

CP-MLR/PLS directed quantitative structure-activity relationship study on the histamine H₃ receptor binding affinity: The cyclohexylamine based series

Shahista Chuhann and Brij Kishore Sharma *

Department of Chemistry, Government College, Bundi-323 001, Rajasthan, India.

World Journal of Biology Pharmacy and Health Sciences, 2023, 13(02), 089–104

Publication history: Received on 28 December 2022; revised on 01 February 2023; accepted on 04 February 2023

Article DOI: <https://doi.org/10.30574/wjbphs.2023.13.2.0074>

Abstract

The histamine H₃ receptor binding affinities of cyclohexylamine derivatives has been analysed with the topological and molecular features from Dragon software. Analysis of the structural features in conjunction with the biological endpoints in combinatorial protocol in multiple linear regression (CP-MLR) led to the identification of 26 descriptors for modelling the activity. The study clearly suggested the role of atomic properties such as mass, electronegativity or charge content, polarizability, atomic van der Waals volume, average valence connectivity index chi-5 and absence of number of acceptor atoms for H-bonds (N, O, F) type functionality to optimise the histamine H₃ receptor binding affinity of titled compounds. The models developed and the participating descriptors advocate that the substituent groups of the cyclohexylamine moiety hold scope for further modification in the optimization of the H₃ receptor binding affinity.

Analysis of these descriptors in partial least squares (PLS) highlighted their relative significance in modulating the biological response. The selected descriptors are enriched with information corresponding to the activity when compared to the remaining ones. 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.

Keywords: QSAR; Histamine H₃ receptor; Binding affinity; Combinatorial protocol in multiple linear regression (CP-MLR) analysis; PLS; Dragon descriptors; Cyclohexylamine derivatives

1. Introduction

The histamine H₃ receptor (H₃R) which is a G-protein-coupled receptor was discovered by Arrang et al. [1] in 1983 and cloned by Lovenberg et al. [2] in 1999. Extensive studies and reviews on H₃R revealed that this receptor mostly controls the histamine biosynthesis and release via a negative-feedback process [3-10]. It is found that the release of several other neurotransmitters such as acetylcholine [11,12], noradrenaline [13], dopamine [14] and serotonin [15] regulated by H₃R. In the brain the release of waking and pro-cognitive action associated histamine is triggered by inverse agonists rather than neutral antagonists of H₃R [16]. Based on these observations H₃R inverse agonists might be useful in several CNS-related disorders that includes narcolepsy, attention deficit hyperactivity disorder (ADHD), schizophrenia, Alzheimer's disease, and excessive daytime sleepiness in obstructive sleep apnea or Parkinson's disease.

In search of potent H₃R inverse agonists, devoid of hERG and CYP450 interactions with favorable CNS penetration and pharmacokinetic profile, cyclohexylamine based derivatives have been reported by Labeeuw and coworkers [17]. The aim of present communication is to establish the quantitative relationships between the reported binding affinities and molecular descriptors unfolding the substitutional changes in titled compounds.

*Corresponding author: Brij Kishore Sharma

2. Material and methods

2.1. Data-set

For present work the reported sixty cyclohexylamine based derivatives have been considered as the data set [17]. The general structure of these compounds is represented in Figure 1 and structural variations are mentioned in Table 1.

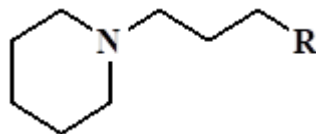
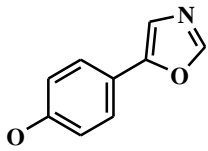
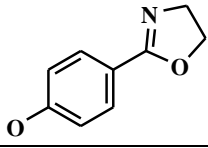
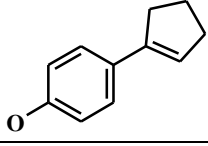
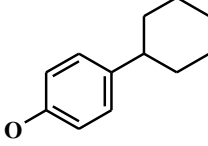
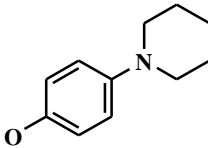
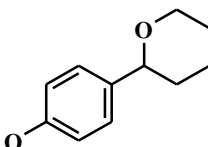
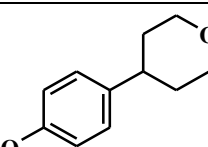
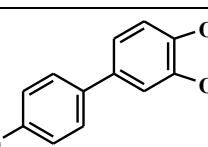
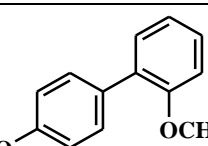
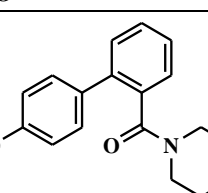
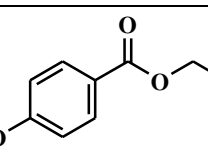


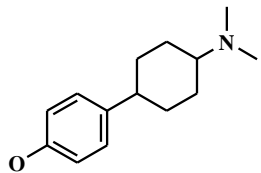
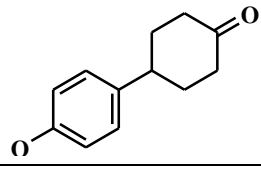
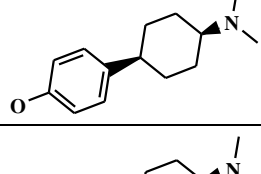
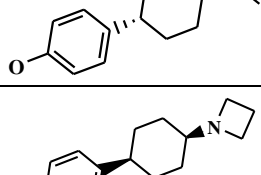
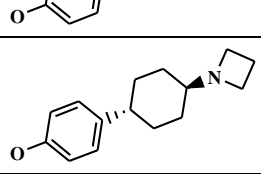
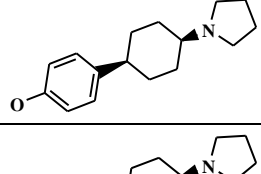
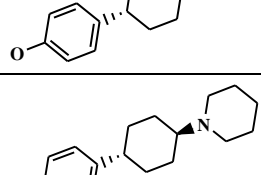
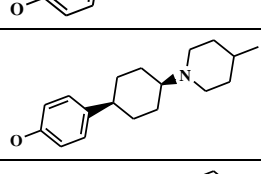
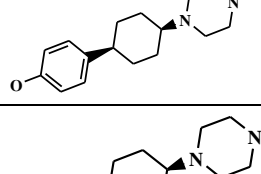
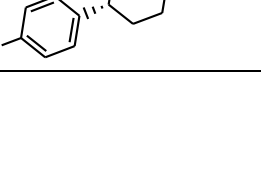

Figure 1 General structure of cyclohexylamine based derivatives

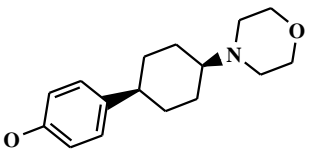
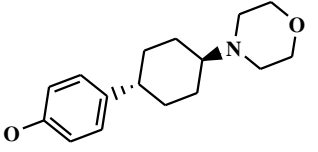
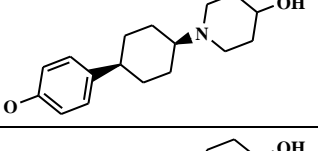
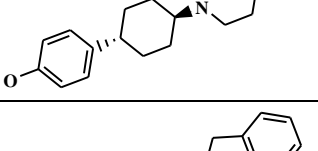
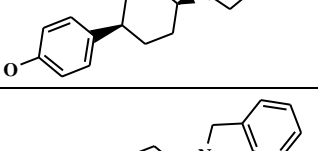
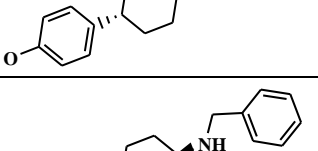
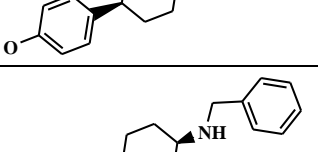
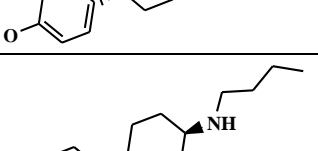
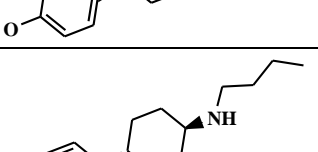
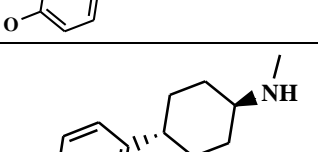

These compounds were evaluated for their histamine H₃ binding affinity by displacement of [¹²⁵I]-iodoproxyfan (IPX) binding to membranes of stably transfected HEK-293 cells [18]. The binding affinity has also been reported in Table 1 [17]. The same is expressed as pK_i on a molar basis and considered as the dependent variable for the present quantitative analysis. The data set was sub-divided into training set to develop models and test set to validate the models externally. The test set compounds which were selected using an in-house written randomization program, are also mentioned in Table 1.

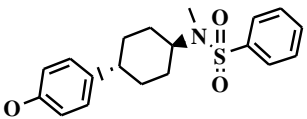
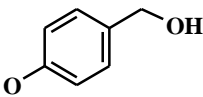
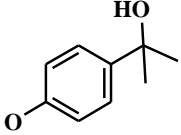
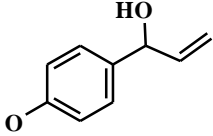
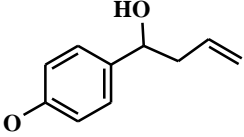
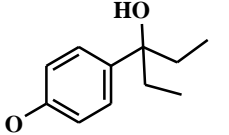
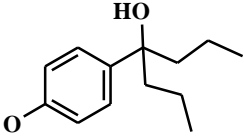
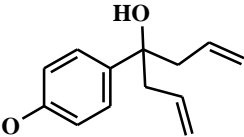
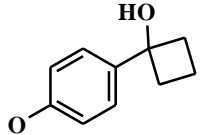
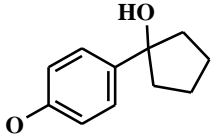
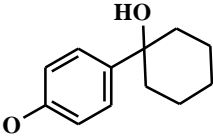
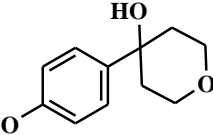
Table 1 Structural variations and observed H₃R binding affinity of cyclohexylamine derivatives

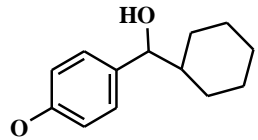
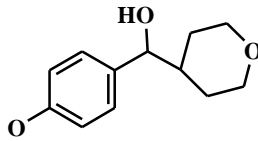
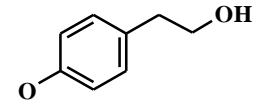
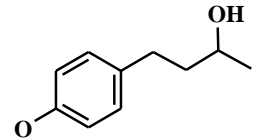
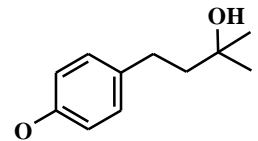
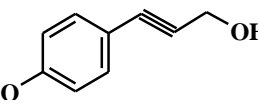
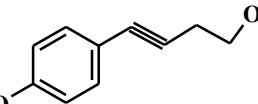
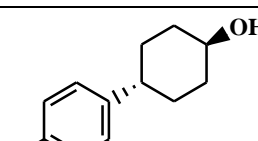
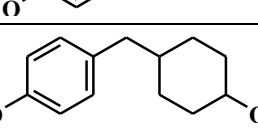
Cpd.	R	pK _i (M) ^a					
		Obsd.	Calculated				
			Eq. (4)	Eq. (5)	Eq. (6)	Eq. (7)	PLS
1		8.85	8.41	8.75	8.59	8.56	8.68
2 ^b		8.30	8.28	8.10	8.16	8.04	8.31
3		8.05	8.69	8.42	8.28	8.32	8.36
4 ^b		8.32	8.16	8.11	8.35	8.25	8.30
5 ^b		8.74	8.42	8.61	8.58	8.70	8.60

6 ^b		8.60	8.08	8.07	8.36	8.19	8.11
7 ^b		8.44	7.88	8.10	8.11	8.22	8.09
8		8.80	9.18	9.17	9.05	8.93	8.98
9		8.55	8.80	8.57	8.23	8.43	8.61
10		8.31	8.38	8.34	8.42	8.53	8.51
11		8.28	8.17	8.36	8.41	8.40	8.30
12		8.08	8.39	8.43	8.50	8.56	8.45
13		8.44	8.53	8.15	8.21	8.26	8.58
14		8.66	8.66	8.69	8.48	8.47	8.92
15 ^b		8.80	9.11	8.84	9.22	8.56	8.84
16		7.98	7.79	7.96	7.95	7.92	7.79

17		9.70	9.55	9.54	9.42	9.53	9.50
18		8.77	8.44	8.35	8.51	8.48	8.58
19		9.52	9.55	9.54	9.42	9.53	9.50
20		9.68	9.55	9.54	9.42	9.53	9.50
21		9.47	9.52	9.29	9.40	9.30	9.37
22		9.60	9.52	9.29	9.40	9.30	9.37
23		9.48	9.52	9.39	9.36	9.38	9.46
24		9.60	9.52	9.39	9.36	9.38	9.46
25		9.19	9.37	9.32	9.13	9.33	9.38
26 ^b		9.25	9.40	9.45	9.20	9.39	9.50
27		9.04	9.10	9.41	9.39	9.57	9.37
28 ^b		9.49	9.10	9.41	9.39	9.57	9.37

29		9.17	9.13	9.22	9.45	9.44	9.10
30		9.39	9.13	9.22	9.45	9.44	9.10
31		9.04	8.99	9.10	9.24	9.40	9.11
32		9.24	8.99	9.10	9.24	9.40	9.11
33		9.39	9.39	9.57	9.57	9.50	9.75
34		9.77	9.39	9.57	9.57	9.50	9.75
35		9.54	9.41	9.29	9.44	9.23	9.40
36 ^b		9.59	9.41	9.29	9.44	9.23	9.40
37		9.28	9.32	9.26	9.16	9.19	9.12
38		9.19	9.32	9.26	9.16	9.19	9.12
39 ^b		8.49	9.20	9.06	9.09	9.10	9.05

40		8.26	8.72	8.69	8.59	8.64	8.44
41		7.64	7.65	7.95	7.81	7.90	7.63
42		8.17	8.35	8.22	8.28	8.17	8.40
43 ^b		8.07	7.86	7.84	7.87	7.82	7.88
44 ^b		8.16	8.35	7.92	8.02	7.77	8.08
45		8.85	8.89	9.03	9.08	9.05	9.06
46		9.13	9.09	9.37	9.09	9.03	9.15
47		9.27	9.08	9.12	9.28	8.89	9.21
48		7.74	7.89	7.80	8.10	7.80	8.01
49		9.09	9.16	9.15	9.33	9.11	9.00
50		8.41	8.54	8.46	8.53	8.42	8.63
51		8.39	8.09	7.85	8.06	8.08	8.16

52		8.54	8.70	8.65	8.57	8.43	8.44
53		7.89	8.13	8.01	7.98	8.08	7.92
54		7.49	7.78	7.88	7.94	7.81	7.66
55		8.48	8.32	8.45	8.49	8.57	8.23
56 ^b		8.40	8.38	8.74	8.54	8.83	8.54
57 ^b		7.82	7.61	8.00	7.80	8.02	7.77
58 ^b		7.72	7.95	7.97	7.84	8.04	7.95
59		8.59	7.95	7.97	7.84	8.04	7.95
60		8.72	8.70	8.63	8.53	8.73	8.60

^aReference [17]; ^bCompound included in test set.

2.2. Molecular descriptors

The structures of the compounds (Table 1), under study, have been drawn in 2D ChemDraw [19] and were converted into 3D objects using the default conversion procedure implemented in the CS Chem3D Ultra. The generated 3D-structures of the compounds were subjected to energy minimization in the MOPAC module, using the AM1 procedure for closed shell systems, implemented in the CS Chem3D Ultra. This will ensure a well defined conformer relationship across the compounds of the study. All these energy minimized structures of respective compounds have been ported to DRAGON software [20] for computing the descriptors corresponding to 0D-, 1D-, and 2D-classes.

2.3. Development and validation of model

The combinatorial protocol in multiple linear regression (CP-MLR) [21-25] and partial least squares (PLS) [26-28] procedures have been used in the present work for developing QSAR models. The CP-MLR is a “filter”-based variable selection procedure, which employs a combinatorial strategy with MLR to result in selected subset regressions for the extraction of diverse structure-activity models, each having unique combination of descriptors from the generated dataset of the compounds under study. The embedded filters make the variable selection process efficient and lead to unique solution. Fear of “chance correlations” exists where large descriptor pools are used in multilinear QSAR/QSPR

studies [29,30]. Furthermore, in order to discover any chance correlations associated with the models recognized in CP-MLR, each cross-validated model has been put to a randomization test [31,32] by repeated randomization of the activity to ascertain the chance correlations, if any, associated with them. For this, every model has been subjected to 100 simulation runs with scrambled activity. The scrambled activity models with regression statistics better than or equal to that of the original activity model have been counted, to express the percent chance correlation of the model under scrutiny.

Validation of the derived model is necessary to test its prediction and generalization within the study domain. For each model, derived by involving n data points, a number of statistical parameters such as r (the multiple correlation coefficient), s (the standard deviation), F (the F ratio between the variances of calculated and observed activities), and Q^2_{LOO} (the cross-validated index from leave-one-out procedure) have been obtained to assess its overall statistical significance. In case of internal validation, Q^2_{LOO} is used as a criterion of both robustness and predictive ability of the model. A value greater than 0.5 of Q^2 index suggests a statistically significant model. The predictive power of derived model is based on test set compounds. The model obtained from training set has a reliable predictive power if the value of the r^2_{Test} (the squared correlation coefficient between the observed and predicted values of compounds from test set) is greater than 0.5.

2.4. 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 [33]. The 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 $\pm x$ (s.d.) 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.

3. Results and discussion

3.1. QSAR results

For the compounds in Table 1, a total number of 501 descriptors belonging to 0D- to 2D- classes of DRAGON have been computed. Prior to model development procedure, all those descriptors that are inter-correlated beyond 0.90 and showing a correlation of less than 0.1 with the biological endpoints (descriptor versus activity, $r < 0.1$) were excluded. This procedure has reduced the total descriptors from 501 to 120 as relevant ones to explain the biological actions of titled compounds and these were subjected to CP-MLR analysis with default “filters” set in it. The descriptors have been scaled between the intervals 0 to 1 [34] to ensure that a descriptor will not dominate simply because it has larger or smaller pre-scaled value compared to the other descriptors. In this way, the scaled descriptors would have equal potential to influence the QSAR models. 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.

The 60 compounds were divided into training-set and test-set. Fifteen compounds (25% of total population) have been selected for test-set. The identified test-set was then used for external validation of models derived from remaining forty five compounds in the training-set. The squared correlation coefficient between the observed and predicted values of compounds from test-set, r^2_{Test} , was calculated to explain the fraction of explained variance in the test-set which is not part of regression/model derivation. It is a measure of goodness of the derived model equation. A high r^2_{Test} value is always good. But considering the stringency of test-set procedures, often r^2_{Test} values in the range of 0.5 to 0.6 are regarded as logical models. Following the strategy to explore only predictive models, CP-MLR resulted into 02, 54 and 55 models in one, two and three descriptors, respectively. The generated models in one, two and three descriptors, all having $r^2_{Test} < 0.5$, for the CDK8 inhibitory activity. The selected models are mentioned in Table 2.

The signs of the regression coefficients have indicated the direction of influence of explanatory variables in above models. The positive regression coefficient associated to a descriptor will augment the activity profile of a compound while the negative coefficient will cause detrimental effect to it.

Table 2 Highest significant models in one, two and three parameters derived for training set through CP-MLR for histamine H₃ receptor binding affinity

Model	r	s	F	Q ² _{L00}	r ² _{Test}	Eq.
pKi = 8.145 +1.227(0.138)BELp5	0.803	0.369	78.578	0.610	0.807	(1)
pKi = 7.663 +1.155(0.115)BELp5 +0.937(0.202)GATS7e	0.875	0.304	68.672	0.724	0.806	(2)
pKi = 8.470 +1.019(0.144)BELp5 +0.854(0.239)JGI6 -1.507(0.247)nHAcc	0.903	0.272	60.811	0.766	0.561	(3)

In above model Eqs., (1-3), the descriptor BELp5 is BCUT class descriptor. The other participating descriptors are GATS7e (2D autocorrelation class), JGI6 (Galvez topological charge index) and nHAcc (functional group). The positive sign of regression coefficients of descriptors BELp5 (atomic polarizabilities weighted lowest eigenvalue n.5 of Burden matrix), GATS7e (atomic Sanderson electronegativities weighted Geary autocorrelation of lag 7) and JGI6 (mean topological charge index of order 6) suggested that a higher value of these descriptors would be beneficial to augment the H₃R binding affinity. On the other hand, absence of number of acceptor atoms for H-bonds (N, O, F) type functionality (descriptor nHAcc) in a molecule would be supportive to the H₃R binding affinity.

Table 3 Identified descriptors^a along with their physical meaning, average regression coefficient and incidence^b, in modeling the H₃R binding affinities

Descriptor class	Descriptor (physical meaning), avg reg coeff (incidence)
Constitutional	Me (mean atomic Sanderson electronegativity scaled on Carbon atom), -1.029(42)
Topological	DELS (molecular electrotopological variation), -0.916(8); X5Av (average valence connectivity index chi-5), -0.661(4); BLI (Kier benzene-likeness index), -0.771(9); CIC5 (complementary information content of 5-order neighbourhood symmetry), 0.429(2); BIC5 (bond information content of 5-order neighbourhood symmetry), -0.379(7); SEigm (Eigenvalue sum from mass weighted distance matrix), -0.668(2); piPC10 (molecular multi path count of order 10), 0.902(14); T(O..O) (sum of topological distances between O..O), -0.505(5)
BCUT	BELp5 (atomic polarizabilities weighted lowest eigenvalue n.5 of Burden matrix), 0.946(59); BELp7 (atomic polarizabilities weighted lowest eigenvalue n.7 of Burden matrix), 0.568(10)
Galvez topological charge indices	GGI1 (topological charge index of order 1), 0.848(2); GGI10 (Galvez topological charge index of order 10), -0.854(2); JGI6 (mean topological charge index of order 6), 0.800(3)
2D autocorrelations	MATS6m (atomic masses weighted Moran autocorrelation of lag 6), 0.440(2); MATS8m (atomic masses weighted Moran autocorrelation of lag 8), -0.316(1); MATS2v (atomic van der Waals volume weighted Moran autocorrelation of lag 2), 0.442(12); MATS2e (atomic Sanderson electronegativities weighted Moran autocorrelation of lag 2), -0.864(1); GATS6v (atomic van der Waals volume weighted Geary autocorrelation of lag 6), -0.288(1); GATS3e (atomic Sanderson electronegativities weighted Geary autocorrelation of lag 3), 0.820(2); GATS5e (atomic Sanderson electronegativities weighted Geary autocorrelation of lag 5), -0.538(14); GATS7e (atomic Sanderson electronegativities weighted Geary autocorrelation of lag 7), 0.547(5); GATS8p (atomic polarizabilities weighted Geary autocorrelation of lag 8), 0.507(8)
Functional groups	nNR2 (number of tertiary aliphatic amines), 0.533(1); nHAcc (number of acceptor atoms for H-bonds (N, O, F), -0.941(15)
Properties	MLOGP (Moriguchi octanol-water partition coefficient, logP), 0.692(5)

^aThe descriptors are identified from the four parameter models for activity emerged from CP-MLR protocol with filter-1 as 0.79, filter-2 as 2.0, filter-3 as 0.896 and filter-4 as $0.3 \leq q^2 \leq 1.0$ with a training set of 45 compounds. ^bThe average regression coefficient of the descriptor corresponding to all models and the total number of its incidence. The arithmetic sign of the coefficient represents the actual sign of the regression coefficient in the models.

Considering the number of observation in the dataset, models with up to four descriptors were explored. It has resulted in 59 four-parameter models with test set $r^2 > 0.50$. These models (with 120 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. These models have shared 26 descriptors among them. All these 26 descriptors along with their brief meaning, average regression coefficients, and total incidence are listed in Table 3, which will serve as a measure of their estimate across these models.

Following are the selected four-descriptor models for the histamine H₃ receptor binding affinities emerged through CP-MLR.

$$pK_i = 7.814 + 1.027(0.122)BELp5 + 0.714(0.176)BELp7 + 0.616(0.169)GATS8p - 1.113(0.205)nHAcc$$

$$n = 45, r = 0.921, s = 0.250, F = 56.114, Q^2_{L00} = 0.799, Q^2_{L50} = 0.804, r^2_{Test} = 0.687 \dots\dots(4)$$

$$pK_i = 8.460 - 0.864(0.207)Me + 1.045(0.123)BELp5 + 0.512(0.138)MATS2v - 0.712(0.244)nHAcc$$

$$n = 45, r = 0.921, s = 0.250, F = 56.017, Q^2_{L00} = 0.796, Q^2_{L50} = 0.791, r^2_{Test} = 0.762 \dots\dots(5)$$

$$pK_i = 8.803 - 1.316(0.204)Me - 0.842(0.244)X5Av + 0.986(0.107)BELp5 + 0.598(0.169)GATS8p$$

$$n = 45, r = 0.919, s = 0.253, F = 54.433, Q^2_{L00} = 0.797, Q^2_{L50} = 0.787, r^2_{Test} = 0.837 \dots\dots(6)$$

$$pK_i = 8.432 - 1.011(0.190)Me + 0.904(0.109)BELp5 + 0.451(0.141)MATS2v - 0.484(0.179)GATS5e$$

$$n = 45, r = 0.919, s = 0.253, F = 54.378, Q^2_{L00} = 0.795, Q^2_{L50} = 0.776, r^2_{Test} = 0.737 \dots\dots(7)$$

These models have accounted for nearly 85% variance in the observed activities. In the randomization study (100 simulations per model), none of the identified models has shown any chance correlation. 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 Eqs. (4) to (7) have been included in Table 1. The models (4) to (7) are validated with an external test set of 15 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 same is 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 2.

The newly appeared descriptors in above models are BELp7 (BCUT class), Me (constitutional class), GATS8p, GATS5e and MATS2v (2D autocorrelation class) and X5Av (topological class). The descriptors BELp7, GATS8p and MATS2v have shown positive correlation to the activity whereas descriptors Me, X5Av and GATS5e have correlated negatively to the activity. The signs of regression coefficients advocated that higher values of atomic polarizabilities weighted lowest eigenvalue n.7 of Burden matrix (descriptor BELp7), atomic polarizabilities weighted Geary autocorrelation of lag 8 (descriptor GATS8p) and atomic van der Waals volume weighted Moran autocorrelation of lag 2 (descriptor MATS2v) and lower values of mean atomic Sanderson electronegativity scaled on Carbon atom (descriptor Me), average valence connectivity index chi-5 (X5Av) and atomic Sanderson electronegativities weighted Geary autocorrelation of lag 5 (descriptor GATS5e) would be helpful to augment the histamine H₃ receptor binding affinity.

A partial least square (PLS) analysis has been carried out on these 26 CP-MLR identified descriptors, mentioned in Table 3, to facilitate the development of a “single window” structure–activity model. For the purpose of PLS, the descriptors have been autoscaled (zero mean and unit SD) to give each one of them equal weight in the analysis. In the PLS cross-validation, four components are found to be the optimum for these 26 descriptors and they explained 88.17% variance in the activity. The MLR-like PLS coefficients of these 26 descriptors are given in Table 4.

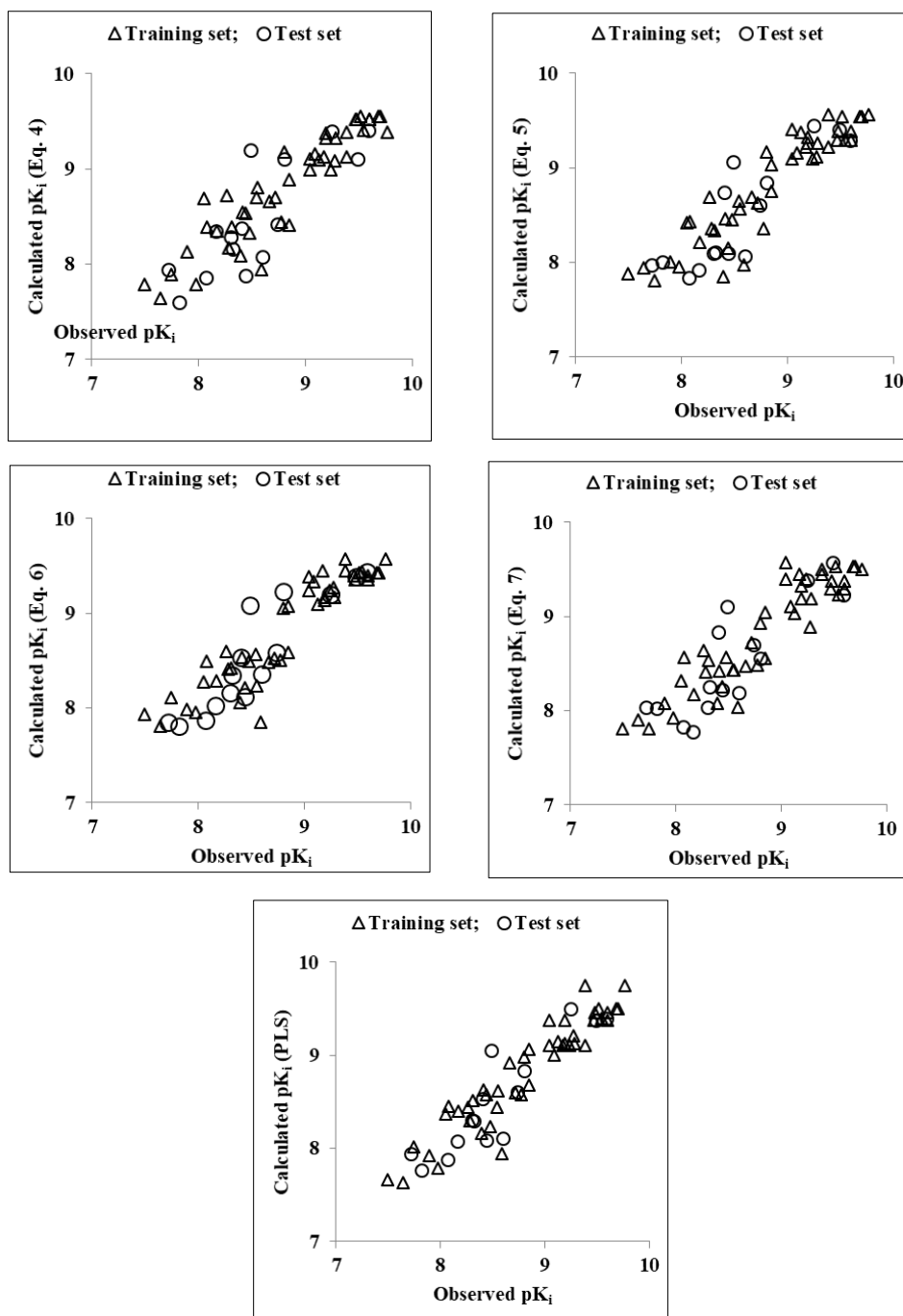


Figure 2 Plot of observed versus calculated pK_i values for training- and test-set compounds for histamine H_3 receptor binding affinity

Table 4 PLS and MLR-like PLS models from the 26 descriptors of four parameter CP-MLR models for histamine H3 receptor binding affinity

A: PLS equation									
PLS components					PLS coefficient (s.e.)^a				
Component-1					-0.194(0.012)				
Component-2					-0.101(0.017)				
Component-3					0.044(0.016)				
Component-4					0.083(0.024)				
Constant					8.816				
B: MLR-like PLS equation									
S. No.	Descriptor	MLR-like coefficient ^b	(f.c.) ^c	Order	S. No.	Descriptor	MLR-like coefficient ^b	(f.c.) ^c	Order
1	Me	-0.484	-0.091	3	14	JGI6	0.174	0.035	10
2	DELS	-0.017	-0.002	26	15	MATS6m	0.366	0.057	5
3	X5Av	-0.146	-0.023	17	16	MATS8m	0.131	0.029	12
4	BLI	-0.174	-0.025	16	17	MATS2v	0.121	0.027	15
5	CIC5	0.101	0.019	19	18	MATS2e	-0.319	-0.042	7
6	BIC5	-0.055	-0.012	25	19	GATS6v	-0.060	-0.013	24
7	SEigm	-0.130	-0.015	20	20	GATS3e	0.231	0.037	9
8	piPC10	0.266	0.033	11	21	GATS5e	-0.131	-0.022	18
9	T(O..O)	-0.140	-0.029	13	22	GATS7e	-0.083	-0.014	23
10	BELp5	0.523	0.156	1	23	GATS8p	0.157	0.027	14
11	BELp7	0.386	0.077	4	24	nNR2	-0.072	-0.014	22
12	GGI1	0.316	0.054	6	25	nHAcc	-0.292	-0.042	8
13	GGI10	-0.081	-0.015	21	26	MLOGP	0.510	0.092	2
					Constant = 7.872				
C: PLS regression statistics					Values				
n					45				
r					0.939				
s					0.220				
F					75.401				
Q ² _{L00}					0.847				
Q ² _{L50}					0.849				
r ² _{Test}					0.829				

^aRegression coefficient of PLS factor and its standard error. ^bCoefficients of MLR-like PLS equation in terms of descriptors for their original values; ^cf.c. is fraction contribution of regression coefficient, computed from the normalized regression coefficients obtained from the autoscaled (zero mean and unit s.d.) data.

For the sake of comparison, the plot showing goodness of fit between observed and calculated activities (through PLS analysis) for the training and test set compounds is also given in Figure 2. Figure 3 shows a plot of the fraction contribution of normalized regression coefficients of these descriptors to the activity.

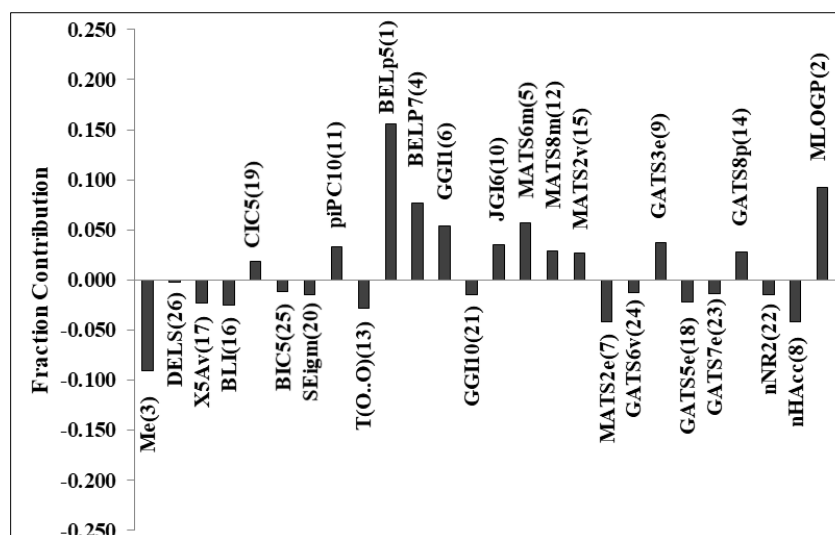


Figure 3 Plot of fraction contribution of MLR-like PLS coefficients (normalized) against 26 CP-MLR identified descriptors (Table 3) associated with histamine H3 receptor binding affinity of cyclohexylamine derivatives

The PLS analysis has suggested BELp5 as the most determining descriptor for modeling the activity of the compounds (descriptor S. No. 10 in Table 4; Figure 3). The other nine descriptors in decreasing order of significance are MLOGP, Me, BELp7, MATS6m, GGI1, MATS2e, nHAcc, GATS3e and JGI6. The descriptors BELp5, BELp7, Me, nHAcc and JGI6 are part of Eqs. (1) to (7) and convey same inference in the PLS model as well. The positive influence of descriptors, MLOGP (Moriguchi octanol-water partition coefficient, logP), MATS6m (atomic masses weighted Moran autocorrelation of lag 6), GGI1 (topological charge index of order 1) and GATS3e (atomic Sanderson electronegativities weighted Geary autocorrelation of lag 3) advocate that higher values of these descriptors would be beneficiary to the binding affinity whereas, a lower value of descriptor MATS2e (atomic Sanderson electronegativities weighted Moran autocorrelation of lag 2) would be helpful for improved activity. It is also observed that PLS model from the dataset devoid of CP-MLR identified 26 descriptors (Table 3) is inferior in explaining the activity of the analogues.

3.2. Applicability domain

On analyzing the applicability domain (AD) for the H₃ receptor binding affinity in the Williams plot (Figure 4) of the model based on the whole data set (Table 5), No any compound has been identified as an obvious 'outlier' for the H₃R binding affinity if the limit of normal values for the Y outliers (response outliers) was set as 3×(standard deviation) units. None of the compound was found to have leverage (h) values greater than the threshold leverage (h*=0.333). 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, all of the compounds were within the applicability domain of the proposed model and were evaluated correctly.

Table 5 Models derived for the whole data set (n = 60) in descriptors identified through CP-MLR for histamine H3 receptor binding affinity

Model	r	s	F	Q ² _{Loo}	Eq.
$pK_i = 7.898 + 1.049(0.121)BELp5 + 0.595(0.168)BELp7 + 0.428(0.160)GATS8p - 0.884(0.194)nHAcc$	0.905	0.267	62.394	0.779	(4a)
$pK_i = 8.414 - 0.727(0.182)Me + 1.063(0.115)BELp5 + 0.416(0.127)MATS2v - 0.573(0.229)nHAcc$	0.912	0.257	68.140	0.797	(5a)
$pK_i = 8.844 - 1.274(0.189)Me - 0.837(0.214)X5Av + 0.982(0.094)BELp5 + 0.476(0.151)GATS8p$	0.918	0.248	74.402	0.810	(6a)
$pK_i = 8.376 - 0.828(0.169)Me + 0.973(0.104)BELp5 + 0.331(0.129)MATS2v - 0.349(0.165)GATS5e$	0.909	0.261	65.733	0.791	(7a)

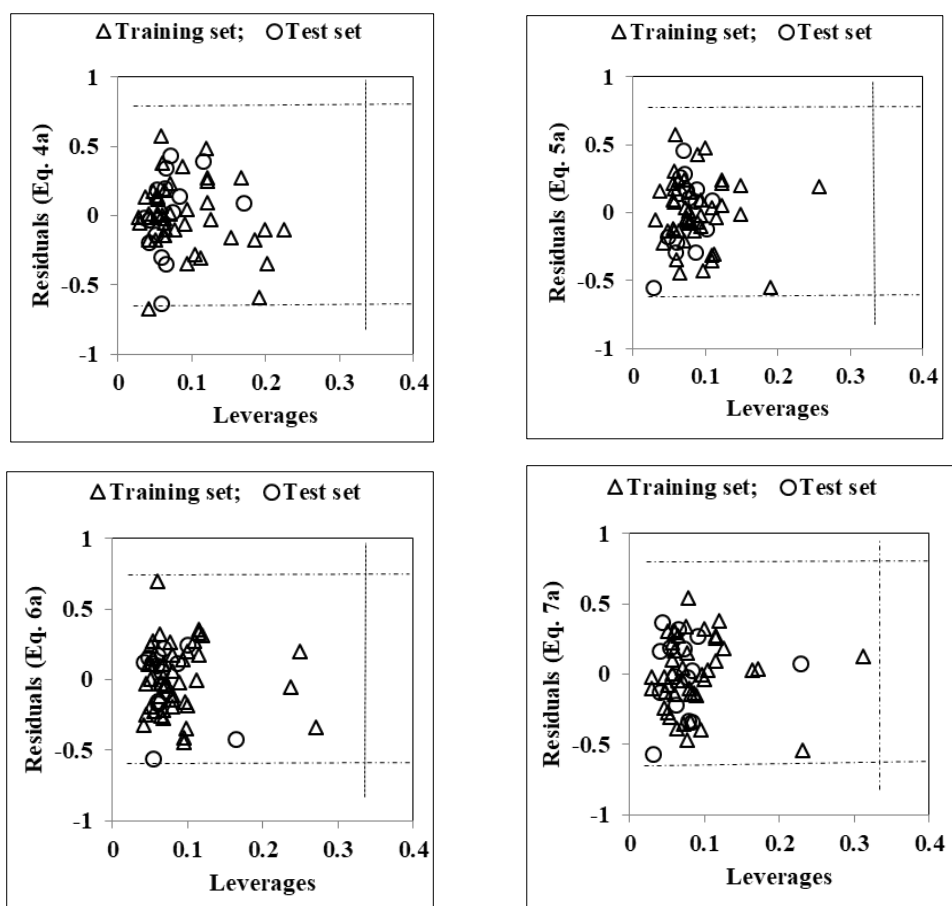


Figure 4 Williams plot for the training-set and test-set for H₃R binding affinity of compounds in Table 1. The horizontal dotted line refers to the residual limit ($\pm 3 \times$ standard deviation) and the vertical dotted line represents threshold leverage h^* ($=0.333$)

4. Conclusion

The histamine H₃ receptor binding affinities of cyclohexylamine derivatives has been analysed with the topological and molecular features from Dragon software. Analysis of the structural features in conjunction with the biological endpoints in combinatorial protocol in multiple linear regression (CP-MLR) led to the identification of 26 descriptors for modelling the activity. The study clearly suggested the role of atomic properties such as mass, electronegativity or charge content, polarizability, atomic van der Waals volume, average valence connectivity index chi-5 and absence of number of acceptor atoms for H-bonds (N, O, F) type functionality to optimise the histamine H₃ receptor binding affinity of titled compounds. The models developed and the participating descriptors advocate that the substituent groups of the cyclohexylamine moiety hold scope for further modification in the optimization of the H₃ receptor binding affinity.

Analysis of these descriptors in partial least squares (PLS) highlighted their relative significance in modulating the biological response. The selected descriptors are enriched with information corresponding to the activity when compared to the remaining ones. 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.

Compliance with ethical standards

Acknowledgments

Authors are thankful to their Institution for providing necessary facilities to complete this work.

Disclosure of conflict of interest

The authors declare no conflict of interest.

References

- [1] Arrang JM, Garbarg M, Schwartz JC. Auto-inhibition of brain histamine release mediated by a novel class (H3) of histamine receptor. *Nature* 1983; 302:832-837.
- [2] Lovenberg TW, Roland BL, Wilson SJ, Jiang X, Pyati J, Huvar A, Jackson MR, Erlander MG. Cloning and functional expression of the human histamine H3 receptor. *Mol Pharmacol.* 1999; 55:1101-1107.
- [3] Gemkow MJ, Davenport AJ, Harich S, Ellenbroek BA, Cesura A, Hallett D. The histamine H3 receptor as a therapeutic drug target for CNS disorders. *Drug Discov Today* 2009; 14:509-515.
- [4] Celanire S, Lebon F, Stark H. In *The Third Histamine H3 Receptor: Selective Ligands as Potential Therapeutic Agents in CNS Disorders*; Vohora DS, Ed.; Taylor & Francis CRC Press Inc., 2008: Boca Raton Fla, 2009; p 103.
- [5] Esbenshade TA, Browman KE, Bitner RS, Strakhova M, Cowart MD, Brioni JD. The histamine H3 receptor: an attractive target for the treatment of cognitive disorders. *Br J Pharmacol.* 2008; 154:1166-1181.
- [6] Sander K, Kottke T, Stark H. Histamine H3 receptor antagonists go to clinics. *Biol Pharm Bull.* 2008; 31:2163-2181.
- [7] Stocking EM, Letavic MA. Histamine H3 antagonists as wake-promoting and pro-cognitive agents. *Curr Top Med Chem.* 2008; 8:988-1002.
- [8] Letavic MA, Barbier AJ, Dvorak CA, Carruthers NI. 5 Recent Medicinal Chemistry of the Histamine H3 Receptor. *Prog Med Chem.* 2006; 44:181-206.
- [9] Celanire S, Wijtmans M, Talaga P, Leurs R, de Esch IJP. Keynote review: Histamine H3 receptor antagonists reach out for the clinic. *Drug Discov Today* 2005; 10:1613-1627.
- [10] Kuhne S, Wijtmans M, Lim HD, Leurs R, de Esch IJP. Several down, a few to go: histamine H3 receptor ligands making the final push towards the market? *Expert Opin Investig Drugs* 2011; 20:1629-1648.
- [11] Arrang J-M, Drutel G, Schwartz J-C. Characterization of histamine H3 receptors regulating acetylcholine release in rat entorhinal cortex. *Br J Pharmacol.* 1995; 114:1518-1522.
- [12] Blandina P, Giorgetti M, Bartolini L, Cecchi M, Timmerman H, Leurs R, Pepeu G, Giovannini MG. Inhibition of cortical acetylcholine release and cognitive performance by histamine H3 receptor activation in rats. *Br J Pharmacol.* 1996; 119:1656-1664.
- [13] Schlicker E, Behling A, Lümmer G, Göthert M. Histamine H3A receptor-mediated inhibition of noradrenaline release in the mouse brain cortex. *Naunyn-Schmiedeberg's Arch Pharmacol.* 1992; 345:489-493.
- [14] Schlicker E, Fink K, Detzner M, Göthert M. Histamine inhibits dopamine release in the mouse striatum via presynaptic H3 receptors. *J Neural Transm Gen Sect.* 1993; 93:1-10.
- [15] Fink K, Schlicker E, Neise A, Göthert M. Involvement of presynaptic H3 receptors in the inhibitory effect of histamine on serotonin release in the rat brain cortex. *Naunyn-Schmiedeberg's Arch Pharmacol.* 1990; 42:513-519.
- [16] Morisset S, Rouleau A, Ligneau X, Gbahou F, Tardivel-Lacombe J, Stark H, Schunack W, Ganellin CR, Schwartz JC, Arrang JM. High constitutive activity of native H3 receptors regulates histamine neurons in brain. *Nature* 2000; 408:860-864.
- [17] (a) Levoine N, Labeeuw O, Calmels T, Poupardin-Olivier O, Berrebi-Bertrand I, Lecomte J-M, Schwartz J-C, Capet M. Novel and highly potent histamine H3 receptor ligands. Part 1: withdrawing of hERG activity. *Bioorg Med Chem Lett.* 2011; 21:5378-5383. (b) Labeeuw O, Levoine N, Poupardin-Olivier O, Calmels T, Ligneau X, Berrebi-Bertrand I, Robert P, Lecomte J-M, Schwartz J-C, Capet M. Novel and highly potent histamine H3 receptor ligands. Part 2: Exploring the cyclohexylamine-based series. *Bioorg Med Chem Lett.* 2011; 21:5384-5388. (c) Labeeuw O, Levoine N, Poupardin-Olivier O, Calmels T, Ligneau X, Berrebi-Bertrand I, Robert P, Lecomte J-M, Schwartz J-C, Capet M. Novel and highly potent histamine H3 receptor ligands. Part 3: An alcohol function to improve the pharmacokinetic profile. *Bioorg Med Chem Lett.* 2013; 23:2548-2554.

- [18] (a) Ligneau X, Garbarg M, Vizuete ML, Diaz J, Purand K, Stark H, Schunack W, Schwartz JC. [125I]iodoproxyfan, a new antagonist to label and visualize cerebral histamine H₃ receptors. *J Pharmacol Exp Ther.* 1994; 271:452-459. (b) Ligneau X, Morisset S, Tardivel-Lacombe J, Gbahou F, Ganellin CR, Stark H, Schunack W, Schwartz JC, Arrang JM. Distinct pharmacology of rat and human histamine H(3) receptors: role of two amino acids in the third transmembrane domain. *Br J Pharmacol.* 2000; 131:1247-1250. (c) Ligneau X, Perrin D, Landais L, Camelin JC, Calmels TP, Berrebi-Bertrand I, Lecomte JM, Parmentier R, Anaclet C, Lin JS, Bertaina-Anglade V, la Rochelle CD, d'Aniello F, Rouleau A, Gbahou F, Arrang JM, Ganellin CR, Stark H, Schunack W, Schwartz JC. BF2.649 [1-{3-[3-(4-Chlorophenyl)propoxy]propyl}piperidine,hydrochloride], a nonimidazole inverse agonist/antagonist at the human histamine H₃ receptor: Preclinical pharmacology. *J Pharmacol Exp Ther.* 2007; 320:365-375.
- [19] Chemdraw ultra 6.0 and Chem3D ultra, Cambridge Soft Corporation, Cambridge, USA. <http://www.cambridgesoft.com>
- [20] Dragon software (version 1.11-2001) by Todeschini R, Consonni V. Milano, Italy. <http://www.taletе.mi.it/dragon.htm>
- [21] 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.
- [22] 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 Inhibn Med Chem.* 2009; 24(1):86-93.
- [23] Sharma BK, Singh P, Paliania P, Sarbhai K, Prabhakar YS. CP-MLR/PLS directed QSAR study on apical sodium-coupled bile acid transporter inhibition activity of benzothiepies. *Mol Divers.* 2011; 5:135-147.
- [24] Sharma S, Prabhakar YS, Singh P, Sharma BK. QSAR study about ATP-sensitive potassium channel activation of cromakalim analogues using CP-MLR approach. *Eur J Med Chem* 2008; 43:2354-2360.
- [25] Sharma S, Sharma BK, Sharma SK, Singh P, Prabhakar YS. Topological descriptors in modeling the agonistic activity of human A₃ adenosine receptor ligands: the derivatives of 2-Chloro-N₆-substituted-4-thioadenosine-5'-uronamide. *Eur J Med Chem.* 2009; 44:1377-1382.
- [26] Wold S. Cross-validated estimation of the number of components in factor and principal components models. *Technometrics.* 1978; 20:397-405.
- [27] Kettaneh N, Berglund A, Wold S. PCA and PLS with very large data sets. *Comput Stat Data Anal.* 2005; 48:69-85.
- [28] Stahle L, Wold S. Multivariate data analysis and experimental design. In: Ellis GP, West WB. Eds., *Biomedical research. Progress in medicinal chemistry.* Elsevier Science Publishers, BV, Amsterdam.1988; 25:291-338.
- [29] Topliss JG, Edwards RP. Chance factors in studies of quantitative structure-activity relationships. *J Med Chem.* 1979; 22:1238-1244.
- [30] Katritzky AR, Dobchev DA, Slavov S, Karelson M. Legitimate utilization of large descriptor pools for QSAR/QSPR models. *J Chem Inf Model.* 2008; 48:2207-2213.
- [31] 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.
- [32] 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.
- [33] Gramatica, P. Principles of QSAR models validation: internal and external. *QSAR Comb Sci.* 2007; 26:694-701.
- [34] Golbraikh A, Tropsha A. Be aware of q²! *J Mol Graph Model.* 2002; 20:269-76.