*Available online at* www.derpharmachemica.com

Scholars Research Library

**Der Pharma Chemica, 2014, 6(6):321-330**
*(http://derpharmachemica.com/archive.html)*

**ISSN 0975-413X**
**CODEN (USA): PCHHAX**

# QSAR rationales for the 5-HT$_6$ antagonistic activity of Epiminocyclohepta[*b*]indoles

**Manju Choudhary and Brij Kishore Sharma***

*Department of Chemistry, Government PG College Bundi, (Rajasthan) India*

---

**ABSTRACT**

*The 5-HT$_6$ receptor binding affinities of the epiminocyclohepta[b]indole derivatives have been quantitatively expressed in terms of topological and molecular features. The analysis revealed that more number of rings (nBnz, nCIC and nR05) and lesser number of rotatable bonds (RBN) in molecular structure are advantageous to improve 5-HT$_6$ receptor binding affinity. A higher value of the molecular topology and symmetry accounting parameters (SIC4, structural information content of 3-order neighborhood symmetry and IC5, information content index of 5-order neighborhood symmetry) is favorable to the activity. A lower value of atomic polarizabilities associated to path length 8 of the Geary autocorrelation (GATS8p) and more hydrophobicity of molecule (MLOGP) are favorable to activity. Presence or absence of certain structural fragments X- -CH..X (descriptor C-033), R- - N- -R or R- -N- - X (descriptor N-075) and more number of hydrogen atoms attached to sp or sp2 or sp3 hybridized carbon atoms (H-047) in a molecular structure are also relevant for the binding affinity. The derived models and participating descriptors in them have suggested that the substituents of epiminocyclohepta[b]indole moiety have sufficient scope for further modification.*

**Key words**: QSAR, epiminocyclohepta[b]indole derivatives, 5-HT$_6$ antagonists, binding affinity, combinatorial protocol in multiple linear regression (CP-MLR).

---

## INTRODUCTION

Serotonin receptor, 5-HT$_6$, is expressed in various regions of the brain including the hypothalamus and pre-frontal cortex [1]. The exclusively location of these receptor within the CNS represent 5-HT$_6$ an attractive target as the modulation of the receptor will not be associated with any of the peripheral side effects which often plague CNS targeted programs. In pre-clinical animal models, the modulation of the 5-HT$_6$ receptor has been shown to induce promising efficacy in conditions associated with sleep, anxiety and depression, epilepsy and pain [2-5]. Most of the 5-HT$_6$ research, particularly antagonists, has centered on weight loss and cognition. 5-HT$_6$ receptor has been defined as a promising new target for weight management as the modulation of this receptor produced significant weight loss in rodent models of obesity [6]. The antagonism of the 5-HT$_6$ receptor has also demonstrated beneficial effects in treating cognitive deficits associated with conditions like Alzheimer's disease and schizophrenia [7]. For treating cognitive deficits associated with Alzheimer's disease [8] a number of 5-HT$_6$ antagonists are currently undergoing clinical evaluation including the compound SB-742457. Henderson et al. [9] have recently reported a series of epiminocyclohepta[*b*]indole derivatives through a process of rational drug design and the use of ligand–receptor pharmacophore models.

In view of the importance of 5-$HT_6$ antagonists in the clinical management of several disorders, a quantitative structure–activity relationship is attempted on the binding affinities of these epiminocyclohepta[*b*]indoles. The present study is aimed at rationalizing the substituent variations of these analogues to provide insight for the future endeavours.

## MATERIALS AND METHODS

### 2.1. Chemical structure database and biological activity

This study comprises a chemical structure database of thirty epiminocyclohepta[*b*]indole derivatives, reported by Henderson Zhao et al. [9]. The binding affinity of these derivatives was determined by displacement of [$^3$H]LSD from human 5-$HT_6$ receptor membranes. The structural variations and the binding affinities of titled compounds have been given in Table 1. The reported activity data on molar basis has been used for subsequent QSAR analyses as the response variables. For the purpose of modeling all 30 analogues have been divided into training and test sets. Out of the 30 analogues, nearly one fourth compounds (7) have been placed in the test set for the validation of derived models. The training and test set compounds are also listed in Table 1.

### 2.2. Theoretical molecular descriptors

The structures of the compounds under study have been drawn in 2D ChemDraw [10]. 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 [11] for the computation of descriptors for the titled compounds (Table 1). This software offers several hundreds of descriptors from different perspectives corresponding to 0D-, 1D-, and 2D-descriptor modules. The outlined modules comprised of ten different classes, namely, the constitutional (CONST), the topological (TOPO), the molecular walk counts (MWC), the BCUT descriptors (BCUT), the Galvez topological charge indices (GALVEZ), the 2D autocorrelations (2D-AUTO), the functional groups (FUNC), the atom-centered fragments (ACF), the empirical descriptors (EMP), and the properties describing descriptors (PROP). For each of these classes the DRAGON software computes a large number of descriptors which are characteristic to the molecules under multi-descriptor environment. The definition and scope of these descriptor's classes is given in Table 2. The combinatorial protocol in multiple linear regression (CP-MLR) [12] 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 483 to 128 descriptors as relevant ones for the binding activity. A brief description of the computational procedure is given below.

### 2.3. Model development

The CP-MLR is a 'filter' based variable selection procedure for model development in QSAR studies [12]. Its procedural aspects and implementation are discussed in some of our recent publications [13-17]. 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 $\leq 0.79$). In this, if two variables are correlated higher than a predefined cutoff value the respective variable combination is forbidden and will be rejected. The second filter is in terms of t-values of regression coefficients of variables associated with a subset (filter-2, default value 2.0). Here, if the ratio of regression coefficient and associated standard error of any variable is less than a predefined cutoff value then the variable combination will be rejected. Since successive additions of variables to multiple regression equation will increase successive multiple correlation coefficient (r) values, square-root of adjusted multiple correlation coefficient of regression equation, r-bar, has been used to compare the internal explanatory power of models with different number of variables. Accordingly, a filter has been set in terms of predefined threshold level of r-bar (filter-3, default value 0.71) to decide the variables' 'merit' in the model formation. Finally, to exclude false or artificial correlations, the external consistency of the variables of the model have been addressed in terms of cross-validated $R^2$ or $Q^2$ criteria from the leave-one-out (LOO) cross-validation procedure as default option (filter-4, default threshold value $0.3 \leq Q^2 \leq 1.0$). All these filters make the variable

selection process efficient and lead to unique solution. In order to collect the descriptors with higher information content and explanatory power, the threshold of filter-3 was successively incremented with increasing number of descriptors (per equation) by considering the r-bar value of the preceding optimum model as the new threshold for next generation.

**2.4. Model validation**

In this study, the data set is divided into training set for model development and test set for external prediction. Goodness of fit of the models was assessed by examining the multiple correlation coefficient (r), the standard deviation (s), the F-ratio between the variances of calculated and observed activities (F). A number of additional statistical parameters such as the Akaike's information criterion, AIC [18,19], the Kubinyi function, FIT [20,21], and the Friedman's lack of fit, LOF [22], (Eqs. 1-3) have also been derived to evaluate the best model.

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

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

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

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

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

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

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

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

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

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

**2.5. Y-randomization**

Chance correlations, if any, associated with the CP-MLR models were recognized in randomization test [23,24] 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, 7 compounds have been included in the test set for the validation of the models derived from 23 training set compounds. A total number of 128 significant descriptors from 0D-, 1D- and 2D-classes have been subjected to CP-MLR analysis with default 'filters' set in it. Statistical models in two and three descriptor(s) have been derived successively to achieve the best relationship correlating 5-$HT_6$ binding affinity. These models (with 128 descriptors) were identified in CP-MLR by successively incrementing the filter-3 with increasing number of descriptors (per equation). For this the optimum r-bar value of the preceding level model has been used as the new threshold of filter-3 for the next generation. A total number of 96 models in three descriptors were obtained. These models shared 60 descriptors. These descriptors along with their physical meaning, average regression coefficients and total incidences are listed in Table 3.

The selected models in two and three descriptors are given below.

$pK_i$ = 1.948(0.358)nBnz – 2.547(0.687)GATS8p + 7.205
n = 23, r = 0.814, s = 0.705, F = 19.686, FIT = 1.458, LOF = 0.634, AIC = 0.647,
$Q^2_{LOO}$ = 0.565, $Q^2_{L5O}$ = 0.538, $r^2_{randY}$(sd) = 0.090(0.076), $r^2_{Test}$ = 0.727 (6)

$pK_i$ = 3.114(0.677)nCIC + 2.172(0.655)MLOGP + 4.444
n = 23, r = 0.805, s = 0.719, F = 18.507, FIT = 1.370, LOF = 0.660, AIC = 0.673,
$Q^2_{LOO}$ = 0.553, $Q^2_{L5O}$ = 0.554, $r^2_{randY}$(sd) = 0.096(0.091), $r^2_{Test}$ = 0.691 (7)

$pK_i$ = 2.498(0.345)SIC4 + 1.923(0.546)C-033 + 3.513(0.426)MLOGP + 3.454
n = 23, r = 0.934, s = 0.444, F = 43.489, FIT = 4.077, LOF = 0.298, AIC = 0.280,
$Q^2_{LOO}$ = 0.821, $Q^2_{L5O}$ = 0.796, $r^2_{randY}$(sd) = 0.127(0.105), $r^2_{Test}$ = 0.654 (8)

$pK_i$ = 2.076(0.286)nBnz + 1.791(0.391)SIC4 + 1.449(0.436)H-048 + 4.562
n = 23, r = 0.918, s = 0.492, F = 34.207, FIT = 3.206, LOF = 0.367, AIC = 0.345,
$Q^2_{LOO}$ = 0.758, $Q^2_{L5O}$ = 0.760, $r^2_{randY}$(sd) = 0.138(0.088), $r^2_{Test}$ = 0.667 (9)

$pK_i$ = -1.060(0.428)RBN + 2.835(0.371)IC5 – 1.531(0.399)N-075 + 6.878
n = 23, r = 0.909, s = 0.517, F = 30.455, FIT = 2.855, LOF = 0.404, AIC = 0.380,
$Q^2_{LOO}$ = 0.771, $Q^2_{L5O}$ = 0.796, $r^2_{randY}$(sd) = 0.133(0.091), $r^2_{Test}$ = 0.547 (10)

$pK_i$ = 0.750(0.294)nR05 + 1.797(0.281)nBnz + 1.718(0.450)SIC4 + 4.886
n = 23, r = 0.903, s = 0.534, F = 28.102, FIT = 2.634, LOF = 0.432, AIC = 0.406,

$Q^2_{LOO}$ = 0.716, $Q^2_{L5O}$ = 0.735, $r^2_{randY}$(sd) = 0.151(0.091), $r^2_{Test}$ = 0.533 (11)

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.

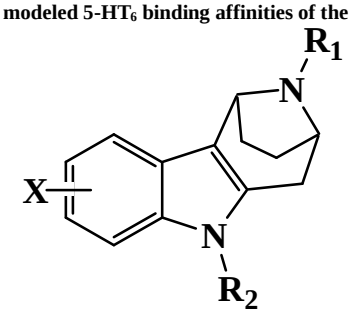
The descriptors nBnz, nCIC, nR05 and RBN belong to CONST class of Dragon descriptors. The constitutional class descriptors are based on simple constitutional facts and are independent from molecular connectivity and conformations. The descriptors nBnz (number of benzene-like rings), nCIC (number of rings) and nR05 (number of 5-membered rings) correlate positively and descriptor RBN (number of rotatable bonds) negatively to the activity suggest that more number of rings and lesser number of rotatable bonds in a molecular structure will be favorable to the binding affinity. The descriptor MLOGP is Moriguchi octanol-water partition coeff (logP) and reflects upon the hydrophobic property of a molecule. The positive contribution of this descriptor to the activity advocates a higher value of hydrophobicity for augmented activity.

The participated descriptors SIC4 and IC5 are from the TOPO class of Dragon descriptors. The TOPO class descriptors are based on a graph representation of the molecule and are numerical quantifiers of molecular topology obtained by the application of algebraic operators to matrices representing molecular graphs and whose values are independent of vertex numbering or labeling. They can be sensitive to one or more structural features of the molecule such as size, shape, symmetry, branching and cyclicity and can also encode chemical information concerning atom type and bond multiplicity. The descriptor SIC4 is the structural information content of 4-order neighborhood symmetry and IC5 is the information content index of 5-order neighborhood symmetry. Both the descriptors contributed positively to the activity. Thus, suggesting that a higher positive value of the structural information content of $4^{th}$ order neighborhood symmetry (SIC4) and the information content index of $5^{th}$ order neighborhood symmetry (IC5) would be beneficiary to the activity.

The descriptor GATS8p, in above models, is lone representative of 2D-AUTO class of Dragon descriptors. The 2D-AUTO descriptors, MATSke and GATSke have their origin in autocorrelation of topological structure of Moran and of Geary [25,26], respectively. 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 [27]. 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 GATS8p (Geary autocorrelation –lag 8/weighted by atomic polarizabilities) correlate negatively to the activity suggesting the unfavorable conditions associated with lag 8 weighted by atomic polarizabilities. The descriptors H-048, C-033 and N-075 emerged in above models are from the ACF (atom centered fragments) class. These molecular descriptors are based on the counting of 120 atom centered fragments as defined by Ghose and Crippen [28]. 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. Descriptors C-033 and H-047 have shown positive and N-075 negative correlation to the activity. Thus presence of X- -CH..X (descriptor C-033) and H attached to C2(sp3)/C1(sp2)/C0(sp) (descriptor H-047) and absence of R- - N- -R/R- -N- -X (descriptor N-075) type fragments in a molecular structure would be beneficiary to the activity.

These models have accounted for up to 87.28 percent variance in the observed activities. The values greater than 0.5 of $Q^2$-index is in accordance to a reasonable robust QSAR model. The $pK_i$ values of training set compounds calculated using Equations (8) to (11) have been included in Table 1. These models are validated with an external test set of seven compounds listed in Table 1. The predictions of the test set compounds based on external validation are found to be satisfactory as reflected in the test set $r^2$ ($r^2_{Test}$) values and the predicted activity values are also reported in Table 1. The plot showing goodness of fit between observed and calculated activities for the training and test set compounds is given in Figure 1.

**Table 1. Structures[a], observed and modeled 5-HT$_6$ binding affinities of the epiminocyclohepta[*b*]indole analogs**




| S. No. | $R_1$ | $R_2$ | X | *pK$_i$* | | | | |
|---|---|---|---|---|---|---|---|---|
| | | | | Obsd. | Eq.8 | Eq.9 | Eq.10 | Eq.11 |
| 1 | H | $CH_3$ | 6-$SO_2$Ph | 6.09 | 5.82 | 5.70 | 6.34 | 5.28 |
| 2 | H | $CH_3$ | 7-$SO_2$Ph | 6.88 | 7.15 | 7.46 | 7.11 | 7.47 |
| 3[b] | H | $CH_3$ | 8-$SO_2$Ph | 8.09 | 7.15 | 7.46 | 7.11 | 7.47 |
| 4 | $CH_3$ | $CH_3$ | 8-$SO_2$Ph | 8.11 | 7.02 | 7.18 | 6.91 | 7.20 |
| 5 | $C_2H_5$ | $CH_3$ | 8-$SO_2$Ph | 7.04 | 7.28 | 7.18 | 7.09 | 7.20 |
| 6 | H | H | 8-$SO_2$Ph | 7.27 | 7.35 | 7.79 | 7.33 | 7.79 |
| 7[b] | H | $C_2H_5$ | 8-$SO_2$Ph | 7.28 | 7.38 | 7.43 | 7.30 | 7.45 |
| 8 | H | $CH(CH_3)_2$ | 8-$SO_2$Ph | 6.46 | 6.53 | 6.64 | 6.30 | 6.68 |
| 9 | H | $CH_3$ | 8-$SO_2$(2-Fluorophenyl) | 8.34 | 8.74 | 8.28 | 8.31 | 8.26 |
| 10[b] | H | $CH_3$ | 8-$SO_2$(3-Fluorophenyl) | 8.39 | 8.74 | 8.28 | 8.31 | 8.26 |
| 11 | H | $CH_3$ | 8-$SO_2$(4-Fluorophenyl) | 7.22 | 7.60 | 7.46 | 7.11 | 7.47 |
| 12[b] | H | $CH_3$ | 8-$SO_2$(2-Chlorophenyl) | 8.32 | 8.87 | 8.28 | 8.31 | 8.26 |
| 13 | H | $CH_3$ | 8-$SO_2$(3-Chlorophenyl) | 8.74 | 8.87 | 8.28 | 8.31 | 8.26 |
| 14 | H | $CH_3$ | 8-$SO_2$(4-Chlorophenyl) | 7.43 | 7.73 | 7.46 | 7.11 | 7.47 |
| 15 | H | $CH_3$ | 8-$SO_2$(3-Trifluoromethylphenyl) | 8.62 | 8.79 | 7.94 | 8.04 | 7.94 |
| 16 | H | $CH_3$ | 8-$SO_2$(3-Trifluoromethoxyphenyl) | 8.14 | 7.90 | 7.97 | 7.93 | 7.96 |
| 17 | H | $CH_3$ | 8-$SO_2$(3-Aminophenyl) | 8.20 | 7.50 | 8.15 | 8.18 | 8.13 |
| 18 | H | $CH_3$ | 8-$SO_2$(3-Hydroxyphenyl) | 7.96 | 7.75 | 8.33 | 8.21 | 8.30 |
| 19 | H | $CH_3$ | 8-$SO_2$(3-Cyanophenyl) | 7.46 | 7.93 | 8.33 | 8.21 | 8.30 |
| 20[b] | H | $CH_3$ | 8-$SO_2$(2-Pyridyl) | 6.72 | 6.75 | 6.18 | 6.52 | 6.43 |
| 21 | H | $CH_3$ | 8-$SO_2$(3-Pyridyl) | 6.81 | 6.75 | 6.18 | 6.52 | 6.43 |
| 22[b] | H | $CH_3$ | 8-$SO_2$(4-Pyridyl) | 5.00 | 5.57 | 5.33 | 5.29 | 5.62 |
| 23 | H | $CH_3$ | 8-$SO_2$(3-Thiophenyl) | 7.29 | 7.43 | 7.55 | 7.50 | 7.11 |
| 24 | H | $CH_3$ | 8-$SO_2$(3-(1-Methyl)pyrazolyl) | 6.00 | 6.02 | 6.44 | 7.25 | 6.74 |
| 25[b] | H | $CH_3$ | 8-$SO_2$(1-Pyrrolo) | 7.49 | 7.73 | 6.70 | 6.53 | 6.30 |
| 26 | H | $CH_3$ | 8-$SO_2$(1-Indolyl) | 9.07 | 9.45 | 9.16 | 9.29 | 9.15 |
| 27 | H | $CH_3$ | 8-$SO_2$(3-Indolyl) | 8.89 | 8.97 | 9.16 | 9.29 | 9.15 |
| 28 | H | $CH_3$ | 8-$SO_2$(5-Indolyl) | 9.49 | 9.00 | 8.95 | 9.29 | 8.96 |
| 29 | H | $CH_3$ | 8-$SO_2$(3-Benzthiophenyl) | 9.48 | 8.87 | 9.10 | 9.06 | 9.10 |
| 30[b] | H | $CH_3$ | 8-$SO_2$(3-(1-Methyl)indolyl) | 9.09 | 8.76 | 8.82 | 9.01 | 8.83 |

[a]*Reference [9].*

**Table 2. Descriptor classes[a] used along with their definition and scope for modeling the binding affinity of epiminocyclohepta[*b*]indole derivatives**

| Descriptor class (acronyms) | Definition and scope |
|---|---|
| Constitutional (CONST) | Dimensionless or 0D descriptors; independent from molecular connectivity and conformations |
| Topological (TOPO) | 2D-descriptor from molecular graphs and independent conformations |
| Molecular walk counts (MWC) | 2D-descriptors representing self-returning walks counts of different lengths |
| Modified Burden eigenvalues (BCUT) | 2D-descriptors representing positive and negative eigenvalues of the adjacency matrix, weights the diagonal elements and atoms |
| Galvez topological charge indices (GALVEZ) | 2D-descriptors representing the first 10 eigenvalues of corrected adjacency matrix |
| 2D-autocorrelations (2D-AUTO) | Molecular descriptors calculated from the molecular graphs by summing the products of atom weights of the terminal atoms of all the paths of the considered path length (the lag) |
| Functional groups (FUNC) | Molecular descriptors based on the counting of the chemical functional groups |
| Atom centered fragments(ACF) | Molecular descriptors based on the counting of 120 atom centered fragments, as defined by Ghose-Crippen |
| Empirical (EMP) | 1D-descriptors represent the counts of non-single bonds, hydrophilic groups and ratio of the number of aromatic bonds and total bonds in an H-depleted molecule |
| Properties (PROP) | 1D-descriptors representing molecular properties of a molecule |

[a]*Reference [11]*

**Table 3. Descriptors[a] identified for modeling the binding affinity of epiminocyclohepta[*b*]indole derivatives along with the average regression coefficient[b], standard deviation and the total incidence**

| Descriptor | Avg reg coeff(sd) total incidence | Descriptor | Avg reg coeff(sd) total incidence | Descriptor | Avg reg coeff(sd) total incidence |
|---|---|---|---|---|---|
| MW | 3.546(0.403)2 | IC2 | 1.759(0.383)2 | MATS3v | -1.944(0.668)2 |
| AMW | 1.510(0.469)5 | TIC2 | 2.380(0.000)1 | MATS1e | -1.457(0.361)5 |
| Me | 1.233(0.050)2 | IC3 | 1.771(0.320)7 | MATS8e | -2.740(0.000)1 |
| nBM | 2.680(0.862)3 | SIC3 | 1.649(0.404)8 | MATS8p | 1.666(0.000)1 |
| nCIC | 2.855(0.607)16 | SIC4 | 2.350(0.383)19 | GATS1e | 1.371(0.000)1 |
| RBN | -1.584(0.742)2 | IC5 | 2.835(0.000)1 | GATS6e | 2.205(0.000)1 |
| nS | 1.463(0.000)1 | D/Dr06 | 2.332(0.274)2 | GATS8p | -1.781(0.152)8 |
| nR05 | 1.275(0.304)18 | T(N..N) | -1.250(0.000)1 | nCaH | 2.178(0.401)4 |
| nR06 | 2.884(0.604)3 | BEHm4 | 2.493(0.204)3 | nHDon | 1.861(0.000)1 |
| nBnz | 1.868(0.334)28 | BELm1 | -2.295(0.244)2 | C-024 | 2.681(0.347)8 |
| HNar | 3.737(0.568)2 | BELm2 | 2.975(0.030)2 | C-025 | 0.794(0.012)2 |
| MSD | -2.182(0.665)3 | BELm3 | -3.840(0.000)1 | C-027 | -1.328(0.298)8 |
| X1A | -2.201(0.513)13 | BELm8 | 1.759(0.000)1 | C-033 | 1.903(0.178)3 |
| X1Av | -1.816(0.000)1 | BELv4 | 1.868(0.000)1 | C-034 | 1.167(0.000)1 |
| PW2 | 2.655(0.445)3 | BELv7 | 1.040(0.000)1 | H-047 | 1.698(0.000)1 |
| PW4 | 2.033(0.000)1 | BEHe7 | 1.416(0.000)1 | H-048 | 2.023(0.724)11 |
| Lop | -2.428(0.227)3 | BEHp6 | 1.777(0.639)2 | H-052 | -2.085(0.648)10 |
| IDDE | 2.094(1.068)2 | BEHp7 | 1.088(0.000)1 | N-075 | -1.711(0.178)25 |
| CIC0 | 1.206(0.000)1 | BELp4 | -2.625(0.000)1 | ARR | 1.464(0.000)1 |
| TIC1 | 2.612(0.000)1 | JGT | 3.787(1.698)3 | MLOGP | 2.854(0.549)24 |

[a]The descriptors are identified from the three parameter models emerged from CP-MLR protocol with filter-1 as 0.79; filter-2 as 2.0; filter-3 as 0.79; filter-4 as $0.3 \leq Q^2 \leq 1.0$; number of compounds in the study are 23; **CONST**: MW, molecular weight; AMW, average molecular weight; Me, mean atomic Sanderson electronegativity (scaled on Carbon atom); nBM, number of multiple bonds; nCIC, number of rings; RBN, number of rotatable bonds; nS, number of sulfur atoms; nR05, number of 5-membered rings; nR06, number of 6-membered rings; nBnz, number of benzene-like rings; **TOPO**: HNar, Narumi harmonic topological index; MSD, mean square distance index (Balaban); X1A, average connectivity index chi-1; X1Av, average valence connectivity index chi-1; PW2, path/walk 2 - Randic shape index; PW4, path/walk 4 - Randic shape index; Lop, Lopping centric index; IDDE, mean information content on the distance degree equality; CIC0, complementary information content (neighborhood symmetry of 0-order); TIC1, total information content index (neighborhood symmetry of 1-order); IC2, information content index (neighborhood symmetry of 2-order); TIC2, total information content index (neighborhood symmetry

of 2-order); IC3, information content index (neighborhood symmetry of 3-order); SIC3, structural information content (neighborhood symmetry of 3-order); SIC4, structural information content (neighborhood symmetry of 4-order); IC5, information content index (neighborhood symmetry of 5-order); D/Dr06, distance/detour ring index of order 6; T(N..N), sum of topological distances between N..N; **BCUT**: BEHm4, highest eigenvalue n. 4 of Burden matrix / weighted by atomic masses; BELm1, BELm2, BELm3 and BELm8, lowest eigenvalue n. 1, 2,3 and 4 of Burden matrix / weighted by atomic masses, respectively; BEHv3 and BEHv6, highest eigenvalue n. 3 and 6 of Burden matrix / weighted by atomic van der Waals volumes, respectively; BEHe7, highest eigenvalue n. 7 of Burden matrix / weighted by atomic Sanderson electronegativities; BELv4 and BELv7, lowest eigenvalue n. 4 and 7 of Burden matrix / weighted by atomic van der Waals volumes; BEHp6, BEHp7, highest eigenvalue n. 6 and 7 of Burden matrix / weighted by atomic polarizabilities, respectively; BELp4, lowest eigenvalue n. 4 of Burden matrix / weighted by atomic polarizabilties; **GALVEZ**: JGT, global topological charge index; **2D-AUTO**: MATS3v, Moran autocorrelation - lag 3/ weighted by atomic van der Waals volumes; MATS1e, MATS8e, Moran autocorrelation - lag 1 and 8, respectively / weighted by atomic Sanderson electronegativities; MATS8p, Moran autocorrelation - lag 8 / weighted by atomic polarizabilities; GATS1e and GATS6e, Geary autocorrelation - lag 1and 6, respectively / weighted by atomic Sanderson electronegativities; GATS8p, Geary autocorrelation - lag 8 / weighted by atomic polarizabilities; **FUNC**: nCaH, number of unsubstituted aromatic C(sp2); nHDon, number of donor atoms for H-bonds (with N and O); **ACF**: C-024, R- - CH- - R; C-025, R- -CR- -R; C-027, R- - CH- -X; C-033, X- - CH..X; C-034, X- - CR..X; H-047, H attached to C1(sp3)/C0(sp2); H-048, H attached to C2(sp3)/C1(sp2)/C0(sp); H-052, H attached to C0(sp3) with one X attached to next C atom; N-075, R- - N- -R/R- -N- -X; **EMP**: ARR, aromatic ratio; **PROP**: MLOGP, Moriguchi octanol-water partition coeff (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.

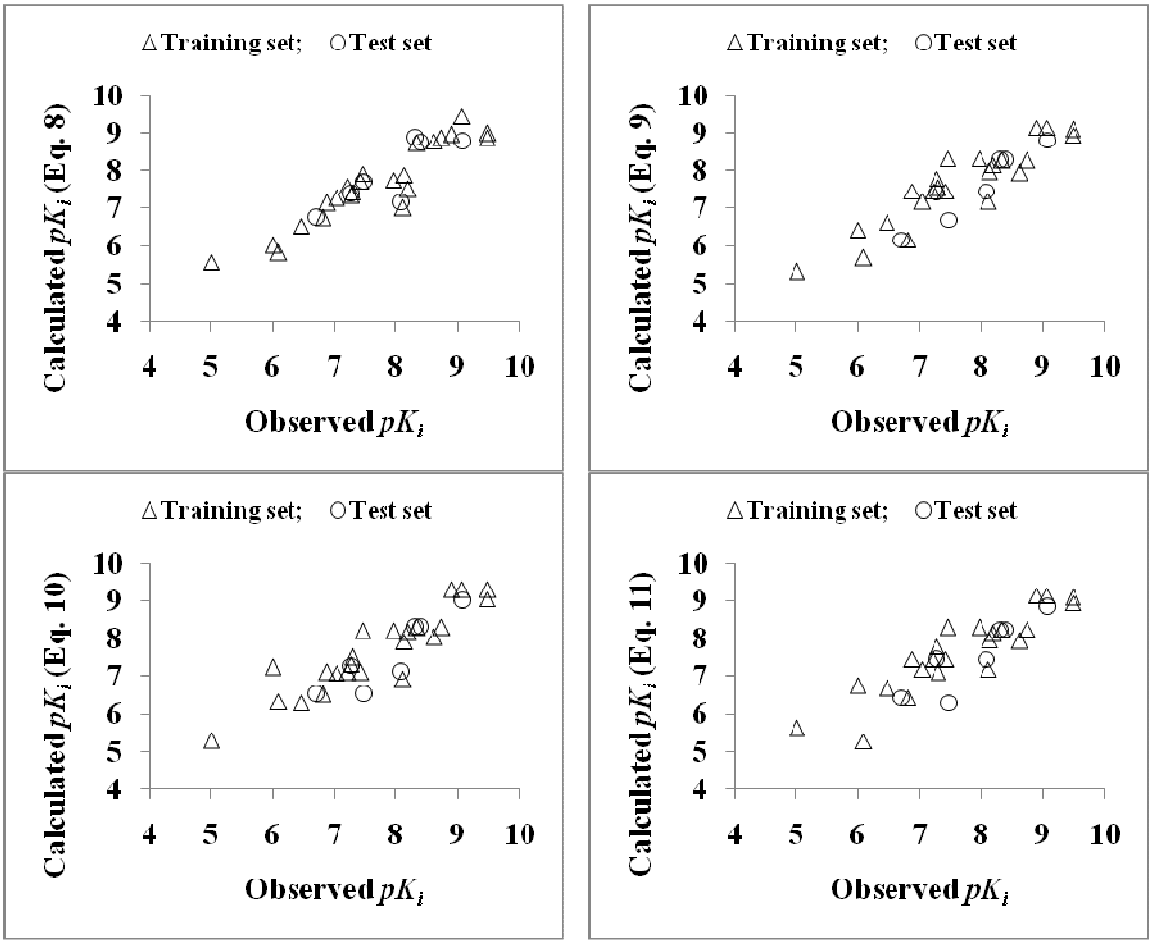



**Figure 1. Plot of the observed versus calculated pKi values of epiminocyclohepta[*b*]indoles**

## CONCLUSION

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

In order to improve the 5-$HT_6$ receptor binding affinity of a compound, more number of rings and lesser number of rotatable bonds in molecular structure are advocated by descriptors nBnz, nCIC and nR05, and RBN, respectively. A higher value of the molecular topology and symmetry accounting parameters (SIC4, structural information content of 3-order neighborhood symmetry and IC5, information content index of $5^{th}$ order neighborhood symmetry) is favorable to the activity. The polarizability associated to path length 8 of the Geary autocorrelation (GATS8p) and hydrophobicity of molecule (MLOGP) have shown their prevalence to explain the binding affinity. Additionally, presence of structural fragment X- -CH..X (descriptor C-033), absence of R- - N- -R or R- -N- -X type fragment (descriptor N-075) and more number of hydrogen atoms attached to sp or sp2 or sp3 hybridized carbon atoms (H-047) in a molecular structure are also relevant for elevated binding affinity. The derived models and participating descriptors in them have suggested that the substituents of epiminocyclohepta[*b*]indole moiety have sufficient scope for further modification.

**Acknowledgements**
Authors are thankful to their Institution for providing necessary facilities to complete this study. MC thanks CSIR, New Delhi, India for financial support in the form of SRF.

## REFERENCES

[1] W.D. Hirst, B. Abrahamsen, F.E. Blaney, A.R. Calver, L. Aloj, G.W. Price, A.D. Medhurst, *Mol. Phamacol.*, **2003**, 64, 1295.
[2] S.R. Morairty, L. Hedley, J. Flores, R. Martin, T.S. Kilduff, *Sleep*, **2008**, 31, 44.
[3] A. Wesolowska, A. Nikiforuk, *Neuropharmacology*, **2007**, 52, 1274.
[4] C. Routledge, S.M. Bromidge, S.F. Moss, G.W. Price, W. Hirst, H. Newman, G. Riley, T. Gager, T. Stean, N. Upton, S.E. Clarke, A.M. Brown, D.M. Middlemiss, *Br. J. Pharmacol.*, **2000**, 130, 1606.
[5] D.P. Finn, K.C. Fone, S.R. Beckett, J.A. Baxter, L. Ansell, C.A. Marsden, V. Chapman,*Eur. J. Pharmacol.*, **2007**, 569, 59.
[6] (a) D.J. Heal, S.L. Smith, A. Fisas, X. Codony, H. Buschmann, *Pharmacol. Therap.*, **2008**, 117, 207.; (b) B.J. Sargent, A.J. Henderson, *Curr. Opin. Pharmacol.*, **2011**, 11, 52.
[7] (a) M.V. King, C.A. Marsden, K.C.F. Fone, *Trends Pharmacol. Sci.*, **2008**, 29, 482.; (b) G. Rosse, H. Schaffhauser, *Curr. Top. Med. Chem.*, **2010**, 10, 207.
[8] G. Maher-Edwards, M. Zvartau-Hind, A.J. Hunter, M. Gold, G. Hopton, G. Jacobs, M. Davy, P. Williams, *Curr. Alzheimer. Res.*, **2010**, 7, 374.
[9] A.J. Henderson, P.R. Guzzo, A. Ghosh, J. Kaur, J.-M. Koo, K. Nacro, S. Panduga, R. Pathak, B. Shimpukade, V. Tan, K. Xiang, J.D. Wierschke, M.L. Isherwood, *Bioorg. Med. Chem. Lett.*, **2012**, 22, 1494.
[10] ChemDraw Ultra 6.0 and Chem3D Ultra, Cambridge Soft Corporation, Cambridge, USA. http://www.cambridgesoft.com
[11] Dragon software (version 1.11-2001) by R. Todeschini, V. Consonni, Milano, Italy. http://www.talete.mi.it/dragon.htm
[12] Y.S. Prabhakar, *QSAR Comb. Sci.*, **2003**, 22, 583.
[13] S. Sharma, Y.S. Prabhakar, P. Singh, B.K. Sharma, *Eur. J. Med. Chem.*, **2008**, 43, 2354.
[14] B.K. Sharma, P. Pilania, P. Singh, *J. Enzy. Inhibn. Med. Chem.*, **2009**, 24, 607.
[15] S. Sharma, B.K. Sharma, P. Pilania, P. Singh, Y.S. Prabhakar, *J. Enzy. Inhibn. Med. Chem.*, **2009**, 24, 1024.
[16] B.K. Sharma, P. Pilania, P. Singh, Y.S. Prabhakar, *SAR QSAR Environ. Res.*, **2010**, 21, 169.
[17] B.K. Sharma, P. Pilania, K. Sarbhai, P. Singh, Y.S. Prabhakar, *Mol. Divers.*, **2010**, 14, 371.
[18] H. Akaike, Information theory and an extension of the minimum likelihood principle. In: B.N Petrov, F. Csaki F (eds) Second international symposium on information theory. Akademiai Kiado, Budapest, **1973**, 267.
[19] H. Akaike, *IEEE Trans Automat Contr* 1974; AC-19: 716-723.
[20] H. Kubinyi, *Quant. Struct.-Act. Relat.*, **1994**, 13, 285.
[21] H. Kubinyi, *Quant. Struct.-Act. Relat.*, **1994**, 13, 393.

[22] J. Friedman, In: Technical report no. 102. Laboratory for computational statistics. Stanford University, Stanford, **1990**.
[23] S.-S. So, M. Karplus, *J. Med. Chem.*, **1997**, 40, 4347.
[24] Y.S. Prabhakar, V.R. Solomon, R.K. Rawal, M.K. Gupta, S.B. Katti, *QSAR Comb. Sci.*, **2004**, 23, 234.
[25] P.A.P. Moran, *Biometrika*, **1950**, 37, 17.
[26] R.C. Geary, *Incorp. Statist*., **1954**, 5, 115.
[27] P. Broto, G. Moreau, C. Vandycke, *Eur. J. Med. Chem.*, **1984**, 19, 66.
[28] V.N. Viswanadhan, A.K. Ghose, G.R. Revankar, R.K. Robins, *J. Chem. Inf. Comput. Sci.*, **1989**, 29, 163.