British Journal of Pharmaceutical Research
8(3): 1-17, 2015, Article no.BJPR.18732
ISSN: 2231-2919

SCIENCEDOMAIN *international*
www.sciencedomain.org

CP-MLR Directed QSAR Rationales for the 1-aryl Sulfonyl Tryptamines as 5-HT_6 Receptor Ligands

Manju Choudhary[1], Shreekant Deshpande[2] and Brij Kishore Sharma[1*]

[1]Department of Chemistry, Government PG College, Bundi-323 001, Rajasthan, India.
[2]Analytical Chemistry Division, Eutech Scientific Services Inc, Highland Park, NJ 08904, USA.

Authors' contributions

This work was carried out in collaboration between all authors. Author MC designed the study, performed the statistical analysis, and wrote the protocol. Author SD wrote the first draft of the manuscript and managed the literature searches. Authors BKS and SD managed the analyses of the study. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/BJPR/2015/18732
Editor(s):
(1) Elena G. Zavyalova, Chemistry Department, Moscow State University, Russia.
Reviewers:
(1) Fatma Kandemirli, Biomediacal Engineer, Kastamonu University, Turkey.
(2) Anonymous, Jiaxing University, China.
(3) Anonymous, University of Chile, Chile.
Complete Peer review History: http://sciencedomain.org/review-history/11208

Original Research Article

Received 7th May 2015
Accepted 26th July 2015
Published 1st September 2015

ABSTRACT

A QSAR study has been carried out to rationalize the 5-HT_6 receptor binding affinities of the 1-aryl sulfonyl tryptamine derivatives using Dragon descriptors. A higher value of molecular symmetry and topology accounting Randic shape index descriptor PW4 (path/walk 4) would be favorable to improve the binding affinity. Presence of more number of bromine atoms (descriptor nBR) and presence of such structural fragment in which a hydrogen atom attached to sp3 hybridized carbon with no hetero atom rather than one hetero atom attached to next carbon atom (descriptors H-046 and H-052) will be supportive to the activity. The prevalence of atomic properties to explain the binding affinity is evident from the associations of polarizability to the path length 7 of Moran autocorrelation (MATS7p), masses to eigenvalues n.2 and 7 of Burden m atrix (BELm2 and BEHm7), Sanderson electronegativity to highest eigenvalue n.2 Burden matrix (BEHe2) and van der Waals volume to path length 8 of Geary autocorrelation (GATS8v) and charge content in terms of topological and mean topological charge indices (GGI3 and JGI2). The dominance of the information content of the descriptors, emerged in CP-MLR models, has also confirmed by the PLS analysis.

**Corresponding author: Email: bksharma_sikar@rediffmail.com;*

The derived QSAR models and descriptors shared in these models revealed that the substituents of tryptamine moiety have sufficient scope for further modification.

Keywords: QSAR; 1-aryl sulfonyl tryptamines; 5-HT$_6$ ligands; binding affinity; combinatorial protocol in multiple linear regression (CP-MLR).

1. INTRODUCTION

The 5-HT$_6$ receptor, a member of 5-Hydroxytryptamine (5-HT; serotonin) receptor family, plays a vital role in the modulation of learning, memory [1-3] and feeding behavior [4,5] related disorders. The 5-HT$_6$ receptor is positively coupled to adenylyl cyclase [6-8]. 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. Many antipsychotic and antidepressant drugs have shown significant affinity for 5-HT$_6$ receptor [9,10]. Over the past few years, 5-HT$_6$ receptor has become an important therapeutic target for schizophrenia, anxiety, impairment of learning, memory and obesity [11-17] due to the specific localization of 5-HT$_6$ receptors in CNS and high affinity of antipsychotic and antidepressant drugs.

SB-742457 [18], SUVN-502 [19], Lu AE58054 [20], SAM-760 [21] and SYN-114 [22] are among the clinically advancing ligands for the 5-HT$_6$ receptor. The one of the most explored chemical class of 5-HT$_6$ receptor ligands is the indole nucleus. MS-245 [23] (the N_1-arylsulfonyl tryptamines), PMDT [24] (2-aryl tryptamines) and carbazole derivatives (conformationally restricted tryptamines) [25] are representative from this class. A series of rigidized side chain tryptamine derivatives have been synthesized and evaluated for the 5-HT$_6$ receptor binding affinity by Nirogi et al. [26]. A QSAR study has been carried out on the binding affinities of these rigidized side chain tryptamine derivatives to rationalize the substituent variations and to provide insight for further modification.

2. MATERIALS AND METHODS

2.1 Chemical Structure Database and Biological Activity

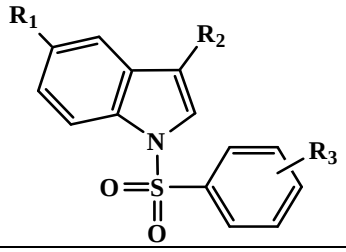
This study comprises a chemical structure database of reported forty three tryptamine derivatives. The *in vitro* binding affinities of these compounds were determined by nonradioactive cell-based assay using a stable CHO cell line expressing recombinant human 5-HT$_6$R. The structural variations and the binding affinities, in terms of K_b, 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 43 analogues have been divided into training and test sets. Out of the 43 analogues, nearly one fourth compounds (10) 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.

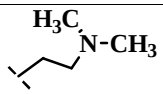
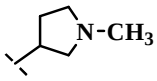
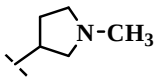
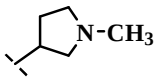
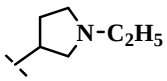
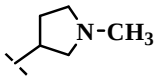
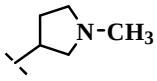
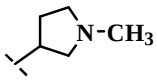
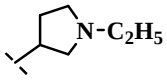
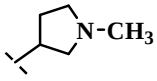
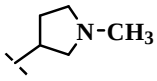
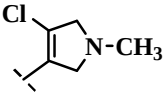
2.2 Theoretical Molecular Descriptors

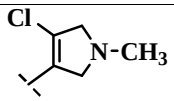
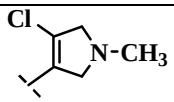
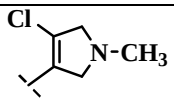
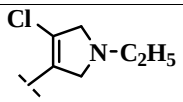
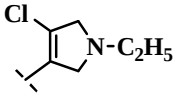
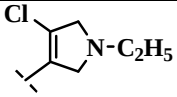
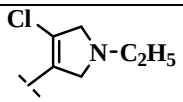
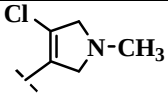
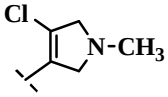
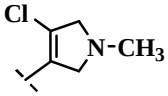
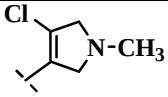
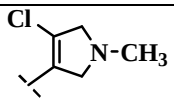
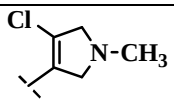
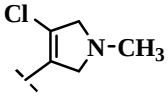
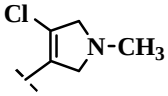
The structures of the compounds under study have been drawn in 2D ChemDraw [27]. The drawn structures were then converted into 3D modules using the default conversion procedure implemented in the CS Chem3D Ultra. The energy of these 3D-structures was minimized in the MOPAC module using the AM1 procedure 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 [28] 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 description of these descriptors is available in manuals [28,29]. The combinatorial protocol in multiple linear regression [30] procedure has

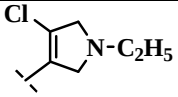
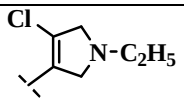
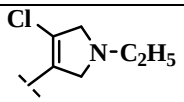
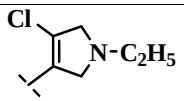
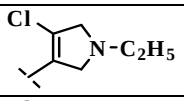
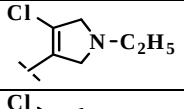
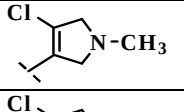
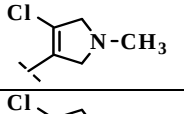
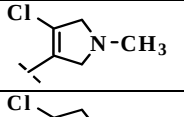
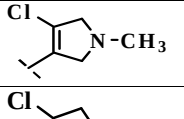
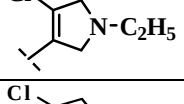
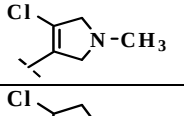
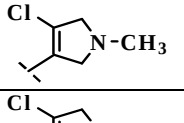
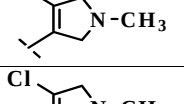
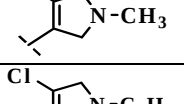
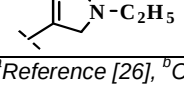
been used in the present work for developing QSAR models. This procedure, as a variable selection or model development has been employed successfully by us [31-33].

Table 1. Structures[a] and observed and modeled binding activities of 1-aryl sulfonyl tryptamines



Cpd.	R_1	R_2	R_3	pK_b		
				Obsd[a].	Calc.	
					Eq. (9)	PLS
1	OCH_3		H	8.24	8.16	8.48
2	H		H	8.82	9.51	9.46
3	H		4-F	9.00	8.77	8.75
4[b]	H		2-Br	9.89	10.45	10.23
5	H		2-Br	10.00	9.46	9.63
6	OCH_3		H	9.05	9.23	9.32
7	OCH_3		2-Br	9.89	9.77	9.83
8	OCH_3		4-$CH(CH_3)_2$	10.00	9.69	9.86
9	OCH_3		2-Br	8.89	9.07	9.38
10	OC_2H_5		2-Br	10.00	9.49	9.62
11	$OCH(CH_3)_2$		2-Br	8.38	9.09	9.13
12	H		H	10.00	10.07	9.78

Cpd.	R_1	R_2	R_3	pK_b		
				Obsd[a].	Calc.	
					Eq. (9)	PLS
13	H		4-Br	9.46	9.29	8.90
14[b]	H		4-F	9.22	9.38	9.06
15	H		4-OCH_3	8.00	8.38	8.45
16[b]	H		2-Br	8.96	9.27	9.48
17	H		4-Br	8.41	8.60	8.36
18	H		4-F	8.24	8.66	8.49
19	H		3-Cl	7.90	8.10	8.28
20	OCH_3		H	10.00	9.41	9.51
21[b]	OCH_3		2-Br	9.52	9.59	9.82
22[b]	OCH_3		4-Br	8.75	8.93	8.75
23	OCH_3		4-F	9.20	8.83	8.86
24[b]	OCH_3		3-Cl	9.08	8.51	8.71
25	OCH_3		4-OCH_3	8.78	8.53	8.51
26	OCH_3		4-$CH(CH_3)_2$	9.60	9.82	9.98
27	OCH_3		2,4,5-Cl	7.73	7.77	7.83

Cpd.	R_1	R_2	R_3	pK_b		
				Obsd[a].	Calc.	
					Eq. (9)	PLS
28	OCH_3		H	9.15	9.00	9.08
29[b]	OCH_3		2-Br	8.28	8.92	9.34
30	OCH_3		4-Br	8.32	8.35	8.29
31[b]	OCH_3		4-F	8.06	8.44	8.43
32	OCH_3		3-Cl	8.62	8.06	8.26
33[b]	OCH_3		4-OCH3	7.84	8.21	8.18
34	OC_2H_5		H	9.00	9.31	9.22
35	OC_2H_5		2-Br	9.40	9.72	9.63
36[b]	OC_2H_5		4-Br	8.35	9.04	8.57
37	OC_2H_5		4-F	8.21	8.64	8.57
38	OC_2H_5		2-Br	9.35	9.14	9.13
39	$OCH(CH_3)_2$		H	9.15	8.92	8.67
40	$OCH(CH_3)_2$		4-F	8.14	8.40	8.07
41	F		H	10.00	9.84	9.52
42	F		2-Br	9.60	9.72	9.83
43	F		3-Cl	8.21	7.94	8.07

[a]Reference [26], [b]Compounds included in test set

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 471 to 89 descriptors as relevant ones for the binding activity. A brief description of the computational procedure is given below.

2.3 Model Development

The combinatorial protocol in multiple linear regression (CP-MLR) is a 'filter' based variable selection procedure for model development in QSAR studies. 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, r-bar, 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 r-bar (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 r-bar value of the preceding optimum model as the new threshold for next generation.

2.4 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 [34,35], the Kubinyi function, FIT [36,37], and the Friedman's lack of fit, LOF [38] (Eqs. 1-3) have also been derived to evaluate the best model.

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

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

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

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

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

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, $Y_{Pred(Test)}$ and $Y_{(Test)}$ indicate predicted and observed activity values, respectively of the test-set compounds, and $\bar{Y}_{(Training)}$ 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.

2.5 Y-randomization

Chance correlations, if any, associated with the CP-MLR models were recognized in randomization test [39,40] 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.

3. RESULTS AND DISCUSSION

The explored QSAR model(s) using Dragon descriptors, in multi-descriptor class environment, may be utilized to correlate the biological actions shown by the compounds. The data set has been divided in training and test set comprising 33 and 10 compounds, respectively. Statistical models up to five descriptors have been derived using 89 significant descriptors (from 0D-, 1D- and 2D-classes) by CP-MLR analysis to correlate the 5-HT6 binding affinity. These models were identified in CP-MLR. In doing so successive increments were made in the filter-3 with increasing number of descriptors and the optimum r-bar value of the preceding level model was used as the new threshold of this filter for the next generation.

The highest significant models in three, four and five variables (descriptors) are presented below.

pK_b = 6.478+1.994(0.413)PW4+1.749(0.403)GATS8v+2.712(0.471)H-046
n = 33, r = 0.768, s = 0.482, F = 13.953, FIT = 0.996, LOF = 0.305, AIC = 0.296,
Q^2_{LOO} = 0.503, Q^2_{L5O} = 0.515, r^2_{randY}(sd) = 0.101(0.070), r^2_{Test} = 0.627 (6)

pK_b = 4.157+2.283(0.421)X4v+4.619(0.615)BELm2+1.924(0.402)MATS7p
–1.575(0.354)H-052
n = 33, r = 0.832, s = 0.424, F = 15.845, FIT = 1.293, LOF = 0.266, AIC = 0.244,
Q^2_{LOO} = 0.510, Q^2_{L5O} = 0.530, r^2_{randY}(sd) = 0.125(0.076), r^2_{Test} = 0.557 (7)

$$pK_b = 2.662–1.618(0.342)X2A+6.024(0.713)BELm2+1.583(0.424)GGI3+1.058(0.342)JGI2 +2.943(0.413)MATS7p$$
$$n = 33, r = 0.861, s = 0.396, F = 15.607, FIT = 1.345, LOF = 0.264, AIC = 0.226,$$
$$Q^2_{LOO} = 0.606, Q^2_{L5O} = 0.571, r^2_{randY}(sd) = 0.161(0.085), r^2_{Test} = 0.605 \quad (8)$$

In above models, the parenthesized values are the standard errors of the regression coefficients. The parameter $r^2_{randY}(sd)$ represents the mean random squared multiple correlation coefficient of the regressions in the activity randomization study with its standard deviation. In hundred simulation runs per model none of the identified models has shown any chance correlation. The inter-correlation among the predictor variables (inter-correlation matrix) of models presented above is provided in Table 3. The signs of numerical values of the regression coefficients propose the direction of influence of explanatory variables in the models.

The participated descriptors PW4, X2A and X4v are from the TOPO class of Dragon descriptors. The TOPO class descriptors are numerical quantifiers of molecular topology represented graphically and are obtainable by the application of algebraic operators to matrices representing molecular graphs and whose values are independent of vertex numbering or labeling. These descriptors are sensitive to structural features of the molecule like size, shape, symmetry, branching, cyclicity and chemical information concerning atom type and bond multiplicity. The descriptor PW4 is the Randic shape index (Path/walk 4). Descriptors X2A and X4v are connectivity indices representing average connectivity index (chi-2) and valence connectivity index (chi-4), respectively. The positive correlation of descriptors PW4 and X4v to the activity advocates that higher values of these descriptors would be beneficiary to activity. On the other hand negative correlation of descriptor X2A to the activity suggests a lower value of it for elevated activity.

The descriptors MATS7p and GATS8v belong to the 2D-AUTO class. The 2D-AUTO descriptors are the autocorrelation of topological structure of Broto-Moreau (ATS), of Moran (MATS) and of Geary (GATS). These descriptors deal with the topology of a molecular structure or parts thereof in association with a selected physicochemical property such as atomic mass, van der Waal's volume, electronegativity and polarizability. In these descriptors' nomenclature, the penultimate character, a number, indicates the number of consecutively connected edges considered in its computation and is called as the autocorrelation vector of lag k (corresponding to the number of edges in the unit fragment). The very last character of the descriptor's nomenclature indicates the physicochemical property considered in the weighting component for its computation. Both of the participated descriptors, MATS7p (atomic polarizabilities weighted Moran autocorrelation of lag 7) and GATS8v (atomic van der Waals volumes weighted Geary autocorrelation of lag 8) correlated positively to the activity and suggest the favorable conditions associated with lag 7 weighted by atomic polarizabilities and lag 8 weighted by atomic van der Waal's volumes.

The descriptors H-046 and H-052 are from the ACF class of Dragon descriptors defined by Ghose and Crippen. These are simple molecular descriptors based on the counting of 120 atom centered fragments as the number of specific atom types in a molecule and evaluated by the knowledge of the molecular composition and atom connectivities. Descriptors H-046 represent the structural fragment in which H attached to C0(sp3) with no hetero atom (X) attached to next carbon atom and H-052 a fragment with H attached to C0(sp3) with one hetero atom (X) attached to next carbon atom. The descriptors H-046 and H-052 have shown positive and negative correlation, respectively, to the activity advocating a structural fragment in which a hydrogen atom attached to sp^3 hybridized carbon with no hetero atom rather than one hetero atom attached to next carbon atom supportive to the activity.

Descriptors BELm2 is BCUT class descriptor. The first 8 highest and lowest absolute eigenvalues (BEHwk and BELwk) of the modified Burden adjacency matrix are represented by the BCUT descriptors, in which w refers to the atomic property and k to the eigenvalue rank. The relevant and useful aspects for similarity searching of molecular structure correspond to the ordered sequence of the highest and the lowest eigen values. The descriptor BELm2 (atomic mass weighted lowest eigenvalue n.2 of Burden matrix) have shown positive correlation to the activity suggesting that a higher value of this descriptor is advantageous for activity.

The descriptors GGI3 and JGI2 belong to GALVEZ class of Dragon descriptors. The GALVEZ descriptors are topological charge

indices representing the first ten eigenvalues of the polynomial of corrected adjacency matrix of the compounds. The first category of GALVEZ class descriptors corresponds to the topological charge index (GGIn) and the second to the mean topological charge index (JGIn). The order of eigenvalue is represent by 'n'. Both the descriptors GGI3 (topological charge index of order 3) and JGI2 (mean topological charge index of order 2) have shown positive correlations to the activity suggesting that a higher value of these descriptors would augment the 5-HT_6 binding activity of titled compounds.

The above discussed models are able to explain 74.13 percent variance in observed activity. Considering the number of observation in the dataset (33), models with up to six descriptors were explored. A total number of 24 models, sharing 29 descriptors among them, were obtained through CP-MLR. To serve as a measure of descriptors' estimate across these models, the brief meaning, average regression coefficients and total incidence of all these 29 descriptors is mentioned in Table 4. The given below is the highest significant six-descriptor model for the activity. Table 3 contains the inter-correlation matrix of variables of this model. This model has accounted for 76.91 percent variance in the observed activities.

$$pK_b = 3.100+0.632(0.215)nBR+1.327(0.319)PW4–1.749(0.448)BEHm7+3.672(0.612)BELm2 +4.024(0.833)BEHe2+2.836(0.625)MATS7p$$

$$n = 33, r = 0.877, s = 0.382, F = 14.452, FIT = 1.256, LOF = 0.284, AIC = 0.225, Q^2_{LOO} = 0.654, Q^2_{L5O} = 0.664, r^2_{randY}(sd) = 0.177(0.092), r^2_{Test} = 0.546 \quad (9)$$

No any chance correlation has been observed for the above model in the randomization study. This model is a reasonable robust QSAR model as evinced from the Q^2-index values greater than 0.5. The pK_b values of training set compounds have been calculated using Equation (9) and are mentioned in Table 1. This model has further validated externally with test set of ten compounds (Table 1). The value greater than 0.5 of test set r^2 (r^2_{Test}) reflects that the predictions of the test set compounds are satisfactory. The predicted activities of test set compounds are also given in Table 1. The goodness of fit between observed and calculated activities for the training and test set compounds is shown in Fig. 1.

Table 2. Dragon descriptor classes[a] used along with their definition and scope for modeling the binding affinity of tryptamine 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

[a]*Reference [28]*

Table 3. Intercorrelation matrix amongst independent variables of equations (6)-(9)

Eq. (6)	**PW4**	**GATS8v**	**H-046**			
PW4	1.000					
GATS8v	0.013	1.000				
H-046	0.204	0.286	1.000			
Eq. (7)	X4v	BELm2	MATS7p	H-052		
X4v	1.000					
BELm2	0.291	1.000				
MATS7p	0.198	0.412	1.000			
H-052	0.217	0.023	0.171	1.000		
Eq. (8)	X2A	BELm2	GGI3	JGI2	MATS7p	
X2A	1.000					
BELm2	0.048	1.000				
GGI3	0.008	0.213	1.000			
JGI2	0.073	0.103	0.061	1.000		
MATS7p	0.013	0.412	0.000	0.000	1.000	
Eq. (9)	nBR	PW4	BEHm7	BELm2	BEHe2	MATS7p
nBR	1.000					
PW4	0.135	1.000				
BEHm7	0.186	0.024	1.000			
BELm2	0.228	0.116	0.262	1.000		
BEHe2	0.055	0.025	0.276	0.011	1.000	
MATS7p	0.408	0.089	0.021	0.412	0.420	1.000

[a]*The matrix elements are the r^2-values*

The descriptors PW4, MATS7p and BELm2, which were emerged in models discussed earlier, have once again shown their importance in this six parameter model and convey same inferences to the activity. The newly emerged BCUT class descriptors are BEHm7 (the highest eigenvalue n.7 of atomic masses weighted Burden Matrix) and BEHe2 (the highest eigenvalue n.2 of atomic Sandersons electronegativities weighted Burden Matrix). The negative and positive signs of regression coefficients of descriptors BEHm7 and BEHe2, respectively, suggest that a lower value of descriptor BEHm7 and a higher value of descriptor BEHe2 would be beneficial to enhance the activity.

The remaining descriptor nBR, representing number of bromine atoms, is from CONST class. The positive correlation of it to the activity advocates that presence of more number of bromine atoms in a molecular structure is advantageous for the binding affinities of tryptamine derivatives. Thus the descriptors identified for rationalizing the activity may offer opportunities to modulate the structure to a desirable biological end point. Equation (9) has further been used to explore some new tryptamine derivatives as 5-HT_6 receptor ligands. The potential structural variations and predicted activity values using Eq. (9) for these compounds are given in Table 6.

A PLS (partial least squares) analysis [41-43] has been carried out on the 13 descriptors which were emerged in Eqs. (6) to (9) to recognize their potential in explaining the 5-HT_6 receptor binding affinities of tryptamine derivatives. This analysis also provides an opportunity to make a comparison of the relative significance among the descriptors. The fraction contributions obtainable from the normalized regression coefficients of the descriptors allow this comparison within the modeled activity. The descriptors have been autoscaled (zero mean and unit s.d.) to give each one of them equal weight in the PLS analysis. The PLS cross-validation found three components to be the optimum for these 13 descriptors and 75.52% variance in the activity has been explained by them.

The MLR-like PLS coefficients of these descriptors are given in Table 5. The calculated activity values of training and test set compounds are in close agreement to that of the observed ones (Table 1). The comparison of goodness of fit between observed and calculated activities of the training and test set compounds (through PLS analysis) is also presented in Fig. 1. The plot of the fraction contribution of normalized

regression coefficients of these descriptors to the activity is shown in Fig. 2.

The BCUT class descriptor BELm2 emerged as the most determining descriptor for modeling the 5-HT_6 binding activity of the titled compounds in PLS analysis. The other descriptors in decreasing order of significance are H-046, GATS8v, PW4, nBR, X4v, BEHm7, H-052, MATS7p, JGI2, GGI3, BEHe2 and X2A (Table 5; Fig. 2).

Table 4. Descriptors[a] identified for modeling the binding affinity of tryptamine derivatives along with the average regression coefficient[b], standard deviation and the total incidence

No	Descriptor	Physical meaning (Class)	Avg reg coeff (sd) incidence
1	RBN	Number of rotatable bonds (CONST)	-2.174(0.629)9
2	nCL	Number of chlorine atoms (CONST)	1.782(0.000)1; -1.121(0.000)1
3	nBR	Number of bromine atoms (CONST)	0.570(0.139)7
4	SPI	Superpendentic index (TOPO)	-1.180(0.239)4
5	TI2	Second Mohar index (TOPO)	-1.361(0.147)7
6	X2A	Average connectivity index chi-2 (TOPO)	-1.554(0.348)4
7	PW4	Path/walk 4- Randic shape index (TOPO)	1.740(0.190)9
8	DECC	Eccentric (TOPO)	-1.606(0.000)1
9	BEHm7	Highest eigenvalue n. 7 of Burden matrix/ weighted by atomic masses (BCUT)	-1.620(0.136)4
10	BEHm8	Highest eigenvalue n. 8 of Burden matrix/ weighted by atomic masses (BCUT)	-0.910(0.171)3
11	BELm2	Lowest eigenvalue n. 2 of Burden matrix/ weighted by atomic masses (BCUT)	3.925(1.126)18
12	BELm6	Lowest eigenvalue n. 6 of Burden matrix/ weighted by atomic masses (BCUT)	0.707(0.000)1
13	BEHv5	Highest eigenvalue n. 5 of Burden matrix/ weighted by atomic van der Waals volumes (BCUT)	-0.891(0.230)3
14	BEHv6	Highest eigenvalue n. 6 of Burden matrix/ weighted by atomic van der Waals volumes (BCUT)	-0.814(0.221)5
15	BELv3	Lowest eigenvalue n. 3 of Burden matrix/ weighted by atomic van der Waals volumes (BCUT)	1.082(0.000)1
16	BEHe2	Highest eigenvalue n. 2 of Burden matrix/ weighted by atomic Sanderson electronegativities (BCUT)	5.281(1.247)4
17	GGI2	Topological charge index of order 2 (GALVEZ)	1.819(0.622)4
18	GGI3	Topological charge index of order 3 (GALVEZ)	1.239(0.063)3
19	GGI10	Topological charge index of order 10 (GALVEZ)	-2.247(0.619)2
20	JGI2	Mean topological charge index of order 2 (GALVEZ)	1.133(0.297)2
21	JGI3	Mean topological charge index of order 3 (GALVEZ)	1.246(0.233)4
22	MATS5m	Moran autocorrelation - lag 5 / weighted by atomic masses (2D-AUTO)	0.974(0.000)1; -1.066(0.000)1
23	MATS8m	Moran autocorrelation - lag 8 / weighted by atomic masses (2D-AUTO)	-1.878(0.742)13
24	MATS7p	Moran autocorrelation - lag 5 / weighted by atomic polarizabilities (2D-AUTO)	2.563(1.062)14
25	GATS2v	Geary autocorrelation - lag 2 / weighted by atomic van der Waals volumes (2D-AUTO)	-1.107(0.000)1
26	GATS8p	Geary autocorrelation - lag 8 / weighted by atomic polarizabilities (2D-AUTO)	1.414(0.340)2
27	C-008	CHR2X (ACF)	-0.792(0.201)4
28	H-046	H attached to C0(sp3) with no X attached to next C (ACF)	2.479(0.384)6
29	H-052	H attached to C0(sp3) with 1X attached to next C (ACF)	-0.963(0.359)5

[a]*The descriptors are identified from the five parameter models emerged from CP-MLR protocol with filter-1 as 0.79; filter-2 as 2.0; filter-3 as 0.83; filter-4 as 0.3 ≤ Q^2 ≤ 1.0; number of compounds in the study are 33*

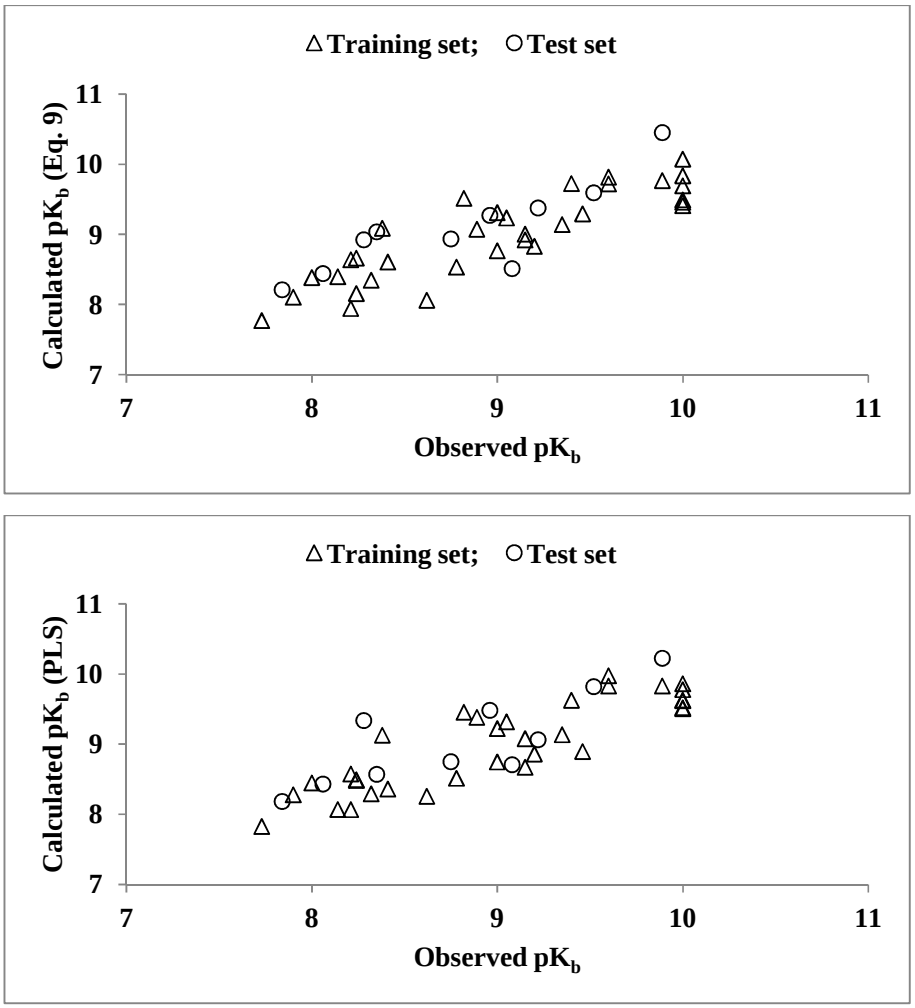


Fig. 1. Plot of observed versus calculated pK$_b$ values of the 1-aryl sulfonyl tryptamines

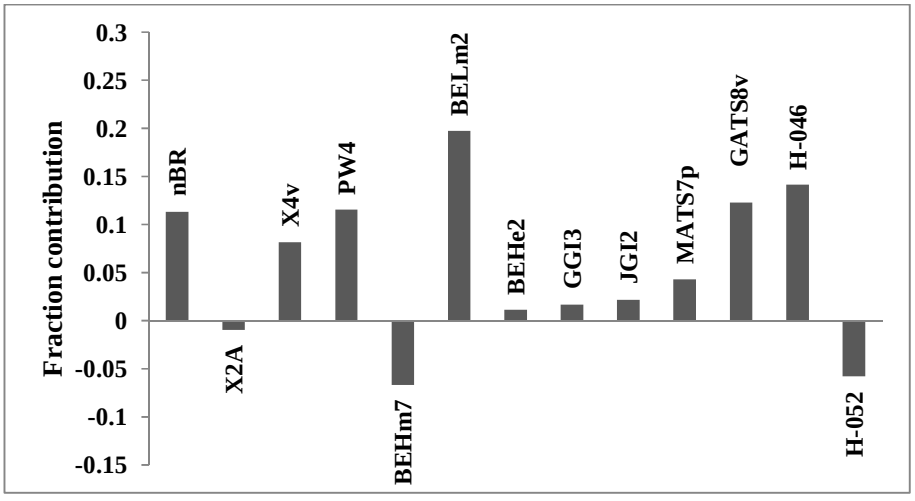


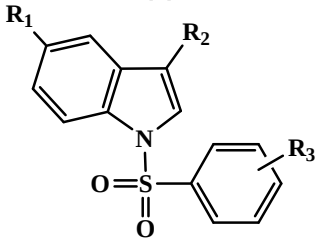
Fig. 2. Plot of fraction contribution of MLR-like PLS coefficients (normalized) of the 13 descriptors (Table 5) to the activity

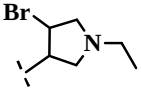
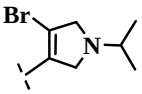
Table 5. PLS and MLR-like PLS models from the descriptors of Eqs. 6-9 for the 5-HT$_6$ binding affinity

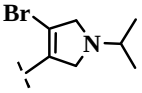
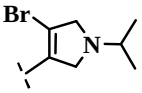
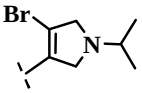
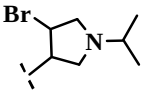
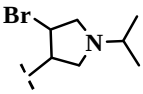
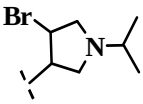
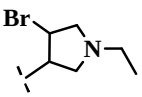
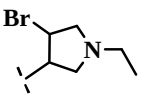
A: PLS equation			
PLS components	PLS coefficient (s.e.)[a]		
Component-1	-0.521(0.060)		
Component-2	-0.130(0.041)		
Component-3	-0.109(0.047)		
Constant	8.992		
B: MLR-like PLS equation			
S. No.	**Descriptor**	**MLR-like coefficient (f.c.)[b]**	**order**
1	nBR	0.346(0.113)	5
2	X2A	-0.029(-0.009)	13
3	X4v	0.250(0.081)	6
4	PW4	0.353(0.115)	4
5	BEHm7	-0.205(-0.067)	7
6	BELm2	0.604(0.197)	1
7	BEHe2	0.035(0.011)	12
8	GGI3	0.051(0.016)	11
9	JGI2	0.067(0.021)	10
10	MATS7p	0.132(0.043)	9
11	GATS8v	0.375(0.122)	3
12	H-046	0.432(0.141)	2
13	H-052	-0.177(-0.57)	8
	Constant	5.489	
C: PLS regression statistics		**Values**	
n		33	
r		0.869	
s		0.372	
F		29.932	
Q^2_{LOO}		0.672	
Q^2_{L5O}		0.687	
r^2_{Test}		0.522	

[a]*Regression coefficient of PLS factor and its standard error.* [b]*Coefficients of MLR-like PLS equation in terms of descriptors for their original values; f.c. is fraction contribution of regression coefficient, computed from the normalized regression coefficients obtained from the autoscaled (zero mean and unit s.d.) data*

Table 6. The structures and predicted activity of compounds based on QSAR model Equation (9)



Cpd.	**R_1**	**R_2**	**R_3**	**pKb**
1	F		3-C_2H_5	11.37
2	F		3-OCH_3	10.19

Cpd.	R_1	R_2	R_3	pKb
3	Br		3-OCH_3	10.52
4	F		3-$CH(CH_3)_2$	12.83
5	Br		3-$CH(CH_3)_2$	13.25
6	OCH_3		3-OCH_3	9.42
7	OC_2H_5		3-$CH(CH_3)_2$	12.13
8	$OCH(CH_3)_2$		3-$CH(CH_3)_2$	12.03
9	F		3-OCH_3	10.16
10	F		3-$CH(CH_3)_2$	12.26

4. CONCLUSIONS

In conclusion, the present study has provided structure–activity relationships of the binding affinities of tryptamine derivatives to 5-HT_6 receptor in terms of structural requirements. The binding affinity has, therefore become the function of the cumulative effect of different structural features which were identified in terms of individual descriptors.

In order to improve the 5-HT_6 receptor binding affinity of a compound higher value of molecular topology and symmetry accounting Randic shape index descriptor PW4 (path/walk 4) is favorable. Presence of more number of bromine atoms (descriptor nBR) and presence of such structural fragment in which a hydrogen atom attached to sp^3 hybridized carbon with no hetero atom rather than one hetero atom attached to next carbon atom (descriptors H-046 and H-052) will be supportive to the activity.

The associations of polarizability to the path length 7 of Moran autocorrelation (MATS7p), masses to eigenvalues n.2 and 7 of Burden matrix (BELm2 and BEHm7), Sanderson electronegativity to highest eigenvalue n.2 Burden matrix (BEHe2) and van der Waals volume to path length 8 of Geary autocorrelation (GATS8v) have shown the prevalence of atomic properties and charge content in terms of topological and mean topological charge indices (GGI3 and JGI2) to explain the binding affinity. The dominance of the information content of the descriptors, emerged in CP-MLR models, has also confirmed by the PLS analysis.

The derived QSAR models and descriptors shared in these models revealed that the substituents of tryptamine moiety have sufficient scope for further modification.

CONSENT

It is not applicable.

ETHICAL APPROVAL

It is not applicable.

ACKNOWLEDGEMENTS

MC thanks CSIR, New Delhi for the financial support provided in the form of SRF.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Rosse G, Schaffhauser H. 5-HT6 receptor antagonists as potential therapeutics for cognitive impairment. Curr Top Med Chem. 2010;10:207-221.
2. Liu KG, Robichaud AJ. 5-HT6 medicinal chemistry. Int Rev Neurobiol. 2010;94:1-34.
3. Alexandre VI, Yan AI, Sergey ET. 5-hydroxytryptamine subtype 6 receptor modulators: A patent survey. Exp Opin Ther Patents. 2010;20:1171-1196.
4. Heal D, Gosden J, Smith S. The 5-HT6 receptor as a target for developing novel antiobesity drugs. Int Rev Neurobiol. 2011; 96:73-109.
5. Heal DJ, Smith SL, Fisas A, Codony X, Buschmann H. Selective 5-HT6 receptor ligands: progress in the development of a novel pharmacological approach to the treatment of obesity and related metabolic disorders. Pharmacol Ther. 2008;117:207-231.
6. Kohen R, Metcalf MA, Khan N, Druck T, Huebner K, Lachowicz JE, Meltzer HY, Sibley DR, Roth BL, Hamblin MW. Cloning, characterization, and chromosomal localization of a human 5-HT6 serotonin receptor. J Neurochem. 1996;66:47-56.
7. Ruat M, Traiffort E, Arrang J-M, Tardivel-Lacombe J, Diaz J, Leurs R, Swartz J-C. A novel rat serotonin (5-HT6) receptor: molecular cloning, localization and stimulation of cAMP accumulation. Biochem. Biophys. Res. Commun. 1993; 193:268-76.
8. Sebben M, Ansanay H, Bockaert J, Dumuis A. 5-HT6 receptors positively coupled to adenylyl cyclase in striatal neurones in culture. Neuroreport. 1994; 5:2553-2557.
9. Arnt J, Olsen CK. 5-HT6 receptor ligands and their antipsychotic potential. Int Rev Neurobiol. 2011;96:141-161.
10. Roth BL, Craig SC, Choudhary MS, Uluer A, Monsma FJ, Shen Y, Meltzer HY, Sibley DR. Binding of typical and atypical antipsychotic agents to 5-hydroxytryptamine-6 and 5-hydroxytryptamine-7 receptors. J Pharmacol Exp Ther. 1994;268:1403-1410.
11. Witty D, Ahmed M, Chuang TT. Advances in the design of 5-HT6 receptor ligands with therapeutic potential. Prog Med Chem. 2009;48:163-224.
12. Hirano K, Piers TM, Searle KL, Miller ND, Rutter AR, Chapman PF. Procognitive 5-HT6 antagonists in the rat forced swimming test: Potential therapeutic utility in mood disorders associated with Alzheimer's disease. Life Sci. 2009; 84:558-562.
13. Emsley R. Drugs in development for the treatment of schizophrenia. Exp Opin Invest Drugs. 2009;18:1103-1118.
14. Geldenhuys WJ, Van der Schyf CJ. Serotonin 5-HT6 receptor antagonists for the treatment of alzheimers disease. Curr Top Med Chem. 2008;8:1035-1048.
15. Fone KCF. An update on the role of the 5-hydroxytryptamine6 receptor in cognitive function. Neuropharmacology. 2008; 55:1015-1022.
16. Rodefer JS, Nguyen TN, Karlsson JJ, Arnt J. Reversal of subchronic PCP-induced deficits in attentional set shifting in rats by sertindole and a 5-HT6 receptor antagonist: Comparison among antipsychotics. Neuropsychopharmacology.2008;33:2657-2666.
17. Jones CA, McCreary AC. Serotonergic approaches in the development of novel antipsychotics. Neuropharmacology. 2008; 55:1056-1065.
18. Zvartau-Hind M, Mather-Edwards G, Hunter J, Gold M, Hopton G, Davy M, Williams P. 11th International Conference of Alzheimer's Disease (ICAD), Chicago. 2008; Abstract 03-04-06.
19. Nirogi R, Kambhampati R, Shinde A, Kandikere V, Mudigonda K, Bhyrapuneni

G, Jayarajan P, Abraham R, Mulla S, Jasti V. 12th International Conference of Alzheimer's Disease (ICAD), Vienna. 2009; Abstract 250.

20. Arnt J, Andersen BB, Bymaster FP, Chohen MP. 2nd Biennial Schizophrenia international Research Conference (SIRS), Florence; 2010.

    Available:http://clinicaltrials.gov/ct2/show/NCT00948662.

21. Synosis Therapeutics. 2008.

    Available:http://www.synosia.com

22. Tsai Y, Dukat M, Slassi A, MacLean N, Demchyshyn L, Savage JE, Roth BL, Hufesein S, Lee M, Glennon RA. N1-(Benzenesulfonyl)tryptamines as novel 5-HT6 antagonists. Bioorg Med Chem Lett. 2000;10:2295-2299.

23. Glennon RA, Lee M, Rangisetty JB, Dukat M, Roth BL, Savage JE, McBride A, Rauser L, Hufeisen L, Lee DKH. 2-Substituted tryptamines: agents with selectivity for 5-HT6 serotonin receptors. J Med Chem. 2000;43:1011-1018.

24. Mooradian A, Dupont PE, Hlavec AG, Aceto MD, Pearl J. 3-Aminotetrahydrocarbazoles as a new series of central nervous system agents. J Med Chem. 1977;20:487-492.

25. Nirogi R, Dwarampudi A, Kambhampati R, Bhatta V, Kota L, Shinde A, Badange R, Jayarajan P, Bhyrapuneni G, Dubey PK. Rigidized 1-aryl sulfonyl tryptamines: Synthesis and pharmacological evaluation as 5-HT6 receptor ligands. Bioorg Med Chem Lett. 2011;21:4577-4580.

26. Chemdraw ultra 6.0 and Chem3D ultra, Cambridge Soft Corporation, Cambridge, USA.

    Available: http://www.cambridgesoft.com

27. Dragon software (version 1.11-2001) by Todeschini R, Consonni V. Milano, Italy.

28. Available:http//www.talete.mi.it/dragon.htm

29. Available:http://www.strandls.com/sarchitect/documents/manual_html/desctheory.html

30. Prabhakar YS. A combinatorial approach to the variable selection in multiple linear regression: Analysis of Selwood, et al. Data Set – A Case Study. QSAR Comb Sci. 2003;22:583-595.

31. Sharma BK, Sharma SK, Singh P, Sharma S, Prabhakar YS. 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. 2009;24:86-93.

32. Sharma BK, Pilania P, Sarbhai K, Singh P, Prabhakar YS. Chemometric descriptors in modeling the carbonic anhydrase inhibition activity of sulfonamide and sulfamate derivatives. Mol Divers. 2010;14:371-384.

33. Sharma BK, Pilania P, Singh P, Prabhakar YS. A QSAR study on 2-(4-Methylpiperazin-1-yl) quinoxalines as human histamine H_4 receptor ligands. J Enzyme Inhib Med Chem. 2011;26:412-421.

34. Akaike H. Information theory and an extension of the minimum likelihood principle. In: Petrov BN, Csaki F. Editors. Second international symposium on information theory. Akademiai Kiado, Budapest. 1973;267-281.

35. Akaike H. A new look at the statistical identification model. IEEE Trans Automat Control. 1974;AC-19:716-23.

36. Kubinyi H. Variable selection in QSAR studies. I. An evolutionary algorithm. Quant Struct-Act Relat. 1994;13:285-94.

37. Kubinyi H. Variable selection in QSAR studies. II. A highly efficient combination of systematic search and evolution. Quant Struct-Act Relat. 1994;13:393-401.

38. Friedman J. In: Technical report no 102. Laboratory for computational statistics Stanford. Stanford University; 1990.

39. So S-S, Karplus M. Three-dimensional quantitative structure-activity relationship from molecular similarity matrices and genetic neural networks. 1. Method and validation. J Med Chem. 1997;40:4347-4359.

40. Prabhakar YS, Solomon VR, Rawal RK, Gupta MK, Katti SB. 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. 2004;23:234-244.

41. Wold S. Cross-validatory estimation of the number of components in factor and principal components models. Technometrics. 1978;20:397–405.

42. Kettaneh N, Berglund A, Wold S. PCA and PLS with very large data sets. Comput Statist Data Anal. 2005;48:69–85.

43. Stahle L, Wold S. Multivariate Data Analysis and Experimental Design, in Biomedical Research, Progress in Medicinal Chemistry, , G.P. Ellis and W.B. West, eds., Elsevier Science Publishers, BV, The Netherlands. 1998;25:291–338.

Peer-review history:
The peer review history for this paper can be accessed here:
http://sciencedomain.org/review-history/11208