[b]thiophen-2-yl	-[d]
24	2	1-Pyrrolidinyl	H	6-Chloro-imidazo[2,1-b]thiazol-5-yl	52.2
25[c]	2	1-Pyrrolidinyl	H	5-Chloro-3-methyl-benzo[b]thiophen-2-yl	108.8
26	2	1-Pyrrolidinyl	H	1-Naphthyl	18.6
27	2	1-Pyrrolidinyl	H	2-Naphthyl	100.8
28	2	1-Pyrrolidinyl	H	1-Ethyl-naphthalene	194.8
Structure 1(B)					
Cpd	n	NR_1R_2		R_3	K_i(nM)[b]
29	1	NMe_2		5-Chloro-3-methyl-benzo[b]thiophen-2-yl	5.3
30	2	NMe_2		6-Chloro-imidazo[2,1-b]thiazol-5-yl	2.2
31[c]	2	NMe_2		5-Chloro-3-methyl-benzo[b]thiophen-2-yl	0.1
32	2	NMe_2		1-Naphthyl	0.8
33	2	NMe_2		5-Chloro-2-Naphthyl	0.2
34	2	NMe_2		p-Biphenyl	0.4
35[c]	2	NEt_2		5-Chloro-3-methyl-benzo[b]thiophen-2-yl	0.3
36	2	NEt_2		1-Naphthyl	3.5
37	2	NEt_2		2-Naphthyl	5.3
38	2	NEt_2		5-Chloro-1- Naphthyl	3.3
39	2	NEt_2		3,5-dichlorophenyl	9.3

40	2	NEt_2	p-Biphenyl	1.0
41[e]	2	NEt_2	2-Naphthyl	40.8
42	3	NEt_2	2-Naphthyl	163.0
43	2	NPr_2	5-Chloro-3-methyl-benzo[b]thiophen-2-yl	0.5
44	2	NPr_2	2-Naphthyl	3.9
45[c]	2	NBu_2	1-Naphthyl	7.2
46	2	1-Pyrrolidinyl	5-Chloro-3-methyl-benzo[b]thiophen-2-yl	4.0
47	2	1-Pyrrolidinyl	1-Naphthyl	4.5
48	2	1-Pyrrolidinyl	2-Naphthyl	4.3
49	2	1-Morpholinyl	5-Chloro-3-methyl-benzo[b]thiophen-2-yl	15.1
50	2	1-Morpholinyl	1-Naphthyl	163.5
51[c]	2	1-Morpholinyl	2-Naphthyl	132.0
52	2	1-Morpholinyl	8-Quinolyl	117.0
Structure 1(C)				
Cpd		NR_1R_2	R_3	K_i(nM)[b]
53		NMe_2	5-Chloro-3-methyl-benzo[b]thiophen-2-yl	-[d]
54		NMe_2	6-Chloro-imidazo[2,1-b]thiazol-5-yl	18.4
55[c]		NEt_2	6-Chloro-imidazo[2,1-b]thiazol-5-yl	224.4
56		NEt_2	1-Naphthyl	-[d]
57		NEt_2	8-Quinolyl	-[d]
Structure 1(D)				
Cpd			R_1	K_i(nM)[b]
58[c]			5-Chloro-3-methyl-benzo[b]thiophen-2-yl	1.0
59			1-Naphthyl	4.7
60[c]			5-Chloro-1-Naphthyl	9.6
61			2-Chloro-5-thienyl	24.3
62			8-Quinolyl	21.2
63			p-Biphenyl	6.8
Structure 1(E)				
Cpd		R_1	R_2	K_i(nM)[b]
64		H	3-Benzothienyl	125.8
65		Me	3-Benzothienyl	124.5
66		H	5-Chloro-3-methyl-benzo[b]thiophen-2-yl	107.4
67		H	1-Naphthyl	51.7
68		Me	1-Naphthyl	246.2
69		Me	N,N-Dimethyl-naphthalen-5-yl-1-amine	336.0
70		H	8-Quinolyl	151.9
71[c]		Me	8-Quinolyl	165.9
72		Me	5-Bromo-2-methoxyphenyl	-[d]
73		Me	2,5-Dimethoxyphenyl	131.3

[a]Compounds 1-3, 4-17 and 18-28 are 4-sulfonamides, 5-sulfonamides and 6-sulfonamides, respectively; [b]Taken from reference [30]; [c]Compound in test set; [d]Compound with uncertain activity and not included in the study; [e]N-methylindole derivative.

Table 2 Descriptor classes used for the quantification of binding affinities of 5-HT_6 receptor ligands.

Descriptor class (acronyms)[a]	Definition and scope
	Dimensionless or 0D descriptors; independent from molecular connectivity and conformations
Constitutional (CONST)	
Topological (TOPO)	2D-descriptor from molecular graphs and independent conformations
Molecular walk counts (MWC)	2D-descriptors representing self-returning walk counts of different lengths
Modified Burden eigenvalues (BCUT)	2D-descriptors representing positive and negative eigenvalues of the adjacency matrix, weights the diagonal elements and atoms
Galvez topological charge indices (GALVEZ)	2D-descriptors representing the first 10 eigenvalues of corrected adjacency matrix
2D- autocorrelations (2D-AUTO)	Molecular descriptors calculated from the molecular graphs by summing the products of atomic weights of the terminal atoms of all the paths of the considered path length (the

	lag)
Functional groups (FUNC)	Molecular descriptors based on the counting of the chemical functional groups
Atom centered fragments (ACF)	Molecular descriptors based on the counting of 120 atom centered fragments, as defined by Ghose-Crippen
Empirical (EMP)	1D- descriptors represent the counts of nonsingle bonds, hydrophilic groups and ratio of the number of aromatic bonds and total bonds in an H-depleted molecule
Properties (PROP)	1D- descriptors representing molecular properties of a molecule

[a]Ref [41].

Table 3 Physical meaning of descriptors identified from five parameter CP-MLR models for the binding affinity of sulfonamide derivatives and their MLR-like coefficients from PLS model.

Class	Descriptor			Significance to activity	
	Sno	Symbol	Structural Information[a]	Reg. Coeff. (fc)[b]	Order[c]
CONST	1	Ms	mean electrotoplogical state	-0.286(-0.014)	21
	2	RBF	rotatable bond fraction	0.730(0.017)	19
	3	nDB	number of double bonds	-0.055(-0.025)	16
TOPO	4	Ram	ramification index; addresses the branching in the molecule	-0.078(-0.040)	10
	5	MAXDN	maximal electrotopological negative variation	0.587(0.031)	13
	6	TIE	E-state topological parameter	-0.0004(-0.009)	25
	7	TIC1	total information content of 1-order neighborhood symmetry	0.002(0.014)	22
	8	VRA2	average Randic-type eigenvector-based index from adjacency matrix	0.005(0.019)	18
	9	T(N..O)	sum of topological distances between N and O in molecule	-0.004(-0.029)	14
BCUT	10	BEHm5	5^{th} highest eigenvalue of Burden matrix weighted by atomic masses (*m*)	0.528(0.033)	11
	11	BEHv8	8^{th} highest eigenvalue of Burden matrix weighted by atomic VDW volumes	-1.069(-0.048)	8
GALVEZ	12	JGI2	mean topological charge index of order (path) 2	-1.788(-0.013)	23
2DAUTO	13	ATS5m	Moreau-Broto 2D-Auto of lag (path) 5 weighted by *m*	-0.013(-0.052)	7
	14	MATS2m	Moran 2D-Auto of lag 2 weighted *m*	-4.355(-0.017)	20
	15	MATS5e	Moran 2D-Auto of lag 5 weighted by atomic electronegativities (*e*)	0.524(0.031)	12
	16	MATS7e	Moran 2D-Auto of lag 7 weighted by *e*	1.249(0.066)	5
	17	GATS2e	Geary 2D-Auto of lag 2 weighted by *e*	-1.680(-0.076)	3
FUNC	18	nCrHR	No. of ring tertiary C(sp3)	-0.108(-0.028)	15
	19	nCaR	No. of substituted aromatic C(sp2)	0.142(0.074)	4
	20	nNHR	No. of secondary aliphatic amines	0.407(0.122)	1
	21	nROR	No. of ethers (aliphatic)	-0.360(-0.045)	9
ACF	22	C-005	CH3X; i.e. methyl group attached to any electronegative atom (O, N, S, P, Se, halogens)	0.025(0.012)	24
	23	C-006	CH2RX, where R represents any group linked through carbon and X any electronegative atom	-0.180(-0.107)	2
	24	H-052	H attached to C0(sp3) with one electronegative atom (X) attached to next C	0.037(0.055)	6
	25	O-058	=O (oxygen bonded through double bond)	-0.093(-0.022)	17

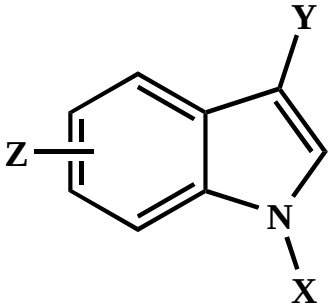
[a]Also see ref.[41]; [b]MLR like regression coefficient of three-component PLS model; (fc) is fraction contribution of the regression coefficient to the activity; the constant term of PLS model is 7.686; number of compounds are 48. PLS regression and validation statistics: r = 0.891, s = 0.383, F = 56.596, Q^2_{LOO} = 0.750, Q^2_{L5O} = 0.744, FIT = 2.979, LOF = 0.176, AIC = 0.174, r^2_{Test} = 0.690; [c]Order indicates the order of their significance in the PLS model.

Table 4 Observed, calculated and residual activity of test set compounds.

	pK_i[a]				
		Eq. (9)		PLS	
Cpd.[b]	Obsd.	Calc.	Res.[c]	Calc.	Res.[c]
2	7.19	7.61	-0.42	7.23	-0.04
3	7.83	7.61	0.22	7.09	0.74
9	6.91	7.51	-0.60	7.61	-0.70
10	7.07	7.54	-0.47	7.50	-0.43
11	7.01	7.32	-0.30	7.26	-0.20
20	7.70	7.62	0.09	7.43	0.27
25	6.96	7.41	-0.45	7.78	-0.82
31	10.00	9.00	1.00	8.84	1.16
35	9.52	8.66	0.86	8.78	0.74
45	8.14	9.04	-0.90	8.30	-0.16
51	6.88	7.23	-0.35	6.82	0.06
55	6.65	7.09	-0.44	7.42	-0.77
58	9.00	8.01	0.99	8.51	0.49
60	8.02	7.92	0.10	8.10	-0.08
71	6.78	6.81	-0.03	6.92	-0.14

[a]On molar basis; [b]As in Table 1; [c]The difference between observed and calculated pK_i valu

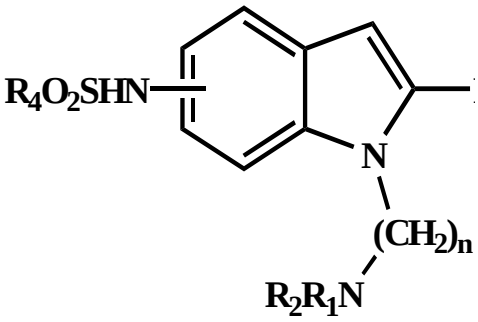
Table 5 The structures and predicted activity of compounds based on QSAR model Eq. (9).



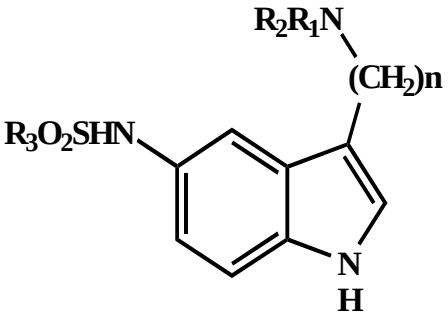
Cpd.	X	Y	Z	pK_i
1	$(CH_2)_2NBu_2$	H	-5-$NHSO_2$-(6-Chloro-imidazo[2,1-b]thiazol-5-yl)	9.66
2	$(CH_2)_2NBu_2$	H	-6-$NHSO_2$-(6-Chloro-imidazo[2,1-b]thiazol-5-yl)	9.51
3	$(CH_2)_3NBu_2$	H	-4-$NHSO_2$-(5-Chloro-3-methyl-benzo[b]thiophen-2-yl)	10.40
4	$(CH_2)_3NBu_2$	H	-5-$NHSO_2$-(5-Chloro-3-methyl-benzo[b]thiophen-2-yl)	11.72
5	$(CH_2)_3NBu_2$	H	-6-$NHSO_2$-(5-Chloro-3-methyl-benzo[b]thiophen-2-yl)	11.46
6	H	$(CH_2)_2NBu_2$	-5-$NHSO_2$-(5-Chloro-3-methyl-benzo[b]thiophen-2-yl)	10.38
7	H	$(CH_2)_2NBu_2$	-5-$NHSO_2$-(6-Chloro-imidazo[2,1-b]thiazol-5-yl)	10.70

Table 6 The RMSE and PRESS values for training and test set, and predictive R^2

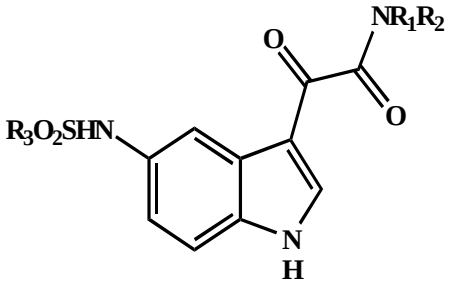
Model	$RMSE_{Training}$	$RMSE_{Test}$	$PRESS_{Training}$	$PRESS_{Test}$	r^2_{Test}
Eq. (8)	0.4422	0.5985	9.7782	5.3730	0.655
Eq. (9)	0.3567	0.5757	6.1084	4.9718	0.680
PLS	0.3671	0.5674	6.4675	4.8288	0.690



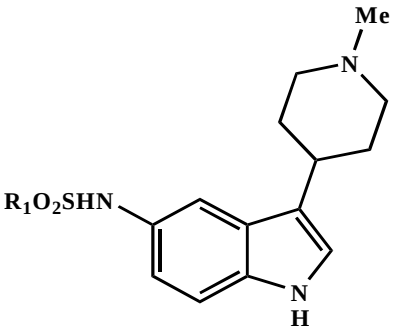
A



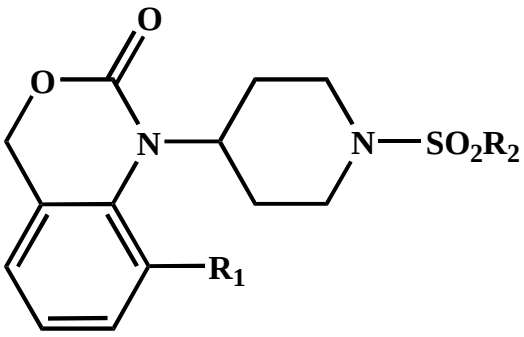
B



C



D



E

Figure 1 Structures of the indolyl (A to D) and piperidinyl (E) sulfonamides.

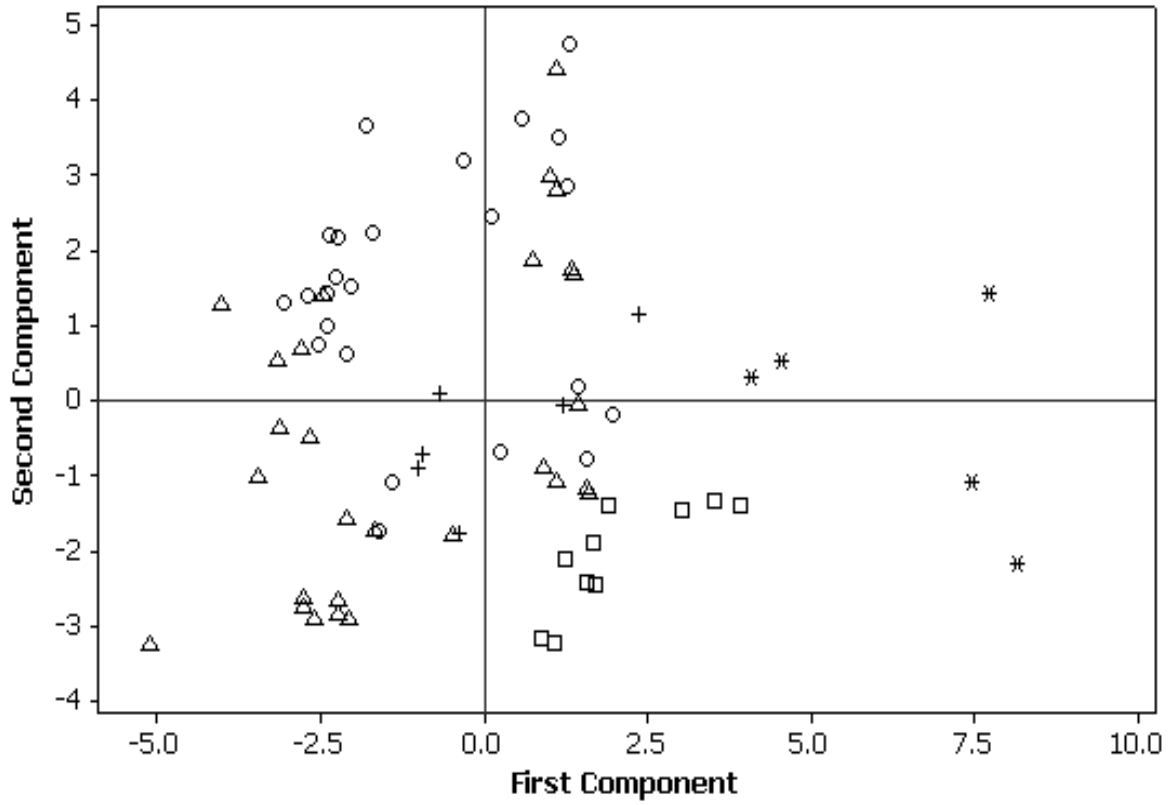


Figure 2. The plot of structure classes A to E (Fig.1A to 1E; Table 1) in terms of first two principal components from the 25 identified descriptors (Table 3). The Locations of structure classes A to E are shown by symbols Δ, O, ∗, +, and □, respectively.

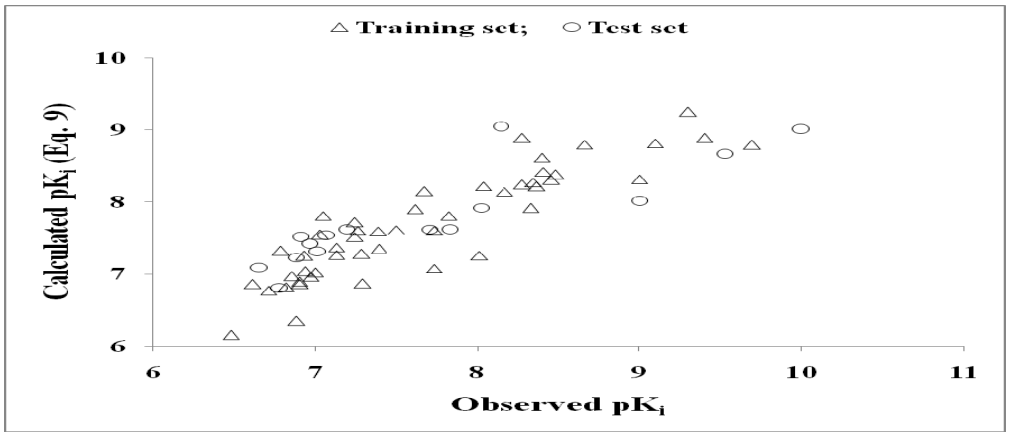


Figure 3 Plot of calculated (from Eq. 9) and observed pK_i values. The calculated pK_is for training and test set compounds are shown by Δ and O, respectively.

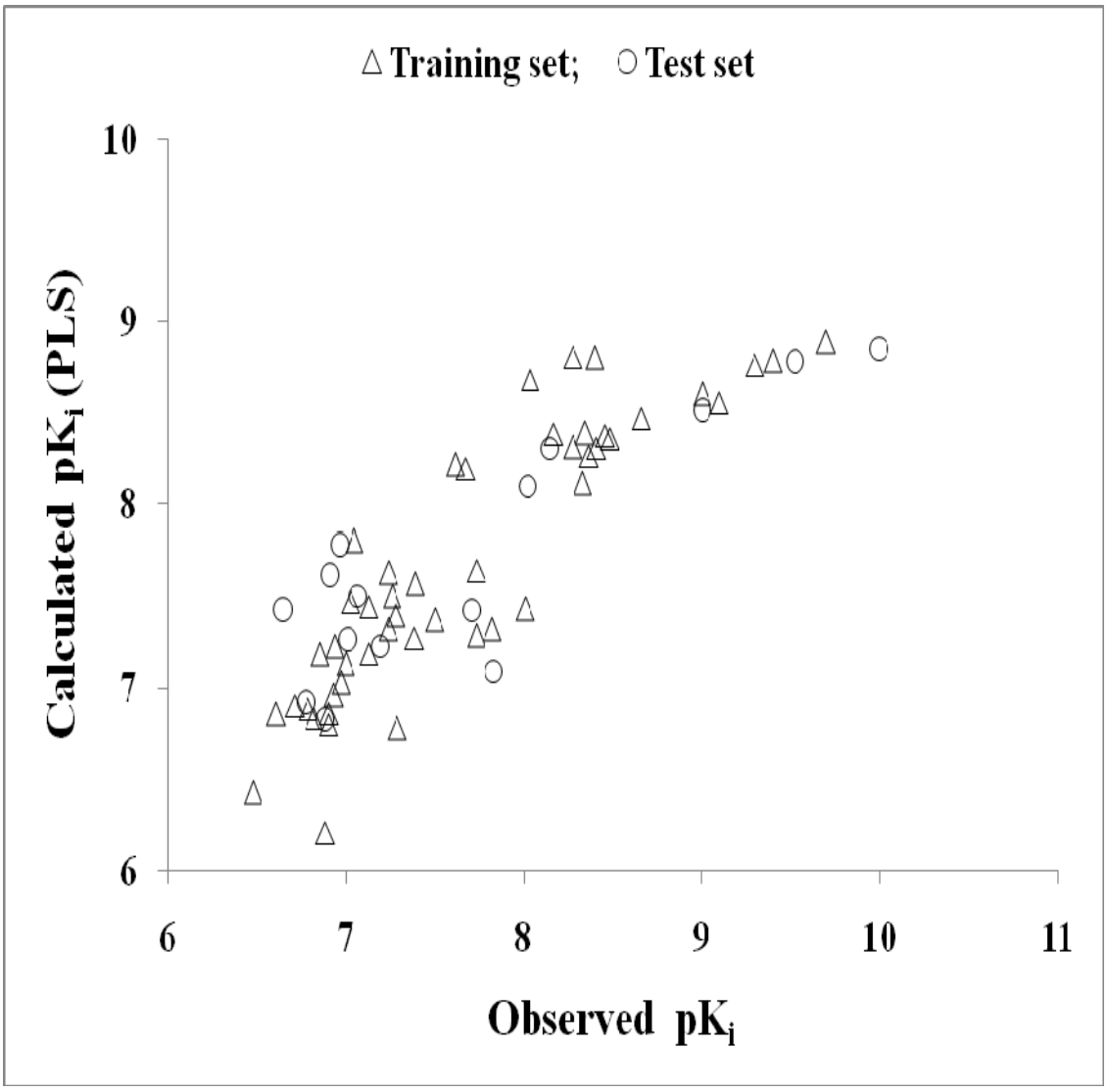


Figure 4 Plot of calculated (through PLS analysis) and observed pK_i values. The calculated pK_is for training and test set compounds are shown by Δ and O, respectively.

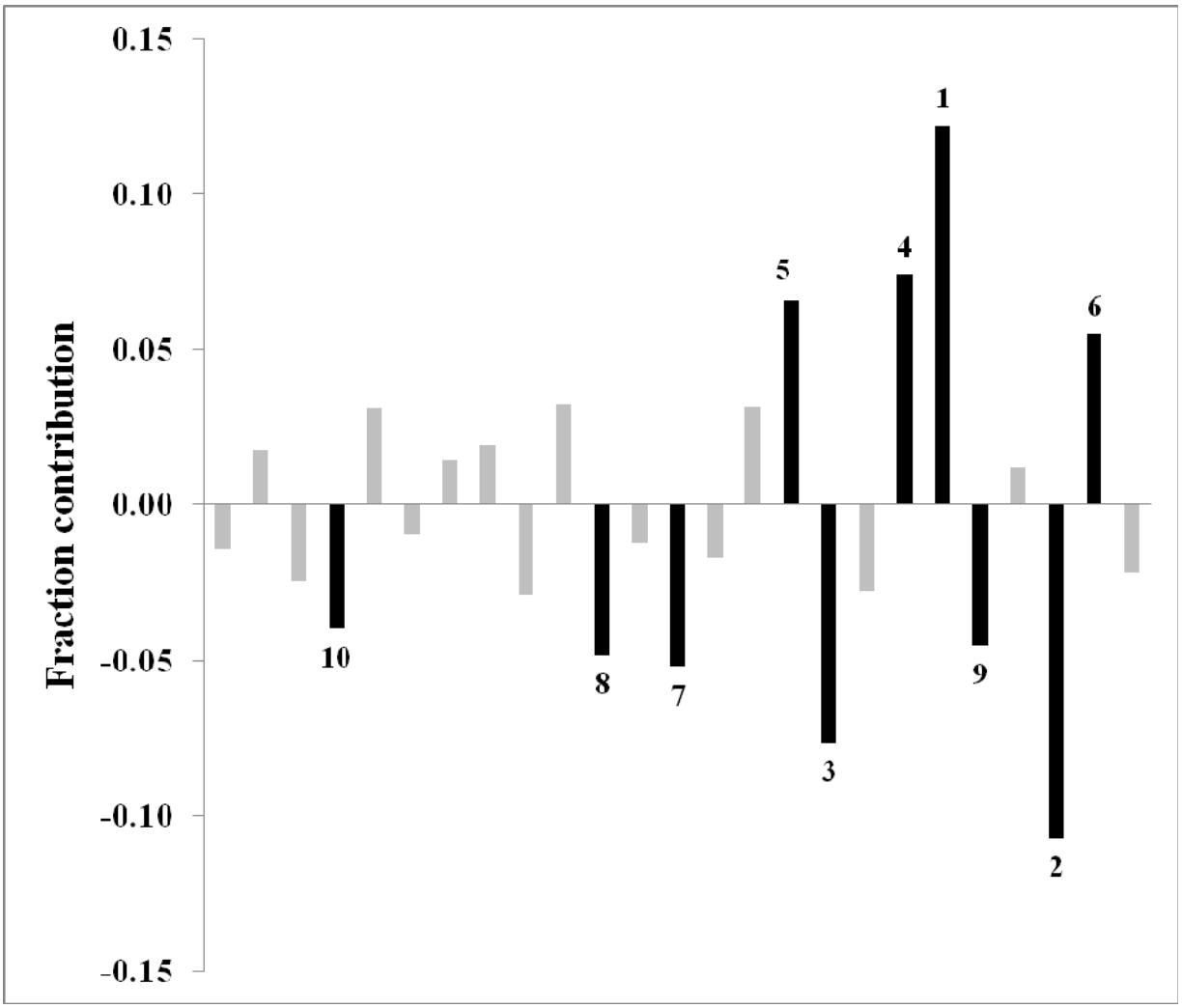


Figure 5 Plot of fraction contribution of MLR-like PLS coefficients (normalized) of the 25 descriptors (Table 3) to the activity. The ten most significant descriptors are identified by black shaded lines. The numbers on the bars correspond to the descriptor serial numbers in Table 3.

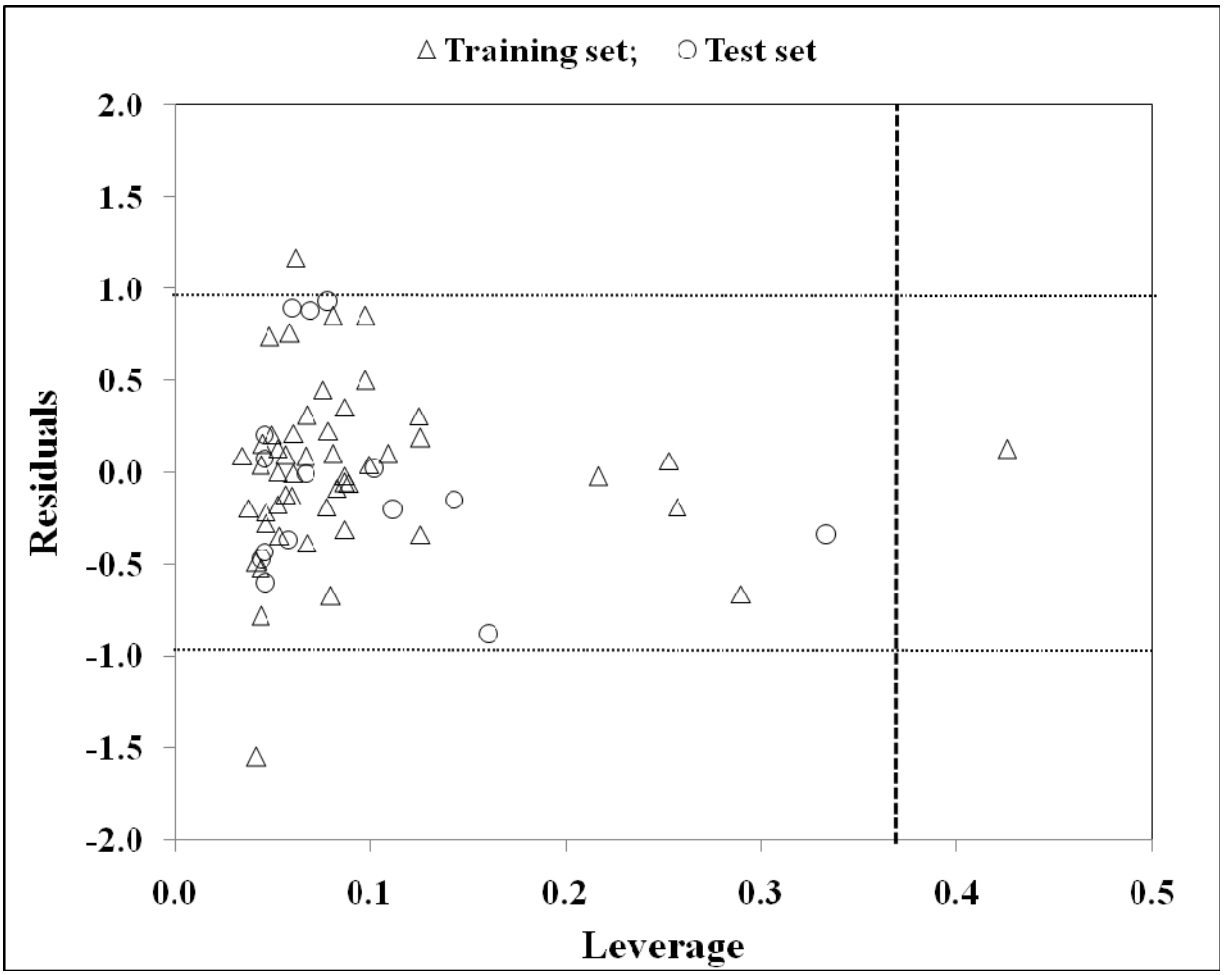


Figure 6 Williams plot for the training set and external prediction set for 5-HT_6 binding affinities of sulfonamide derivatives. The residuals for training and test set compounds are shown by Δ and O, respectively. The horizontal dotted line refers to the residual limit (±2× standard deviation) and the vertical dotted line represents threshold leverage h* (= 0.360).