Available on line www.jocpr.com

Journal of Chemical and Pharmaceutical Research

J. Chem. Pharm. Res., 2010, 2(2): 147-162

ISSN No: 0975-7384

QSAR analysis and validation studies on substituted spiropiperidines as GlyT1 inhibitors

Ajay Babu Padavala[a*], V. Vishnu Prasanth[a], Sankar Aditya Jayanthi[b], Agasteswar Vadlamani[b] and Sashikanth Chitti[a]

[a] *Bioinformatics Division, Translational Research Institute of Molecular Sciences (TRIMS), Visakhapatnam, India*
[b]*Department of Biotechnology, Anil Neerukonda Institute of Technology and Sciences, Tagarapuvalasa, Visakhapatnam, India*

Abstract

QSAR (Quantitative Structure Activity Relationship) studies were carried out on a set of 61 *N*-(2-aryl-cyclohexyl) and *N*-(2-hydroxy-2-aryl-cyclohexyl) substituted spiropiperidines as GlyT1 inhibitors (Glycine Transporter1) using multiple regression procedure. The activity contributions of these compounds were determined from regression equation and the validation procedures such as external set cross-validation r^2 ($R^2_{cv,ext}$) and the regression of observed activities against predicted activities and vice versa for validation set were described to analyze the predictive ability of the QSAR model. An accurate and reliable QSAR model involving six descriptors was chosen based on the FIT Kubinyi function. Applicability domain of QSAR model such as leverages, y-randomization test and RMSE (Root Mean Square Error) of training and validation set were reported.

Keywords: Multivariate analysis; Molecular diversity; Physico-chemical properties.

Introduction

Several new antipsychotic drugs have been introduced in the last decade that promised to treat schizophrenia than others without unwanted side effects [1]. Schizophrenia is characterized by failures in nearly all aspects of higher-order behavior such as disruption of information processing and sensory perception, abnormal mood and personal hygiene, cognitive impairments including attention, short-term memory, and behavioral flexibility and certain movement abnormalities as well [2]. It has been postulated that enhancement of glutamate transmission, in

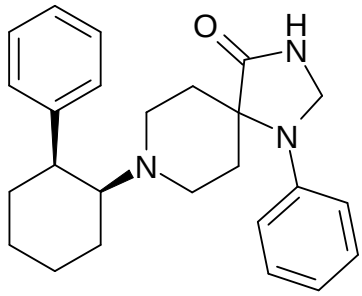
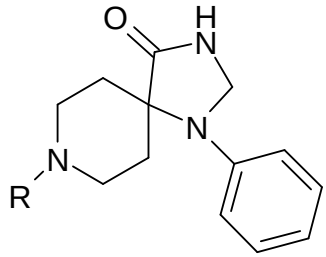
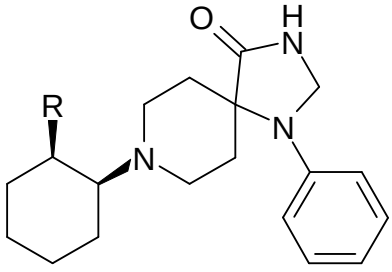
particular, N-methyl-D-aspartate (NMDA) receptor activation, produces both anti-psychotic and cognitive enhancing effects, thus constituting a potential therapeutic target for the treatment of schizophrenia, psychoses and cognitive impairment [3-7]. Some of the potential targets for pharmacological intervention in schizophrenia include the glycine and D-serine binding site on the NR1 subunits of the NMDA receptor, the Glycine Transporter (Gly T), and potentiators of metabotropic glutamate (mGlu) receptors, in particular mGlu5, which positively modulates NMDA receptors through activation of the G protein Gq, and mGlu2/3, which regulate the release of glutamate [2]. NMDA receptors are complex, abundant, ubiquitously distributed throughout the brain and activation requires both glutamate and glycine binding to open the ion channel and permit the calcium entry. Glutamate, released from pre-synaptic terminals, has the neurotransmitter role whereas glycine which is present in the extracellular fluid acts as a modulator. Evidence indicates that potentiating NMDA receptor should be beneficial for treating cognitive disorders and schizophrenia [8]. Moreover, activation of the glycine site has shown some clinical benefit. Apart from normal antipsychotic therapy, addition of glycine or other glycine site agonists, D-seine and D-cycloserine were reported to show efficacy in treating schizophrenia [9]. However, few companies have focused on another approach to increase extracellular levels of glycine by blocking glycine re-uptake into neurons through the inhibition of GlyT1 transporter [10]. GlyT1 is the only sodium chloride dependent glycine transporter in the forebrain, co-expressed with the NMDA receptor. GlyT1 is thought to be responsible for control of extracellular level of glycine at the synapse. Several groups have focused their efforts in developing selective GlyT1 inhibitors [11] and a variety of non-amino acid GlyT1 inhibitors were reported [12].

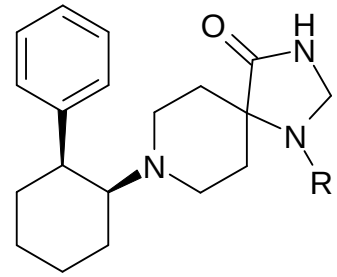
Quantitative Structure activity relationship (QSAR) studies delineate the structural requirements for potency of inhibitors. QSAR studies have been investigated on the basis of the fact that the biological activity of the compound is a function of its physicochemical properties. From literature, it was observed that very few attempts were made to build QSAR models of GlyT1 inhibitors. In this paper, we report QSAR studies on N-(2-aryl-cyclohexyl) substituted spiropiperidines as GlyT1 inhibitors to investigate the influence of molecular structure on biological activity. Several validations were reported which state the robustness and domain applicability of the model.

Materials and Methods

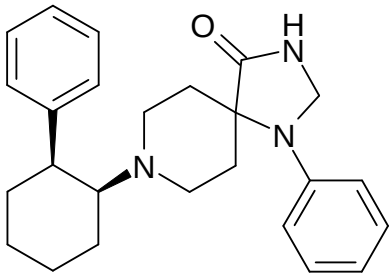
To obtain a reliable and robust QSAR model, it is desirable to consider a large data set that covers reasonable chemical diversity and biological activity. Hence, a set of 61 compound biological data was taken from 2 references [13, 14]. The structures along with bioactivities are given in Table I. The inhibitory activities of these derivatives reported in terms of IC_{50} in μM were transformed into their corresponding concentration values in order to overcome overlapping data. Therefore, to guarantee the linear distribution of data, the enzyme inhibition was converted to negative logarithmic values and then used for subsequent QSAR analysis. The structures were sketched using ISIS Draw 2.3 (www.mdli.com) software and the descriptors were calculated using Tsar 3.3 software (www.accelrys.com). Before the calculation of descriptors, three dimensional structures of all molecules were generated using Corina 3D package, charges were derived and the geometries optimized using cosmic module of Tsar.

Table I: Structure and biological activities of compounds

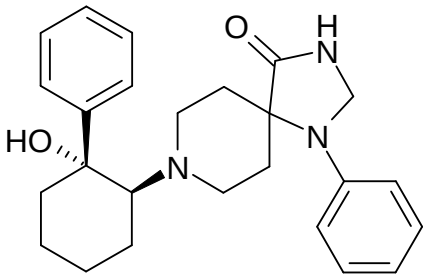
ID	Compound name	EC_{50} (μM)	Molecular Mass	Molar conc. log(C)
				
1	(cis,rac)	0.026	389.59	-7.176
2	(trans,rac)	0.073	389.59	-6.727
3	(cis, 1R,2R)	0.004	389.59	-7.988
4	(cis,1S,2S)	0.380	389.59	-6.010
				
5	cis-2-phenyl-cyclopent-1yl	1.1	375.56	-5.533
6	cis-2-phenyl-cyclohept-1yl	1.7	403.57	-5.375
7	1,1-dimethyl-2-phenyl-ethyl	9.3	363.55	-4.592
8	2-methyl-2-phenyl-propyl	7.3	363.55	-4.697
				
9	4-F-Ph	0.023	407.58	-7.248
10	4-Cl-Ph	0.040	424.03	-7.025
11	4-Me-Ph	0.085	403.62	-6.676
12	4-MeO-Ph	0.610	419.62	-5.837
13	3-Cl-Ph	0.130	424.03	-6.513
14	3,4-Cl_2-Ph	0.067	458.47	-6.835
15	3-CF_3,4-Cl-Ph	1.8	492.03	-5.436
16	2-Me-Ph	0.510	403.62	-5.898
17	2-Py	6.6	390.53	-4.772



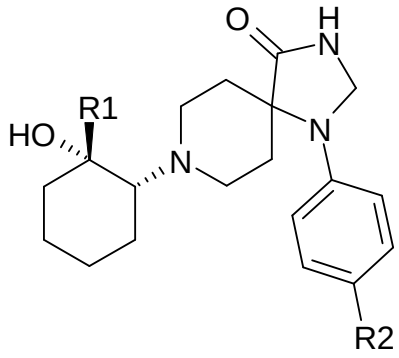
18	4-F-Ph	0.024	407.58	-7.232
19	4-Cl-Ph	0.027	424.03	-7.196
20	4-CF_3-Ph	0.065	457.59	-6.847
21	4-MeO-Ph	0.066	419.62	-6.803
22	4-Me-Ph	0.055	403.62	-6.865
23	Et	6.6	341.55	-4.713
24	nPr	0.450	355.58	-5.897
25	nPent	0.034	383.64	-7.052
26	nHex	0.049	397.67	-6.909
27	cPr	3.3	353.56	-5.029
28	cBu	1.5	367.59	-5.389
29	cPent	0.63	381.62	-5.782
30	cHex	0.75	395.65	-5.722
31	CH_2-cHex	0.065	409.68	-6.799
32	CH_2CH_2-cHex	0.025	423.71	-7.229
33	CH_2-Ph	0.258	403.62	-6.194
34	CH_2CH_2-Ph	0.025	417.65	-7.222
35	CH_2CH_2-OMe	5.3	371.58	-4.845



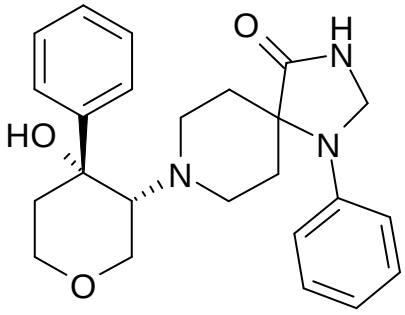
36	(cis,rac)	0.026	389.59	-7.175
37	(trans,rac)	0.073	389.59	-6.727



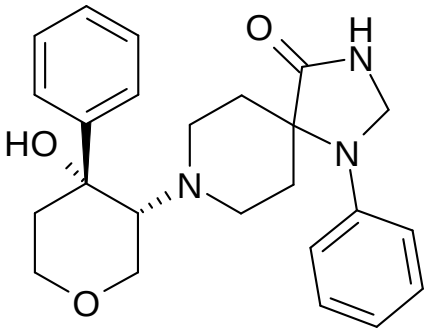
38	(trans,rac)	0.056	405.59	-6.859
39	(cis,rac)	0.044	405.59	-6.964



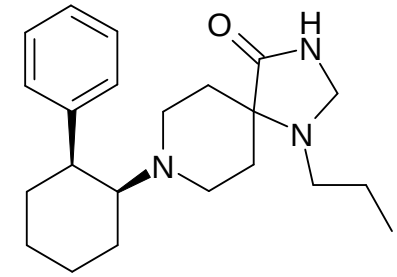
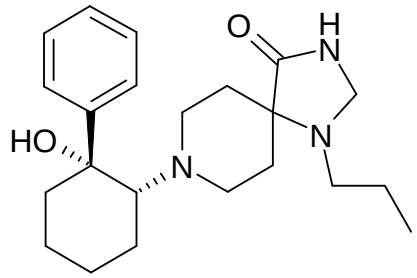
40	4-MeO-Ph	H	0.261	435.62	-6.222
41	4-Me-Ph	H	0.140	419.62	-6.479
42	4-Cl-Ph	H	0.080	440.03	-6.740
43	3,4-Cl_2 –Ph	H	0.173	474.47	-6.438
44	4-F-Ph	H	0.040	423.58	-7.024
45	2-Me-Ph	H	0.050	419.62	-6.923
46	3-Cl-Ph	H	0.130	440.03	-6.529
47	3-Me-Ph	H	0.130	435.62	-6.525
48	3-MeO-Ph	H	0.130	435.57	-6.525
49	2-Py	H	0.130	406.58	-6.495
50	3-Py	H	0.110	406.58	-6.567
51	4-Py	H	0.062	406.58	-6.816
52	Me	H	9	343.52	-4.581
53	t-Bu	H	29	385.61	-4.123
54	4-F-Ph	F	0.024	441.57	-7.264
55	4-F-Ph	Cl	0.015	458.02	-7.484
56	4-Cl-Ph	F	0.024	458.02	-7.280
57	4-Cl-Ph	MeO	0.099	470.06	-6.676



58	(cis,rac)		0.058	407.56	-6.846



59	(cis,rac)		0.090	407.56	-6.655

				
60	-	0.45	355.58	-5.897
				
61	-	0.40	371.58	-5.967

Multivariate Regression Analysis

QSAR models were constructed on complete and training sets, respectively. Validation was done internally using leave-one-out (LOO) technique and externally by predicting the activities of validation set. The relationship between dependent variable (logC) and independent variables was established by linear multiple regression analysis using Tsar. Significant descriptors were chosen based on the statistical data of analysis. Statistical quality of the generated QSAR equation was judged based on the parameters like correlation coefficient (r), standard error of estimate (s), F-value, cross-validation r^2 (q^2) and predictive residual sum of squares (PRESS). Cross-validation was calculated using leave-one-out (LOO) technique over 2 random trials with F to leave and F to enter being 2 in F stepping to include the most significant variables in generating the QSAR model.

Thirty five molecular descriptors were selected for the study: topological, shape and connectivity indices, total dipole and lipole, molecular weight, h-bond donors, h-bond acceptors, logP and rotatable bond counts. A semi-empirical molecular orbital package was used to calculate thermodynamic property like heat of formation and electrostatic properties like HOMO (Highest Occupied Molecular Orbital) and LUMO (Lowest Unoccupied Molecular Orbital).

Predictive Ability of QSAR model

Predictive ability of the generated model was estimated externally by predicting the activities of validation set. This criterion may not be sufficient for a QSAR model to be truly predictive [15]. An additional condition for high predictive ability of QSAR model is based on external set cross-validation r^2, ($R^2_{cv,ext}$) and the regression of observed activities against predicted activities and vice versa for validation set, if the following conditions are satisfied [15, 16].

$$R^2_{cv,ext} > 0.5 \quad (1)$$

$$R^2 > 0.6 \quad (2)$$

$$(R^2 - R_0^2) / R^2 < 0.1 \text{ or } (R^2 - R_0'^2) / R^2 < 0.1 \quad (3)$$

$$0.85 \leq k \leq 1.15 \text{ or } 0.85 \leq k' \leq 1.15 \quad (4)$$

Calculations relating to $R^2_{cv,ext}$, R_0^2 and the slopes, k and k' are based on regression of observed values against predicted values and vice versa [15].

Applicability Domain of QSAR Model

Applicability domain of a QSAR model must be defined if the model is to be used for screening new compounds. Predictive ability of the model may be considered reliable for compounds that fall into this domain [16]. One simple approach is based on y-randomization and the calculation of leverage for each chemical compound used in the study.

Y-randomization

This test ensures the robustness of a QSAR model [17] and to assess the multiple linear regression models obtained by descriptor selection [18]. In y-randomization test, the dependent variable or y-data is randomly shuffled and a new QSAR model is developed keeping X-data intact. The new models are expected to have low R^2 and Q^2 values, which determine the statistical significance of the original model. Moreover, if the model development includes F-stepping, then it is necessary to shuffle both dependent and independent variables to indicate that the original model is not because of chance correlation.

Leverage Test

Leverage values refers to the diagonal elements of the hat matrix H = (X (X'X)–1X'). A given diagonal element (h [ii]) represents the distance between the X value for the i^{th} observation and the means of all X values. Leverages measure the distance of an observation from the centre of a set of X observations [19]. A leverage value, hii, greater than 3p/n is usually considered large (where p is the number of parameters in the model plus constant and n is the number of observations). If high leverage points fit the model well (i.e. have small residuals), they are called "good high leverage points" or good influence points. Such points stabilize the model and make it more precise. High leverage points, which do not fit the model (i.e. have large residuals) are called "bad high leverage points" or bad influence points [19].

Results and Discussion

Multivariate regression analysis with F stepping (F to enter and F to leave being 2) and cross-validation by leaving-out-one row, to test the predictive power, resulted in Kappa2 index, Kier chi4 (cluster index), Kier chiV3 (cluster index), H-bond donors and H-bond acceptors as the most significant descriptors. Equation 5 represents the linear QSAR model from a complete set of 61 GlyT1 inhibitors.

log (C) = + 1.044 *Kier chiV3
+ 3.652 * Kier chi4
- 0.725 * Kappa2 index
+ 0.614 * H-bond acceptors
- 1.533 * H-bond donors
- 1.131

r =0.876, r2 = 0.767, q2 =0.865, F = 33.616, n = 61,
PRESS = 5.170, s = 0.418 (5)

A new QSAR model was attempted by dividing the set as a 50 molecule training set and a 7 molecule validation set (Table II). More specifically, the selection of molecules in the training set was made according to the biological action and molecular structure, so that representatives

of a wide range of structures with different substituents, atoms and activity were included. The distribution of activity values for the validation set follows the similar distribution of the activity values for the training set [20].

Table II: Logarithmic molar concentration values of complete, training and validation set and descriptor values of the proposed QSAR model (Eq. 7).

ID	Activity log(C)	Complete Set[b] log (C)	Training Set[c] log (C)	Validation Set[d] log (C)	Balaban topological index	Molecular Refractivity	Kappa 2 index	H-bond Donors	H-bond Acceptors	Kier Chi4
1	-7.176	-6.738	-6.812	----	1.161	116.481	9.24	1	2	0.083
2	-6.727	-6.738	-6.812	----	1.161	116.481	9.24	1	2	0.083
3	-7.988	-7.189	-7.012	----	1.161	116.481	9.24	1	2	0.083
4	-6.011	-6.738	-6.812	----	1.161	116.481	9.24	1	2	0.083
5	-5.533	-6.292	-6.350	----	1.175	111.88	8.63	1	2	0.083
6	-5.375	-5.154	-5.241	----	1.196	121.08	9.86	1	2	0.055
7[a]	-4.592	-4.674	----	-4.637	1.320	109.324	8.39	1	2	0.287
8	-4.697	-4.595	-4.747	----	1.301	109.138	8.39	1	2	0.287
9	-7.248	-6.838	-7.049	----	1.172	116.697	9.47	1	2	0.083
10	-7.025	-6.696	-6.611	----	1.172	121.286	9.47	1	2	0.083
11	-6.676	-6.730	-6.588	----	1.172	121.522	9.47	1	2	0.083
12	-5.838	-6.671	-6.628	----	1.162	122.944	10.09	1	3	0.083
13	-6.513	-6.696	-6.628	----	1.168	121.286	9.47	1	2	0.083
14	-6.835	-6.711	-6.485	----	1.168	126.09	9.70	1	2	0.083

15	-5.437	-6.129	-6.030	----	1.158	127.259	10.40	1	2	0.372
16	-5.898	-6.767	-6.496	----	1.191	121.522	9.47	1	2	0.083
17	-4.772	-4.845	-4.235	----	1.161	114.007	9.24	1	3	0.055
18	-7.230	-6.838	-7.119	----	1.157	116.697	9.47	1	2	0.083
19	-7.196	-6.696	-6.681	----	1.157	121.286	9.47	1	2	0.083
20	-6.847	-6.128	-6.226	----	1.147	122.454	10.17	1	2	0.372
21	-4.697	-6.671	-6.689	----	1.149	122.944	10.09	1	3	0.083
22[a]	-7.248	-6.730	-----	- 6.659	1.157	121.522	9.47	1	2	0.083
23	-7.025	-5.221	-5.015	----	1.316	101.451	7.94	1	3	0.083
24	-6.676	-5.679	-5.446	----	1.314	105.975	8.57	1	3	0.083
25	-5.838	-6.625	-6.442	----	1.292	115.177	9.87	1	3	0.083
26	-6.513	-7.113	-6.997	----	1.275	119.778	10.54	1	3	0.083
27	-6.835	-4.748	-4.777	----	1.176	103.914	7.44	1	3	0.083
28[a]	-5.437	-5.173	-----	-5.129	1.177	108.515	8.02	1	3	0.083
29[a]	-5.898	-5.609	-----	-5.531	1.172	113.116	8.63	1	3	0.083
30	-4.772	-6.054	-5.977	----	1.161	117.717	9.24	1	3	0.083
31	-7.230	-6.410	-6.494	----	1.136	122.448	9.87	1	3	0.083
32	-7.196	-6.874	-7.034	----	1.112	127.126	10.51	1	3	0.083
33	-6.847	-6.500	-6.602	----	1.137	121.316	9.87	1	3	0.083

34	-7.222	-6.964	-7.135	----	1.112	126.07	10.51	1	3	0.083
35	-4.846	-5.533	-5.476	----	1.306	107.747	9.21	1	4	0.083
36	-7.176	-6.738	-6.769	----	1.170	116.481	9.24	1	2	0.083
37	-6.727	-6.738	-6.731	----	1.178	116.481	9.24	1	2	0.083
38	-6.860	-6.921	-6.936	----	1.195	117.778	9.09	2	3	0.201
39	-6.965	-6.921	-6.936	----	1.195	117.778	9.09	2	3	0.201
40	-6.222	-6.843	-6.702	----	1.202	124.241	9.93	2	4	0.201
41	-6.480	-6.918	-6.696	----	1.211	122.819	9.33	2	3	0.201
42	-6.740	-6.884	-6.719	----	1.211	122.583	9.33	2	3	0.201
43	-6.438	-6.904	-6.595	----	1.208	127.387	9.57	2	3	0.201
44	-7.025	-7.026	-7.156	----	1.211	117.994	9.33	2	3	0.201
45	-6.924	-6.953	-6.602	----	1.231	122.819	9.33	2	3	0.201
46	-6.530	-6.884	-6.736	----	1.208	122.583	9.33	2	3	0.201
47	-6.525	-6.843	-6.734	----	1.196	124.241	9.93	2	4	0.201
48	-6.525	-6.011	-6.025	----	1.184	125.33	9.93	2	4	0.108
49[a]	-6.495	-6.326	-----	-6.361	1.216	115.249	9.09	2	4	0.201
50	-6.567	-6.307	-6.328	----	1.216	115.589	9.09	2	4	0.201
51	-6.817	-6.307	-6.328	----	1.216	115.589	9.09	2	4	0.201
52	-4.582	-5.061	-5.009	----	1.372	97.9131	7.20	2	3	0.287

53	-4.124	-3.997	-3.956	----	1.387	111.307	7.92	2	3	0.435
54	-7.264	-7.132	-7.573	----	1.187	118.211	9.57	2	3	0.201
55	-7.485	-6.990	-7.136	----	1.187	122.799	9.57	2	3	0.201
56[a]	-7.281	-6.990	-----	-7.136	1.187	122.799	9.57	2	3	0.201
57	-6.677	-6.805	-6.676	----	1.178	129.046	10.17	2	4	0.201
58	-6.847	-6.307	-6.486	----	1.195	114.975	9.09	2	4	0.201
59	-6.656	-6.307	-6.420	----	1.209	114.975	9.09	2	4	0.201
60	-5.898	-5.679	-5.446	----	1.315	105.975	8.57	1	3	0.083
61[a]	-5.968	-5.842	-----	-5.466	1.363	107.272	8.39	2	4	0.201

[a] *Validation set molecules. ;* [b] *Calculated values from Equation 5;* [c] *Calculated values from Equation 7*
[d] *Predicted values from Equation 7*

Cross-validation was performed using leave-one-out (LOO) technique over 2 random trials with F to enter and F to leave being 2 in F stepping to include the most significant variables in generating the QSAR model. The results obtained from the multiple linear regression procedure with varied number of descriptors are shown in Table III with their statistics (Eqs. 6-8).

Table III: Descriptor data and statistical values of 5, 6 and 7 variable model equations

Descriptor	**Coefficient**		
	5-variable model	**6-variable model**	**7-variable model**
Balaban index	-	+4.803	+5.257
Molecular refractivity	-	+0.095	+0.131
Kappa2 index	-1.210	-1.355	-1.498
H-bond donors	-1.672	-1.872	-2.024
H-bond acceptors	+0.673	+0.717	+0.771
Kier ChiV2 path	-	-	-0.271
Kier ChiV0 atoms	+0.334	-	-
Kier Chi4 index	+4.860	+4.634	+5.001
Intercept	-1.373	-10.93	-11.871
Statistics			
R	0.859	0.884	0.888
r^2	0.737	0.780	0.788
q^2	0.717	0.693	0.682
F	24.686	25.527	22.418

N	54	54	54
PRESS	9.185	9.943	10.32
s	0.440	0.407	0.404
Equation No.	6	7	8

FIT Kubinyi function

All the three models passed the conditions for validation sets (Eqs.1-4, Table IV). Further, to define the statistical quality of activity prediction, the number of variables that enter in a QSAR model are compared by using FIT Kubinyi function (Eq. 9), a criteria closely related to F value was proven to be useful [21].

$$FIT = R^2 (n – k – 1) / (n + k^2) (1 – R^2) \qquad (9)$$

Where n is the number of compounds in training set and k is the number of variables in the QSAR equation.

Table IV: Predictive ability of validation sets consisting of 5, 6 and 7 variables

Var[a]	**$R^2_{cv,ext}$**	**R^2**	***k***	***K'***	**Eq[b]**	**Eq[c]**
5	0.636	0.887	1.038	0.963	0.013	0.001
6	0.792	0.967	1.035	0.966	0	0.001
7	0.828	0.975	1.033	0.968	0.001	0.004

[a] number of significant variables; [b] $(R^2 – R_0^2) / R^2$
[c] $(R^2 – R_0'^2) / R^2$

The main feature of F value is its sensitivity to changes in k, if k is small and its lower sensitivity if k is large. The FIT criterion has a low sensitivity towards changes in k values, as long as they are small numbers, and a substantially increasing sensitivity for large k values [22]. The best model will be the one that possess a high value of this function. Hence, QSAR models with five, six and seven variables are generated (Table V) to choose the best among them.

Table V: Statistical parameters of the regression models obtained for five, six and seven variables.

Var[a]	***r^2***	***s***	***F***	**FIT**	**Eq No.**
5	0.737	0.440	24.686	1.702	6
6	0.780	0.407	25.527	1.851	7
7	0.788	0.404	22.419	1.660	8

[a] *number of significant variables*

According to the statistical values of the models reported in Table V, we chose the model with six variables since this showed high FIT than others. The observed, calculated and predicted values of the statistically significant six parameter QSAR model (Eq. 7) are presented in Table II.

Equation 7 accounts for the significant correlation of descriptors with biological activity and displayed good internal predictivity as shown by q^2 value of 0.693 and was able to explain 78.0 % variance of inhibitory activities of derivatives. The predictive residual sum of squares and the standard error of estimate are 9.943 and 0.407 respectively. Observed verses predicted values of molecules in training and validation set are shown graphically in Fig. 1. The proposed QSAR model Eq. 7 illustrated the predictive ability of Eqs. 1-4 and depicted graphically in Figs. 2 and 3.

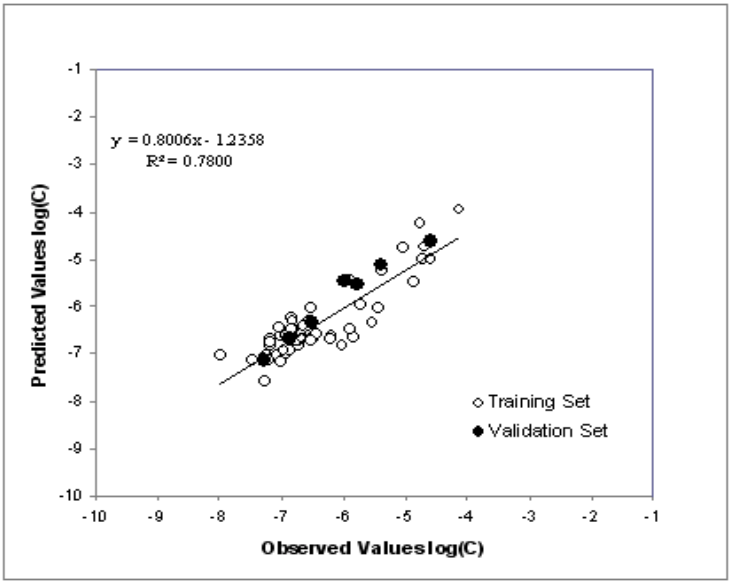


Fig. 1: Observed and predicted values of molecules in training and validation set

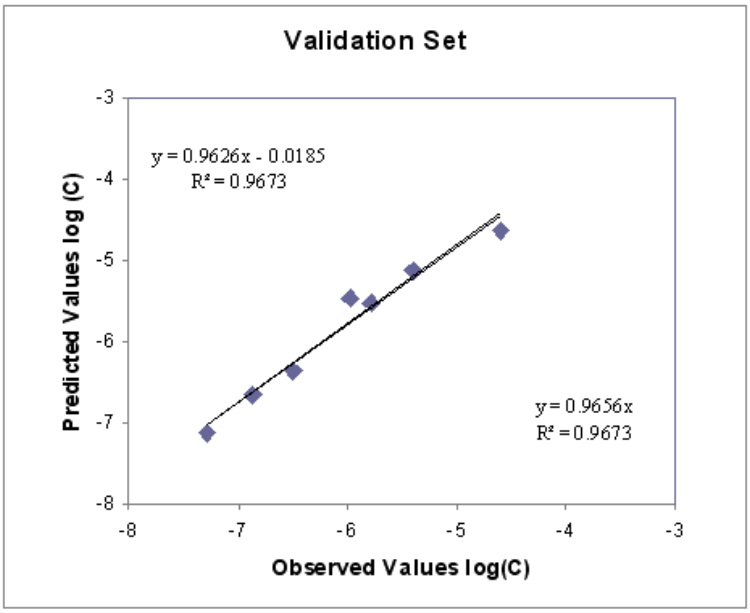


Fig. 2: Regression plot between observed vs. predicted values of compounds from validation set justifying the predictive ability of QSAR model Eq. 7

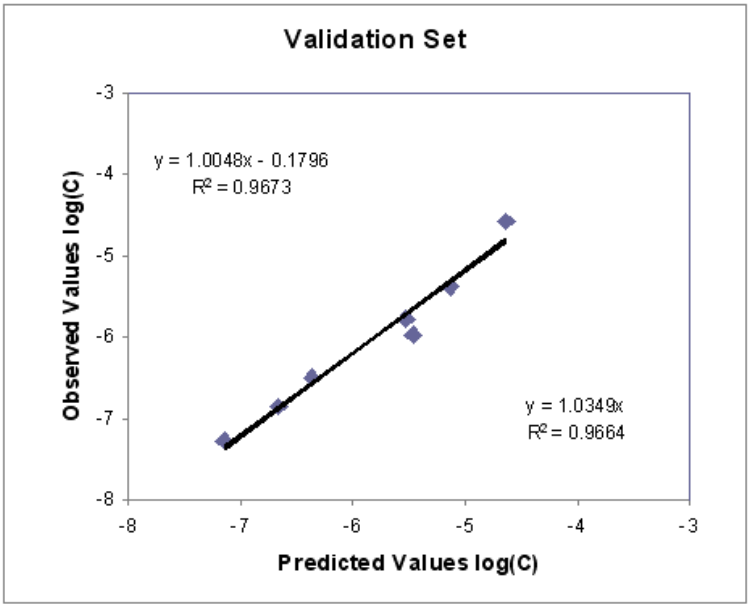


Fig. 3: Regression plot between predicted vs. observed values of compounds from validation set justifying the predictive ability of QSAR model Eq. 7

The model was further validated by applying the y-randomization test. As the model selection included F-stepping, several random shuffles of the dependent as well as independent variables were performed and the results are shown in Table VI. The low R^2 and Q^2 values indicate that the results obtained in our original model (Eq. 7) are not due to chance correlation. Alternatively, Table VII given below represents the leverage values of training and validation sets. RMSE of training and validation sets were calculated based on Euclidean distance method using Ambit software and the values are within the limits, RMSE for training set is 0.37 and RMSE for validation set is 0.04, respectively.

Interpretation of Descriptors

A brief explanation of the descriptors that were utilized to generate the statistical QSAR model:
From Eq. 7 it can be observed that the Balaban topological index, molecular refractivity, H-bond acceptors and Kier Chi4 index properties has positive contribution towards GlyT1 inhibition. Topological index can be used to evaluate structural similarity and diversity and describes the nature of atoms and bond multiplicity such as atomic order (Z), relative eletronegativity (X), length of covalent radius (Y), atomic mass (A), atomic and adjacent hydrogen mass (AH), atomic polarity (P), atomic radius (R), and atomic eletronegativity (E) [23, 24]. The Balaban index, J, is the average-distance sum connectivity [25] and measures the ramification which tends to increase with molecular ramification [26].

Molar refractivity is a measure of steric factors, a constitutive-additive property, may be treated as the additive sum of contributions of constituent structural fragments, and a measure of the volume occupied by a group of atoms. Molar refractivity increases as the formula weight and molar volume increases, indicating the concurrent increase in steric effects [27]. Therefore, Eq. 7 suggests that better GlyT1 inhibition can be achieved with an increase in molecular refractivity and Balaban index with an increase in molecular structure by substituting groups that increase steric property on spiropiperidine moiety.

Most of the studied inhibitors contain halogen groups like Cl, F and are known to act by forming H-bonds with the acceptor residues within the active site region. Negative correlation of H-bond donor term with activity indicates lower the number of H-bond donor groups in the molecule, more active it would be. On the other hand, Eq. 7 defines that increase in acceptor groups is favorable for Gly T1 activity. Therefore, designing new analogs with decrease in H-bond donor atoms with introduction of few acceptor atoms on the basic scaffold would increase Gly T1 inhibition.

On the other hand, a high value of Kappa2 index represents a negative contribution to the activity. The Kappa index [28] is a molecule shape index based on the assumption that the shape of a molecule is a function of the number of atoms and their bonding relationship. Kappa 1 shows the degree of complexity of a bonding pattern. Kappa 2 indicates the degree of linearity of bonding patterns. Kappa 3 indicates the degree of branching at the centre of a molecule, larger for predominantly linear molecules with branching at the ends. Equation 7 suggests that a high value of kappa2 index decreases the activity, in other words, the degree of linearity of bonding patterns should be modified by introducing groups that enhance H-bond acceptors and by reducing H-bond donor property with a concomitant increase in molecular structure on spiropiperidines increases Gly T1 inhibition.

Hall and Kier [29] have developed molecular connectivity indices (Chi) that reflect the atom identities, bonding environments and number of bonding hydrogens. Hall and Kier defined four series of fragment categories: Path, Cluster, Path/Cluster, and Ring. Moreover, the size, branching, unsaturation, cyclic and chemical nature of various chemical species are determined by molecule connectivity. Therefore, an increase in Kier Chi4 path/cluster index can be achieved by increasing the size and branching order on the basic skeleton.

Conclusion

The generated QSAR model on the data set with reasonable chemical diversity and biological activity demonstrated a promising method and the six descriptors [Balaban Topological index, Molecular refractivity, Kappa2 index, Kier chi4 (cluster) index, H-bond donors and H-bond acceptors] were found to be important in describing the GlyT1 inhibition. The predictive abilities and the internal and external validation procedures illustrated the accuracy of the model. This work indicates that accurate predictions can be achieved with few computational efforts in a relatively short time and the procedure described can be extended to study the receptor-ligand interactions based on QSAR. Various validation procedures described in the paper demonstrate the robustness of the QSAR model.

References

[1] S.E. Hyman and W.S. Fenton. *Science*, **2003,** 299, 350-351.
[2] B. Moghaddam. *Neuron,* **2003**, 40, 881-884.
[3] D.C. Javitt, A. Balla, H. Sershen and A. Lajtha. *Journal of Biological Psychiatry,* **1999**, 45**,** 668-679.
[4] A.R. Mohn, R.R. Gainetdinov, M.G. Caron and B.H. Koller. *Cell,* **1999,** 98, 427-436.
[5] T.V. Bliss and G.L. Collingridge. *Nature*, **1993**, 361, 31-39.
[6] Y.P. Tang, E. Shimizu, G.R. Dube, C. Rampon, G.A. Kerchner, M. Zhuo, G. Liu and J.Z. Tsien. *Nature*, **1999**, 401, 63-69.

[7] H. Scherk and P. Falkai. *Current Opinion in Psychiatry*, **2006,** 19, 145-150.
[8] J.A. Kemp and R.M. McKernan. *Nature Neuroscienc*, **2002,** 5, 1039-1042.
[9] C.A. Tamminga. *Critical Reviews in Neurobiology*, **1998,** 12**,** 21-36.
[10] G. Tsai, P. Yang, L.C. Chung, N. Lange, and J.T. Coyle. *American Journal of Psychiatry*, **2002,** 159, 480-482.
[11] C. Sur and G.G. Kinney, *Expert Opinion on Investigational Drugs*, **2004,** 13, 515-521.
[12] A. Slassi and I. Egle. *Expert Opinion on Investigational Drugs*, **2004,** 14, 201-214.
[13] E. Pinard, S.M. Ceccarelli, H. Stalder, and D. Alberati. *Bioorgonic Medicinal Chemistry Letters*, **2006**, 16, 349–353.
[14] S.M. Ceccarelli, E. Pinard, H. Stalder and D. Alberati. *Bioorgonic Medicinal Chemistry Letters*, **2006**, 16, 354–357.
[15] A. Golbraikh and A. Tropsha. *Journal of Molcular Graphics Modelling*, **2002,** 20, 269-276.
[16] A. Afantitis, G. Melagraki, H. Sarimveis, P.A. Koutentis, J. Markopoulos and O. Igglessi Markopoulou. *Bioorgonic Medicinal Chemistry Letters*, **2006**, 14, 6686-6694.
[17] A. Tropsha, P. Gramatica and V.K. Gombar. *QSAR and Combinatorial Science*, **2003**, 22**,** 69-77.
[18] D.J. Livingstone and D.W. Salt. *Journal of Medicinal Chemistry*. **2005,** 48, 661-663.
[19] J. Jaworska, N. Nikolova-Jeliazkova and T. Aldenberg. *Alternatives to Laboratory Animals*, **2005,** 33, 445-459.
[20] A. Afantitis, G. Melagraki, H. Sarimveis, P.A. Koutentis, J. Markopoulos and O. Igglessi Markopoulou. *Journal of Computer-Aided Molecular Design*, **2006**, 20, 83-95.
[21] H. Kubinyi. *Quantitative Structure-Activity Relationship*, **1994,** 13**,** 393-401.
[22] H. Kubinyi. *Quantitative Structure-Activity Relationship*, **1994,** 13, 285-294.
[23] O. Ivanicuc, O. Ivanicuc, A.T. Balaban and "Vertex-and edge-weighted molecular graphs and derived structural descriptors;" in: *Topological Indices and Related Descriptors in QSAR and QSPR*, Eds. A.T. Balaban and J. Devillers, Gordon and Breach Science Publishers, Amsterdam, **1999**, 169-220.
[24] A.T. Balaban, "Chemical graphs 48. Topological index *J* for heteroatom-containing molecules taking into account periodicities of element properties". *Mathematics in Chemistry*, **1986,** 21, 115-122.
[25] A.T. Balaban. *Chemical Physics Letters*, **1982,** 89, 399-404.
[26] Eliasi, Mehdi, Taeri and Bijan. *Journal of Computational Theoretical Nanoscience*, **2007,** 4, 1174-1178.
[27] R. Bartzatt and L. Donigan. *AAPS Pharm Sci Tech*, **2006,** 7, 30-37.
[28] L.B. Kier and L.H. Hall. "The molecular connectivity chi indexes and kappa shape indexes in structure-property modeling," in *Reviews in Computational Chemistry*, Eds. K.B. Lipkowitz and D.B. Boyd, John Wiley & Sons, New York, **1992**.
[29] L.B. Kier and L.H. Hall. *Journal of Pharmaceutical Sciences*, **1983**, 72, 1170-1173.