Research Journal of Pharmaceutical, Biological and Chemical Sciences

QSAR Rationales for the Isoindolone Derivatives as 5-HT$_{2C}$ Receptor Antagonists.

Manju Choudhary, and Brij Kishore Sharma[*].

Department of Chemistry, Government PG College Bundi-323 001(Rajasthan) India

ABSTRACT

The 5-HT$_{2C}$ receptor binding affinities of the isoindolone derivatives have been quantitatively analyzed in terms of Dragon descriptors. The derived QSAR models have provided rationales to explain the binding affinity of titled compounds. The associations of atomic mass to the highest eigenvalue n.2 and lowest eigenvalue n.5 of the Burden matrix (BEHm2 and BELm5, respectively), atomic van der Waals volume to path length 2 and 8 of the Moran autocorrelations (MATS2v and MATS8v) and path length 2 of Geary autocorrelation (GATS2v), polarizability to the highest eigenvalue n.1 and n.6 of the Burden matrix (BEHp1 and BEHp6) and Sanderson electronegativity to path length 4 of Geary autocorrelation (GATS4e) have shown the prevalence of atomic properties to explain the binding affinity. Absence of R--CR--R type structural fragment (C-025) in a molecular structure and a lower values of path/walk 4 - Randic shape index (PW4) and Balaban Y index (Yindex) are favorable to the activity. The derived models and participating descriptors in them have suggested that the substituents of isoindolone moiety have sufficient scope for further modification.

Keywords: QSAR, isoindolones, 5-HT$_{2C}$ ligands, binding affinity, combinatorial protocol in multiple linear regression (CP-MLR).

*Corresponding author

INTRODUCTION

The 5-HT$_{2C}$ receptor is one of 14 distinct serotonin receptor subtypes. Two receptors (the 5-HT$_{2A}$ and 5-HT$_{2B}$) are closely related to the 5HT$_{2C}$ and share considerable sequence homology. The activation of central 5-HT$_{2A}$ receptors is a cause for a number of adverse central nervous system effects of nonselective serotonergic drugs such as changes in perception and hallucination. It is hypothesized that activation of 5-HT$_{2B}$ receptors which are located in the cardiovascular system result in the heart valve disease and pulmonary hypertension associated with the use of drugs that act via serotonergic mechanisms. The modulation of the 5-HT$_{2C}$ receptor is of considerable interest in the area of neuropsychiatric disorders [1]. The 5-HT$_{2C}$ receptor class GSK ligands which are based on several alternative templates have been investigated and disclosed. [2-4]. A novel series of isoindolone derivatives has been reported by Hamprecht et.al [5] as potent and selective 5-HT$_{2C}$ antagonists. The reported isoindolone derivatives are a result of two independent approaches. From a medicinal chemistry point of view the isoindolone template allowed the incorporation of the carbon–carbon double bond of an earlier dihydropyrrolone series in an aromatic system within a comparatively simple and compact motif. Furthermore, the *in silico* screening approach of the corporate database using a 5-HT$_{2C}$ pharmacophore model resulted in the identification of a related structure containing this template.
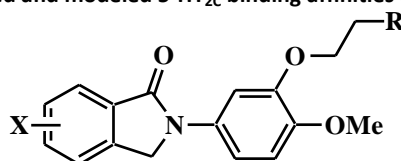
In view of the importance of 5-HT$_{2C}$ antagonists in the clinical management of several disorders, a quantitative structure–activity relationship is attempted on the binding affinities of these isoindolone derivatives. The present study is aimed at rationalizing the substituent variations of these analogues to provide insight for the future endeavours.
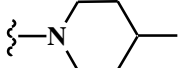
MATERIALS AND METHODS
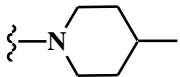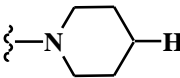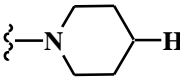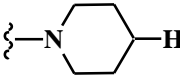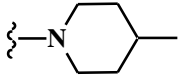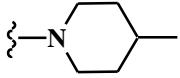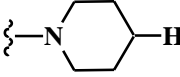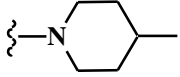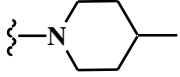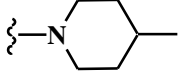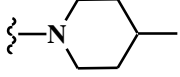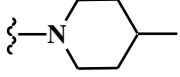
Chemical structure database and biological activity
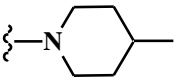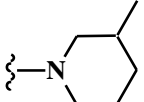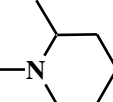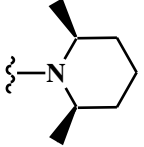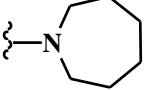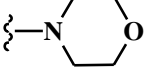
This study comprises a chemical structure database of reported thirty one isoindolone derivatives. The binding affinities (pK$_i$ data) of these compounds were determined using [^{3}H]mesulergine displacement in HEK293 cells with stable expression of human 5-HT$_{2C}$ receptor in analogy to the procedure described by Wood *et al.* [6]. The structural variations and the binding affinities 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 31 analogues have been divided into training and test sets. Out of the 31 analogues, nearly one fifth compounds (06) 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.

Table 1: Structures[a], observed and modeled 5-HT$_{2C}$ binding affinities of the isoindolone derivatives

Cpd.	X	R	pK$_i$				
			Obsd[a]	Eq. (14)	Eq. (15)	Eq. (16)	Eq. (17)
1[b]	H	⎰—N⟨ ⟩—H	7.3	6.73	6.71	6.79	6.64
2	4-Cl	⎰—N⟨ ⟩—H	6.6	6.73	6.83	7.07	6.82
3	4-Cl	⎰—N⟨ ⟩—	6.9	7.31	7.45	7.39	7.33
4	5-Cl	⎰—N⟨ ⟩—H	7.3	7.17	7.14	7.11	7.28

No.	Substituent	Structure					
5	5-Cl		7.6	7.79	7.78	7.55	7.83
6	6-Cl	(—N...H)	7.8	7.32	7.19	7.27	7.34
7	6-Cl		7.7	7.84	7.78	7.57	7.83
8[b]	7-Cl	(—N...H)	8.4	7.80	7.64	7.99	7.59
9	7-Cl		8.8	8.31	8.21	8.25	8.05
10	7-F	(—N...H)	7.1	7.67	7.53	7.57	7.44
11	7-Br		9.1	8.71	9.20	8.84	9.15
12	7-Me		8.6	8.26	8.23	8.21	8.13
13	$7\text{-}CF_3$		8.8	8.97	8.85	9.02	9.11
14	7-CN		8.0	8.41	8.44	8.54	8.09
15[b]	7-OMe	(—N...H)	6.7	6.77	6.88	7.42	6.72
16	3-Me, 7-Cl		8.9	9.10	9.12	9.10	8.73
17[b]	$4\text{-}F,7\text{-}CF_3$		8.2	8.34	8.50	8.71	8.76
18	$4,5\text{-}F_2,7\text{-}Cl$		8.7	8.52	8.49	8.55	8.31
19	$5,6\text{-}Cl_2$		8.3	8.21	8.16	8.00	8.34
20	$5,7\text{-}Cl_2$		8.8	9.06	8.86	8.99	8.71
21	$6,7\text{-}Cl_2$		9.0	8.77	8.58	8.79	8.61
22	$5,6\text{-}Cl_2,7\text{-}Me$		8.6	8.47	8.62	8.73	8.62
23[b]	$5,7\text{-}Cl_2,6\text{-}Me$		8.9	8.65	8.15	8.43	8.35
24	$5\text{-}F,7\text{-}CF_3$		9.4	9.08	9.04	9.20	9.37

25	5,7-Me$_2$		8.5	8.90	8.57	8.35	8.71
26	5,7-Cl$_2$		8.6	8.14	8.24	8.39	8.20
27	5,7-Cl$_2$		8.0	8.25	8.30	8.33	8.50
28	5,7-Cl$_2$		8.2	7.98	8.32	8.01	8.53
29	5,7-Cl$_2$		8.7	8.98	9.07	8.93	8.95
30 [b]	5,7-Cl$_2$		8.6	8.71	8.72	8.83	8.91
31	5,7-Cl$_2$		7.9	7.97	7.89	8.12	7.92

[a]Reference [5], [b]Compounds included in test set.

Theoretical molecular descriptors

Table 2: Dragon descriptor classes used along with their definition and scope for modeling the binding affinity of isoindolone 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

The structures of the compounds under study have been drawn in 2D ChemDraw [7]. 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 [8] for the computation of descriptors for the titled compounds (Table 1). This software offers several hundreds of descriptors from different perspectives corresponding to 0D-, 1D-, and 2D-descriptor modules. The outlined modules comprised of ten different classes, namely, the constitutional (CONST), the topological (TOPO), the molecular walk counts (MWC), the BCUT descriptors (BCUT), the Galvez topological charge indices (GALVEZ), the 2D autocorrelations (2D-AUTO), the functional groups (FUNC), the atom-centered fragments (ACF), the empirical descriptors (EMP), and the properties describing descriptors (PROP). For each of these classes the DRAGON software computes a large number of descriptors which are characteristic to the molecules under multi-descriptor environment. The definition and scope of these descriptor's classes is given in Table 2. The combinatorial protocol in multiple linear regression [9] procedure has been used in the present work for developing QSAR models. Before the application of CP-MLR procedure, all those descriptors which are intercorrelated beyond 0.90 and showing a correlation of less than 0.1 with the biological endpoints (descriptor *vs.* activity, r < 0.1) were excluded. This has reduced the total dataset of the compounds from 450 to 98 descriptors as relevant ones for the binding activity. A brief description of the computational procedure is given below.

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.

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 [10,11], the Kubinyi function, FIT [12,13], and the Friedman's lack of fit, LOF [14] (Eqs. 1-3) have also been derived to evaluate the best model.

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

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

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

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

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

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

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

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

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.

Y-randomization

Chance correlations, if any, associated with the CP-MLR models were recognized in randomization test [15,16] 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.

RESULTS AND DISCUSSION

In multi-descriptor class environment, exploring for best model equation(s) along the descriptor class provides an opportunity to unravel the phenomenon under investigation. In other words, the concepts embedded in the descriptor classes relate the biological actions revealed by the compounds. For the purpose of modeling study, 06 compounds have been included in the test set for the validation of the models derived from 25 training set compounds. A total number of 98 significant descriptors from 0D-, 1D- and 2D-classes have been subjected to CP-MLR analysis with default 'filters' set in it. Statistical models in two, three and four descriptor(s) have been derived successively to achieve the best relationship correlating 5-HT$_{2C}$ binding affinity. These models (with 98 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.

The selected models in two, three and four descriptors are presented below.

pK_i = 8.398 +1.468(0.417)BEHp1 −1.619(0.417)GATS4e
n = 25, r = 0.779, s = 0.476, F = 17.052, FIT = 1.176, LOF = 0.282, AIC = 0.288,
Q^2_{LOO} = 0.496, Q^2_{L5O} = 0.478, r^2_{randY}(sd) = 0.102(0.090), r^2_{Test} = 0.544 (6)

pK_i = 7.006 +1.132(0.390)BEHm6 +1.358(0.424)BEHe4
n = 25, r = 0.730, s = 0.519, F = 12.577, FIT = 0.867, LOF = 0.336, AIC = 0.343,
Q^2_{LOO} = 0.402, Q^2_{L5O} = 0.385, r^2_{randY}(sd) = 0.069(0.069), r^2_{Test} = 0.566 (7)

pK_i = 7.827 +1.053(0.455)BEHv5 +1.044(0.422)BEHp1 −1.674(0.382)GATS4e
n = 25, r = 0.829, s = 0.435, F = 15.397, FIT = 1.358, LOF = 0.275, AIC = 0.261,
Q^2_{LOO} = 0.560, Q^2_{L5O} = 0.502, r^2_{randY}(sd) = 0.121(0.085), r^2_{Test} = 0.518 (8)

pK_i = 7.938 +1.330(0.386)BEHp1 +0.912(0.396)MATS8v −1.553(0.383)GATS4e
n = 25, r = 0.828, s = 0.435, F = 15.345, FIT = 1.354, LOF = 0.275, AIC = 0.261,
Q^2_{LOO} = 0.524, Q^2_{L5O} = 0.491, r^2_{randY}(sd) = 0.121(0.093), r^2_{Test} = 0.536 (9)

pK_i = 9.358 +3.629(0.684)BEHp1 −1.509(0.471)BEHp6 −1.969(0.628)MATS2v
    −2.714(0.467)GATS4e
 n = 25, r = 0.874, s = 0.387, F = 16.198, FIT = 1.580, LOF = 0.259, AIC = 0.224,
Q^2_{LOO} = 0.599, Q^2_{L5O} = 0.541, r^2_{randY}(sd) = 0.189(0.116), r^2_{Test} = 0.865 (10)

pK_i = 8.510 +1.943(0.424)BEHp1 −1.396(0.561)MATS2v +1.155(0.368)MATS8v
    −2.167(0.422)GATS4e
n = 25, r = 0.872, s = 0.390, F = 15.893, FIT = 1.550, LOF = 0.263, AIC = 0.228,
Q^2_{LOO} = 0.628, Q^2_{L5O} = 0.590, r^2_{randY}(sd) = 0.163(0.105), r^2_{Test} = 0.575 (11)

pK_i = 9.679 −0.927(0.315)PW4 +2.244(0.448)BEHp1 −1.560(0.592)MATS2v
    −2.501(0.457)GATS4e
n = 25, r = 0.866, s = 0.398, F = 15.070, FIT = 1.470, LOF = 0.274, AIC = 0.237,
Q^2_{LOO} = 0.642, Q^2_{L5O} = 0.609, r^2_{randY}(sd) = 0.161(0.093), r^2_{Test} = 0.546 (12)

pK_i = 7.753 −0.754(0.283)BELm6 +1.727(0.395)BEHp1 −1.574(0.370)GATS4e
    +0.967(0.362)MLOGP
n = 25, r = 0.865, s = 0.399, F = 14.932, FIT = 1.456, LOF = 0.275, AIC = 0.239,
Q^2_{LOO} = 0.570, Q^2_{L5O} = 0.617, r^2_{randY}(sd) = 0.173(0.091), r^2_{Test} = 0.595 (13)

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

The descriptors BEHp1, BEHe4, BEHv5, BEHm6, BELm6 and BEHp6 belong to BCUT class of Dragon descriptors. The BCUT descriptors are the first 8 highest and the lowest absolute eigenvalues, BEHwk and BELwk, respectively, for the modified Burden adjacency matrix. Here w refers to the atomic property and k to the eigenvalue rank. The ordered sequence of the highest and the lowest eigen values reflect upon the relevant aspects of molecular structure, useful for similarity searching. The signs of regression coefficients of BCUT descriptors BEHp1 (the highest eigenvalue n.1 of Burden Marix/weighted by atomic polarizabilities), BEHe4 (the highest eigenvalue n.4 of Burden Marix/weighted by atomic Sanderson electronegativities), BEHv5 (the highest eigenvalue n.5 of Burden Marix/weighted by atomic van der Waals volumes) and BEHm6 (the highest eigenvalue n.6 of Burden Marix/weighted by atomic masses), delineate that a higher value of these descriptors and a lower value of descriptors BEHp6 (the highest eigenvalue n.6 of Burden Marix/weighted by atomic polarizabilities) and BELm6 (the lowest eigenvalue n.6 of Burden Marix/weighted by atomic masses) is advantageous for activity. The participated descriptor PW4 is from the TOPO class of Dragon descriptors. The TOPO class descriptors are based on a graph representation of the molecule and are numerical quantifiers of molecular topology obtained by the application of algebraic operators to matrices representing molecular graphs and whose values are independent of vertex numbering or labeling. They can be sensitive to one or more structural features of the molecule such as size, shape, symmetry, branching and cyclicity and can also encode chemical information concerning atom type and bond multiplicity. The negative correlation of descriptor PW4 (path/walk 4 - Randic shape index) to the binding activity advocate that a lower value of path/walk 4 - Randic shape index would be beneficiary to the activity.

The descriptors MATS2v, MATS8v and GATS4e, in above models, are representative of 2D-AUTO class of Dragon descriptors. The 2D-AUTO descriptors have their origin in autocorrelation of topological structure of Broto-Moreau (ATS), of Moran (MATS) and of Geary (GATS). The computation of these descriptors involves the summation of different autocorrelation functions corresponding to the different fragment lengths and lead to different autocorrelation vectors corresponding to the lengths of the structural fragments. Also a weighting component in terms of a physicochemical property has been embedded in these descriptors. As a result, these descriptors address the topology of the structure or parts thereof in association with a selected physicochemical property. 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. The participated descriptor MATS8v (Moran autocorrelation - lag 8 weighted by atomic van der Waals volumes) correlate positively to the activity and descriptors MATS2v (Moran autocorrelation - lag 2 weighted by atomic van der Waals volumes) and GATS4e (Geary autocorrelation - lag 4 weighted by atomic Sanderson electronegativities) contribute negatively to the activity. The positive correlation suggest the favorable conditions associated with lag 8 weighted by atomic van der Waals volumes and negative contribution hints at the unfavorable conditions associated with lag 2 and lag 4 weighted respectively, by atomic van der Waals volumes and Sanderson electronegativities.

The other participated descriptor, MLOGP, in above models belong to PROP class. Descriptors from this class represent the molecular properties of a molecule. Descriptor MLOGP is Moriguchi octanol-water partition coefficient (logP) of a molecule. It reflects upon the hydrophobic nature of a compound. The positive contribution of descriptor MLOGP to the activity recommends more hydrophobic character of a molecule for elevated binding affinity of titled compounds.

The four descriptor models could estimate 76.38 percent variance in observed activity of the compounds. Considering the number of observation in the dataset, models with up to five descriptors were explored. A total number of 22 models, sharing 24 descriptors among them, were obtained through CP-MLR. All these 24 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. The given below are

some five-descriptor models for the activity. These models have accounted for up to 81.00 percent variance in the observed activities

pK_i = 9.928 −0.688(0.316)Yindex +3.515(0.630)BEHp1 −1.355(0.438)BEHp6
    −2.329(0.599)MATS2v − 2.753(0.429)GATS4e
n = 25, r = 0.900, s = 0.355, F = 16.335, FIT = 1.633, LOF = 0.266, AIC = 0.206,
Q^2_{LOO} = 0.668, Q^2_{L5O} = 0.728, r^2_{randY}(sd) = 0.225(0.111), r^2_{Test} = 0.795   (14)

pK_i = 8.399 +0.858(0.410)BEHm2 +1.984(0.393)BEHp1 −1.553(0.524)MATS2v
    +1.125(0.341)MATS8v −2.179(0.391)GATS4e
n = 25, r = 0.897, s = 0.360, F = 15.735, FIT = 1.573, LOF = 0.274, AIC = 0.212,
Q^2_{LOO} = 0.640, Q^2_{L5O} = 0.661, r^2_{randY}(sd) = 0.204(0.108), r^2_{Test} = 0.577   (15)

pK_i = 6.437 −1.416(0.370)BELm5 +4.745(0.891)BEHp1 +2.372(0.602)GATS2v
    −2.164(0.374)GATS4e − 2.369(0.688)C-025
n = 25, r = 0.897, s = 0.361, F = 15.674, FIT = 1.567, LOF = 0.275, AIC = 0.213,
Q^2_{LOO} = 0.666, Q^2_{L5O} = 0.620, r^2_{randY}(sd) = 0.205(0.108), r^2_{Test} = 0.613   (16)

pK_i = 9.555 −0.935(0.287)PW4 +0.930(0.411)BEHm2 +2.292(0.409)BEHp1
    −1.749(0.546)MATS2v −2.522(0.416)GATS4e
n = 25, r = 0.896, s = 0.362, F = 15.547, FIT = 1.554, LOF = 0.277, AIC = 0.214,
Q^2_{LOO} = 0.708, Q^2_{L5O} = 0.657, r^2_{randY}(sd) = 0.184(0.105), r^2_{Test} = 0.530   (17)

Table 3: Descriptors[a] identified for modeling the binding affinity of cyclic guanidines along with the average regression coefficient[b], standard deviation and the total incidence.

Descriptor	Avg reg coeff(sd) total incidence	Descriptor	Avg reg coeff(sd) total incidence	Descriptor	Avg reg coeff(sd) total incidence
Sv	-4.538(0.000)1	nCL	-0.649(0.127)4	nX	1.731(0.580)2
Jhetp	-1.089(0.000)1	PW4	-0.918(0.242)3	Yindex	-0.724(0.030)3
BEHm2	1.115(0.292)5	BEHm6	1.208(0.268)4	BELm5	-1.143(0.217)4
BELm6	-0.569(0.000)1	BEHv3	2.438(2.050)2	BELv6	-1.181(0.169)4
BEHe1	0.739(0.000)1	BEHp1	2.482(1.024)21	BEHp6	-1.222(0.188)2
MATS2v	-1.713(0.310)13	MATS8v	1.125(0.000)1	MATS4e	1.377(0.210)4
MATS8e	-1.166(0.330)4	GATS2v	1.645(0.493)4	GATS4e	-2.372(0.263)18
GATS8e	0.954(0.395)2	C-025	-1.515(0.739)3	MLOGP	0.927(0.166)3

[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 0.84; filter-4 as 0.3 ≤ Q^2 ≤ 1.0; number of compounds in the study are 25; **CONST**: Sv, sum of atomic van der Waals volumes (scaled on Carbon atom); nCL, number of chlorine atoms; nX, number of halogen atoms; **TOPO**: Jhetp, Balaban-type index from polarizability weighted distance matrix; PW4, path/walk 4 - Randic shape index; Yindex, Balaban Y index; **BCUT**: BEHm2, highest eigenvalue n. 2 of Burden matrix/ weighted by atomic masses; BEHm6, highest eigenvalue n. 6 of Burden matrix/ weighted by atomic masses; BELm5, lowest eigenvalue n. 5 of Burden matrix/ weighted by atomic masses; BELm6, lowest eigenvalue n. 6 of Burden matrix/ weighted by atomic masses; BEHv3, highest eigenvalue n. 3 of Burden matrix/ weighted by atomic van der Waals volumes; BELv6, lowest eigenvalue n. 6 of Burden matrix/ weighted by atomic van der Waals volumes; BEHe1, highest eigenvalue n. 1 of Burden matrix/ weighted by atomic Sanderson electronegativities; BEHp1, highest eigenvalue n. 1 of Burden matrix / weighted by atomic polarizabilities; BEHp6, highest eigenvalue n. 6 of Burden matrix / weighted by atomic polarizabilities; **2D-AUTO**: MATS2v, Moran autocorrelation - lag 2/ weighted by atomic van der Waals volumes; MATS8v, Moran autocorrelation - lag 8 / weighted by atomic van der Waals volumes; MATS4e, Moran autocorrelation - lag 4 / weighted by atomic Sanderson electronegativities; MATS8e, Moran autocorrelation - lag 8/ weighted by atomic Sanderson electronegativities; GATS2v, Geary autocorrelation - lag 2/ weighted by atomic van der Waals volumes; GATS4e, Geary autocorrelation - lag 4 / weighted by atomic Sanderson electronegativities; GATS8e, Geary autocorrelation - lag 8/ weighted by atomic Sanderson electronegativities; **ACF**: C-025, R--CR--R; **PROP**: MLOGP, Moriguchi octanol-water partition coefficient (logP). [b]The average regression coefficient of the descriptor corresponding to all models and the total number of its incidences; the arithmetic sign represents the sign of the regression coefficient in the models.

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 Equations (14) to (17) have been included in Table 1. These models are validated with an external test set of fifteen compounds listed in Table 1. The predictions of the test set compounds based on external validation are found to be satisfactory as reflected in the test set r^2 (r^2_{Test}) values and the predicted activity values are also reported in Table 1. The plot showing

goodness of fit between observed and calculated activities for the training and test set compounds is given in Figure 1.
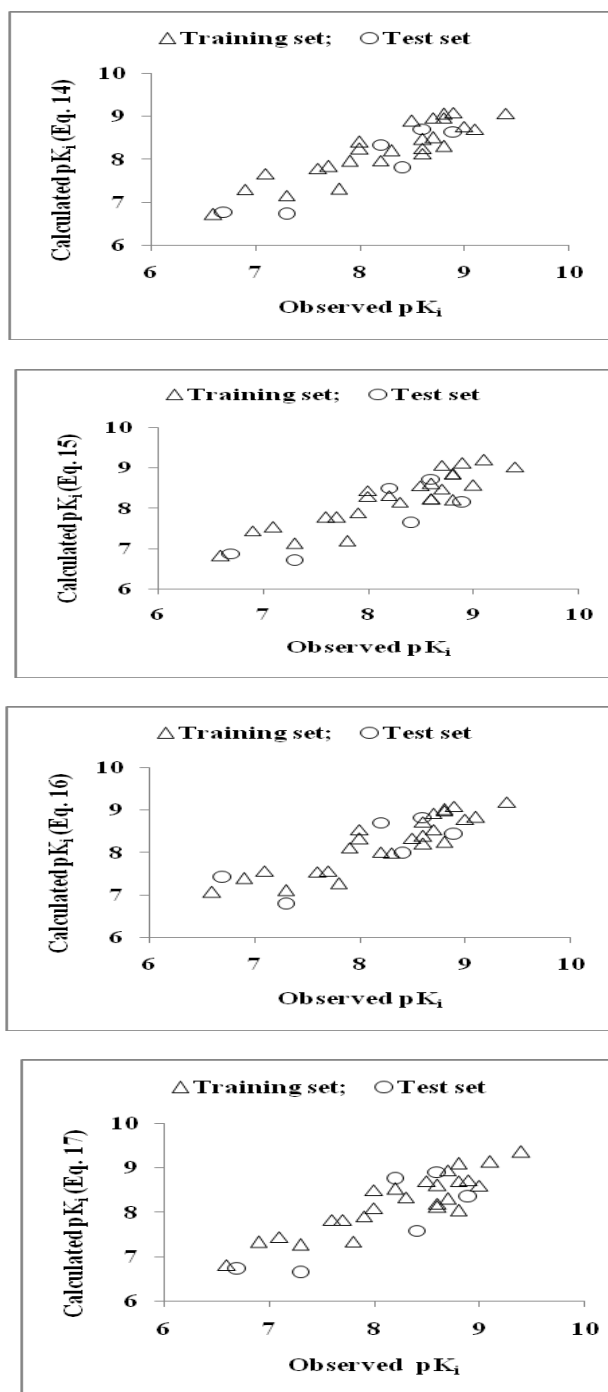

Figure 1: Plot of observed versus calculated pK$_i$ values of the isoindolone derivatives.

The descriptors PW4, MATS2v, MATS8v, GATS4e, BEHp1 and BEHp6, which were emerged in models discussed earlier, have once again shown their importance in five parameter models and convey same inferences to the activity. The newly emerged descriptors, BEHm2 and BELm5, GATS2v, Yindex and C-025 are from BCUT, 2D-AUTO, TOPO and ACF (atom centered fragments) classes of Dragon descriptors, respectively. The ACF class descriptors are based on the counting of 120 atom centered fragments as defined by Ghose and Crippen. These are simple molecular descriptors defined as the number of specific atom types in a molecule. They are calculated by knowing the molecular composition and atom connectivities. Descriptor C-025 has shown negative correlation to the activity. Thus absence of R--CR--R (descriptor C-025) type structural

fragment in a molecular structure would be beneficiary to the activity. The descriptors, GATS2v (Geary autocorrelation - lag 2 weighted by atomic van der Waals volumes), and BEHm2 (the highest eigenvalue n.2 of Burden Marix/weighted by atomic masses) contributed positively to the activity suggesting that higher values of these descriptors would be supportive to enhance the activity. The descriptors Yindex (Balaban Y index) and BELm5 (the lowest eigenvalue n.5 of Burden Marix/weighted by atomic masses) have shown negative contribution to the activity advocating a lower value of these descriptors to augment the binding activity of isoindolone derivatives.

CONCLUSIONS

In conclusion, the present study has provided structure–activity relationships of the binding affinities of isoindolone derivatives to 5-HT$_{2C}$ 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$_{2C}$ receptor binding affinity of a compound, absence of R--CR--R type structural fragment in a molecular structure is advocated by descriptor C-025. A lower value of the molecular topology and symmetry accounting parameters, path/walk 4 - Randic shape index (descriptor PW4) and Balaban Y index (descriptor Yindex) are favorable to the activity. The associations of atomic mass to the highest eigenvalue n.2 and lowest eigenvalue n.5 of the Burden matrix (BEHm2 and BELm5, respectively), atomic van der Waals volume to path length 2 and 8 of the Moran autocorrelations (MATS2v and MATS8v) and path length 2 of Geary autocorrelation (GATS2v), polarizability to the highest eigenvalue n.1 and n.6 of the Burden matrix (BEHp1 and BEHp6) and Sanderson electronegativity to path length 4 of Geary autocorrelation (GATS4e) have shown the prevalence of atomic properties to explain the binding affinity. The derived models and participating descriptors in them have suggested that the substituents of isoindolone moiety have sufficient scope for further modification.

ACKNOWLEDGEMENTS

REFERENCES

[1]   (a) Wood MD, Thomas DR, Watson JM. Expert Opin Investig Drugs 2002; 11: 457.
         (b) Wood MD, Heidbreder C, Reavill C, Ashby CR Jr., Middlemiss DN. Drug Dev Res 2001; 54: 88.
         (c) Barnes NM, Sharp T. Neuropharmacol 1999; 38:1083.
         (d) Keppel Hesselink JM, Sambunaris A. Int Rev Psychiatr 1995; 7: 41.
         (e) Kunovac JL, Stahl SM. Int Rev Psychiatr 1995; 7: 55.
[2]   (a) Bromidge SM, Dabbs S, Davies DT, Davies S, Duckworth DM, Forbes IT, Gaster LM, Ham P, Jones GE, King FD, Mulholland KR, Saunders DV, Wyman PA, Blaney FE, Clarke SE, Blackburn TP, Holland V, Kennett GA, Lightowler S, Middlemiss DN,   Trail B, Riley GJ, Wood MD. J Med Chem 2000; 43: 1123.
         (b) Bromidge SM, Davies S, Duckworth DM, Forbes IT, Jones GE, Jones J, King FD, Blackburn TP, Holland V, Kennett GA, Lightowler S, Middlemiss DN, Riley GR, Trail B, Wood MD. Bioorg Med Chem Lett 2000; 10: 1867.
         (c) Goodacre CJ, Bromidge SM, Clapham D, King FD, Lovell PJ, Allen M, Campbell LP, Holland V, Riley GJ, Starr KR, Trail BK, Wood MD. Bioorg Med Chem Lett 2005;15: 4989.
[3]   Micheli F, Pasquarello A, Tedesco G, Hamprecht D, Bonanomi G, Checchia A, Jaxa-   Chamiec   A, Damiani F, Davalli S, Donati D, Gallotti C, Petrone M, Rinaldi M, Riley G, Terreni S, Wood M. Bioorg Med Chem Lett 2006; 16: 3906.
[4]   Hamprecht D, Micheli F, Tedesco G, Donati D, Petrone M, Terreni S, Wood M. Bioorg Med Chem Lett 2007; 17: 424-427.
[5]   Hamprecht D, Micheli F, Tedesco G, Checchia A, Donati D, Petrone M, Terrenia S, Wood M. Bioorg Med Chem Lett 2007; 17: 428–433.

[6]   Wood MD, Reavill C, Trail B, Wilson A, Stean T, Kennett GA, Lightowler S, Blackburn TP, Thomas D, Gager DL, Riley G, Holland V, Bromidge SM, Forbes IT, Middlemiss   DN. Neuropharmacol 2001; 41: 186.
[7]   ChemDraw Ultra 6.0 and Chem3D Ultra, Cambridge Soft Corporation, Cambridge, USA. Available from: http://www.cambridgesoft.com.
[8]   Dragon software (version 1.11-2001) by R. Todeschini; V. Consonni, Milano, Italy. Available from: http://www.talete.mi.it/dragon.htm.
[9]   Prabhakar YS. QSAR Comb Sci 2003; 22:583-95.
[10]   Akaike H. Information theory and an extension of the minimum likelihood   principle. In: Petrov BN, Csaki F. (eds) Second international symposium on information theory. Akademiai Kiado, Budapest, 1973; 267-281.
[11]   Akaike H. IEEE Trans Automat Contr 1974; AC-19:716-723.
[12]   Kubinyi H. Quant Struct-Act Relat 1994; 13:285-294.
[13]   Kubinyi H. Quant Struct-Act Relat 1994; 13:393-401.
[14]   Friedman J. In: Technical report no. 102. Laboratory for computational statistics. Stanford University, Stanford, 1990.
[15]   So S-S, Karplus M. J Med Chem 1997; 40:4347-4359.
[16]   Prabhakar YS, Solomon VR, Rawal RK, Gupta MK, Katti SB. QSAR Comb Sci 2004; 23:234-244.