



Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

A QSAR study on 5-HT₆ receptor antagonists: the 3,4-dihydro-2H-benzo[1,4]oxazines

Manju Choudhary, Pradeep Pilonia and Brij Kishore Sharma*

Department of Chemistry, Government P. G. College, Bundi, Rajasthan, India

ABSTRACT

This study has provided a rational approach for the development of new 3,4-dihydro-2H-benzo[1,4]oxazine derivatives as potential 5-HT₆ antagonists. The descriptors identified in CP-MLR analysis have highlighted the role of atomic van der Waals volume in terms of BEHv6, path/walk ratio 3 and 4- Randic shape index (PW3 and PW4), number of rotatable bonds (RBN) and superpendentic index (SPI) to explain the biological actions of 3,4-dihydro-2H-benzo[1,4]oxazine derivatives as potential 5-HT₆ antagonists. Applicability domain analysis revealed that the suggested model matches the high quality parameters with good fitting power and the capability of assessing external data and almost all of the compounds was within the applicability domain of the proposed model and were evaluated correctly.

Key words: QSAR, 3,4-dihydro-2H-benzo[1,4]oxazines, 5-HT₆ antagonists, binding affinity, combinatorial protocol in multiple linear regression (CP-MLR).

INTRODUCTION

5-Hydroxytryptamine (5-HT; serotonin) plays a vital role in various central nervous system (CNS) disorders [1]. At present, 14 distinct 5-HT receptor subclasses have been reported in the 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-HT₁₋₇) [3]. In 5-HT receptors, all except the 5-HT₃ receptor are G protein-coupled receptors. The 5-HT₆ 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 the amygdale, hypothalamus, substantia nigra, cerebellum, or cerebral cortex. The distribution and high affinity of 5-HT₆ receptors for several tricyclic antipsychotics suggest that it may be involved in memory disorders, psychosis, depression and appetite control [7-9]. Over the past few years, 5-HT₆ receptor has become an important therapeutic target [7] because antagonists of it increase extracellular concentrations of neurotransmitters important for cognition. A new series of 5-HT₆ receptor antagonists, the 3,4-dihydro-2H-benzo[1,4]oxazine derivatives have recently been synthesized and evaluated for binding activities for the 5-HT₆ receptor by Zhao et al. [10]. In view of the importance of 5-HT₆ receptor antagonists in the clinical management of several disorders, a quantitative structure-activity relationship is attempted on the 5-HT₆ receptor binding affinity of these benzoxazine derivatives. The present study is aimed at rationalizing the substituent variations of these analogues to provide insight for the future endeavors.

EXPERIMENTAL SECTION

Chemical structure database and biological activity

This study comprises a chemical structure database of twenty eight 3,4-dihydro-2*H*-benzo[1,4]oxazine derivatives, reported by Zhao *et al.* [10]. The *in vitro* binding activity of these derivatives was determined by displacement of [³H]LSD in HEK293 cells expressing the recombinant human 5-HT₆ receptor. The structural variations and the binding activities on molar basis of titled compounds have been given in Table 1. The reported activity data has been used for subsequent QSAR analyses as the response variables. For the purpose of modeling all 28 analogues have been divided into training and test sets. Out of the 28 analogues, one fourth compounds (7) have been placed in the test set for the validation of derived models. The training and test set compounds are also listed in Table 1.

Theoretical molecular descriptors

The structures of the compounds under study have been drawn in 2D ChemDraw [11]. The drawn structures were then converted into 3D modules using the default conversion procedure implemented in the CS Chem3D Ultra. The energy of these 3D-structures was minimized in the MOPAC module using the AM1 procedure for closed shell systems. This will ensure a well defined conformer relationship among the compounds of the study. All these energy minimized structures of respective compounds have been ported to DRAGON software [12] for the computation of descriptors for the titled compounds (Table 1). This software offers several hundreds of descriptors from different perspectives corresponding to 0D-, 1D-, and 2D-descriptor modules. The outlined modules comprised of ten different classes, namely, the constitutional (CONST), the topological (TOPO), the molecular walk counts (MWC), the BCUT descriptors (BCUT), the Galvez topological charge indices (GALVEZ), the 2D autocorrelations (2D-AUTO), the functional groups (FUNC), the atom-centered fragments (ACF), the empirical descriptors (EMP), and the properties describing descriptors (PROP). For each of these classes the DRAGON software computes a large number of descriptors which are characteristic to the molecules under multi-descriptor environment. The definition and scope of these descriptor's classes is given in Table 2. The combinatorial protocol in multiple linear regression (CP-MLR) [13] procedure has 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 has reduced the total dataset of the compounds from 473 to 109 descriptors as relevant ones for the binding activity. A brief description of the computational procedure is given below.

Model development

The CP-MLR is a 'filter' based variable selection procedure for model development in QSAR studies [13]. Its procedural aspects and implementation are discussed in some of our recent publications [14-18]. It involves selected subset regressions. In this procedure a combinatorial strategy with appropriately placed 'filters' has been interfaced with MLR to result in the extraction of diverse structure-activity models, each having unique combination of descriptors from the dataset under study. In this, the contents and number of variables to be evaluated are mixed according to the predefined confines. Here the 'filters' are significance evaluators of the variables in regression at different stages of model development. Of these, filter-1 is set in terms of inter-parameter correlation cutoff criteria for variables to stay as a subset (filter-1, default value 0.3 and upper limit ≤ 0.79). In this, if two variables are correlated higher than a predefined cutoff value the respective variable combination is forbidden and will be rejected. The second filter is in terms of t-values of regression coefficients of variables associated with a subset (filter-2, default value 2.0). Here, if the ratio of regression coefficient and associated standard error of any variable is less than a predefined cutoff value then the variable combination will be rejected. Since successive additions of variables to multiple regression equation will increase successive multiple correlation coefficient (r) values, square-root of adjusted multiple correlation coefficient of regression equation, \bar{r} , has been used to compare the internal explanatory power of models with different number of variables. Accordingly, a filter has been set in terms of predefined threshold level of \bar{r} (filter-3, default value 0.71) to decide the variables' 'merit' in the model formation. Finally, to exclude false or artificial correlations, the external consistency of the variables of the model have been addressed in terms of cross-validated R^2 or Q^2 criteria from the leave-one-out (LOO) cross-validation procedure as default option (filter-4, default threshold value $0.3 \leq Q^2 \leq 1.0$). All these filters make the variable selection process efficient and lead to unique solution. 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 \bar{r} value of the preceding optimum model as the new threshold for next generation.

Model validation

In this study, the data set is divided into training set for model development and test set for external prediction. Goodness of fit of the models was assessed by examining the multiple correlation coefficient (r), the standard deviation (s), the F-ratio between the variances of calculated and observed activities (F). A number of additional statistical parameters such as the Akaike's information criterion, AIC [19, 20], the Kubinyi function, FIT [21, 22], and the Friedman's lack of fit, LOF [23], (Eqs. 1-3) have also been derived to evaluate the best model.

$$AIC = \frac{RSS \times (n + p')}{(n - p')^2} \quad (1)$$

$$FIT = \frac{r^2 \times (n - k - 1)}{(n + k^2) \times (1 - r^2)} \quad (2)$$

$$LOF = \frac{RSS/n}{\left[1 - \frac{k(d+1)}{n}\right]^2} \quad (3)$$

where, RSS is the sum of the squared differences between the observed and the estimated activity values, k is the number of variables in the model, p' is the number of adjustable parameters in the model, and d is the smoothing parameter. 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 (Fisher ratio), was proved to be a useful parameter for assessing the quality of the models. The main disadvantage of the F-value is its sensitivity to changes in k (the number of variables in the equation, which describe the model), if k is small, and its lower sensitivity if k is large. The FIT criterion has a low sensitivity toward changes in k -values, as long as they are small numbers, and a substantially increasing sensitivity for large k -values. The model that produces the minimum value of AIC and the highest value of FIT is considered potentially the most useful and the best. The LOF takes into account the number of terms used in the equation and is not biased, as are other indicators, toward large numbers of parameters. A minimum LOF value infers that the derived model is statistically sound.

The internal validation of derived model was ascertained through the cross-validated index, Q^2 , from leave-one-out and leave-five-out procedures. 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 has been removed once only. Then one model is developed for each reduced data set, and the response values of the deleted observations are predicted from these models. The squared differences between predicted and actual values are added to give the predictive residual sum of squares, PRESS. In this way, PRESS will contain one contribution from each observation. The cross-validated Q^2_{LOO} value may further be calculated as

$$Q^2_{LOO} = 1 - \frac{PRESS}{SSY} \quad (4)$$

where, SSY represents the variance of the observed activities of molecules around the mean value. In leave-five-out procedure, a group of five compounds is randomly kept outside the analysis each time in such a way that all the compounds, for once, become the part of the predictive groups. A value greater than 0.5 of Q^2 -index hints toward a reasonable robust model.

The external validation or predictive power of derived model is based on test set compounds. The squared correlation coefficient between the observed and predicted values of compounds from test set, r^2_{Test} , has been calculated as

$$r^2_{Test} = 1 - \frac{\sum (Y_{Pred(Test)} - Y_{(Test)})^2}{\sum (Y_{(Test)} - \bar{Y}_{(Training)})^2} \quad (5)$$

where, $Y_{\text{Pred(Test)}}$ and $Y_{\text{(Test)}}$ indicate predicted and observed activity values, respectively of the test-set compounds, and $\bar{Y}_{\text{(Training)}}$ indicate mean activity value of the training set. r^2_{Test} is the squared correlation coefficient between the observed and predicted data of the test-set. A value greater than 0.5 of r^2_{Test} suggests that the model obtained from training set has a reliable predictive power.

Y-randomization

Chance correlations, if any, associated with the CP-MLR models were recognized in randomization test [24, 25] by repeated scrambling of the biological response. The data sets with scrambled response vector have been reassessed by multiple regression analysis (MRA). The resulting regression equations, if any, with correlation coefficients better than or equal to the one corresponding to the unscrambled response data were counted. Every model has been subjected to 100 such simulation runs. This has been used as a measure to express the percent chance correlation of the model under scrutiny.

Applicability domain

The utility of a QSAR model is based on its accurate prediction ability for new compounds. 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 assessed by the leverage values for each compound [26]. 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 the model. In this plot, the applicability domain is established inside a squared area within $\pm x$ (standard deviations) and a leverage threshold h^* . The threshold h^* 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 . Prediction must be considered unreliable for compounds with a high leverage value ($h > h^*$). 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.

RESULTS AND DISCUSSION

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. For the purpose of modeling study, 7 compounds have been included in the test set for the validation of the models derived from 21 training set compounds. A total number of 109 significant descriptors from 0D-, 1D- and 2D-classes have been subjected to CP-MLR analysis with default 'filters' set in it. Statistical models in two and three descriptor(s) have been derived successively to achieve the best relationship correlating 5-HT₆ binding affinity. These models (with 109 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. A total number of 70 models in three descriptors were obtained. These models shared 52 descriptors. These descriptors along with their physical meaning, average regression coefficients and total incidences are listed in Table 3.

The selected models in three descriptors are given below.

$$\begin{aligned} \text{pK}_i &= 5.484(0.413)\text{S2K} + 3.005(0.289)\text{PW4} - 3.337(0.318)\text{BEHv6} + 3.943 \\ n &= 21, r = 0.967, s = 0.294, F = 81.586, Q^2_{\text{LOO}} = 0.913, Q^2_{\text{L50}} = 0.922, \\ r^2_{\text{randY}}(\text{sd}) &= 0.165(0.120), \text{AIC} = 0.127, \text{FIT} = 8.159, \text{LOF} = 0.137, r^2_{\text{Test}} = 0.637 \end{aligned} \quad (6)$$

$$\begin{aligned} \text{pK}_i &= 5.827(0.564)\text{S2K} + 3.086(0.424)\text{PW3} - 2.926(0.422)\text{BEHv6} + 3.483 \\ n &= 21, r = 0.940, s = 0.392, F = 43.301, Q^2_{\text{LOO}} = 0.764, Q^2_{\text{L50}} = 0.737, \\ r^2_{\text{randY}}(\text{sd}) &= 0.164(0.112), \text{AIC} = 0.226, \text{FIT} = 4.330, \text{LOF} = 0.244, r^2_{\text{Test}} = 0.542 \end{aligned} \quad (7)$$

$$\begin{aligned} \text{pK}_i &= -2.487(0.349)\text{SPI} + 5.243(0.555)\text{S2K} - 2.566(0.431)\text{BEHv6} + 6.237 \\ n &= 21, r = 0.938, s = 0.398, F = 41.864, Q^2_{\text{LOO}} = 0.835, Q^2_{\text{L50}} = 0.825, \\ r^2_{\text{randY}}(\text{sd}) &= 0.168(0.113), \text{AIC} = 0.233, \text{FIT} = 4.186, \text{LOF} = 0.251, r^2_{\text{Test}} = 0.627 \end{aligned} \quad (8)$$

$$\text{pK}_i = -2.649(0.428)\text{RBN} + 5.179(0.614)\text{S2K} - 2.669(0.476)\text{BEHv6} + 6.398$$

$$n = 21, r = 0.924, s = 0.440, F = 33.059, Q^2_{\text{LOO}} = 0.799, Q^2_{\text{LSO}} = 0.795, r^2_{\text{randY}}(\text{sd}) = 0.146(0.110), \text{AIC} = 0.285, \text{FIT} = 3.305, \text{LOF} = 0.308, r^2_{\text{Test}} = 0.502 \quad (9)$$

In above regression equations, the values given in the parentheses are the standard errors of the regression coefficients. The $r^2_{\text{randY}}(\text{sd})$ 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. The signs of the regression coefficients suggest the direction of influence of explanatory variables in the models.

The participated descriptors S2K, PW3, PW4 and SPI are from the TOPO class of Dragon descriptors. The TOPO class descriptors are based on a graph representation of the molecule and are numerical quantifiers of molecular topology obtained by the application of algebraic operators to matrices representing molecular graphs and whose values are independent of vertex numbering or labeling. They can be sensitive to one or more structural features of the molecule such as size, shape, symmetry, branching and cyclicity and can also encode chemical information concerning atom type and bond multiplicity. The descriptor S2K is 2-path Kier alpha-modified shape index and SPI is super pendent index. PW3 (path/walk ratio 3) and PW4 (path/walk ratio 4) are Randic's molecular shape descriptor.

The descriptors S2K, PW3 and PW4 contributed positively to the activity whereas SPI negatively to the activity. Thus, suggesting that a higher positive values of 2-path Kier alpha-modified shape index (S2K) and Randic's shape indices (PW3 and PW4) would be beneficiary to the activity. On the other hand, a lower value of super pendent index (SPI) would augment the activity.

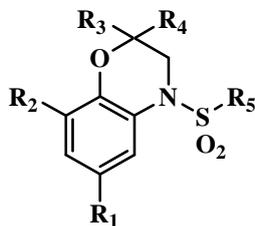
The descriptor BEHv6, in above models, is lone representative of BCUT class of Dragon descriptors. The BCUT descriptors are the first 8 highest and the lowest absolute eigenvalues, BEHwk and BELwk, respectively, for the modified Burden adjacency matrix. Here w refers to the atomic property and k to the eigenvalue rank. The ordered sequence of the highest and the lowest eigenvalues reflect upon the relevant aspects of molecular structure, useful for similarity searching. The negative contribution of descriptor BEHv6 to the activity advocates that a higher value of this descriptor would be detrimental to the activity.

The other participated descriptor RBN, belong to CONST class of Dragon descriptors. The constitutional class descriptors are based on simple constitutional facts and are independent from molecular connectivity and conformations. Descriptor RBN represents the number of rotatable bonds in a molecular structure. The negative sign of regression coefficient of this descriptor suggests that less number of rotatable bonds in a molecular structure would be favorable to the activity.

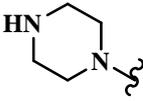
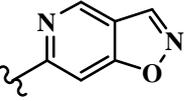
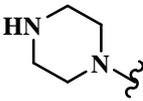
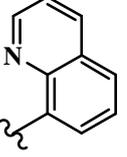
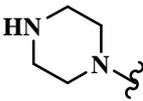
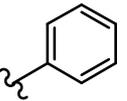
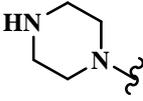
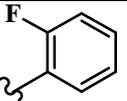
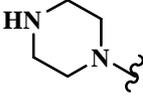
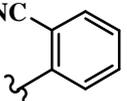
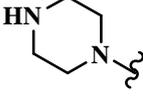
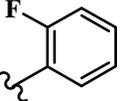
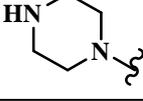
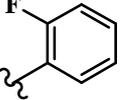
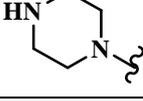
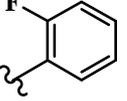
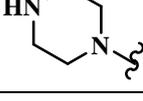
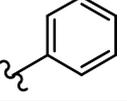
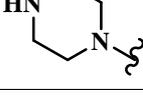
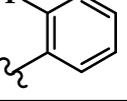
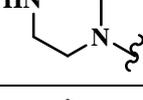
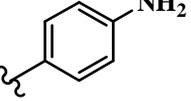
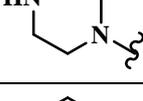
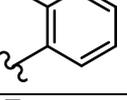
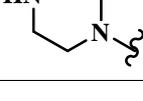
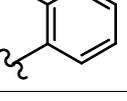
These models have accounted for up to 93.51 percent variance in the observed activities. The values greater than 0.5 of Q^2 -index is in accordance to a reasonable robust QSAR model. The pK_i values of training set compounds calculated using Equations (6) to (9) have been included in Table 1. These models are validated with an external test set of seven compounds listed in Table 1. The predictions of the test set compounds based on external validation are found to be satisfactory as reflected in the test set r^2 (r^2_{Test}) values and the predicted activity values are also reported in Table 1. The plot showing goodness of fit between observed and calculated activities for the training and test set compounds is given in Figure 1.

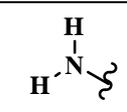
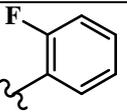
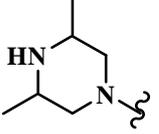
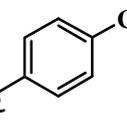
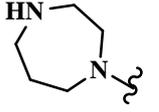
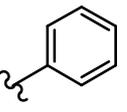
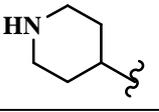
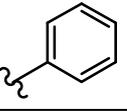
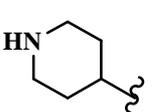
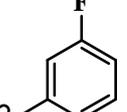
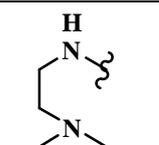
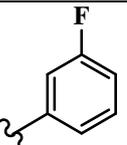
Applicability domain

On analyzing the model applicability domain (AD) in the Williams plot (Fig. 2) of the model based on the whole data set (Table 4), it has appeared that none of the compounds were identified as an obvious outlier for the 5-HT₆ binding affinity if the limit of normal values for the Y outliers (response outliers) was set as 2.5(standard deviation) units. Two compounds (23 and 28) were found to have leverage (h) values greater than the threshold leverages (h^*) and may be treated as structurally influential chemicals. 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.

Table 1: Structures, observed and modeled 5-HT₆ binding affinities 3,4-dihydro-2H-benzo[1,4]oxazines.

Cpd.	R ₁	R ₂	R ₃ /R ₄	R ₅	pK _i				
					Obs. ^a	Eq. (6)	Eq. (7)	Eq. (8)	Eq. (9)
1	H		H/H		10.2	9.91	10.53	9.68	9.96
2 ^b	H		H/H		7.9	8.24	8.03	8.42	8.46
3	H		H/H		8.9	8.97	8.83	8.74	8.84
4 ^b	H		H/H		8.5	9.36	9.25	9.12	9.21
5 ^b	H		H/H		8.7	8.80	8.73	9.11	9.21
6	H		H/H		8	8.41	8.32	8.74	8.84
7 ^b	H		H/H		6.7	6.59	6.06	7.08	7.30
8	H		H/H		6.5	6.58	6.57	6.39	6.66
9	Cl		H/H		9.1	9.02	8.71	8.81	8.82

10 ^b	Cl		H/H		9.3	9.71	9.89	9.77	9.85
11	Cl		H/H		9.2	9.07	8.74	8.83	8.85
12	Me		H/H		8.6	8.33	7.97	8.59	8.33
13	Me		H/H		9.4	9.05	8.77	8.84	8.71
14	Me		H/H		8.7	9.06	8.76	8.70	8.23
15	F		H/H		9.2	9.29	8.97	9.01	9.23
16	Cl		H/H		9.4	9.68	9.38	9.38	9.59
17	OMe		H/H		9.8	10.08	10.16	9.93	9.49
18	OMe		H/H		9.4	9.37	9.44	9.70	9.11
19	<i>t</i> -Bu		H/H		8.1	7.73	8.62	7.73	7.95
20 ^b	F		H/H		8.9	8.37	8.56	9.02	8.92
21 ^b	H		Me/Me		9.5	8.69	8.48	8.59	8.52
22	H		-CH ₂ CH ₂ CH ₂ -		9.2	8.54	9.28	9.01	9.12

23	H		H/H		6.2	6.24	6.57	6.21	6.07
24	Me		H/H		7.4	7.58	7.97	7.88	7.82
25	Cl		H/H		8.4	8.49	8.30	9.16	9.19
26	F		Me/Me		8.4	8.34	7.93	8.72	8.56
27	F		Me/Me		8.3	8.52	8.30	8.13	8.95
28	H		Me/Me		6.9	7.05	7.17	7.11	6.98

^aTaken from reference [10]; ^bCompounds included in test set.

Table 2: Descriptor classes^a used along with their definition and scope for modeling the binding affinity of 3,4-dihydro-2H-benzo[1,4]oxazine derivatives.

Descriptor class (acronyms)	Definition and scope
Constitutional (CONST)	Dimensionless or 0D descriptors; independent from molecular connectivity and conformations
Topological (TOPO)	2D-descriptor from molecular graphs and independent conformations
Molecular walk counts (MWC)	2D-descriptors representing self-returning walks 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 atom 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 non-single 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

^aReference [12]

Table 3: Descriptors^a identified for modeling the binding affinity of 3,4-dihydro-2H-benzo[1,4]oxazine derivatives along with the average regression coefficient^b, standard deviation and the total incidence.

Descriptor	Avg reg coeff(sd) total incidence	Descriptor	Avg reg coeff(sd) total incidence	Descriptor	Avg reg coeff(sd) total incidence
MW	4.917(0.291)2	Sv	3.506(0.966)2	Ms	-1.951(0.640)4
nCIC	1.865(0.271)2	RBN	-2.408(0.858)8	nO	-1.150(0.000)1
nCl	1.522(0.000)1	nX	1.095(0.000)1	nR06	1.845(0.589)2
ZM2V	1.607(0.595)2	MSD	-1.509(0.000)1	SPI	-2.418(0.523)13
Rww	-2.876(0.000)1	JhetZ	-3.504(1.022)5	X1Av	-2.416(0.621)11
X2Av	-2.538(0.537)11	S2K	3.597(1.081)21	PW2	-1.886(0.000)1
PW3	3.086(0.000)1	PW4	2.325(0.973)5	PJI2	1.424(0.260)3
SEigm	1.434(0.000)1	T(N..S)	-3.163(0.000)1	BEHm5	2.511(0.000)1
BELm3	3.202(0.363)11	BELm4	3.095(0.674)8	BEHv3	-2.953(0.000)1
BEHv6	-2.744(0.675)6	BELv7	1.817(0.000)1	BEHp1	-1.285(0.000)1
GGI9	-2.383(0.821)10	JGI3	-2.254(0.061)2	JGI5	2.383(0.000)1
ATS3m	2.950(0.722)5	MATS1v	-2.770(1.413)3	MATS3v	2.614(0.591)3
MATS4v	1.336(0.000)1	MATS8v	1.575(0.000)1	MATS1e	-2.492(0.935)4
MATS2e	-2.479(0.376)18	MATS4e	1.511(0.274)10	MATS2p	1.419(0.000)1
GATS1e	1.657(0.000)1	GATS2e	2.728(0.000)1	GATS4e	-2.133(0.844)2
GATS6e	1.996(0.746)2	nCrH2	3.097(0.470)5	nCaR	1.253(0.000)1
C-005	-1.874(0.899)4	C-006	2.174(0.948)4	C-028	1.839(0.000)1
H-052	2.476(0.000)1				

^aThe descriptors are identified from the three parameter models emerged from CP-MLR protocol with filter-1 as 0.79; filter-2 as 2.0; filter-3 as 0.76; filter-4 as $0.3 \leq Q^2 \leq 1.0$; number of compounds in the study are 21; **CONST**: MW, molecular weight; Sv, sum of atomic van der Waals volumes (scaled on Carbon atom); Ms, mean electrotopological state; nCIC, number of rings; RBN, number of rotatable bonds; nO, number of oxygen atoms; nCl, number of chlorine atoms; nX, number of halogen atoms; nR06, number of 6-membered rings; **TOPO**: ZM2V, second Zagreb index by valence vertex degrees; MSD, mean square distance index (Balaban); SPI, superpendent index; Rww, reciprocal hyper-detour index; JhetZ, Balaban-type index from Z weighted distance matrix (Barysz matrix); X1Av, average valence connectivity index chi-1; X2Av, average valence connectivity index chi-2; S2K, 2-path Kier alpha-modified shape index; PW2, path/walk 2 - Randic shape index; PW3, path/walk 3 - Randic shape index; PW4, path/walk 4 - Randic shape index; PJI2, 2D Petitjean shape index; SEigm, Eigen value sum from mass weighted distance matrix; T(N..S), sum of topological distances between N..S; **BCUT**: BEHm5, highest eigenvalue n. 5 of Burden matrix / weighted by atomic masses; BELm3 and BELm4, lowest eigenvalue n. 3 and 4 of Burden matrix / weighted by atomic masses, respectively; BEHv3 and BEHv6, highest eigenvalue n. 3 and 6 of Burden matrix / weighted by atomic van der Waals volumes, respectively; BELv7, lowest eigenvalue n. 7 of Burden matrix / weighted by atomic van der Waals volumes; BEHp1, highest eigenvalue n. 1 of Burden matrix / weighted by atomic polarizabilities; **GALVEZ**: GGI9, topological charge index of order 9; JGI3 and JGI5, mean topological charge index of order 3 and 5, respectively; **2D-AUTO**: ATS3m, Broto-Moreau autocorrelation of a topological structure - lag 3 / weighted by atomic masses; MATS1v, MATS3v, MATS4v and MATS8v, Moran autocorrelation - lag 1, 3, 4 and 8, respectively / weighted by atomic van der Waals volumes; MATS1e, MATS2e and MATS4e, Moran autocorrelation - lag 1, 2 and 4, respectively / weighted by atomic Sanderson electronegativities; MATS2p, Moran autocorrelation - lag 2 / weighted by atomic polarizabilities; GATS1e, GATS2e, GATS4e and GATS6e, Geary autocorrelation - lag 1,2,4 and 6, respectively / weighted by atomic Sanderson electronegativities; **FUNC**: nCrH2, number of ring secondary C(sp³); nCaR, number of substituted aromatic C(sp²); **ACF**: C-005, CH3X; C-006, CH2RX; C-028 R--CR--X; H-052, H attached to C0(sp³) with 1X attached to next C atom. ^bThe average regression coefficient of the descriptor corresponding to all models and the total number of its incidences; the arithmetic sign represents the sign of the regression coefficient in the models.

Table 4: Resultant models for the whole data set (n=28) in descriptors of training set models.

Model	r	s	F	Q ² _{L00}	Q ² _{L50}	Eq.
pK _i = 5.389(0.492)S2K + 2.716(0.339)PW4 - 3.261(0.355)BEHv6 + 4.168	0.939	0.373	59.228	0.851	0.845	6a
pK _i = 5.473(0.587)S2K + 2.560(0.413)PW3 - 2.766(0.429)BEHv6 + 4.038	0.912	0.443	39.625	0.729	0.607	7a
pK _i = -2.329(0.368)SPI + 5.258(0.576)S2K - 2.608(0.430)BEHv6 + 6.166	0.914	0.437	40.788	0.789	0.773	8a
pK _i = -2.361(0.455)RBN + 5.212(0.646)S2K - 2.775(0.477)BEHv6 + 6.310	0.891	0.490	30.795	0.726	0.711	9a

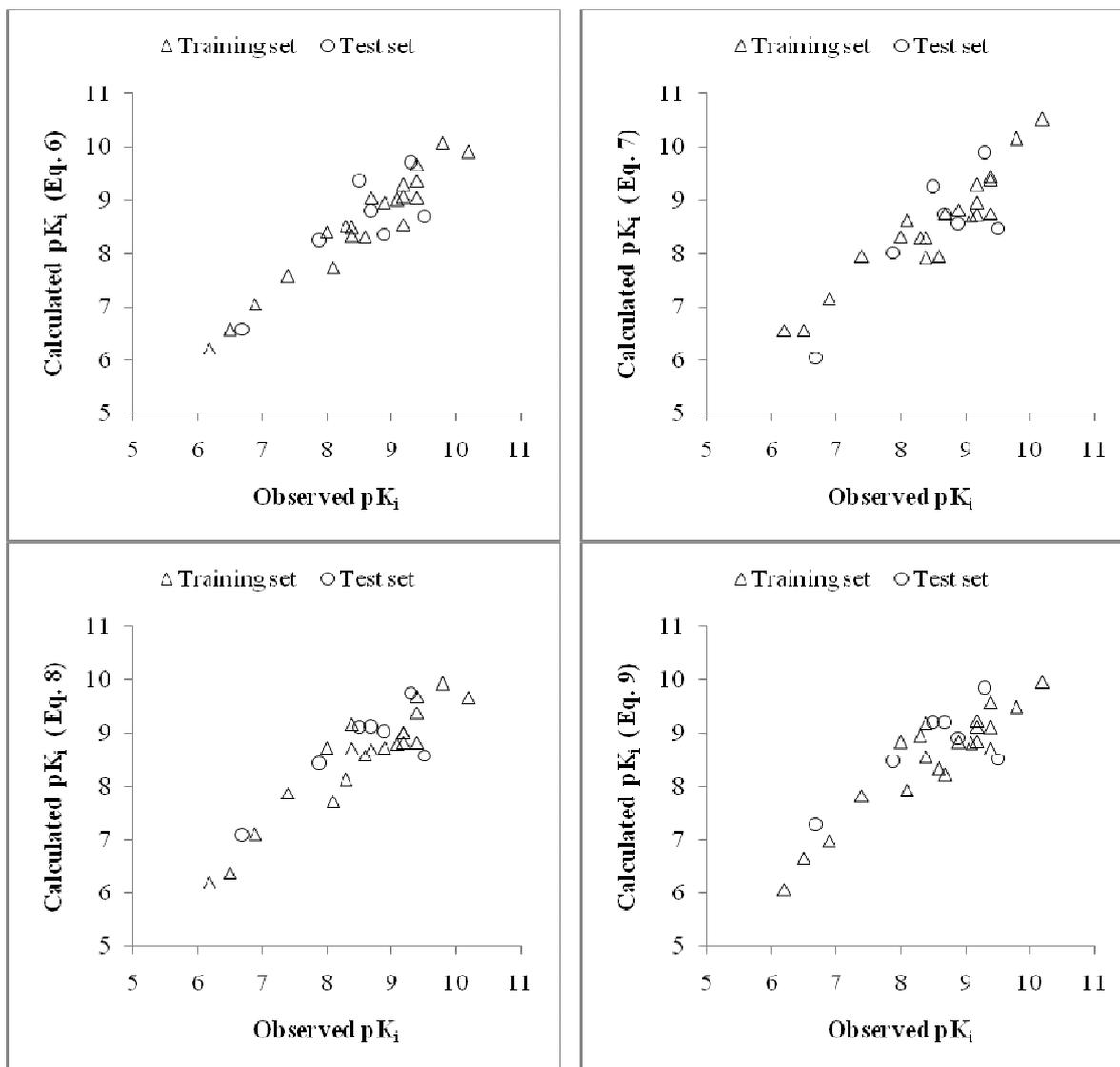


Fig. 1: Plots between observed versus calculated pK_i values.

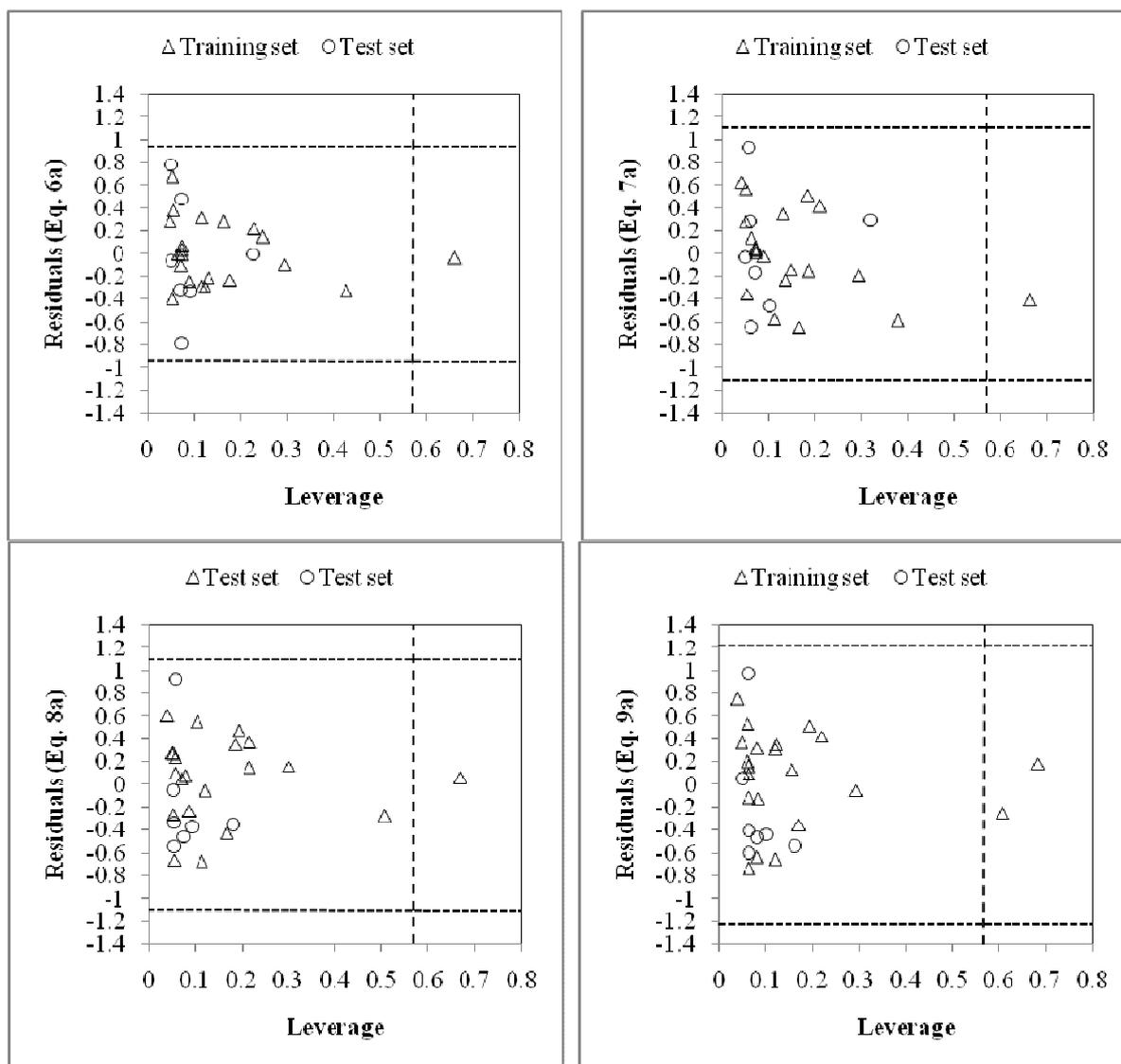


Fig. 2: Williams plot for the training and test set for the 5-HT₆ binding affinity of 3,4-dihydro-2H-benzo[1,4]oxazine derivatives.

CONCLUSION

This study has provided a rational approach for the development of new 3,4-dihydro-2H-benzo[1,4]oxazine derivatives as potential 5-HT₆ antagonists. The descriptors identified in CP-MLR analysis have highlighted the role of atomic van der Waals volume in terms of BEHv6, path/walk ratio 3 and 4- Randic shape index (PW3 and PW4), number of rotatable bonds (RBN) and super pendent index (SPI) to explain the biological actions of 3,4-dihydro-2H-benzo[1,4]oxazine derivatives as potential 5-HT₆ antagonists. Applicability domain analysis revealed that the suggested model matches the high quality parameters with good fitting power and the capability of assessing external data and all of the compounds was within the applicability domain of the proposed model and were evaluated correctly.

Acknowledgements

The financial support in the form of SRF and RA for the authors MC and PP provided by CSIR, New Delhi is thankfully acknowledged. Authors are also thankful to the Institution for providing necessary facilities to complete this study.

REFERENCES

- [1] A Slassi; M Isaac; A O'Brien, *Expert Opin. Ther. Pat.*, **2002**, 12, 513-27.
- [2] D Hoyer; GR Martin, *Neuropharmacology*, **1997**, 36, 419-28.
- [3] D Hoyer; DE Clarke; JR Fozard; PR Hartig; GR Martin; E Mylecharane; PR Saxena; PPA Humphrey, *Pharmacol. Rev.*, **1994**, 46, 157-204.
- [4] R Kohen; MA Metcalf; N Khan; T Druck; K Huebner; JE Lachowicz; HY Meltzer; DR Sibley; BL Roth; MW Hamblin, *J. Neurochem.*, **1996**, 66, 47-56.
- [5] M Ruat; E Traiffort; J-M Arrang; J Tardivel-Lacombe; J Diaz; R Leurs; J-C Swartz, *Biochem. Biophys. Res. Commun.*, **1993**, 193, 269-76.
- [6] M Sebben; H Ansanay; J Bockaert; A Dumuis, *Neuroreport*, **1994**, 5, 2553-2557.
- [7] SL Davies; JS Silvestre; X Guitart, *Drugs Fut.*, **2005**, 30, 479-88.
- [8] ES Mitchell; JF Neumaier, *Pharmacol. Ther.*, **2005**, 108, 320-33.
- [9] A Bourson; E Borroni; RH Austin; FJ Monsma; AJ Sleight, *J. Pharmacol. Exp. Ther.*, **1995**, 274, 173-80.
- [10] S-H Zhao; J Berger; RD Clark; SG Sethofer; NE Krauss; JM Brothers, *Bioorg. Med. Chem. Lett.*, **2007**, 17, 3504-3507.
- [11] ChemDraw Ultra 6.0 and Chem3D Ultra, Cambridge Soft Corporation, Cambridge, USA. Available from: <http://www.cambridgesoft.com>.
- [12] Dragon software (version 1.11-2001) by R Todeschini; V Consonni, Milano, Italy. Available from: <http://www.taletе.mi.it/dragon.htm>.
- [13] YS Prabhakar, *QSAR Comb. Sci.*, **2003**, 22, 583-95.
- [14] S Sharma; YS Prabhakar; P Singh; BK Sharma, *Eur. J. Med. Chem.*, **2008**, 43, 2354-2360.
- [15] BK Sharma; P Pilania; P Singh, *J. Enzy. Inhibn. Med. Chem.*, **2009**, 24, 607-15.
- [16] S Sharma; BK Sharma; P Pilania; P Singh; YS Prabhakar, *J. Enzy. Inhibn. Med. Chem.*, **2009**, 24, 1024-1033.
- [17] BK Sharma; P Pilania; P Singh; YS Prabhakar, *SAR QSAR Environ. Res.*, **2010**, 21, 169-85.
- [18] BK Sharma; P Pilania; K Sarbhai; P Singh; YS Prabhakar, *Mol. Divers.*, **2010**, 14, 371-384.
- [19] H Akaike. Information theory and an extension of the minimum likelihood principle. In: BN Petrov; F Csaki, (Editors). Second international symposium on information theory. Budapest: Akademiai Kiado, **1973**, 267-281.
- [20] H Akaike, *IEEE Trans. Automat. Contr.*, **1974**, AC-19, 716-723.
- [21] H Kubinyi, *Quant. Struct.-Act. Relat.*, **1994**, 13, 285-294.
- [22] H Kubinyi, *Quant. Struct.-Act. Relat.*, **1994**, 13, 393-401.
- [23] J Friedman, In: Technical report no. 102. Laboratory for computational statistics. Stanford University, Stanford, **1990**.
- [24] S-S So; M Karplus, *J. Med. Chem.*, **1997**, 40, 4347-4359.
- [25] YS Prabhakar; VR Solomon; RK Rawal; MK Gupta; SB Katti, *QSAR Comb. Sci.*, **2004**, 23, 234-244.
- [26] P Gramatica, *QSAR Comb. Sci.*, **2007**, 26, 694-701.