ISSN: 0975-5772
Rituparna Sarma et al / Journal of Pharmaceutical Science and Technology Vol. 3 (1), 2011,477-493

An Approach towards Quantitative Structure-Activity Relationship Studies for the Affinity of Diphenyl-pyridines Towards CB1 Receptor

Rituparna Sarma, S.K. Mahmood
Bioinformatics Division, Environmental Microbiology Lab, Department of Botany, Osmania University, Hyderabad 500007, A.P., India

Abstract:
Quantitative structure activity relationship (QSAR) has been established for 87 diphenyl-pyridine analogues having inhibitory activity against human cannabinoid receptor 1. The genetic algorithm and multiple linear regressions were used to generate the relationship between biological activity and calculated descriptors. The validation of the model was done by cross validation, randomization and external test set prediction. The best model was developed using seven descriptors having r^2 value of 0.906, PRESS value 6.569, BS r^2 value 0.844, BS r^2 Err = 0.236, XV r^2 value of 0.952, and external validation using test set (r^2_{pred} = 0.836).

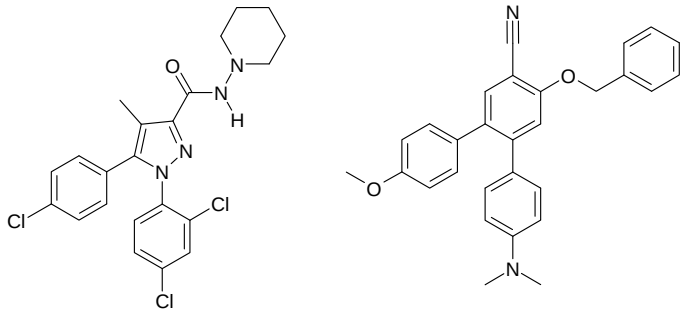
Keywords: QSAR; diphenyl-pyridine derivatives, Antagonist, correlation coefficient (r^2)

1. INTRODUCTION

Modulation of the cannabinoid receptor 1 (CB1; predominantly found in the central nervous system) has been demonstrated to have powerful effects on feeding behavior in both humans and other animals[1]. Initially, the search for CB1 antagonists/inverse agonists was based on the structure of known agonists like Delta^9-THC[2-5], the first potent and selective hCB1 inverse agonist 1 (SR141716, Fig. 1) was reported in 1994 by researchers at Sanofi-Synthelabo and belonged to a new family of CB1 ligands based on a 1,5-diphenylpyrazole structure [2-4,6]. CB1 agonists, Delta^9-THC stimulate food intake, while CB1 inverse agonists such as SR141716A[6] suppress food intake demonstrating the utility of CB1 inhibition for the treatment of obesity[7-10].

Quantitative structure-activity relationships (QSAR) use the concept of correlating structural or property descriptors of compounds with their activities. QSAR is a useful method for the design of bioactive compounds and the prediction of activity from the parameters calculated from chemical structure of compound. There are many examples available in literature where QSAR models have been used for screening of compounds from the chemical databases [11–14].

Herein, we report a QSAR study to rationalize the physico-chemical and structural features among the reported diphenyl-pyridine analogues as CB1 inverse agonists. Ever since the work of Hansch, QSAR has been utilized as a major tool in the field of drug discovery to explore ligand-receptor/enzyme interactions, especially when the structure of the target is not known. In the present study, we have collected IC_{50} values for a large set of CB1 ligands, and employing QSAR methods, used the data to create QSAR models to eventually show predictive power.



1. (SR141716) hCB1 IC50 = 6 nM **2. (hcb1 IC50 = 530 nM)**

Figure 1: Structure of SR141716 and the Merck lead structure

Table 1: 5 best QSAR models generated using various combinations of descriptors

Activity_1 = 2.63749 – 0.39853 * "CHI-2" + 0.032977 * "Jurs-DPSA-3" – 0.768911 * "Atype_Cl_90" + 0.026303 * "MW" - 1.13257 * "Atype_Unknown" - 37.7663 * "Jurs-FPSA-3" – 0.037164 * "MR"

Activity_2 = 5.91448 + 0.457079 * "Hbond acceptor" – 1.45535 * "Atype_Unknown" – 0.781137 * "Atype_Cl_90" – 34.0323 * "Jurs_FPSA_3" + 0.0006555 * "Jurs-DPSA-2" + 0.035297 * "S_sCl" – 0.307141 "Rule of 5 violation" – 0.157525 * "Atype-C-26"

Activity_3 = 5.53995 – 1.45476 * "Atype_Unknown" – 53.1148 * "Jurs-FPSA-3" – 0.039103 * "S_tN" – 0.027645 * "MR" – 0.304223 * "Hbond acceptor" + 0.328116 * "CHI-V-1" – 0.63362 * " Atype_Cl_90" + 0.035493 * " Jurs-PPSA-3"

Activity_4 = 5.557 + 0.304324 * "Hbond acceptor" + 0.32616 * "CHI-V-1" – 0.633127 * "Atype-CI-90" + 0.035526 * "Jurs-PPSA-3" – 1.45466 * "Atype_Unknown" – 53.1779 * "Jurs-FPSA-3" – 0.027527 * "MR" – 0.38082 * "Atype_N_74"

Activity_5 = 4.85497 + 0.51058 + * "CHI-V-1" – 0.619722 * "Atype_CI-90" + 0.332245 * "Hbond acceptor" – 1.3507 *"Atype-Unknown" – 47.5835 "Jurs-FPSA-3" – 0.020589 * "MR" – 0.197291 * "CHI_2" + 0.025927 * " Jurs–DPSA-3"

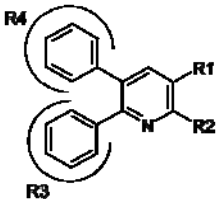
2. MATERIALS AND METHODS:

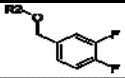
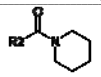
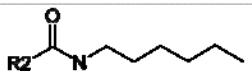
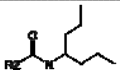
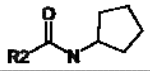
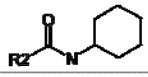
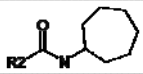
2.1. Data set and Molecular modeling

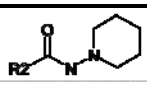
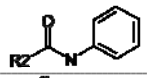
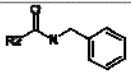
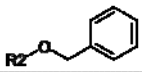
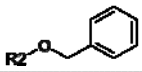
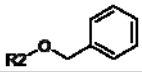
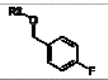
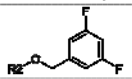
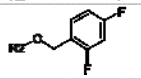
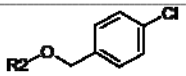
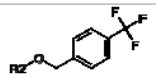
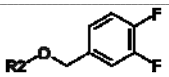
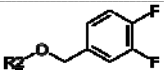
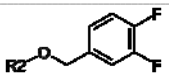
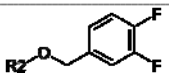
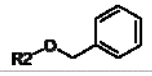
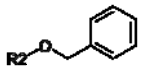
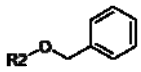
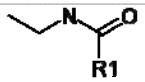
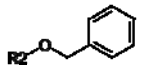
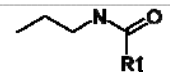
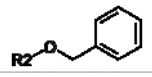
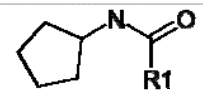
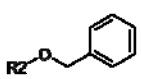
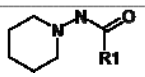
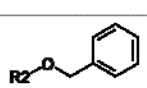
The inhibitory activity of the 87 diphenyl-pyridine[15-17] derivatives was taken from literature in terms of IC50 against human cannabinoid receptor 1. These compounds were collected from literature on the basis of two main criteria, firstly there should same scaffold and till date there is no QSAR work done in this set of compounds. Chemical structures and their respective biological properties (divided in 2 sets of test set and training set) are listed in table 3, 4. The compounds were selected based on a wide range of their activities (ranging from 0.91 nM to 2800 nM) measured under the same experimental conditions. The IC_{50} values were then converted to their pIC_{50} values to get the linear relationship in the equation using the following formula:

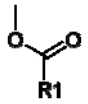
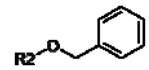
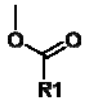
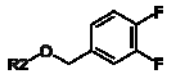
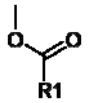
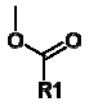
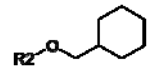
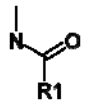
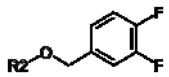
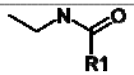
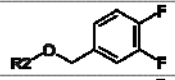
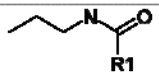
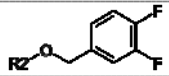
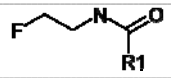
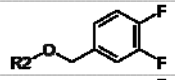
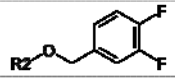
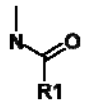
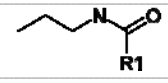
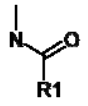
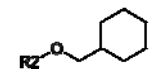
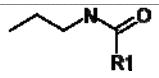
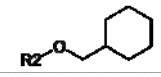
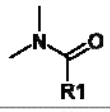
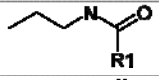
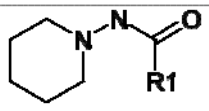
pIC50 = - log IC50.

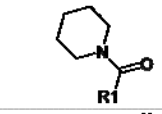
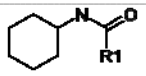
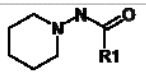
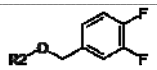
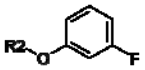
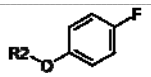
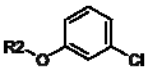
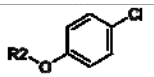
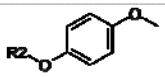
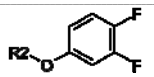
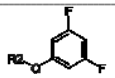
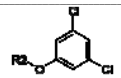
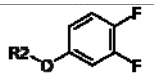
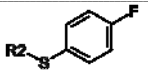
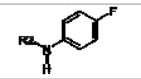
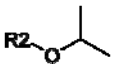
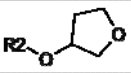
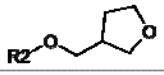
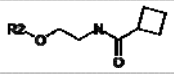
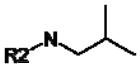
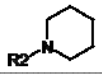
Table 2: Chemical structures of all compounds with observed activity included in the study

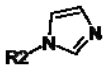
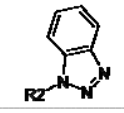
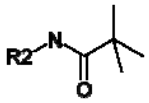
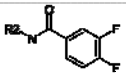
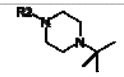
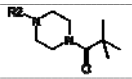
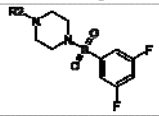
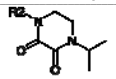
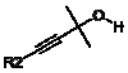
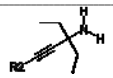
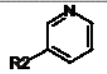
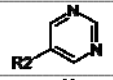
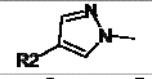
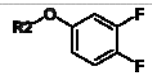
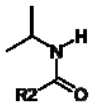
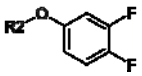
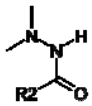
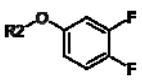
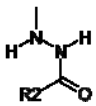
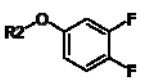


Compound no	*R1*	*R2*	*R3*	*R4*	*IC50*
1	CN		2,4-diCl	4-Cl	526
2	-		2,4-diCl	4-Cl	340
3	-		2,4-diCl	4-Cl	43
4	-		2,4-diCl	4-Cl	32
5	-		2,4-diCl	4-Cl	53
6	-		2,4-diCl	4-Cl	95
7	-		2,4-diCl	4-Cl	35

Compound no	*R1*	*R2*	*R3*	*R4*	*IC50*
8	-		2,4-diCl	4-Cl	56
9	-		2,4-diCl	4-Cl	210
10	-		2,4-diCl	4-Cl	31
11	CN		4-NMe_2	OMe	2800
12	CN		4-Cl	-	11
13	CN		2,4-diCl	4-Cl	3.1
14	CN		2,4-diCl	4-Cl	1.8
15	CN		2,4-diCl	4-Cl	2.7
16	CN		2,4-diCl	4-Cl	5.8
17	CN		2,4-diCl	4-Cl	3.4
18	CN		2,4-diCl	4-Cl	18
19	CN		4-Cl	4-Cl	7.2
20	CN		2,4-diCl	4-F	6.3
21	CN		2,4-diCl	4-Me	66
22	CN		2,4-diF	4-F	18
23			2,4-diCl	4-Cl	7
24			2,4-diCl	4-Cl	22
25			2,4-diCl	4-Cl	1.4
26			2,4-diCl	4-Cl	6
27			2,4-diCl	4-Cl	18
28			2,4-diCl	4-Cl	41
29			2,4-diCl	4-Cl	3.5

Compound no	*R1*	*R2*	*R3*	*R4*	*IC50*
30			2,4-diCl	4-Cl	1.3
31			2,4-diCl	4-Cl	1.9
32			2,4-diCl	4-Cl	2.9
33			2,4-diCl	4-Cl	3.1
34			2,4-diCl	4-Cl	1.9
35			2,4-diCl	4-Cl	1.7
36			2,4-diCl	4-Cl	1.5
37			2,4-diCl	4-Cl	1.8
38			2,4-diCl	4-Cl	32
39			2,4-diCl	4-Cl	17
40			2,4-diCl	4-Cl	21
41			2,4-diCl	4-Cl	3.4
42			2,4-diCl	4-Cl	5.2
43			2,4-diCl	4-Cl	1300
44		Cl	2,4-diCl	4-Cl	52
45		Cl	2,4-diCl	4-Cl	58
46		Cl	2,4-diCl	4-Cl	400
47		Cl	2,4-diCl	4-Cl	460

Compound no	*R1*	*R2*	*R3*	*R4*	*IC50*
48		-	2,4-diCl	4-Cl	65
49		-	2,4-diCl	4-Cl	210
50		-	2,4-diCl	4-Cl	480
51		-	2,4-diCl	4-Cl	1.3
52	CN		2-Cl	4-Cl	8
53	CN		2,4-diCl	4-Cl	11
54	CN		2,4-diCl	4-Cl	14
55	CN		2,4-diCl	4-Cl	4
56	CN		2,4-diCl	4-Cl	7
57	CN		2,4-diCl	4-Cl	26
58	CN		2,4-diCl	4-Cl	3.7
59	CN		2,4-diCl	4-Cl	0.91
60	CN		2,4-diCl	4-Cl	3.9
61	CN		2-Cl	4-Cl	25
62	CN		2,4-diCl	4-Cl	154
63	CN		2,4-diCl	4-Cl	50
64	CN		2,4-diCl	4-Cl	10
65	CN		2,4-diCl	4-Cl	11
66	CN		2,4-diCl	4-Cl	6.6
67	CN		2,4-diCl	4-Cl	5.7
68	CN		2,4-diCl	4-Cl	89
69	CN		2,4-diCl	4-Cl	19

Compound no	*R1*	*R2*	*R3*	*R4*	*IC50*
70	CN		2,4-diCl	4-Cl	12
71	CN		2,4-diCl	4-Cl	5.7
72	CN		2,4-diCl	4-Cl	29
73	CN		2,4-diCl	4-Cl	20
74	CN		2-Cl	4-Cl	108
75	CN		2-Cl	4-Cl	37
76	CN		2-Cl	4-Cl	3.8
77	CN		2-Cl	4-Cl	12
78	CN		2-Cl	4-Cl	5.9
79	CN		2-Cl	4-Cl	44
80	CN		2-Cl	4-Cl	39
81	CN		2-Cl	4-Cl	27
82	CN		2-Cl	4-Cl	38
83	CN		2-Cl	4-Cl	165
84			2,4-diCl	4-Cl	11
85			2,4-diCl	4-Cl	3.5
86			2,4-diCl	4-Cl	5.6
87			2,4-diCl	4-Cl	80

From the original list of 87 selected compounds as CB1 inverse agonists, 31 were listed as test set compounds (table 4) while the rest 56 were used as training set compounds (table 3). It was taken into consideration that compounds with wide range of activities were included in both the training and the test sets. The structures of compounds used in the study along with observed IC50 values are provided in Table 2.

The three-dimensional structures were constructed by using Cerius2 programming package version 4.11[18]. Energy minimization was performed using Universal force field and further geometric optimization of these compounds was done using the semi-empirical program MOPAC 6.0 and applying the AM1 Hamiltonian. The MOPAC charges were used for entire calculations.

Table 3: Actual, Predicted and residual activities of Training set compounds

Compound No	*pIC50*	*G/PLS Predicted Activity*	*G/PLS Residuals Activity*
1	6.279	6.348	-0.069
3	7.367	7.394	-0.027
4	7.495	7.484	0.011
6	7.022	7.014	0.008
8	7.252	6.886	0.366
10	7.509	7.428	0.081
13	8.509	8.291	0.218
15	8.569	8.641	-0.072
17	8.469	9.239	-0.770
18	7.745	7.806	-0.061
20	8.201	8.072	0.129
21	7.180	7.790	-0.610
23	8.155	7.683	0.472
24	7.658	7.777	-0.119
26	8.222	8.045	0.177
27	7.745	7.864	-0.119
29	8.456	8.206	0.250
30	8.886	9.181	-0.295
32	8.538	8.047	0.491
33	8.509	8.369	0.140
34	8.721	8.747	-0.026
36	8.824	9.171	-0.347
37	8.745	8.472	0.273
38	7.495	7.407	0.088
40	7.678	7.837	-0.159
41	8.469	7.920	0.549
43	5.886	6.215	-0.329
44	7.284	7.296	-0.012
46	6.398	6.146	0.252
47	6.337	7.199	-0.862
49	6.678	7.053	-0.375
50	6.319	6.215	0.104
52	8.097	7.899	0.198
53	7.959	8.269	-0.310
55	8.398	8.420	-0.022
56	8.155	8.458	-0.303
58	8.432	8.609	-0.177
59	9.041	8.450	0.591
61	7.602	7.643	-0.041
62	6.812	7.195	-0.383
63	7.301	7.887	-0.586
65	7.959	8.078	-0.119
66	8.180	8.112	0.068
67	8.244	8.161	0.083
69	7.721	7.640	0.081
70	7.921	7.790	0.131
71	8.244	8.170	0.074
74	6.967	6.751	0.216
75	7.432	7.131	0.301
77	7.921	7.965	-0.044
78	8.229	7.994	0.235
79	7.357	7.419	-0.062
81	7.569	7.373	0.196
82	7.420	7.343	0.077
85	8.456	8.509	-0.053
87	7.097	7.453	-0.356

Table 4: Actual, Predicted and residual activities of Test Set compounds

Compound No	*pIC50*	*G/PLS Predicted Activity*	*G/PLS Residuals Activity*
2	6.469	7.161	-0.692
5	7.276	7.106	0.170
7	7.456	7.256	0.200
9	6.678	7.124	-0.446
11	5.553	5.813	-0.260
12	7.959	7.971	-0.012
14	8.745	8.661	0.084
16	8.237	8.255	-0.018
19	8.143	8.451	-0.308
22	7.745	7.659	0.086
25	8.854	7.969	0.885
28	7.387	6.902	0.485
31	8.721	8.415	0.306
35	8.770	8.825	-0.055
39	7.770	7.944	-0.174
42	8.284	8.136	0.148
45	7.237	7.477	-0.240
48	7.187	7.307	-0.120
51	8.886	8.754	0.132
54	7.854	8.197	-0.343
57	7.585	8.125	-0.540
60	8.409	8.832	-0.423
64	8.000	7.639	0.361
68	7.051	7.528	-0.477
72	7.538	7.661	-0.123
73	7.699	8.699	-1.000
76	8.420	8.797	0.377
80	7.409	7.316	0.093
83	6.783	7.296	-0.513
84	7.959	8.288	-0.329
86	8.252	8.031	0.221

2.2 Descriptor calculation

Conventional QSAR studies require the calculation of molecular descriptors, such as connectivity indices, 2D autocorrelation descriptors, and Burden eigen values, which are used as independent variables in QSAR modeling. The Cerius2 software was used to generate the descriptors for the QSAR studies. This procedure afforded more than 100 descriptors which were subjected to the following selection strategy. Different descriptor classes such as E-state indices, electronic, spatial, structural, thermodynamic and topological descriptors were calculated for the molecules in the dataset[19-21]. Description of descriptors included in the model is listed below.

Different types of descriptors:

Type	***Descriptors***
Structural	Molecular weight, number of chiral centers, number of rotatable bonds, number of hydrogen-bond acceptors, number of hydrogen-bond donors
Electronic	Sum of atomic polarizabilities, sum of partial charges, sum of formal charges, dipole moment, energy of highest occupied orbital (HOMO), energy of lowest unoccupied orbital (LUMO), superdelocalizability
Topological	Kier and Hall molecular connectivity index, Wiener index, Zagreb index, Hosoya index, Balaban indices
E-state indices	Electrotopological-state indices
Spatial	Jurs descriptors, radius of gyration, PMI, area, shadow indices, density, Vm
Thermo dynamic	Molar refractivity, heat of formation, log of the partition coefficient, log of the partition coefficient atom type value, desolvation free energy of water, desolvation free energy of octanol

More than 100 descriptors were calculated and some were rejected because they contain a value of zero for all the compounds. The inter-correlation of descriptors was taken into account and highly correlated descriptors were

grouped together and the descriptor with the highest correlation with biological activity was taken from the group. The descriptors which remained after removing the ones which had zero values for all compounds were subjected to genetic function approximation (GFA) for selection of variables to obtain the QSAR models using genetic algorithm principles.

2.3 Regression analysis

GFA is a genetics based method of variable selection, which combines Holland's genetic algorithm (GA) with Friedman's multivariate adaptive regression splines (MARS) [22, 23]. The GFA method works by generating equations (set at 100 by default in the Cerius2 software) randomly. Then pairs of ''parent'' equations are chosen for ''crossover'' operations from this set of 100 equations randomly. The number of crossing over was set by default at 5000. The goodness of each progeny equation is assessed by Friedman's lack of fit (LOF) score which is described by the following formula:

$$LOF = LSE/1\{1-(c + dp)/m^2\}$$

Where LSE is the least-squares error, c is the number of basic functions in the model, d is smoothing parameter, p is the number of descriptors and m is the number of observations in the training set[21]. The smoothing parameter that controls the scoring bias between equations of different sizes was set at default value of 1.0 and the new term was added with a probability of 50%. Only the linear equation terms were used for model building. The best equation out of the 100 equations was taken based on the statistical parameters such as regression coefficient, adjusted regression coefficient, regression coefficient cross validation and F-test values.

2.4 Statistical analysis

The statistical model generated in our study was investigated using the PLS leave-one-out (LOO) method. The predictive ability of the model was assessed by their LOO cross-validated (q^2) values. This method was used to determine the optimum number of PLS components and the stability of the model. The number of components used in the final non validated model was optimized to give the highest q^2 value and the lowest standard error of prediction. The non cross-validated model was assessed by the conventional correlation coefficient r^2 and F-value. The external validation process can be considered the most reliable validation method, as cross-validation procedures may lead to very optimistic statistics[24, 25]. External validation was performed with a test set of 31 compounds, which were not included in the training set during the process of QSAR model generation.

3. Results and Discussion

3.1 **Equations of Quantitative Structure Activity Relationship**:

It is a computer-based statistical model that correlates descriptor variations to quantitative changes in biological activity. A set of 87 inverse agonists of CB1 were considered for the studies, the molecules were divided into training set (56) and test set (31) on the basis of structural diversity. This could give light to the structural features important for CB1 binding as well as help to club the most correlated descriptors which could be further explored in QSAR modeling. However the predictive model should have all ranges of activity in training set so that the model can categorize the new molecules according to activity, further for proper validation of the model the test set should also contain all ranges of activity. Accordingly we have considered all ranges of activity in training and test set. 130 descriptors were calculated for each molecule and were used for generating QSAR model. A QSAR model was constructed by correlating the activity of the inverse agonists with descriptors using G/PLS regression method in Cerius2. A population of 100 randomly generated equations was initially built by utilizing the default settings and the built-in descriptors. These equations were then run for 50,000 generations, which means that, for each generation, two better scoring equations are selected as parents to continue further equation generation. Parts of each parent equation are

then pooled into a child equation. Only improved scoring child equations will be incorporated in the parent population and replace for the worst of the parents. This will result in the best 100 equations for the next evolution iteration. The scoring of the equations was authenticated by calculating the predictive residual sum of squares (PRESS) and through the cross validation test. The best QSAR model was chosen on the basis of the highest q^2 and the lowest difference between q^2 and r^2. Furthermore, this equation was validated by assessing its predictive power on a set of 31 test compounds. The plot between actual and predicted pIC50 values for training set and test set is shown in Fig.2 and Fig. 3.

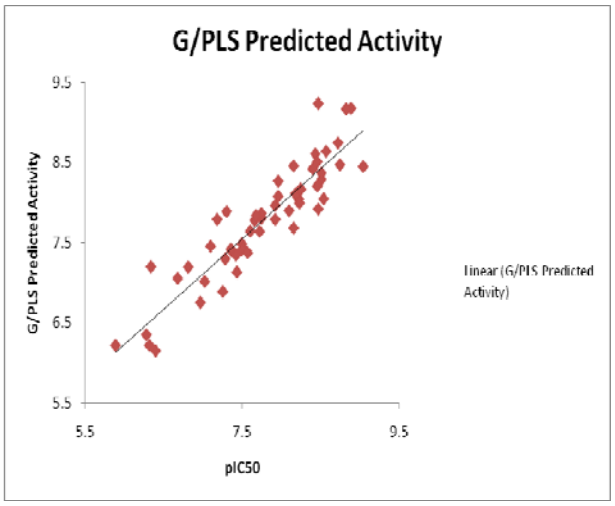


Figure 2: Plot of actual versus predicted pIC50 of the training set molecules

Number of descriptors necessary and sufficient for the QSAR equation was first determined. The molecules 86, 76 in the dataset occasionally disturbed the robustness of the model and it was kept in the test set. The same molecules turned out to be an outlier with a residual of 1.221, 1.377 (Fig. 4). The probable reason for the high residual of this two compound is due to its very low activity (IC50 = 3.8, 5.6) in comparison to other compounds.
Statistical analysis was done by applying the genetic partial least square (PLS) procedure to the appropriate columns of the QSAR table and using the standard scaling method. Leave-one-out (LOO) method (one compound is removed from the dataset and its activity is predicted using the model derived from the rest of the dataset and r^2 value (cross validated r^2 or q^2) is computed using the predicted values of the omitted molecules) was used for the cross validation of the equation, with a minimum column filtering value (F) of 20 kcal/mol to hasten the analysis and reduce noise. Optimum number of components obtained from the leave one out cross-validation analysis was utilized for the non cross-validation to obtain conventional r^2. The predictive credibility of the model was analyzed by group cross-validation Leave-one-out (LOO) method with 10 groups and bootstrapping analysis.

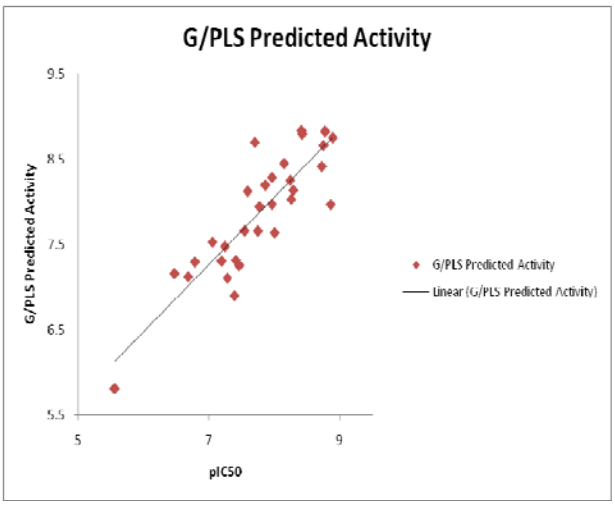


Figure 3: Plot of actual versus predicted pIC50 of the test set molecules

Basing on the test set compounds, the predictive correlation co-efficient (r^2_{pred}) is computed using the following equation:

$$r^2_{pred} = (SD - PRESS)/SD \qquad (1)$$

Where SD is the sum of squared deviations between the biological activities of each molecule and the mean activity of the training set molecules and PRESS is the sum of squared deviations between the predicted and actual activity values for every molecule in the test set.

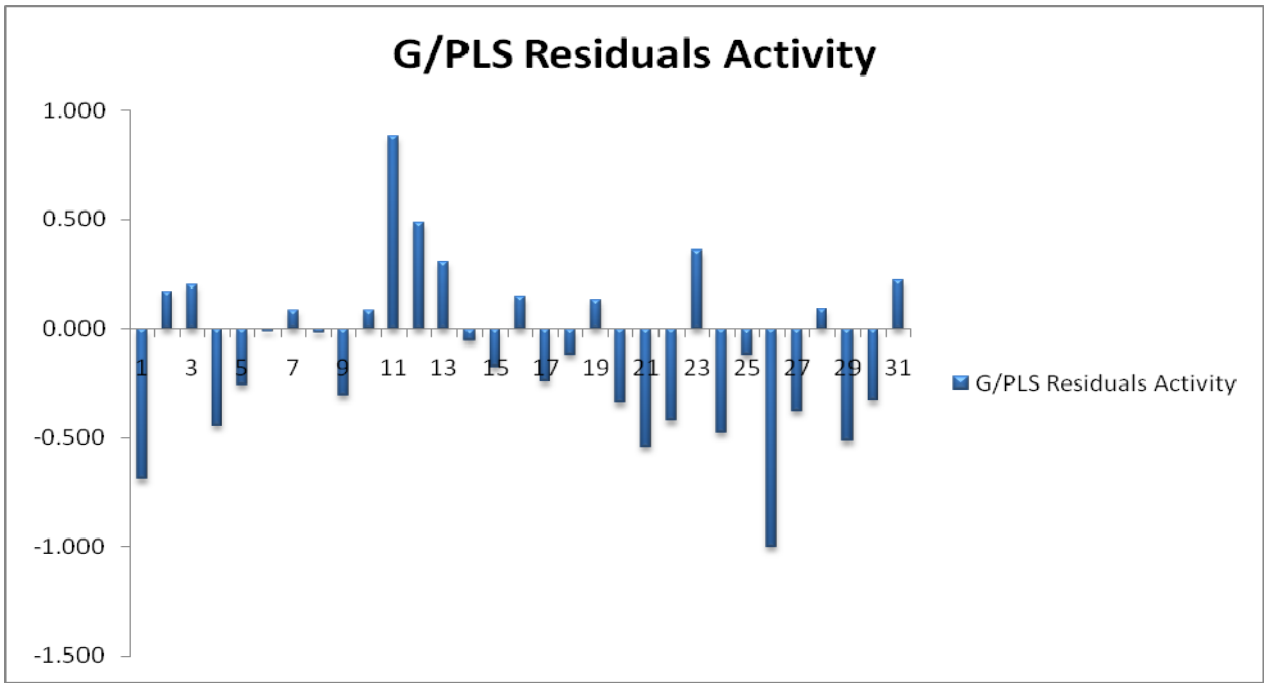


Figure 4: Histogram of residuals of the test set molecules

3.2 Validation Test

All 87 molecules were used for modeling in order to incorporate as much structural information as possible. Apart from Equation 1, we have also described 4 other next best equations with varying number of descriptors. These equations are statistically evaluated with their varied descriptors and can be compared with the best equation, describing their potential activities as CB1 inverse agonist. The top 5 equations/models generated by the G/PLS statistical method for a training set of 56 cb1 inverse agonists are represented in Table 1. Out of these 5 equations the equation I having the particular combination of 7 descriptors was having the best predictive ability compared to the rest of the equations, confirming the contributions of these descriptors towards the biological activity. The statistical parameters for all the 5 equations are described in Table 5. On comparison of these statistical parameters, the best model equation I is having better internal validation (r^2 = 0.906, XV r^2 = 0.952, BS r^2 = 0.844, BS r^2 Err = 0.236) and external validation using test set (r^2_{pred} = 0.836). Due to large number of descriptors available, they were selected based on their correlation to biological property ($r^2 < 0.10$) or that are more than (r^2_{pred}) were discarded. This could give light to the structural features important for CB1 binding as well as help to club the most correlated descriptors which could be further explored in QSAR modeling.

Table 5: Values of different statistical parameters

Parameters	***Value***	***Value***	***Value***	***Value***	***Value***
r^2	0.906	0.905	0.904	0.904	0.904
XV r^2	0.952	0.951	0.951	0.951	0.951
BS r^2	0.844	0.838	0.832	0.831	0.831
LSE	0.048	0.049	0.049	0.049	0.049
r^2_{pred}	0.836	0.772	0.736	0.714	0.746

Where r^2 is Predictive correlation co-efficient, XV r^2 is cross validation correlation co-efficient, BS r^2 is Bootstrapping correlation coefficient, LSE is least square error, $r2_{pred}$ is the predicted correlation coefficient calculated from the predicted activity of the test set compounds

3.3 Statistical description:

The statistical models generated in our studies were investigated using the PLS leave-many-

out (LMO) method. The predictive ability of the models was assessed by their full cross-validated r^2 (q^2) values. Internal validation methods leave-one-out and LMO (with 10 groups of compounds) were used to determine the optimum number of PLS components and the stability of the models. The number of components used in the final non-validated model was optimized to give the highest q^2 value and the lowest standard error of prediction. The non-cross-validated models were assessed by the conventional correlation coefficient r^2 and F-values.

QSAR's most general mathematical form is:

Activity = f(physiochemical properties and/or structural properties).

The inter correlation of descriptors with each other and with the biological activity are given in the table. The table shows that descriptors are independent of each other. The important descriptors appeared in the equation I is shown in the table 1. The calculated correlations among the seven descriptors used in the equation and activity are represented in Table 6. Except Atype_Cl_90 and Jurs-FPSA-3 descriptor, other all the descriptors shows positive correlation with activity. Different types of descriptors are none other than number of molecular properties which are calculated in a new model of QSAR relationships. The 1st descriptor CHI-2 comes is molecular connectivity indices representing topological descriptors. In the best equation, this descriptor shows positive and good correlation with activity.

Kier & Hall valence-modified connectivity index (CHI-2): CHI-2 is Kier and Hall molecular connectivity index of order 2. The connectivity indices belong to topological class of descriptors and are single valued parameters that can be calculated from the 2D graph representation of molecules. They characterize structures according to their size, degree of branching and overall shape[26]. This index is a refinement of the molecular connectivity index, where a vertex sub graph valence d is enhanced to d^v to take into account electron configuration of the atom represented by the vertex:

$$\delta^v = (Z^v - h)/(Z - Z^v - 1) \quad (2)$$

where Z^v is the number of valence electrons in the atom, Z is its atomic number, and h is the number of hydrogens bound to it. This formula is designed to reproduce the unmodified molecular connectivity index for saturated hydrocarbons, for which d^v = d. However, d^v distinguishes between multiple and single bonds. The denominator introduces further distinction between element rows due to the presence of the atomic number Z[27, 28].

Table 6: Correlation matrix for the biological activity and the descriptors used in model 1

	CHI-2	***Jurs-DPSA-3***	***Atype_Cl_90***	***MW***	***Atype_ Unknown***	***Jurs-FPSA-3***	***MR***	***Activity***
CHI-2	1							
Jurs-DPSA-3	0.537	1						
Atype_Cl_90	-0.092	0.925	1					
MW	-0.053	-0.053	-0.553	1				
Atype_Unknown	0.207	0.208	0.207	0.207	1			
Jurs-FPSA-3	0.199	0.199	0.199	0.199	0.199	1		
MR	0.373	0.373	0.373	0.373	0.373	0.373	1	
Activity	0.505	0.086	-0.133	0.919	0.118	-0.047	0.773	1

The connectivity indices belong to topological class of descriptors and are single valued parameters that can be calculated from the 2D graph representation of molecules. They characterize structures according to their size degree of branching and overall shape. However, while similar molecules are expected to exert similar activities, there is no rigorous or unambiguous method for defining and calculating their similarity. It may be expected that, in certain cases, overall molecular similarity will produce similar activity, whereas, in other cases, only the molecular similarity of certain (active) regions of the molecules will give rise to similar activities. These results suggest that by modifying the shape of the molecule which produces the subgraphs of order 2 would enhance the activity of the CB1 inverse agonists. Further, the descriptors Jurs-DPSA-3 and Atype_Cl_90 reveal that the overall shape, size flexibly and branching in the molecules is important for activity and should be considered during lead optimization and drug designing.

Atype_C_90 is one of the various atom type AlogP descriptors that are used to estimate the logP of molecule which characterizes the lipophilicity. Ghose and Crippen[29] developed these descriptors that describe the individual atom type contribution towards the overall hydrophobicity of molecules. Heavy atoms like carbon, oxygen, nitrogen, sulfur, halogens along with hydrogens are categorized into 110 atom types. The number of atom types has elevated to 120 [30] after further revisions. Each AlogP atom-type value represents the number of atoms of that type in the molecule.

In the present set of CB1 inverse agonist molecules, the carbon type Atype_C_90 is surrounded by aromatic ring connected to halogens or oxy-alkyls or alkyls. Hydrophobicity associated with the aromatic ring environment surrounding the C atom is favorable for the CB1 inverse agonistic activity. As this descriptor is having positive coefficient, increasing the hydrophobic nature of the neighboring entities encompassing C atom would increase the activity of CB1 receptor inverse agonists

In the present QSAR equation, the AlogP descriptor (Atype_C_90) which is an indicator of lipophilicity of the molecule is having a negative coefficient. This implies that as the lipophillic nature of the molecule increases further the molecule passes through the cell membrane more effectively and thus increases the activity towards the receptor.

On the contrary, the molecular weight and Atype_C_90 descriptor is having a negative coefficient value. Based on these two statistical descriptors, it can be suggested that a balance between size of the molecule and addition of lipophillic groups has to be maintained to optimize the activity, as larger increase in the weight of the molecule may hinder its penetration through the cell membrane and its interaction with the receptor. This finding may be important as the CB1 receptor falls in GPCR family of proteins which are trans-membrane proteins characterized by seven trans-membrane domains. As shown in table 3, except for Atype_Cl_90, MW is having positive correlation with all the other descriptors (CHI-2, Jurs-DPSA-3, Atype_Unknown, Jurs-FPSA-3 and MR) and activity.

The 115 atom types defined in the calculation of AlogP90 are now available as descriptors. To calculate them, select the entry AlogP_atypes in the Thermodynamic family in the descriptor table. Each AlogP90 atom-type value represents the number of atoms of that type in the molecule. An additional atom type called Unkown_Type can also be added to the table, together with the other AlogP90 atom types. A value greater than zero for this descriptor indicates the presence of atoms that couldn't be classified as any of the defined AlogP90 atom types. The AlogP atom Types control panel allows you to select the elements to be taken into account.

Further, the descriptors Jurs-DPSA-3 and Jurs-FPSA-3 reveal that the overall shape, size flexibly and branching in the CB1 inverse agonist molecules is important for activity and should be considered during lead optimization and drug designing. Jurs descriptors mainly based on partial charges mapped on surface

area. This set of descriptors[31] combines shape and electronic information to characterize molecules. The descriptors are calculated by mapping atomic partial charges on solvent-accessible surface areas of individual atoms. A total of 30 different descriptors are included in this set. Difference in atomic charge weighted surface areas (DPSA-3) where atomic charge weighted positive solvent-accessible surface area minus atomic charge weighted negative solvent-accessible surface area. As shown in table 3, except for Atype_Cl_90 and MW, is also having positive correlation with all the other descriptors (CHI-2, Atype_Unknown, Jurs-FPSA-3, MR) and activity. Jurs-FPSA-3 also comes under Jurs descriptors.

Jurs-FPSA-3 stands for fractional charged partial surface areas is set of six descriptors obtained by dividing descriptors 1 to 6 by the total molecular solvent-accessible surface area. This descriptor shows negative correlation for predicted activity and positive correlation over all other descriptors and activity.

MW is molecular weight of the compound which comes under structural descriptors. Molecular weight is an important parameter that signifies the size of the molecule. In the dataset indicated good QSARs could be developed utilizing molecules with molecular volume and molecular weight providing good quantification of molecular size. It is useful to illustrate the influence of the shape and structural features of a molecule.

One of the most important chemico-physical properties used in QSAR studies is the molar refractivity (MR). It has been shown to be related to lipophilicity, molar volume and steric bulk. MR is Molar refractivity calculated as

$$MR = \left(\frac{(n^2-1)}{(n^2+2)}\right)\frac{(MW)}{d}$$

Where n is the refractive index, MW is the molecular weight, and d is the compound density. So, molar refractivity depends on molecular weight and density of the compound. MR comes under thermodynamic descriptors. The inter correlation of various descriptors was checked and the correlation matrices are given in Table 3, MR showing positive correlation with all the descriptors. As this descriptor is having positive coefficient, increasing lipophilicity, molecular weight and density would show positive to the activity of CB1 receptor inverse agonists.

The atom-type E-state indices are molecular descriptors encoding topological and electronic information related to particular atom types in the molecule[20, 32]. They are calculated by summing the E-state values of all atoms of the same atom type in the molecule. Each atom type is first defined by atom identity, based on the atomic number Z, and valence state, itself identified by the valence state indicator (VSI). Each atom type E-state symbol is a composite of 3 parts. The first part is “S” which refers to the sum of the E-states of all atoms of the same type. The second part is a string representing the bond types associated with the atom (“s”, “d”, “t”, “a” for single, double, triple and aromatic bonds, respectively). The third part is the symbol identifying the chemical element and, if any, bonded hydrogens, such as CH3, CH2, F etc. The E-state indices encode not only the information about the topological environment of an atom, but also the electronic interactions from other atoms in the molecule. Thus, E-state is able to provide useful information on structure features that mostly relate to the property to be modeled[33].

While understanding the SAR of CB1 inverse agonists in the present study basing on the QSAR equation, it is possible to prioritize the contributions provided by the descriptors incorporated in the model. Out of the 7 descriptors, structural descriptors showing highest positive coefficients as compared to other descriptors. Molecular weight is the main criteria for the compounds to predict activities in the study. MW showing highest positive coefficient value and it has good impact in predicting activities. Another chemico-physical properties used in QSAR studies is the molar refractivity (MR). It has been shown to be related to lipophilicity. This descriptor is also showing good correlation coefficient with activity. The next descriptor in terms of

contributory priority to SAR of CB1 inverse agonists in the present study is Kier & Hall valence-modified connectivity indices CHI-2 with positive coefficient is having the better weightage. Especially, CHI-2 index with good coefficient has to be considered while optimizing the molecule. From the above three descriptors contributions we can suggest that the shape, size and lipophilicity of the molecules is an important factor effecting CB1 inverse agonistic activity.

The next priority in considering the descriptors contributing to the activity as per the QSAR equation is for Atype_unknown descriptor with a positive coefficient, which is followed by Jurs-DPSA-3 and Atype_Cl_90 with Jurs-FPSA-3 descriptors having positive and negative coefficients, respectively. The inverse relationship between Atype_Cl_90 with Jurs-FPSA-3 descriptors indicate that the hydrophobic property and fractional charged partial surface areas of the molecules should be provided by incorporation of such groups which do not much increase the size but increase the activity. At the same time when CHI-2 descriptors contribution is taken into consideration, it can be implicated that the small hydrophobic groups having link with chlorine atom on the R3 and R4 positions of the scaffold may increase the activity of the CB1 inverse agonists.

4. CONCLUSIONS

The QSAR model of cannabinoid receptor 1 inhibitory activity have been developed based on topological, E-state, spatial and thermodynamic descriptors to estimate and predict relative inverse agonistic activity of 87 diphenyl-pyridine derivatives. The predictive ability of model was demonstrated by using leave one out cross validation technique, randomization test as well as external test set prediction. The results presented above show that these descriptors can be used to describe the structure activity relationship of cb1 inverse agonist and its performance based on statistical parameters is satisfying.

The top 5 equations/models generated by the G/PLS statistical method using a training set of 56 cannabinoid receptor 1 inverse agonists and validated these models with 31 test set compounds. Out of these 5 equations the equation I having the particular combination of 7 descriptors was having the best predictive ability compared to the rest of the equations. The model signifies the importance of Hydrophobicity associated with the aromatic ring environment surrounding the C atom is favourable for the CB1 inverse agonistic activity. The small hydrophobic groups having link with chlorine atom on the R3, R4 positions of the scaffold may increase the activity of the CB2 agonists. Further, overall shape size flexibly and branching in the CB1 inverse agonist molecules is important for activity and should be considered during lead optimization and drug designing. The structural and topological descriptors were found to play a major role in determining inhibitory activity for Cb1 receptor. The structural descriptors highlight the spatial importance for designing inhibitors for Cb1. Thus the present work may helpful in optimizing the CB1 inverse agonists.

References:

[1]. D. Cota, G. Marsicano, M. Tschöp, Y. Grübler, C. Flachskamm, M. Schubert, D. Auer, A. Yassouridis, C. Thöne-Reineke, S. Ortmann, F. Tomassoni, C. Cervino, E. Nisoli, A. C. Linthorst, R. Pasquali, B. Lutz, G. K. Stalla, and U. Pagotto. The endogenous cannabinoid system affects energy balance via central orexigenic drive and peripheral lipogenesis. J. Clin. Invest. 112 (3): 423-431 (2003).

[2]. G. Di Carlo, A. A. Izzo. Cannabinoids for gastrointestinal diseases: potential therapeutic applications. Expert Opin. Invest. Drugs 12 (1): 39-49 (2003).

[3]. R. G. Pertwee. Cannabinoid receptor ligands: clinical and neuropharmacological considerations, relevant to future drug discovery and development. Expert. Opin. Investig. Drugs 9 (7): 1553-1571 (2000).

[4]. J. N. Xiang, J. C. Lee. Pharmacology of cannabinoid receptot agonists and antagonists. Annu. Rep. Med. Chem. 34: 199-208 (1999).

[5]. F. Barth, M. Rinaldi-Carmona. The development of cannabinoid antagonists. Curr. Med. Chem. 6(8): 745-755 (1999).

[6]. M. Rinaldi-Carmona, F. Barth, M. Héaulme, D. Shire, B. Calandra, C. Congy, S. Martinez, J. Maruani, G. Néliat, D. Caput, P. Ferrara, P.

Soubrie, J. C. Breliere, G. Lefur. SR141716A, a potent and selective antagonist of the brain cannabinoid receptor. FEBS. Lett. 350 (2-3): 240-244 (1994).

[7]. G. G. Muccioli, D. M. Lambert. Current knowledge on the antagonists and inverse agonists of cannabinoid receptors. Curr. Med. Chem. 12 (12): 1361-1394 (2005).

[8]. D. L. Hertzog. Recent advances in the cannabinoids. Expert Opin. Ther. Pat. 14 (10): 1435-1452 (2004).

[9]. J. H. M. Lange, C. G. Kruse. Recent advances in CB_1 cannabinoid receptor antagonists. Curr. Opin. Drug Discov. Dev. 7: 498-506 (2004).

[10]. J. H. M. Lange, C. G. Medicinal chemistry strategies to CB_1 cannabinoid receptor antagonists. Drug Discov. Today 10: 693-702 (2005).

[11]. T. J. Hou, J. M. Wang, N. Liao, X. J. Xu. Applications of genetic algorithms on the structure–activity relationship analysis of some cinnamamides. J. Chem. Inf. Comput. Sci. 39: 775–781 (1999).

[12]. S. Oloff, R. B. Mailman, A. Tropsha. Application of validated QSAR models of D1 dopaminergic antagonists for database mining. J. Med. Chem. 48 (23): 7322–7332 (2005).

[13]. J.L. Medina-Franco, A. Golbraikh, S. Oloff, R. Castillo, A. Tropsha. Quantitative Structure-Activity Relationship Analysis of Pyridinone HIV-1 Reverse Transcriptase Inhibitors Using The k Nearest Neighbor Method and QSAR-Based Database Mining. J. Comput. Aided Mol. Des. 19 (4): 229–242 (2005).

[14]. M. Shen, C. Beguin, A. Golbraikh, J. P. Stables, H. Kohn, A. Tropsha. Application of Predictive QSAR Models to Database Mining: Identification and Experimental Validation of Novel Anticonvulsant Compounds. J. Med. Chem. 47 (9): 2356–2364 (2004).

[15]. L. C. Meurer, P. E. Finke, S. G. Mills, T. F. Walsh, R. B. Toupence, J. S. Debenham, M. T. Goulet, J. Wang, X. Tong, T. M. Fong, J. Lao, M. T. Schaeffer, J. Chen, C. P. Shen, D. Sloan Stribling, L. P. Shearman, A. M. Strack, L. H. Van der Ploeg. Synthesis and SAR of 5,6-diarylpyridines as human CB1 inverse agonists.Bioorg. Med. Chem. Lett 15 (3): 645-651 (2005).

[16]. J. S. Debenham, C. B. Madsen-Duggan, T. F. Walsh, J. Wang, X. Tong, G. A. Doss, J. Lao, T. M. Fong, M. T. Schaeffer, J. C. Xiao, C. R. Huang, C. P. Shen, Y. Feng, D. J. Marsh, D. S. Stribling, L. P. Shearman, A. M. Strack, D. E. MacIntyre, L. H. Van der Ploeg, M. T. Goulet. Synthesis of functionalized 1,8-naphthyridinones and their evaluation as novel, orally active CB1 receptor inverse agonists. Bioorg Med Chem Lett. 16 (3): 681-685 (2006).

[17]. C. B. Madsen-Duggan, J. S. Debenham, T. F. Walsh, R. B. Toupence, S. X. Huang, J. Wang, X. Tong, J. Lao, T. M. Fong, M. T. Schaeffer, J. C. Xiao, C. R. Huang, C. P. Shen, D. S. Stribling, L. P. Shearman, A. M. Strack, D. E. MacIntyre, L. H. Van der Ploeg, M. T. Goulet. Lead optimization of 5,6-diarylpyridines as CB1 receptor inverse agonists. Bioorg. Med. Chem. Lett. 17 (7): 2031-2035 (2007).

[18]. Cerius2 version 4.11, Accelrys Inc., 6985 Scranton Road, San Diego, CA, USA.

[19]. C. De Gregorio, L. B. Kier, L. H. Hall. QSAR modeling with the electrotopological state indices: corticosteroids. J. Comput. Aided Mol. Des. 12 (6): 557-561 (1998).

[20]. Kier LB, Hall LH. An electrotopological-state index for atoms in molecules. Pharm. Res. 7 (8): 801-807 (1990).

[21]. E. Estrada, E. Uriarte. Recent advances on the role of topological indices in drug discovery research. Curr. Med. Chem. 8 (13): 1573-1588 (2001).

[22]. D. Rogers, A. J. Hopfinger. Application of Genetic Function Approximation to Quantitative Structure-Activity Relationships and Quantitative Structure-Property Relationships. J. Chem. Inf. Comput. Sci. 34 (4): 854-866 (1994).

[23]. L. M. Shi, F. Yi, T. G. Myers, P. M. O'Connor, K. D. Paull, S. H. Friend, J. N. Weinstein. Mining the NCI anticancer drug discovery databases: genetic function approximation for the QSAR study of anticancer ellipticine analogues. J. Chem. Inf. Comput. Sci. 38 (2): 189-199 (1998).

[24]. A. Golbraikh, A. Tropsha. Beware of q2!. J Mol Graph Model. 20 (4): 269-276 (2002).

[25]. A. Afantitis, G. Melagraki, H. Sarimveis, P. A. Koutentis, J. Markopoulos, O. Igglessi-Markopoulou. Investigation of substituent effect of 1-(3,3-diphenylpropyl)-piperidinyl phenylacetamides on CCR5 binding affinity using QSAR and virtual screening techniques. J Comput Aided Mol Des. 20 (2): 83-95 (2006).

[26]. J. Liu, L. Yang, Y. Li, D. Pan, A. J. Hopfinger. Prediction of plasma protein binding of drugs using Kier-Hall valence connectivity indices and 4D-fingerprint molecular similarity analyses. J Comput. Aided Mol Des. 19 (8): 567-583 (2005).

[27]. L. B. Kier, L. H. Hall. In: Molecular Connectivity in Chemistry and Drug Research. Academic. Press: New York. 14: 257 (1976).

[28]. L. B. Kier, L. H. Hall. Molecular Connectivity in Structure-Activity Analysis. In: (1st ed.), Chemometric Series, Research Studies Press LTD., England. 9:262 (1985).

[29]. A. K. Ghose; G. M. Crippen. Atomic Physicochemical Parameters for Three-

Dimensional Structure-Directed Quantitative Structure-Activity Relationships I. Partition Coefficients as a Measure of Hydrophobicity. J. Comput. Chem. 7 (4): 565-577 (1986).

[30]. V. N. Viswanadhan, A. K. Ghose, G. R. Revankar, R. K. Robins. Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. J. Chem. Inf. Comput. Sci. 29 (3): 163–172 (1989).

[31]. D. T. Stanton, P. C. Jurs. Development and Use of Charge Partial Surface Area Structural Descriptors in Computer-Assisted Quantitative Structure-Property Relationship Studies. Anal. Chem. 62 (21): 2323-2329 (1990).

[32]. L. B. Kier, L. H. Hall. Molecular Structure Description: The Electrotopological State, Academic Press, San Diego, CA. ISBN 0-12-406555-4 (1999).

[33]. K. Rose, L. H. Hall. Modeling blood-brain barrier partitioning using the electrotopological state. J Chem Inf Comput Sci. 42 (3): 651-666 (2002).