ejpmr, 2021,8(7), 331-344

SJIF Impact Factor 6.222

EUROPEAN JOURNAL OF PHARMACEUTICAL AND MEDICAL RESEARCH

www.ejpmr.com

Research Article

ISSN 2394-3211

EJPMR

COMPARATIVE QSAR STUDY ON ALKYL AND ALKOXY SUBSTITUTED 1, 4 DIHYDROQUINOXALINE-2, 3-DIONES AND 5-(*N*-OXYAZA)-7-SUBSTITUTED-1, 4-DIHYDROQUINOXALINE-2, 3-DIONES AS GLYCINE/NMDA SITE ANTAGONISTS

Dr. Neelam Khan[1*] and Dr. Manish Sharma[2]

[1]Associate Professor, Oriental University, Indore, India.
[2]Head & Professor MM University, India.

***Corresponding Author: Dr. Neelam Khan**

Associate Professor, Oriental University, Indore, India.

Article Received on 22/04/2021 **Article Revised on 12/05/2021** **Article Accepted on 02/06/2021**

ABSTRACT

In this article demonstrated for glycine site, the inhibition potency was found to be well correlated linearly with lipophilicity π and Hammett constant σ parameters of substituent. The models were internally (Q^2) and externally (Q^2_{ext}) validated and all the statistical parameters met with like leave-one-out cross-validation, an external test set and *Y*-randomization test. Applicability domain was verified by the standardization approach. Similarly it was also found that any substitution other than CH_3 and Cl at R_6 will prove to be unfruitful for receptor inhibition. It suggests that there exists a linear relationship with positive slope between π /σ_m and enzyme inhibition, i.e., increasing the value of π and σ will increase the potency of the compounds. The coefficient of π is 1.11, which is well between the optimum range of 0.4-1.4. A positive coefficient for π indicates that lipophilic substituents are favorable for inhibiting the receptor. Simultaneously, the substituents should be electron withdrawing also, which is indicated by the positive sign of σ coefficient.

KEYWORDS: QSAR, Glycine, NMDA.

INTRODUCTION

N-methyl-D-aspartate receptors (NMDARs) are heterooligomeric assemblies of NR1 subunit plus one or more NR2A-D subunits. Different lines of evidences are supporting the fact that misregulation of NMDAR subtypes is responsible for precipitation of multiple central nervous system (CNS) disorders. Growing research in this area is suggesting a role of subsite selectivity in tapping the therapeutic potential of NMDARs like glycine site in epilepsy, NR2B in pain or NR3A in white matter injury. However, the nature of the functional domains is poorly understood and thus puts deterrence in designing site-specific drugs. We therefore undertook a comparative QSAR study on alkyl and alkoxy substituted 1, 4 Dihydroquinoxaline-2,3-diones and 5-aza and 5-(*N*-oxyaza) dihydroquinoxaline-2,3-diones as glycine site antagonists.[1]

In this article demonstrated for glycine site, the inhibition potency was found to be well correlated linearly with lipophilicity π and Hammett constant σ parameters of substituent. The models were internally (Q^2) and externally (Q^2_{ext}) validated and all the statistical parameters met with like leave-one-out cross-validation, an external test set and *Y*-randomization test. Applicability domain was verified by the standardization approach.

Amazingly, the regression coefficients of π (1.114) and σ(1.364) were almost similar suggesting a common binding mechanism and implying that equations have been laterally validated which is the only accurate method for validating a molecular model. The conclusion of this study has been discussed thoroughly and is summarized as Glycine-quinoxaline-2, 3-diones binding involves hydrophobic and electronic effects. Electronic effect of substituents in the benzene ring modulates the binding of the weakly acidic amide. Newer substituents of interest have been proposed. We believe that in contrast to previous findings, these predictive models would narrow down the synthetic challenges in above mentioned series and newer leads in order to yield more specific drugs.

N-methyl-D-aspartate receptors (NMDARs), family members of ionotropic glutamate receptors (iGluRs), are enunciate in the central nervous system (CNS) and play critical roles in differ physiological processes, such as motor function, synaptic plasticity, learning, memory, and neuronal development.

Different subunit compositions have distinct biophysical, pharmacological, and signaling properties6. In addition to this the therapeutic agents that interact with NMDARs at glycine sites, glutamate

binding sites, channel blockers and positive allosteric modulators (PAMs) or negative allosteric modulators (NAMs), can also modulate NMDAR activity. The complexity in subunit combinations leads to manifold physiological uses besides their roles in neurological diseases.[1]

Moreover, the Gly/NMDA receptor has been investigated as potential target for the management of neurodegenerative diseases.

Now, computational chemistry methods are concern to develop a model, which allows one to predict the activity of potent competitive NMDA antagonists. First, various molecular parameters are calculated for a series of competitive NMDA antagonists with known activity values and those parameters are used to make a regression analysis which provides a model that relates the computationally calculated parameters to experimentally determined activity values. By the quantitative structure activity relationship (QSAR) model developed here, it is possible to predict the activity of a potent drug before its synthesis since only theoretically determined molecular parameters are used for the prediction. [3-5]

Molecular Descriptors

The compounds in the series were sketched using Chem Draw module ChemOffice and the sketched structures were subsequently used for the calculation of molecular descriptors available in QSAR software (received from BRNCOP, India) generates equations.

Constitutional, topological, empirical and functional group descriptors for all molecules were calculated using QSAR software.

Biological activity

The biological activity IC_{50} of the compounds was collected from the literature and converted into molar concentrations.

A negative logarithm of biological activity was used provide better correlations with parameters and avoid clustering of data points while generating QSAR regression lines; –logIC50 therefore becomes dependent variable in subsequent equations. The data table depicts various biological activities viz. observed by experimentation (Obs.), calculated by equation (Cal.), and externally predicted (Ext. Pred.). Δ indicates difference between Obs. and Cal. or Ext. Pred. activities.

Biological activity was taken as dependent variable and the calculated descriptors were taken as independent variables. A correlation analysis was performed between dependent and independent variables.[6-7]

Regression Analysis

QSAR models were generated through multiple regression analysis and performed with Hansch approach on glycine site inhibiting activity of NMDA receptor series of quinoxalines, produce 50% inhibition of [3H] glycine binding to rat brain cortical membranes.

Multiple linear regression based software generates QSAR equations and provides correlation coefficient (*r*), standard deviation (*s*), and ratio between variance of calculated and observed activities (*F*). *F* value indicates true relationship or level of significance of QSAR equation and he figures in the parentheses are 95% confidence intervals. n is the number of data points . The software also gives intercorrelation matrix among the descriptors.

For internal validation, *r*2 and *r*2A were calculated to assess "goodness of fit" of the equation. *r* 2 value is variance between observed and calculated biological activity. The value of R^2 should be ≥ 0.7. R^2_A is the adjusted squared values of r, also called explained variance (EV). It can be calculated as[2]

$$\mathbf{R^2_A = r^2(1\text{-}1/F)}$$

*Q*2 is cross-validated leave-one-out (L-O-O) indicates less probability of chance correlation between biological activity and parameters [. To validate a QSAR model, most of the researches apply the leave-one-out method (LOO) or Leave-some-out method (LSO).. High Q^2 (≥ 0.6) is as a proof of the high predictive power of QSAR models. Thus, only model of the best quality are retained in population undergoing evaluation (validation) procedure.

This technique ensures the robustness of a QSAR model. The proposed models are also checked for reliability and robustness by y- scrambling: new models are calculated for 1 model (after several repetitions) is expected to have low R^2 and Q^2_{LOO} values. When checked by the y-scrambling procedure our QSAR suggested models verify this condition. (Mean scrambled R^2 value: 0.02)[3]

R^2_{ext} is variance between observed and predicted activity of a test series. For testing the validity of a model from test set. A QSAR equation was generated on remaining compounds to predict activity of test set molecules.

Compounds were deemed to be outliers on the basis of their difference between observed and calculated activities, which should be greater than 2*s*. Biological activity of outliers was calculated from the final equation.

Outliers in the generated equation were removed to improve the quality of the model. To find out the AD of a QSAR model, the plot of cross-validated residuals verses leverage (HAT diagonal) values i.e. William's plot can be used. In the Williams plot, it is possible to identify that compounds as outliers. These compounds are outliers only in the Y-response space, since they are inside the X descriptor applicability domain of the

model. Where the vertical line is h* = 3(p +1) / n, the warning value for the the X descriptor space and the horizontal line are ±2σ, that was the cut off value for the Y space.[4]

The outliers were identified by cross-validated leave-one-out jackknife procedure, William's plot (plot between jackknifed residuals and HAT matrix leverage) and applicability domain estimation.

Leverage used as a quantitative measure of the model applicability domain.
Critical leverage: 3(p +1) / n

Here,
p = no. of parameters, n = no of compounds

Applicability domain is an indication of the correct application of a model and the reduced uncertainty of a prediction. This uncertainty can be expressed as the root mean squared error (RMSE), confidence intervals.

RMSE is the Root Mean Square and measure to define to define the accuracy in the domain and out of domain of the proposed QSAR that summaries the overall error of the model. It is calculated as the square root of the sum of the squared errors in prediction divided by theirs total number.

Applicability domain is one of the methods for validating the model. Validation principles recommended that a model should be within its applicability domain (AD).

Applicability domains of QSAR models were estimated wherever necessary by software *AMBIT*[5]

All the squared differences between the true responses and the predicted responses of the compounds in the training set are expressed in the predictive residual sum of squares (PRESS). The uncertainty of the model is expressed by the residual errors.[8-10]

A good QSAR equation is indicated by r ≥ 0.9 (in vitro), ≥ 0.8 (in vivo), $R^2 \geq 0.6$, $Q^2_{LOO} \geq 0.5$, $S_{press} \geq 0.4$ and Y-scrambled $R^2 \leq 0.02$-0.2. A low s; high F and R^2_{adj} is important for better 'goodness of fit'. An R^2_{ext} value is greater than 0.5and is necessary for good external validation.

Experimental Section

1) QSAR analysis on alkyl and alkoxy substituted 1, 4 Dihydroquinoxaline-2, 3-diones as Glycine/NMDA site antagonists.

MATERIALS AND METHODS

The compound series subjected to QSAR analysis is alkyl and alkoxy substituted 1, 4 Dihydroquinoxaline-2, 3-diones studied by Cai, S.X. et.al.[7] The series is listed in Table 1. The IC_{50} in the table refers to the concentration (μM) of compounds required to produce 50% inhibition of 5,7-dichloro-kynurenic acid ([^{3}H] DCKA) binding to rat brain cortical membranes. *In vitro* logki values were converted to –logki in order to bring out better linear correlations and reduce clustering of compounds while generating QSAR regression lines**.**

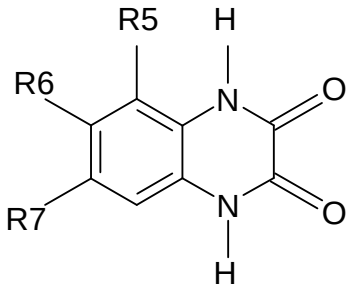


Fig 1: Alkyl- and Alkoxy-substituted 1, 4-Dihydroquinoxaline-2, 3-diones.

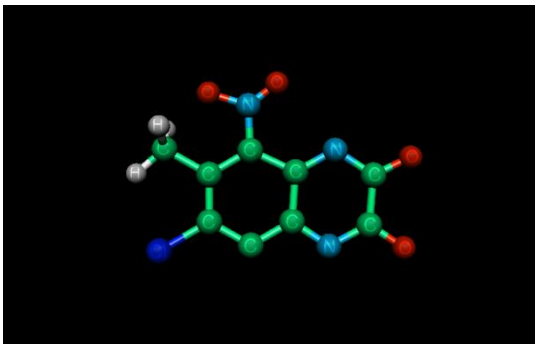


Fig 2: Three-dimensional structure of most potent compound 1, 4 –dihydro-7-Chloro-6-methyl-5-nitroquinoxaline-2, 3- dione Table1(Compound 18).

Multiple linear regression analysis (MLR) was adopted for QSAR study using Hansch approach.

The values of substituent parameters like Hammett constant (σm, σp), E_s, MR, L_1, B_1, B_5 were taken from literature[165,166] other descriptor like vander waal's volumeVw, electrotopological state indices (E-state) were calculated using the classical procedures.[8] Various validation methods were also done to authenticate the mathematical model as described in introduction section.

Table 1: SAR on alkyl and alkoxy substituted 1, 4 Dihydroquinoxaline-2, 3-diones as Glycine / NMDA receptor antagonists: Training set.

Comp.no.	R_5	R_6	R_7	^{3}[H] DCKA IC_{50} (□m)
1	NO_2	Cl	Cl	0.0059 ± 0.0010
2	H	Me	Me	3.3 ±0.7
3	H	Cl	Et	1.4 ± 0.2
4	H	Et	Et	1.8 ± 0.2
5	H	OMe	OMe	17 ± 1
6	H	OCH_2O		>100
7	H	CH=CHCH=CH		0.78 ± 0.06
8	H	$CH_2CH_2CH_2$		0.029 ± 0.003
9	NO_2	H	Me	9.3 ± 1.9
10	H	OMe	Br	1.0 ± 0.2
11	H	Cl	Me	0.78 ± 0.06
12	NO_2	Me	Me	0.029 ± 0.003
13	NO_2	Et	Et	0.16 ± 0.03
14	NO_2	CH2CH2CH2		3.1 ±□0.5
15	NO_2	OMe	Br	0.064 ± 0.003
16	NO_2	Br	Et	0.082 ± 0.019
17	NO_2	Cl	Et	0.029 ± 0.002
18	NO_2	Et	Cl	0.13 ± 0.01
19	NO_2	Me	CN	0.073 ± 0.010
20	NO_2	Me	F	0.095 ± 0.024
21	NO_2	Me	Br	0.0087 ± 0.0004
22	NO_2	Me	Cl	0.0047 ± 0.0006
23	NO_2	Cl	Me	0.045 ± 0.008
24	NO_2	H	CN	2.8 ± 0.4
25	NH_2	Me	Cl	0.15 ± 0.03
26	NO_2	NH2	F	11 ± 1
27	NO_2	OMe	F	1.9 ± 0.1
28	NO_2	OEt	F	3.6 ± 0.7
29	NO_2	$O(CH_2)_3Ph$	F	4.0 ± 0.5
30	NO_2	OBu-n	F	4.3 ± 0.5
31	NO_2	OMe	Cl	0.15 ± 0.02
32	NO_2	Set	Cl	0.64 ± 0.13
33	OCOMe	OMe	OMe	53 ± 16
34	OH	OMe	OMe	17 ± 2
35	H	Cl	Cl	0.13 ± 0.03

Numbers of compounds in the series were 36. Four compounds **(6, 7, 8, 14** in **Table 1**) were not selected in the study as they deviated from the parent structure and hence do not become part of the congeners. One compound **(29)** was also not included in the study, as substituent values could not be found for it. Compound **(6)** did not have exact IC_{50} value and therefore was also not included in the study.

Table 2: External validation of 1, 4-Dihydroquinoxaline-2, 3-diones at the NMDA Receptor Glycine Site: Test set.

Comp.no.	R_5	R_6	R_7	^{3}[H] DCKA IC_{50} (μm)
1	CO_2H	H	Me	13± 3
2	CO_2H	H	NO_2	10 ± 3
3	CO_2H	H	Cl	6.4 ± 0.5
4	CO2Me	H	Cl	65 ± 9
5	COPh	H	Cl	6.7 ± 0.8
6	$CONH_2$	H	Cl	3.1 ± 0.2
7	$CONMe_2$	H	Cl	13 ± 2
8	$CONEt_2$	H	Cl	24 ± 3
9	CO_2Me	Me	Br	0.4 ± 0.08
10	C(O)	CH_2CH_2	Br	4.6 ± 0.5
11	C(O)	OCH_2	Cl	2.9 ± 0.4
12	CN	H	Cl	0.067 ± 0.009
13	CN	Cl	Cl	0.032 ± 0.003
14	CN	Me	Cl	0.026 ± 0.007
15	NO_2	Cl	Cl	0.0059 ± 0.001
16	NO_2	Me	Cl	0.0047 ± 0.0006
17	NO_2	Me	Br	0.0087 ± 0.0004
18	NO_2	Cl	Cl	0.65 ± 0.04

The series in **Table 2** was used as test set to externally validate the model generated for training set in **Table 1**. Here also two compounds **(10, 11** in **Table 1**) were not selected in the study as they deviated from the parent structure and hence do not become part of the congeners.

RESULTS AND DISCUSSION

Using Hansch approach, we correlated the activity of alkyl and alkoxy substituted 1, 4 Dihydroquinoxaline-2, 3-diones **(Table 1)** with various physicochemical, electronic and steric parameters. After many trials **Equation 1** was found to be promising.

$$-\log IC_{50} = 1.284(0.959)\ R_7\sigma p + 1.177(0.497)\ R_7\pi + 1.087(0.396)\ R_6I + 0.921\ (0.445)\ R_5I - 1.408(0.476)$$

$$n = 30\ r = 0.892\ s = 0.481\ F = 24.436 \ldots\ldots(1)$$

Table 3: Training set compounds with their physicochemical parameters values for derivation of QSAR equation (1) *outliers.

Com. No.	R_5I	R_6I	$R_7\pi$	$R_7\sigma_p$	Obs. $-\log IC_{50}$	Calc. $-\log IC_{50}$	Calc. Resid	Pred $-\log IC_{50}$	Pred. Resid	Lev (0.535)
1	1	1	0.71	0.23	2.229	1.731	0.498	1.6731	0.5559	0.105
2	0	1	0.56	-0.17	-0.518	0.12	-0.638	0.2296	-0.7476	0.146
3	0	1	1.02	-0.15	-0.146	0.687	-0.833	0.8789	-1.0249	0.187
4	0	0	1.02	-0.15	-0.255	-0.4	0.145	-0.4327	0.1777	0.185
5	0	0	-0.02	-0.27	-1.23	-1.778	0.548	-1.939	0.709	0.227
6	1	0	0.56	-0.17	-0.477	-0.046	-0.431	0.0116	-0.4886	0.120
7	0	0	0.86	0.23	0	-0.1	0.1	-0.1357	0.1357	0.260
8	0	1	0.56	-0.17	0.107	0.12	-0.013	0.122	-0.015	0.146
9	1	1	0.56	-0.17	1.537	1.041	0.496	0.9438	0.5932	0.162
10	1	0	1.02	-0.15	0.795	0.521	0.274	0.4699	0.3251	0.155
11	1	0	0.86	0.23	1.193	0.82	0.373	0.7795	0.4135	0.127
12	1	0	1.02	-0.15	1.086	0.521	0.565	0.4175	0.6685	0.155
13	1	1	1.02	-0.15	1.537	1.608	-0.071	1.6247	-0.0877	0.189
14	1	0	0.71	0.23	0.886	0.644	0.242	0.6137	0.2723	0.109
15	1	1	-0.57	0.66	1.136	0.777	0.359	0.3589	0.7771	0.538
16	1	1	0.14	0.06	1.022	0.842	0.18	0.8164	0.2056	0.133
17	1	1	0.86	0.23	2.06	1.908	0.152	1.8856	0.1744	0.122
18	1	1	0.71	0.23	2.327	1.731	0.596	1.6599	0.6671	0.105
19	1	1	0.56	-0.17	1.346	1.041	0.305	0.9811	0.3649	0.162

20	*1	0	0.71	0.23	-0.447	0.644	-1.091	0.7777	-1.2247	0.109
21	0	1	0.71	0.23	0.823	0.81	0.013	0.8065	0.0165	0.203
22	*1	0	0.14	0.06	-0.278	0.842	-1.12	1.0142	-1.2922	0.133
23	1	0	0.14	0.06	-0.556	-0.245	-0.311	-0.2057	-0.3503	0.111
24	1	0	0.14	0.06	-0.633	-0.245	-0.388	-0.1969	-0.4361	0.111
25	1	0	0.14	0.06	-0.602	-0.245	-0.357	-0.2005	-0.4015	0.111
26	1	0	0.71	0.23	0.823	0.644	0.179	0.6219	0.2011	0.109
27	1	0	0.71	0.23	0.193	0.644	-0.451	0.6991	-0.5061	0.109
28	0	0	-0.02	-0.27	-1.724	-1.778	0.054	-1.7943	0.0703	0.227
29	0	0	-0.02	-0.27	-1.23	-1.778	0.548	-1.9397	0.7097	0.227
30	0	1	0.71	0.23	0.886	0.81	0.076	0.7906	0.0954	0.203

- **William's Plot**

For detecting the outliers in the training set, applicability domain of the model was analyzed by the William plot. Two data points **(20,22)** were not included in finalizing the model for training set in **Table 1** as they were outside the cut off value of Y space.

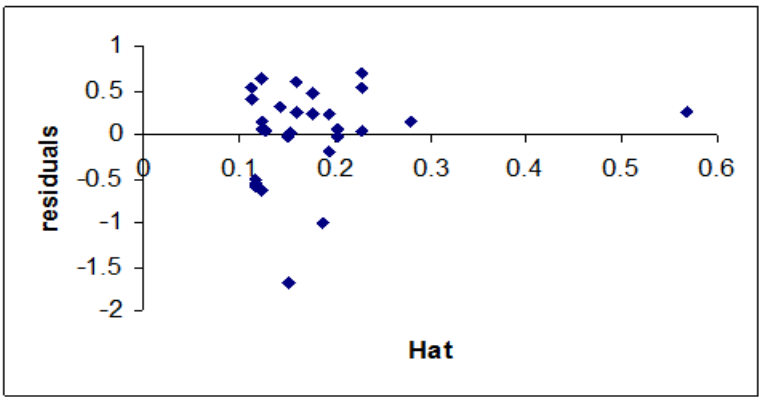


Fig 3: Williams plot of training set compounds h* = 0.3.

$-\log IC_{50}$ = 1.364(0.786) $R_7\sigma_p$ +1.114(0.411)$R_7\pi$ +1.104(0.334)R_6I
+1.033(0.364)R_5I -1.376(0.383)
n = 28 r = 0.935 s = 0.383 F = 40.264..... (2)
R^2 = 0.875, Q^2_{LOO} = 0.765, R^2_{adj} = 0.853, RMSE $_{pred}$ =0. 310,S_{press} = 0.582,

The resultant equation (2) indicated good r = 0.935, low s=0.383 and high F=40.26. Equation (2) was checked for the importance of each parameter by eliminating them one at a time and generating the resultant equation. The resultant equations were checked for their statistical validity. Following was the output.

$-\log IC_{50}$ = 1.106(0.552)$R_7\pi$ +1.175(0.437)R_6I +1.141(0.462)R_5I
-1.511(0.525)
n = 30 r = 0.857 s = 0.549 F = 23.962.....(3)

$-\log IC_{50}$ = 1.043(1.304)$R_7\sigma_p$ +1.122(0.542)R_6I +0.968(0.608)R_5I
-0.810(0.552)
n = 30 r = 0.776 s = 0.672 F = 13.133.....(4)

$-\log IC_{50}$ = 1.719(1.397)$R_7\sigma_p$ +1.227(0.733)$R_7\pi$+0.793(0.654)R_5I
-0.858(0.637)
n = 30 r = 0.732 s = 0.726 F = 10.024.....(5)

$-\log IC_{50}$ = 2.018(1.146)R7σ_p +1.223(0.638)R7π +1.001(0.507)R_6I
-0.805(0.484)
n = 30 r = 0.805 s = 0.632 F = 15.987..... (6)

It was found that in each equation (3-6), s was found to be very high with other poor statistical relationships. Hence it is shown that the presence of each parameter present in equation (2) is important and their removal is detrimental for the validity of the model.

Table 4: Training set compounds with their physicochemical parameters values for derivation of QSAR equation 2.

S.No.	R_5I	R_6I	$R_7\pi$	$R_7\sigma_p$	Obsd -logIC_{50}	Calc -logIC_{50}	Cal. Resid	LOO -logIC_{50}	Pred. Resid
1	1	1	0.71	0.23	2.229	1.865	0.364	1.819	0.41
2	0	1	0.56	-0.17	-0.518	0.12	-0.638	1.162	-1.68
3	0	1	1.02	-0.15	-0.146	0.659	-0.805	0.845	-0.991
4	0	0	1.02	-0.15	-0.255	-0.445	0.19	-0.492	0.237
5	0	0	-0.02	-0.27	-1.23	-1.767	0.537	-1.77	0.54
6	1	0	0.56	-0.17	-0.477	0.048	-0.525	-1.12	0.643
7	0	0	0.86	0.23	0	-0.105	0.105	-0.146	0.146
8	0	1	0.56	-0.17	0.107	0.12	-0.013	0.122	-0.015
9	1	1	0.56	-0.17	1.537	1.153	0.384	1.07	0.467
10	1	0	1.02	-0.15	0.795	0.588	0.207	0.545	0.25
11	1	0	0.86	0.23	1.193	0.928	0.265	0.882	0.311
12	1	0	1.02	-0.15	1.086	0.588	0.498	0.489	0.597
13	1	1	1.02	-0.15	1.537	1.692	-0.155	1.729	-0.192
14	1	0	0.71	0.23	0.886	0.761	0.125	0.742	0.144
15	1	1	-0.57	0.66	1.136	1.026	0.11	0.88	0.256
16	1	1	0.14	0.06	1.022	0.998	0.024	0.993	0.029
17	1	1	0.86	0.23	2.06	2.032	0.028	2.02	0.04
18	1	1	0.71	0.23	2.327	1.865	0.462	1.805	0.522
19	1	1	0.56	-0.17	1.346	1.153	0.193	1.102	0.244
20	0	1	0.71	0.23	0.823	0.832	-0.009	0.834	-0.011
21	1	0	0.14	0.06	-0.556	-0.106	-0.45	-0.045	-0.511
22	1	0	0.14	0.06	-0.633	-0.106	-0.527	-0.035	-0.598
23	1	0	0.14	0.06	-0.602	-0.106	-0.496	-0.039	-0.563
24	1	0	0.71	0.23	0.823	0.761	0.062	0.751	0.072
25	1	0	0.71	0.23	0.193	0.761	-0.568	0.84	-0.647
26	0	0	-0.02	-0.27	-1.724	-1.767	0.043	-1.77	0.046
27	0	0	-0.02	-0.27	-1.23	-1.767	0.537	-1.925	0.695
28	0	1	0.71	0.23	0.886	0.832	0.054	0.818	0.068

Table 5: Correlation matrix between descriptors employed for generating equation (2).

	$R_7\sigma_p$	$R_7\pi$	R_6I	R_5I
R_7Sp	1.000	-0.105	0.166	0.343
R_7pi		1.000	0.073	0.021
R_6I			1.000	-0.053
R_5I				1.000

Regression lines of the cross-validated QSAR equation (2).

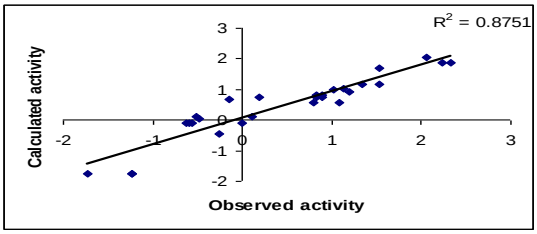
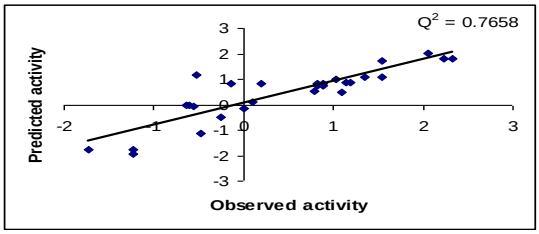


Fig 4(a): **Plot of calculated versus observed-logIC_{50}.**



(b): **Plot of predicted versus observed -logIC_{50}.**

Table 6: Compounds with their physicochemical parameters values for external validation of QSAR equation 2: Test set.

Comp. No.	R_5I	R_6I	$R_7\pi$	$R_7\ \sigma_p$	Obs. -logIC_{50}	Pred. -logIC_{50}	Pred. Residual	Critical Lev. (0.535)
1	0	0	0.56	0.23	-1.113	-0.984	-0.129	0.963
2	0	0	-0.28	0.78	1	-0.889	1.126	0.980
3	0	0	0.71	0.23	-0.806	0.79	-0.533	0.129
4	0	0	0.71	0.23	-1.812	-0.273	-1.539	0.129
5	0	0	0.71	0.23	-0.826	-0.273	-0.553	0.129
6	0	0	0.71	0.23	-0.491	-0.273	-0.218	0.129
7	0	0	0.71	0.23	-1.113	-0.273	-0.84	0.129
8	0	0	0.71	0.23	-1.38	-0.273	-1.107	0.129
9	0	1	0.86	0.23	0.397	-0.273	-0.602	0.305
10	0	0	0.71	0.23	1.173	0.999	1.446	0.129
11	0	1	0.71	0.23	1.494	0.831	0.663	0.272
12	0	1	0.71	0.23	1.585	0.831	0.754	0.272
13	1	1	0.71	0.23	2.229	1.86	0.369	0.282
14	1	1	0.71	0.23	2.327	1.86	0.467	0.282
15	1	1	0.86	0.23	2.06	2.03	0.03	0.285
16	1	1	0.71	0.23	0.187	0.76	-0.573	0.446

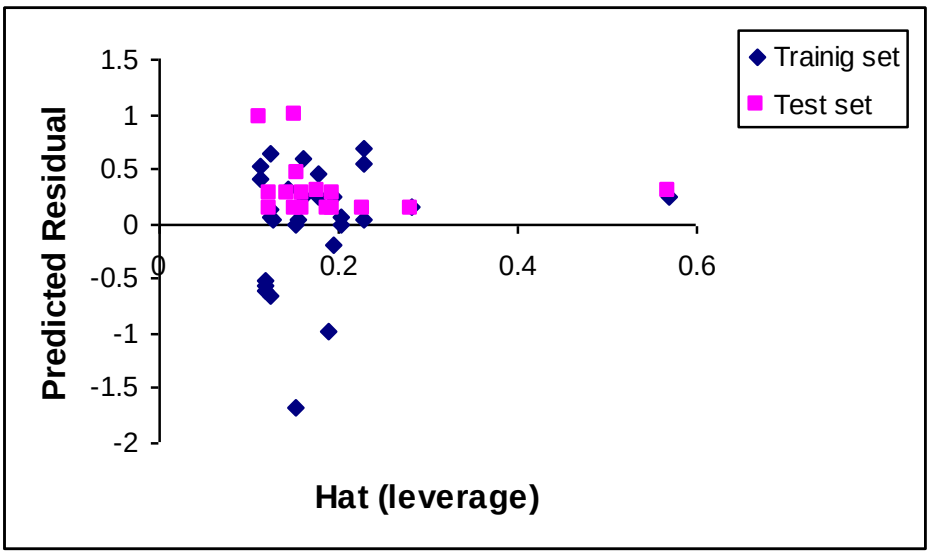


Fig 5: William's plot of training set and test set (outliers: compounds 1, 2) in Table 5.

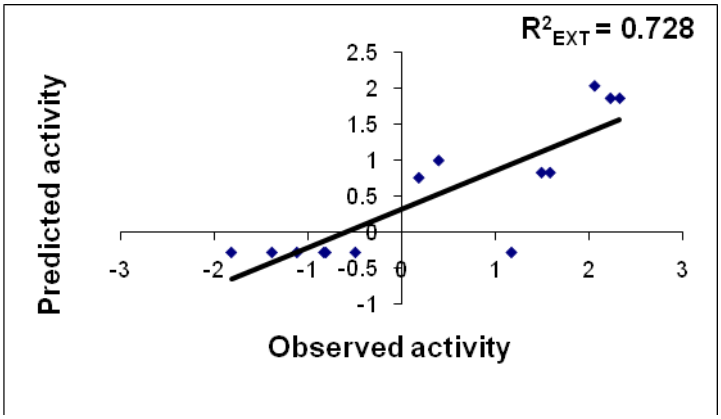


Fig 7: Plot of observed versus predicted -logIC_{50}.

- **Y-scrambling**

Table 7: Y- randomization test.

Iterations	R^2
1	0.187
2	0.124
3	0.224
4	0.081
5	0.275
6	0.135
7	0.064
8	0.124
9	0.079
10	0.209
SUM	1.502
Mean	**0.1502**

Y – Randomization test was performed 10 times iterations. Finally R^2 values are observed 0.150.

The descriptors selected for generating the model spread through a wide range, which is indicated in **Fig 7, 8.** A good spreadability of the descriptor values of the substituents reduces bias in the QSAR equation and makes the activity prediction power of the model much more accurate. **Table 4** gives the observed, calculated and LOO biological activities of compounds in **Table 1**. Moreover **Table 5** gives the inter correlation matrix of the descriptors used in equation 2. **Fig 4(a), 4(b)** indicates the closeness of the observed with calculated and predicted -logIC_{50} respectively (obtained from equation 2), which indicates the accuracy of the model.

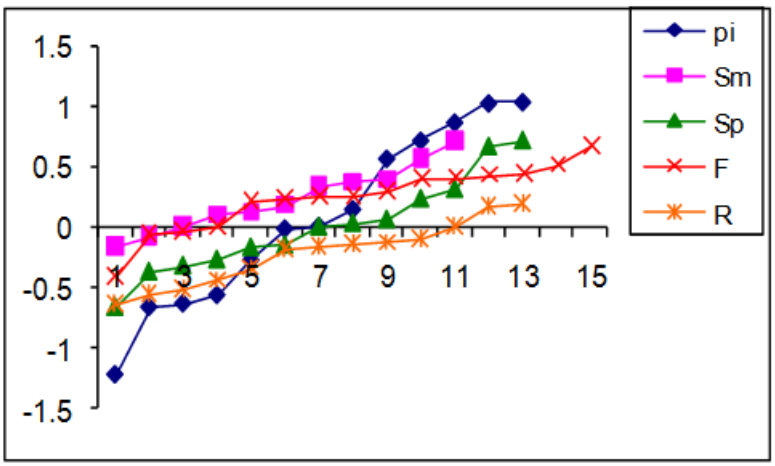


Fig 7: Spreadability of the substituent descriptor values (π, σ_m, σ_p, F, R).

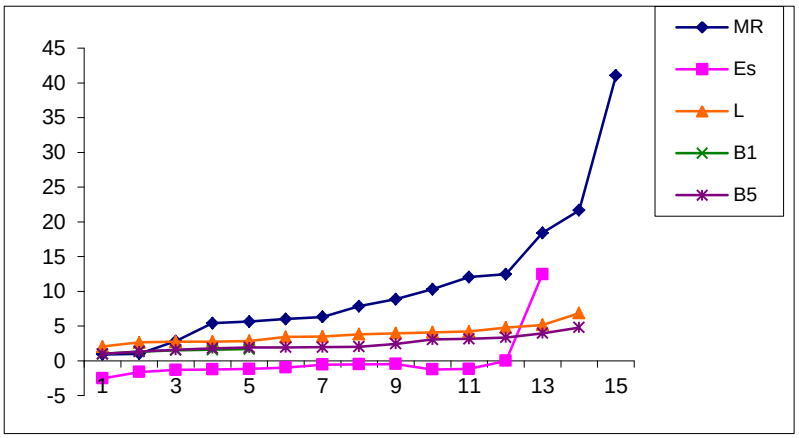


Fig 8: Spreadability of the substituent descriptor values (MR, Es, L, B1, and B5) Uniformity and Normality.

Distribution tests

1) Jark Bera test

Table 8: Jarkbera test- for π.

Sample average	**0.55**
S	**0.41**
Sample variance	**0.17**
Data are Normal?	**TRUE**
JB Statistic	**5.14501**
Critical Value	**5.99**
P-Value	**0.08**

Table 9: Jark Bera Normality test for σ_p.

Sample average	**0.03**
Sample standard deviation	**0.23**
Sample variance	**0.05**
Data are Normal?	**TRUE**
JB Statistic	**1.68053**
Critical Value	**5.99**
P-Value	**0.43**

2) Kolmogorove-Smirnov uniform distribution test: Non-uniform distribution of R_7 σ_p and R_7 π were tested.

Table 10: For π.

Mean	**0.579**
S	0.3360789
Test statistic	0.2775194

Table 11: For σ_p.

Mean	0.926667
S	0.2044168
Test statistic	0.4619048

Applicability Domain

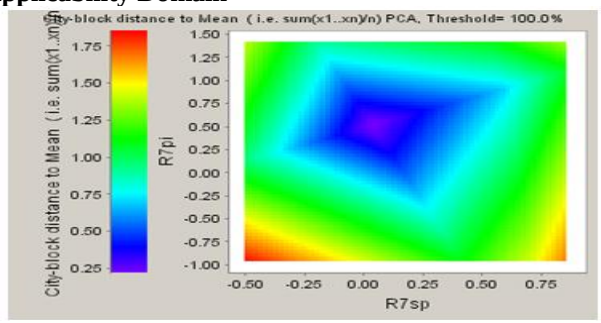


Fig 9(a): City Block distance method.

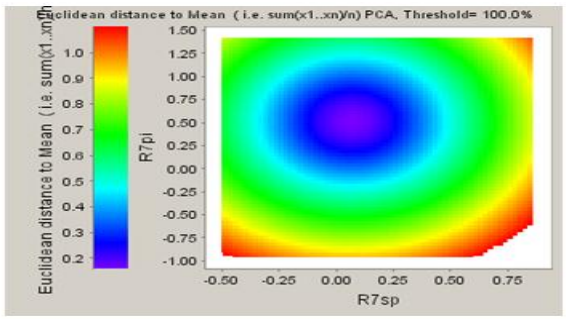


Fig 9(b): Euclidean distance method.

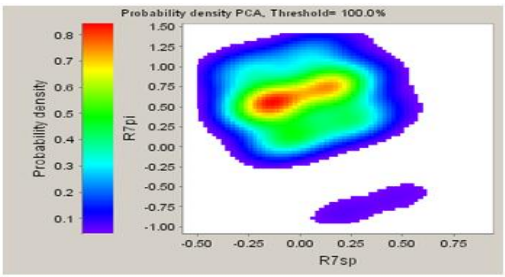


Fig 9(c): Probability density method.

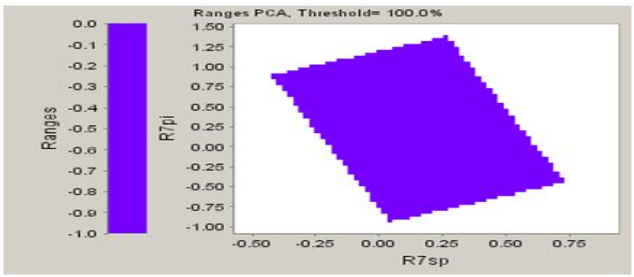


Fig 9(d): Ranges method.

AD between R_7Sp and R_7pi parameters

Fig 9(a, b, c, d): Representation of the Applicability Domain (AD) of the training set by different methods.

II) QSAR on 5-(*N*-Oxyaza)-7-substituted-1, 4-dihydroquinoxaline-2, 3-diones as the glycine/NMDA receptor antagonists.

Materials and methods

The compound series subjected to QSAR analysis is alkyl and alkoxy substituted 1, 4 Dihydroquinoxaline-2, 3-diones studied by Cai, S.X. et.al.[9] The series is listed in **Table 12**. The IC_{50} in the table refers to the concentration (μM) of compounds required to produce 50% inhibition of 5,7 -dichloro-kynurenic acid ([^{3}H] DCKA) binding to rat brain cortical membranes.

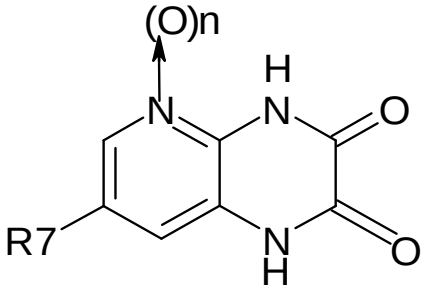


Fig 10: 5-(*N*-Oxyaza)-7-substituted-1,4-dihydroquinoxaline-2,3-diones.

Table 12: SAR of 5-(*N*-Oxyaza)-7-substituted-1, 4-dihydroquinoxaline-2, 3-dione.

Comp. No.	A	*n*	R7	IC_{50} [^{3}H] DCKA
1	N	0	Cl	2.2 ± 0.4
2	N	0	Br	4.1 ± 0.5
3	N	0	Me	41 ± 2
4	N	0	CF3	1.9 ± 0.3
5	N	0	NO2	1.1 ± 0.1
6	N	1	Cl	0.82 ± 0.11
7	N	1	Br	1.0 ± 0.2
8	N	1	Me	5.9 ± 1.1
9	N	1	CF3	1.05 ± 0.03
10	N	1	NO2	0.69 ± 0.06
11	CH	0	Cl	1.8 ± 0.3
12	CH	0	NO2	3.8 ± 0.5

RESULTS AND DISCUSSION

Using Hansch approach, we correlated the activity of alkyl and alkoxy substituted 1, 4 Dihydroquinoxaline-2, 3-diones **(Table 12)** with various physicochemical, electronic and steric parameters. After many trials **Equation 7** was found to be promising.

$$-\log IC50 = 1.181(0.478)\sigma_m + 0.488(0.308)nI - 0.904(0.249)$$
$$n = 12 \quad r = \mathbf{0.906} \quad s = 0.209 \quad F = 20.50 \quad \text{.....(7)}$$

Table 13: Training set compounds with their physicochemical parameters values for derivation of QSAR equation 7.

Comp.no.	Sm	nI	obs-logIc50	Cal-logIc50	Cal. Resd	Pred.-logIC50	Pred. Resd.	Critical Leverage (0.75)
1	0.3	0	-0.342	-0.55	0.208	-0.749	0.407	0.143063
2	0.37	0	-0.612	-0.467	-0.145	-0.599	-0.013	0.145308
3	-0.37	0	-1.612	-1.341	-0.271	0.367	-1.979	0.533921
4	0.39	0	-0.278	-0.444	0.166	0.063	-0.341	0.147447
5	0.61	0	-0.041	-0.184	0.143	-0.947	0.906	0.214885
6	0.3	1	0.086	-0.062	0.148	0.027	0.059	0.201331
7	0.37	1	0	0.021	-0.021	-0.985	0.985	0.210063
8	-0.37	1	-0.77	-0.853	0.083	-0.832	0.062	0.530098
9	0.39	1	-0.021	0.045	-0.066	-0.472	0.451	0.214056
10	0.61	1	0.161	0.305	-0.144	-1.031	1.192	0.301882
11	0.3	0	-0.255	-0.55	0.295	-0.443	0.188	0.143063
*12	0.61	0	-0.579	-0.184	-0.395	-0.584	0.005	0.214885

William's Plot

For detecting the outliers in the training set, applicability domain of the model was analyzed by the William plot. One data point **(12)** in **Table 13** were not included in finalizing the model for training set in **Table 13** as they were outside the cut off value of Y space.

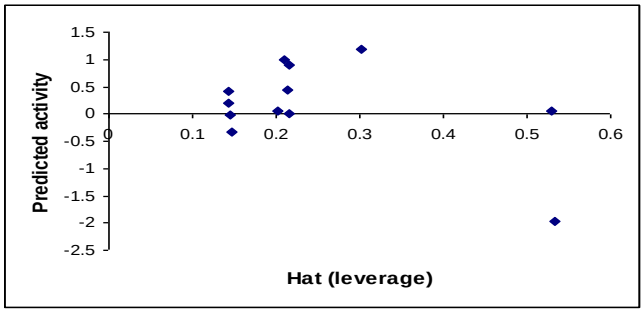


Fig 11: Williams plot of training set compounds h* = 0.75.

-logIC50 = 1.305(0.414)σ_m +0.423(0.263)n - 0.871(0.209)
n = 11 r = 0.945 s = 0.168 F = 33.081.....(8)
R^2 = 0.892, Q^2= 0.796, R^2_{adj} = 0.865, S_{press}= 0.332, RMSE=0.178

Equation (7) was found to suggest a good regression coefficient of 0.906 with less standard deviation 0.209, F-ratio of 20.50 was also fair. The resultant equation (8) indicated good regression coefficient of 0.945, low standard deviation of 0.168 and high F value of 33.08.

Table 14: Correlation matrix between descriptors employed for generating equation.[8]

	R_7 σ_m	nI
$R_7\sigma_m$	1.000	-0.010
nI		1.000

Table 15: Training set compounds with their physicochemical parameters values for derivation of QSAR equation.[8]

Comp.No.	R_7 Sp	nI	-logIC50	Calc.-logIC50	Cal. Residual	Pred. -logIC50	Pred. residual
1	0.3	0	-0.342	-0.48	0.138	-0.508	0.166
2	0.37	0	-0.612	-0.389	-0.223	-0.304	-0.308
3	-0.37	0	-1.612	-1.354	-0.258	-1.054	-0.558
4	0.39	0	-0.278	-0.362	0.084	-0.38	0.102
5	0.61	0	-0.041	-0.075	0.034	-0.088	0.047
6	0.3	1	0.086	-0.057	0.143	-0.094	0.18
7	0.37	1	0	0.035	-0.035	0.044	-0.044
8	-0.37	1	-0.77	-0.931	0.161	-1.136	0.366
9	0.39	1	-0.021	0.061	-0.082	0.083	-0.104
10	0.61	1	0.161	0.348	-0.187	0.432	-0.271
11	0.3	0	-0.255	-0.48	0.225	-0.525	0.27

Regression lines of the cross-validated QSAR equation (8).

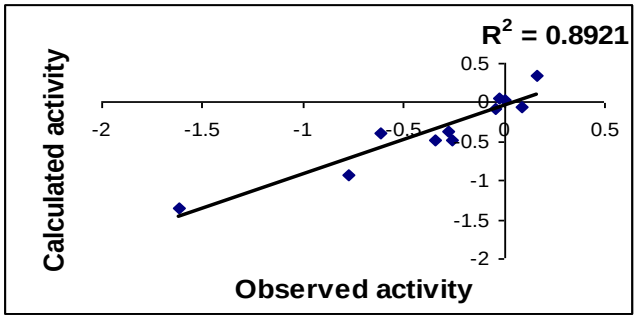


Fig 12(a): **Plot of calculated versus observed-logIC$_{50}$.**

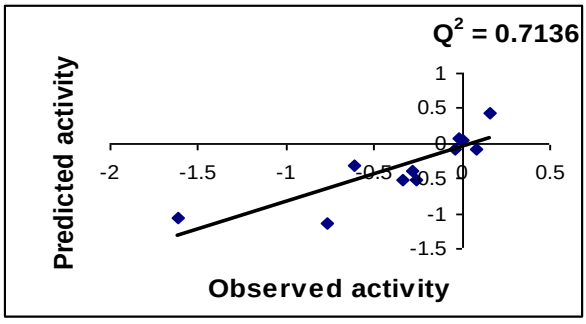


Fig 12(b): **Plot of Predicted versus observed -logIC$_{50}$.**

The descriptors used in generating the model, spread through a wide range, which is indicated in **Fig 13.**

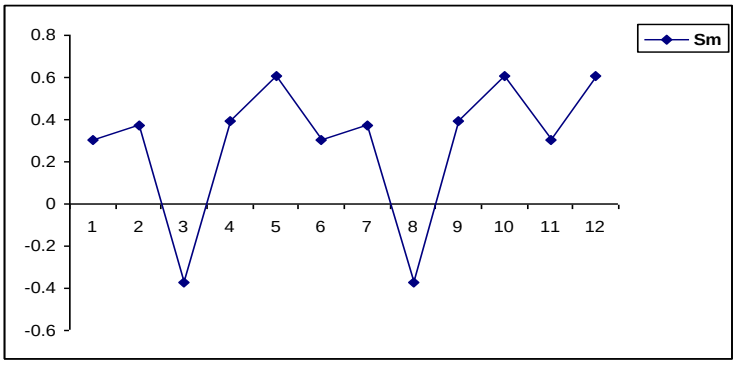
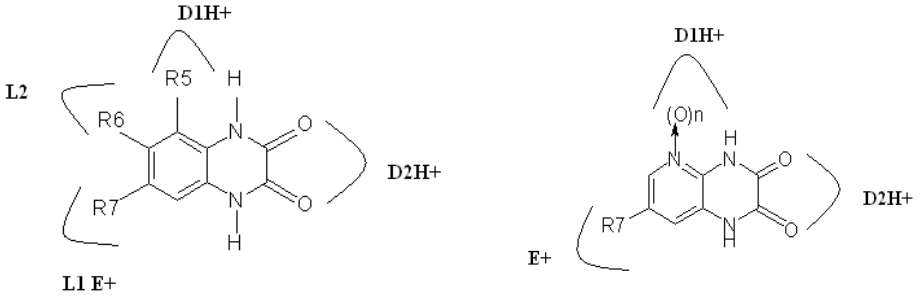


Fig 13: Spreadability of the used substituent descriptor value in equation.[8]

Receptor Mapping



(a) **(b)**

Fig 14 (a): A model of interaction of Alkyl- and Alkoxy-substituted 1, 4-Dihydroquinoxaline-2,3-diones with glycine/NMDA receptor. (b): A model of interaction of 5-(*N*-Oxyaza)-7-substituted-1,4-dihydroquinoxaline-2,3-diones with glycine/NMDA receptor

CONCLUSION

The glycine /NR1/NMDA inhibitory action of **Table 1** compounds was found to have a very good linear relation with hydrophobic constant π and Hammett constant σ of R_7 of the compounds along with two indicator variables (Eq 2). It was also found that the intercorrelation between the descriptors is not significant.

The indicator variable R_5I was used with a value of unity for presence of NO_2 functional group in R_5 and it was given a value of zero for absence of NO_2. Similarly R6I was given a value of one for the presence of CH_3 and Cl substituents at R_6 and zero when they are absent.

The equation (2) shows that presence of NO_2 group is necessary for maintaining the potency of the compounds and its removal will be detrimental for the receptor inhibition. This might be because of the presence of some electropositive center in the receptor, which interacts with the electronegative NO_2 group. But this might be true for other electronegative substituents also; but it may be possible because of the reason that the active site corresponding to R_5 has limited size pocket containing the electropositive centre and the molecular size of NO_2 is optimum enough to get inserted into that pocket and bind electronically. Any large size substituents won't be able to insert into that pocket. Even small size electronegative substituents will not be able to bind properly to the electropositive centre owing to their small atomic/molecular volume rendering them inaccessible to the electropositive centre.

Similarly it was also found that any substitution other than CH_3 and Cl at R_6 will prove to be unfruitful for receptor inhibition. This might be due to the reason that there exists in glycine site a small pocket where small sized substituents like CH_3 and Cl only can gain access. It is interesting to note that both the substituents are having almost same vander Waals volume. Moreover, unlike R_5 substituents, in R_6, we find that the substituents will bind inside the pocket irrespective of the charge on the substituents.

Equation (2) also suggests that for R_7, hydrophobic constant π and Hammett constant σ are important for the enzyme inhibition. It suggests that there exists a linear relationship with positive slope between π/σ_m and enzyme inhibition, i.e., increasing the value of π and σ will increase the potency of the compounds. The coefficient of π is 1.11, which is well between the optimum range of 0.4-1.4. A positive coefficient for π indicates that lipophilic substituents are favorable for inhibiting the receptor. Simultaneously, the substituents should be electron withdrawing also, which is indicated by the positive sign of σ coefficient. The equation suggests that for optimum activity a balance should be

present between lipophilicity and electronegativity. Increasing the lipophilicity may decrease the electronegativity, which would be detrimental for receptor inhibition.

REFERENCES

1. Tropsha, A.Recent trends in Quantitative Structure-Activity Relationships. In *Burger's Medicinal Chemistry and Drug Discovery*, 6th ed.; Abraham, D.J.,Eds.; John Wiley & Sons, **2003;** (1): 49-76.
2. Tropsha, A.Recent trends in Quantitative Structure-Activity Relationships. In *Burger's Medicinal Chemistry and Drug Discovery*, 6th ed.; Abraham, D.J.,Eds.; John Wiley & Sons, **2003;** (1): 49-76.
3. Bloom, F.E. Neurotransmission and the Central Nervous System. Joel G.Hardman, Lee E. Limbard. In *The Pharmacological Basis of Therapeutics*, 9th edition; Gilman, A.G., Goodman, L.S., Rall, T.W., Murad, F., Eds.; Macmillan Publishing Co.: New York, 1985; 304-314.
4. Tropha A.; Gramatica P.; Gombar V.J. The importance of being Earnest: Validation is the Absolute Essential for successful Application and Interpretation of QSAR Models. *Comb. Sci*, **2003**; *22:* 69-77.
5. Jaworska, J.; Nikolova-Jeliazkova, N.; Aidenberg, T. QSAR applicability domain estimation by projection of the training set in descriptor space: a review. *ATLA*, **2005**; *33*: 445-459.
6. Papa E.; Dearden J.C.; Gramatica P. Linear QSAR regression models for the prediction of bioconcentration factors by physicochemical properties and structural theoretical molecular descriptors. *Chemosphere*, **2007**; *67:* 351- 358,
7. Cai, S.X.; Kher, S.M.; Zhou, Z.L.; Ilyin, V.; Espitia, S.A.;Tran, M.; Hawkinson, J.E.; Woodward, R.M.; Weber, E; Keana J.F.W. Structure-Activity Relationships of Alkyl- and Alkoxy-Substituted1,4-Dihydroquinoxaline-2,3-diones: Potent and Systemically Active Antagonists for the Glycine Site of the NMDA Receptor. *J. Med. Chem*, **1997;** *40:* 730-738
8. Gao, H.; Denny W.A.; Garg, G.; Hansch C. Quantitative structure-activity relationships (QSAR) for 9-anilinoacridines: a comparitive analysis. *Chemico-Biological interaction*, **1998**; *116*: 157-180.
9. Zhao, Y.H.; Abraham, M.H.; Zissimos A.M. Fast Calculation of van der Waals Volume as a Sum of Atomic and Bond Contributions and Its Application to Drug Compounds. *J.Org. chem*, **2003**; *68*: 7368-7373.
10. Cai, S.X.; Huang, J.C.; Espitia, S.A.; Tran, M.; Ilyin, V.I.;Hawkinson, J.E.; Woodward, R.M.; Weber, E.; Keana, J.F.W. 5-(*N*-Oxyaza)-7-substituted-1,4-dihydroquinoxaline-2,3-diones: Novel, Systemically Active and Broad Spectrum Antagonists for NMDA/glycine, AMPA, and Kainate Receptors. *J. Med. Chem*, **1997;** *40:* 3679-3686.