December 2006 Volume 2 Issue 5-6

Trade Science Inc.

Organic CHEMISTRY

An Indian Journal

Full Paper

OCAIJ, 2(5-6), 2006 [131-139]

QSAR Study On 1,4-Dihydropyridine-Based Calcium Channel Antagonist: Dominating Role Of Molar Refractivity

Corresponding Author

Abhilash Thakur
Department of Chemistry,
Rishiraj Institute of Technology,
Indore (INDIA)
Ph.: 91-98260-43140
E-Mail: abhilashthakur@yahoo.com

Received: 25th November, 2006
Accepted: 10th December, 2006

Web Publication Date : 28th December, 2006

Co-Authors

Mamta Thakur[1], Padmakar Khadikar[2], Arpan Bharadwaj[3]
[1]Department of Chemistry, Kailash Chand Bansal Academy of Technology, Indore (INDIA)
[2]Research Division, Laxmi Fumigation & Pest control, Pvt. Ltd, 3, Khatipura, Indore (INDIA)
[3]Department of Chemistry, Govt. Madhav Science College, Ujjain, (INDIA)

ABSTRACT

QSAR study on 1, 4-dihydropyridine-based calcium channel antagonists has been carried out using heterogeneous set of molecular descriptors which includes some distance-based topological indices and adhoc physicochemical descriptors. The results have shown that molecular refractivity plays a dominating role in this regard. This single parameter in combination with indicator parameters yielded excellent model. The results are critically discussed with a variety of statistics including neural network.

KEYWORDS

QSAR;
Regression analysis;
Molar refractivity;
Physicochemical properties.

INTRODUCTION

Over the last several years bask group used neural network as well as multivariate methods to analyze the relation between structure and activity. Recently[1] they have reported a comparison between neural network based QSAR and multiple linear regression based QSAR using 1,4-dihydropyridine based calcium channel antagonist. This set of 1,4-dihydropyridines is well known as calcium channel blockers and also as inhibitor of P-glycoprotein and associated with multidrug resistance to anticancer agents[1]. This set of compounds is, therefore, useful as multidrug resistance reversal agents. It was argued that lipophilic, electronic and steric parameters (discriptor) are needed to develop QSAR model[1].

They have indicated that the Hammette constant for the substituent at meta position is of specifics importance. In addition, they have also used three sterimol parameters. Basak and co-workers[1] used a large set of 90 topological indices for developing the QSAR model for this set of calcium channel antago-

Full Paper

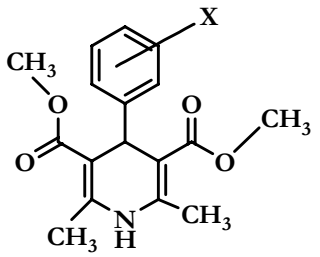

Figure 1: The skeleton of 1,4-dihydropyridine derivative structure with substitutions/changes indicated by X on the phenyl ring for details see TABLE 1

TABLE 1: Statistical results of basak and coworkers[1] for the calculated (predicted) values for 45 compounds predicted by various models

Method	Correl.	RMS	Max.dev.	Min.dev.	R^2_{Pred}	n
MLR	0.55	1.30	5.15	0.00	0.19	45
PCANN (2PE)	0.65	1.15	3.63	0.05	0.36	45
(3PE)	0.61	1.22	3.79	0.01	0.28	45
(4PE)	0.71	1.05	3.63	0.04	0.47	45
(8PE)	0.73	1.02	2.67	0.08	0.50	45
(10PE)	0.68	1.16	3.60	0.01	0.35	45

nist (Figure 1, TABLE 2)

In their study, Basak and coworkers[1] have then employed neural network based as well as multiple linear regression-based QSAR approaches and arrived at the following statistical results.

While reinvestigating the aforementioned study, we came across the excellent review of Hansch group[2] in that it was stated that 'over the years we have been impressed by the great importance of hydrophobic effects in chemical biological interactions as brought out by quantitative structure activity relationships QSAR from mechanistic organic chemistry for comparison, if seems timely to examine those instances where hydrophobic terms are not significant.'

Promoted by this we have reinvestigate the work of Basak[1] to see if we can still further develop QSAR model without hydrophobic term. For this we have used a larger set of molecular descriptors including topological indices and indicator parameters, distance based topological indices viz Wiener(W)[3], first order connectivity($^1\chi$)[4] Balaban(J)[5] and Szeged(Sz)[6-8] indices, Molar Refractivity(MR), Molar Volume (MV), Parachor(Pc), Refractive index(η), surface tension(ST), density(d), polarizability(Pol); indicator parameters I_X, I_A, I_2, I_3, I_4, for halogen substitution at X, alkyl substitution at X, substitution X at 2-, 3-, and 4- positions respectively. We obtained very interesting results indicating that a single molecular distributor(MR) along with some indicator parameters yielded QSAR model with excellent statistics and that is no need of any hydrophobic parameter for this purpose. Surprisingly our result were found still better then those of the earlier worker[1].

At this stage it is worthy to communicate on the molar refraction(MR) which played a dominating role in over study. It is in essence is an adjustable molecular volume which contains an eligible contribution and thus it equals the molar polarizability. Since molar refraction(MR) to a large extent is a measure of volume, it may measure the contribution due to size and shape of organic compounds acting as drugs. To test the hypothesis that MR may be related to steric, hinders, Dunn[9] carried out a regression analysis of MR as a function of the Hancock steric parameter, ES[10,11]. This study indicated that a negative coefficient with MR might indicate that steric effects are involved in the correlation under study. In addition to the inter relationship between MR and ES, the MR of a substituent may be related to the corresponding lipophilisity as expressed by π. In fact, although for a proper set of subsituants MR and π are orthogonal, both parameters depend to a certain extent on molecular volume(MV). Thus the co-linearity between those vectors may be high, depending upon the choice of substitution. Thus we can conclude that MR is a thermodynamic descriptor of a substituent and is a combined measure of substituent's size and polarizability. Also that both steric and lipophilic effects are contained in it. This is, therefore, the main reason for the dominating role played by MR in our present study. In developing QSAR models using this dominate molecular descriptor(MR) we have used maximum-R2 method and followed step-wise regression analysis[11]. The predictive power of the models is determined on the basis of cross validation as well as their statistical method[12,13].

TABLE 2: Structural details of the activity and indicator parameters

Comp. No	X	log 1/ EC_{50} (Exp.)	I_X	I_A	I_2	I_3	I_4
1	3-Br	8.89	1	0	0	1	0
2	2-CF_3	8.82	1	0	1	0	0
3	2-Cl	8.66	1	0	1	0	0
4	3-NO_2	8.40	0	0	0	1	0
5	2-CHdCH_2	8.35	0	1	1	0	0
6	2-NO_2	8.29	0	0	1	0	0
7	2-CH_3	8.22	0	1	1	0	0
8	2-Et	8.19	0	1	1	0	0
9	2-Br	8.12	1	0	1	0	0
10	2-CN	7.80	0	0	1	0	0
11	3-Cl	7.80	1	0	0	1	0
12	3-F	7.68	1	0	0	1	0
13	H*	7.55	0	0	0	0	0
14	3-CN	7.46	0	0	0	1	0
15	3-I	7.38	1	0	0	1	0
16	2-F	7.37	1	0	1	0	0
17	2-I	7.33	1	0	1	0	0
18	2-OCH_3	7.24	0	1	1	0	0
19	3-CF_3	7.13	1	0	0	1	0
20	3-CH_3	6.96	0	1	0	1	0
21	2-OEt	6.96	0	1	1	0	0
22	3-OCH_3	6.72	0	1	0	1	0
23	3-N Me_2	6.05	0	1	0	1	0
24	3-OH	6.00	0	0	0	1	0
25	3-NH_2	5.70	0	0	0	1	0
26	3-OAc	5.22	0	1	0	1	0
27	3-O-COPh	5.20	0	1	0	1	0
28	2-NH_2	4.40	0	0	1	0	0
29	4-F	6.89	1	0	0	0	1
30	4-Br	5.40	1	0	0	0	1
31	4-I	4.64	1	0	0	0	1
32	4-NO_2	5.50	0	0	0	0	1
33	4-N Me_2	4.00	0	1	0	0	1
34	4-CN	5.46	0	0	0	0	1
35	4-Cl	5.09	1	0	0	0	1
36	2,6-Cl_2	8.72	1	0	1	0	0
37	F_5	8.36	1	0	1	1	1
38	2-F,6-Cl	8.12	1	0	1	0	0
39	2,3-Cl_2	7.72	1	0	1	1	0
40	2-Cl,5-NO_2	7.52	1	0	1	0	0
41	3,5-Cl_2	7.03	1	0	0	1	0
42	2-OH,5-NO_2	7.00	0	0	1	0	0
43	2,5-Me_2	7.00	0	1	1	0	0
44	2,4-Cl_2	6.40	1	0	1	0	1
45	2,4,5-$(OCH_3)_3$	3.00	0	1	1	0	1

I_X =1 if Substitution on X position is halogen atoms.
I_A =1 if Substitution on X position is alkyle group.
I_2 =1 if Substitution on X position is on 2nd atom.
I_3 =1 if Substitution on X position is on 3rd atom.
I_4 =1 if Substitution on X position is on 4th atom.

EXPERIMENTAL

Biological activity

The biological activity for the set of 45 compounds used in the present study, as expressed by log1/EC_{50}, were taken from the literature[1].

Topological indices

All the topological indices were calculated from the hydrogen suppressed molecular graphs using the software of lokovits.

Wiener index (W)[3]-Wiener index W=W(G) of G is defined as the half sum of the elements of the distance matrix.

$$W = W(G) = \tfrac{1}{2} \sum_{i=1,\, j=1} \sum (D)_{ij}$$

Where, $(D)_{ij}$ is the ijth element of the distance matrix which denotes the shortest graph-theoretical distance between sites i and j of G.

The connectivity index($^1\chi$)[4]- the connectivity index $^1\chi = {}^1\chi(G)$ of G is defined by Randic as

$$^1\chi = {}^1\chi(G) = \sum_{i,j} [d(i) \times d(j)]^{-0.5}$$

Balaban index (J)[5] - The Balaban index J=J(G) of G is defined as

$$J = M/\mu + 1 \sum_{bonds} (d_i \times d_j)^{-0.5}$$

Where M is the number of bonds in G, μ is the cyclomatic number of G, and d_i (i =1,2,3,….N; N is the number of vertices in G) is the distance sum.

The cyclomatic number $\mu = \mu(G)$ of a cyclic graph G is equal to the minimum number of edges necessary to be erased from G in order to transform it into the related acyclic graph. In case of monocyclic graph μ=1 otherwise it is calculated by means of the fol-

Full Paper

TABLE 3: Topological descriptors for 1, 4-dihydropyridine derivatives used in present study

Comp.No.	W	$^1\chi$	J	Sz
1	1044	10.90028	2.33682	1586
2	1409	12.12844	2.47876	2073
3	1028	10.91712	2.37450	1554
4	1328	11.81097	2.34299	1994
5	1153	11.45512	2.39903	1725
6	1280	11.82780	2.43426	1898
7	1028	10.91712	2.37450	1554
8	1153	11.45512	2.39903	1725
9	1028	10.91712	2.37450	1554
10	1153	11.45512	2.39903	1725
11	1044	10.90028	2.33682	1586
12	1044	10.90028	2.33682	1586
13	926	10.50644	2.31407	1406
14	1185	11.43829	2.33281	1789
15	1044	10.90028	2.33682	1586
16	1028	10.91712	2.37450	1554
17	1028	10.91712	2.37450	1554
18	1153	11.45512	2.39903	1725
19	1473	12.11161	2.36505	2201
20	1044	10.90028	2.33682	1586
21	1302	11.95512	2.39644	1920
22	1185	11.43829	2.33281	1789
23	1328	11.81097	2.34299	1994
24	1044	10.90028	2.33682	1586
25	1044	10.90028	2.33682	1586
26	1496	12.34897	2.33241	2224
27	2604	14.86661	1.75500	3894
28	1028	10.91712	2.37450	1554
29	1060	10.90028	2.30394	1618
30	1060	10.90028	2.30394	1618
31	1060	10.90028	2.30394	1618
32	1376	11.81097	2.26482	2090
33	1376	11.81097	2.26482	2090
34	1217	11.43829	2.27539	1853
35	1060	10.90028	2.30394	1618
36	1134	11.32780	2.43926	1708
37	1538	12.57669	2.53864	2330
38	1134	11.32780	2.43926	1708
39	1149	11.32780	2.40541	1738
40	1447	12.22165	2.41474	2168
41	1166	11.29413	2.36861	1772
42	1447	12.22165	2.41474	2168
43	1151	11.31097	2.40199	1742
44	1166	11.31097	2.37258	1772
45	1763	13.33566	2.46464	2639

*W=Wiener index, $^1\chi$=First -order conectivity index, J=Balaban index, Sz=Szeged index

lowing expression

M=M-N+1

Szeged index (Sz)[6-8]- the Szeged index, Sz= Sz(G), is calculated according to the following expression:

$$Sz = Sz(G) = \sum_{edges} n_u \cdot n_v$$

Where n_u is the number of vertices lying closer to one end of the edge e=uv; the meaning of n_v is analogous. Edges equidistance from both the ends of an edges, e=uv are not taken into account.

Physicochemical parameter

In present study molar refractivity (MR), molar volume(MV), parachor(Pc), index of refraction(η), surface tension(ST), density(D), and polarizability (Pol) shown in TABLE 3 are tested and calculated from computer software acdlabs(Chem Skech5.0).

Regression analysis

Only regression analyses were made using maximum R[2] method[12] adopting step-wise regression. Here also the regress-1 program prepared Lukovits.

Neural network

From neuralware predict demo version

RESULTS AND DISCUSSION

A set 45 1,4-dihydropyridines, there activity expressed as log1/EC_{50} and indicator parameters(I_X, I_A, I_2, I_3, I_4) are present in TABLE 2. This TABLE 2 shows that a low level degeneracy in present in the activity value(log1/EC_{50}) are that activity follow the following sequence:

1>2>36>2\3>4>37>5>6>7>8>9=38>10=11>39>12>13>40>14>16>17>18>19>41> 42=43>20=29>22>44>23>24>25>32>34>30>26>27>35>31> 28>33>45 (1)

Thus, 1,4-dihydropyridine with substitution of Br at 3 position exhibits highest activity, while the one with 2,4,5-$(OCH_3)_3$ has the lowest activity. However, this sequence of activity failed to demonstrate any structure-activity relationships. We have there-

TABLE 4: Physicochemical descriptors for 1, 4-dihydropyridine derivatives used in present study

Comp No.	MR	MV	Pc	η	ST	d	Pol
1.	89.04	276.6	703.4	1.557	41.8	1.374	35.30
2.	86.33	293.9	714.3	1.499	34.8	1.256	34.22
3.	86.25	272.3	689.5	1.546	41.0	1.232	34.19
4.	87.90	272.2	709.4	1.558	46.0	1.271	34.84
5.	92.28	286.3	719.9	1.557	39.9	1.143	36.58
6.	87.90	272.2	709.4	1.558	46.0	1.271	34.84
7.	86.18	276.7	690.6	1.535	38.8	1.139	34.16
8.	90.90	293.2	730.7	1.532	38.5	1.123	36.03
9.	89.04	276.6	703.4	1.557	41.8	1.374	35.30
10.	86.46	261.6	700.0	1.574	51.2	1.24	34.27
11.	86.25	272.3	689.5	1.546	41.0	1.232	34.19
12.	81.35	264.6	659.7	1.526	38.6	1.206	32.25
13.	81.35	260.4	652.4	1.537	39.3	1.157	32.25
14.	86.46	261.6	700.0	1.574	51.2	1.24	34.27
15.	94.26	282.5	726.3	1.581	43.6	1.512	37.37
16.	81.35	260.4	652.4	1.537	39.3	1.157	32.25
17.	94.26	282.5	726.3	1.581	43.6	1.512	37.37
18.	88.03	284.4	711.0	1.531	39.0	1.164	34.90
19.	86.33	293.9	714.3	1.499	34.8	1.256	34.22
20.	86.18	276.7	690.6	1.535	38.8	1.139	34.16
21.	92.67	300.9	751.0	1.528	38.8	1.147	36.73
22.	88.03	284.4	711.0	1.531	39.0	1.164	34.90
23.	95.67	298.4	757.0	1.554	41.4	1.154	37.92
24.	83.24	258.8	667.6	1.556	44.2	1.225	32.99
25.	85.59	262.7	680.3	1.565	44.9	1.204	33.93
26.	92.70	298.0	756.2	1.534	41.4	1.205	36.74
27.	113.00	343.8	890.0	1.571	44.8	1.225	44.79
28.	85.59	262.7	680.3	1.565	44.9	1.204	33.93
29.	81.35	264.6	659.7	1.526	38.6	1.206	32.25
30.	89.04	276.6	703.4	1.557	41.8	1.374	35.30
31.	94.26	282.5	726.3	1.581	43.6	1.512	37.37
32.	87.90	272.2	709.4	1.558	46.0	1.271	34.84
33.	90.96	279.8	718.9	1.563	43.5	1.180	36.06
34.	86.46	261.6	700.0	1.574	51.2	1.24	34.27
35.	86.25	272.3	689.5	1.546	41.0	1.232	34.19
36.	91.14	284.3	726.6	1.554	42.6	1.302	36.13
37.	81.32	281.4	689.1	1.489	35.9	1.390	32.24
38.	86.24	276.5	696.8	1.536	40.2	1.279	34.19
39.	91.14	284.3	726.6	1.554	42.6	1.302	36.13
40.	92.80	284.2	746.6	1.566	47.6	1.339	36.78
41.	91.14	284.3	726.6	1.554	42.6	1.302	36.13
42.	89.78	270.7	724.6	1.577	51.3	1.338	35.59
43.	91.00	292.9	728.9	1.533	38.3	1.124	36.07
44.	91.14	284.3	726.6	1.554	42.6	1.302	36.13
45.	101.39	332.4	828.2	1.522	38.5	1.177	40.19

TABLE 4 Continued

*MR = Molar Refractivity, MV=Molar Volume, Pc=Parachor, η=Index of refraction ST=Surface Tension, D=Density, Pol=Polarizability

for, used topological indices(TABLE 3), adhoc molecular descriptors(TABLE 4) and the indicator parameters(TABLE 2) for obtaining statistically significant QSAR model. As stated above, we have used maximum-R^2 method and followed stepwise regression to obtained statistically significant model.

A perusal of TABLE 2 shows that high to low degeneracy is observed in the topological indices used. This is obvious because W and Sz indices are the first generation indices with both $^1\chi$ and J are the second-generation indices. Balaban[14-15] has shown that the indices in spite of their degeneracy can be used successfully it developing statistically significant QSAR models.

Before a multivariate analysis is undertaken it is convenient to normalize the data in certain ways in order to make the detection of significant correlation career. Normally, it is sufficient to preprocess the data or means of autoscaling and mean centering the variable auto scaling gives each variable unit and variance and hence the same chance to contribute to a calculated model, while mean scaling facilitates interpretation. This can be achieved by obtaining correlation matrix. Such a correlation matrix obtained in the present case is shown in TABLE 5, which shows that except Balaban index(J)[5] other distance-based topological indices are highly linearly correlated. Also, that none of the molecular descriptors (topological indices, adhoc molecular descriptors, indicator parameters) correlate well with the activity (log $1/EC_{50}$). That is not possible to obtained one variable model for modeling log$1/EC_{50}$ and hence we need to go for multi-variable analysis.

For symbols see TABLES 1-3

The preliminary regression analysis has indicated that MR plays a dominating role and that use of indicator parameter I_4 (substitution of X at 4-position) is important for developing multivariate model. Thus, combination of MR and I_4 yielded with fair statistics.

Full Paper

TABLE 5: Inter correlation between descriptors and their correlation with biological activity in form of correlation matrix

	log 1/EC_{50}	W	$^1\chi$	J	Sz	MR	MV	Pc	
log 1/EC_{50}	1.00000								
W	-.30237	1.00000							
$^1\chi$	-.25836	.98395	1.00000						
J	.38071	-.52041	-.37660	1.00000					
Sz	-.31664	.99934	.97943	-.53441	1.00000				
MR	-.35801	. 69485	. 67387	-.51121	. 69433	1.00000			
MV	-.25997	.76404	.77660	-.31232	.75778	.84923	1.00000		
Pc	-.34458	.82628	.82893	-.43963	.82288	.95930	.92119	1.00000	
η	-.21236	-.08535	-.14055	-.35639	-.07577	.33325	-.21376	.13335	
ST	-.17167	.06448	.04826	-.22944	.07101	.14540	-.31582	.07643	
D	.08725	-.06280	-.08817	.06042	-.05393	.14738	-.06915	.04602	
Pol	-.35798	.69446	.67343	-.51115	.69396	1.00000	.84916	.95911	
I_X	.30692	-.24156	-.27209	.19314	-.23465	-.16179	-.09653	-.20766	
I_A	-.24473	.33667	.36992	-.18358	.32908	.45380	.59486	.48966	
I_2	.35717	-.02345	.08986	.53157	-.04107	.03034	.12494	.07915	
I_3	.06780	-.04923	-.06942	-.04572	-.04755	-.19247	-.14948	-.19952	
I_4	-.55815	.09290	.07298	-.07002	.11369	-.00556	.01357	.01916	
	η	**ST**	**D**	**Pol**	**I_X**	**I_A**	**I_2**	**I_3**	**I_4**
η	1.00000								
ST	.84381	1.00000							
D	.40257	.27560	1.00000						
Pol	.33336	.14506	.14812	1.00000					
I_X	-.13562	-.28116	.59852	-.16101	1.00000				
I_A	-.22779	-.32862	-.59716	.45351	-.59621	1.00000			
I_2	-.15424	-.10855	-.02527	.03040	.06535	.06321	1.00000		
I_3	-.10096	-.11809	-.12971	-.19240	.16071	-.09172	-.03564	1.00000	
I_4	-.02916	.01717	.18944	-.00526	.14286	-.10483	-.20198	-.17857	1.00000

$\log 1/EC_{50}$=-0.0910(±0.0290)MR-1.8756(±0.3860) I_4+ 15.4455 (2)

n=45, Se=1.0764, R=0.6648, R^2_A=0.4154, F=16.630

As stated earlier, the negative coefficient of MR in the above equation 2 indicates that steric effect is dominating in the exhibition of the activity(log1/EC_{50}). Also, that the negative coefficient of I_4 term indicates that substituted X at 4-position is not favorable in the exhibition of the activity.

When the indicator parameter I_x(effect due to halogen substitution at X) is added to the model expressed for eq. (2) there is quite good improvement in the statistics:

$\log 1/EC_{50}$=-0.0770(±0.0266)MR-2.0393(±0.3525)I_4+ 0.9612(±0.2916)I_x +13.7892 (3)

n=45, Se=0.9727 R=0.7451, R^2_A =0.5225, F=17.052

Contribution due to MR and I_4 in the exhibition of the activity is the same as discussed above. The positive coefficient of the I_x term in eq (3) indicates that presence of halogen at X is favorable for the exhibition of the activity(log1/EC_{50}).

TABLE 6: Statistics due to deletion of outleir

n	Outlier	Se	R	R^2_A	F
45	1	0.9610	0.7788	0.5672	15.410
44	2	0.7776	0.8412	0.6777	23.604
41	1	0.6506	0.8696	0.7292	27.923
40	0	0.5939	0.8893	0.7669	33.069

The step-wise regression has indicated that when an indicator parameter I_2 (effect due to substitution at 2-position) is added further improvement in the statistics is observed:

$\log 1/EC_{50}$ =-0.0797(±0.0253)MR-1.8704(±0.3436)I_4+ 0.8937(±0.2549)I_X +0.6489(±0.2836)I_2 +13.7059 (4)

TABLE 7: Observed and calculated log1/EC_{50} values along their residuals

Comp No.	log1/ EC_{50}(Obs)	log1/ EC_{50}[a]	Resid.	log1/ EC_{50}[b]	Resid.
1.	8.88	7.50	1.39	7.09*	1.80
2.	8.82	8.37	0.45	8.53	0.29
3.	8.66	8.37	0.29	8.53	0.13
4.	8.40	6.70	1.70	6.55*	1.86
5.	8.35	7.00	1.35	7.40	0.95
6.	8.29	7.35	0.94	7.76	0.53
7.	8.22	7.49	0.73	7.90	0.32
8.	8.19	7.11	1.08	7.51	0.68
9.	8.12	8.15	0.03	8.31	-0.19
10.	7.80	7.46	0.34	7.88	-0.08
11.	7.80	7.73	0.08	7.32	0.48
12.	7.68	8.12	-0.44	7.72	-0.04
13.	7.55	7.22	0.33	7.08	0.47
14.	7.46	6.82	0.65	6.66	0.80
15.	7.38	7.09	0.29	6.67	0.72
16.	7.37	8.77	-1.40	8.94*	-1.56
17.	7.33	7.74	-0.41	7.88	-0.55
18.	7.24	7.34	-0.10	7.75	-0.51
19.	7.13	7.72	-0.59	7.31	-0.18
20.	6.96	6.84	0.12	6.69	0.27
21.	6.96	6.97	-0.01	7.37	-0.40
22.	6.72	6.69	0.03	6.53	0.19
23.	6.05	6.08	-0.03	5.91	0.14
24.	6.00	7.07	-1.07	6.93	-0.93
25.	5.70	6.88	-1.18	6.73	-1.03
26.	5.22	6.31	-1.10	6.15	-0.93
27.	5.20	4.70	0.50	4.49	0.71
28.	4.40	7.53	-3.13	7.95*	-3.55
29.	6.89	6.25	0.64	6.20	0.69
30.	5.40	5.63	-0.23	5.57	-0.17
31.	4.64	5.22	-0.58	5.15	-0.51
32.	5.50	4.83	0.67	5.03	0.47
33.	4.00	4.59	-0.59	4.78	-0.78
34.	5.46	4.95	0.52	5.15	0.32
35.	5.09	5.86	0.77	5.80	-0.71
36.	8.72	7.99	0.74	8.13	0.59
37.	8.36	6.90	1.46	7.42	0.94
38.	8.12	8.40	-0.26	8.54	-0.41
39.	7.72	7.99	-0.27	8.13	-0.41
40.	7.52	7.85	-0.33	8.00	-0.48
41.	7.03	7.34	-0.31	6.92	0.11
42.	7.00	7.20	-0.20	7.61	-0.60
43.	7.00	7.10	-0.10	7.51	-0.51
44.	6.40	6.11	0.27	6.62	-0.26
45.	3.00	4.40	-1.40	5.14*	-2.14

[a]Calculated log1/EC_{50} from eq. (4); [b]Calculated log1/EC_{50} from eq. (5)
*Data point not included in calculation

n=45, Se=0.9261, R=0.7788, R^2_A =0.5672, F=15.418

The physical significance of MR, I_4, and I_x terms are the same as discussed for the model expressed by eq (3). The positive coefficient of I_2 in eq (4) indicated that presence of substitution at the 2- position is favorable for the exhibition of activity. We observed that this model containing 45 compounds has one outlier. Deletion of this outlier from the regression procedure resulted into a model with considerable improved statistics. However, unfortunately this new model contained two outliers, the removal of which generated another model with better statistics, but with one outlier. Finally, removing this outlier also we obtained a model with no outlier. This final model is discussed below. The results due to successive deletion of the outlier are given in TABLE 6.

As stated above successive regression yielded a model expressed by eq (4) which has five compounds **(1,4,16,28,45)** with high residue(difference between observed and calculated activity) indicating thereby that are to be treated as outliers and deleted from regression procedure. When we did so we observed a large increase in the statistics giving the following model:

$$\log 1/EC_{50} = -0.0818(\pm 0.0178)MR + 0.6402(\pm 0.1956)I_x + 1.2139(\pm 0.1981)I_2 - 1.5158(\pm 0.2427)I_4 + 13.7351 \quad (5)$$

n = 40, se = 0.5939, R = 0.8893, R^2_A = 0.7669, F = 33.069

The model, eq (5), does not have any outlier. Also, that in eq (5) the physical significance of MR, I_x, I_2 and I_4 terms are the same as discussed for the earlier model.

In order to confirm our findings we have calculated log1/EC_{50} values using equations (4) and (5) and compared them with the observed value of log1/EC_{50}. Such a comparison is shown in TABLE 7 which shows that the calculated value of log1/EC_{50} using eq (5) are very close to the observed value of log1/EC_{50}. This indicate that the model expressed by eq (5) is the best model for modeling the activity. Further evidences in our favour are obtained by calculating predictive correlation coefficient from the plot

Full Paper

TABLE 8: Cross-validation parameters

Models (eq.)	n a	PRESS	SSY	PRESS/SSY	r^2_{cv}	SPRESS	PSE
1(4)	45(1)	34.3052	52.8915	0.6486	0.3514	0.9261	0.8730
2(5)	40(3)	12.3450	46.6562	0.2646	0.7354	0.7736	0.5335
3(6)	44(2)	23.5790	57.0839	0.4130	0.5869	0.6506	0.7320
4(7)	41(1)	15.2384	47.2774	0.3223	0.6777	0.5939	0.6096

a in column second; the no in the parenthesis is for the outleir

of observed against calculated log1/EC_{50} (Figure 2 and 3) which shows the model with eq (5) is the best model.

It is worthy to mention that a model with excellent statistics may or may not have excellent predictive power. In view of this we have calculated cross-validated parameters (TABLE 8). The parameters calculated being PRESS(predictive residual error sum of squares), SSY(sum of the squares of response value), R^2_{CV}(the overall predictive ability), S_{PRESS}(which is used in deciding uncertainty of prediction) and PSE (used for the prediction of correlation predictability of the models)

Comparison of these parameters indicates that the model expressed by eq(5) is not only statistically excellent but it has excellent predictive power too.

Now, it will be interesting to comment upon the adjustable R^2(R^2_A). The values of this parameter i.e., R^2_A takes into account of adjustment of R^2. Therefore, if a variable is added that does not contribute its fair share, the R^2_A will actually decline. Where as, R^2 will always increase when an independent variable is added. R^2_A will decrease if the added variable does not reduce the unexplained variation enough to offset the loss of degree of freedom. The examination of eq (2) to (5) indicates that as we start from the two variable models to four variable models there is consistent increase in the magnitude of R^2_A. That is, successive increase in the independent variable resulted into increase in the value of R^2_A. Therefore, in each case the added variable has its fair contribution to the resulting models.

The examination of figure 2 and 3 indicates that predictive correlation coefficients(R^2_{pred}) of the model expressed by eq (4) and (5) are found as 0.7562 and 0.7909 respectively. The comparison of this and other parameters for these models with those of Basak and Coworkers[1] indicates that better results are obtained in the present case. This is probably due to the fact that the parameter MR contains three

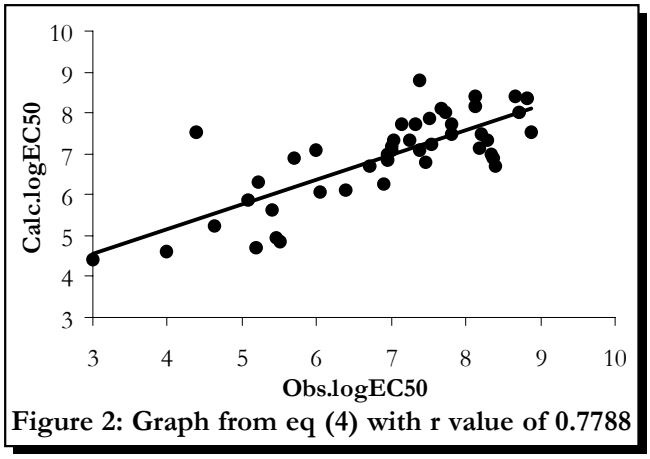

Figure 2: Graph from eq (4) with r value of 0.7788

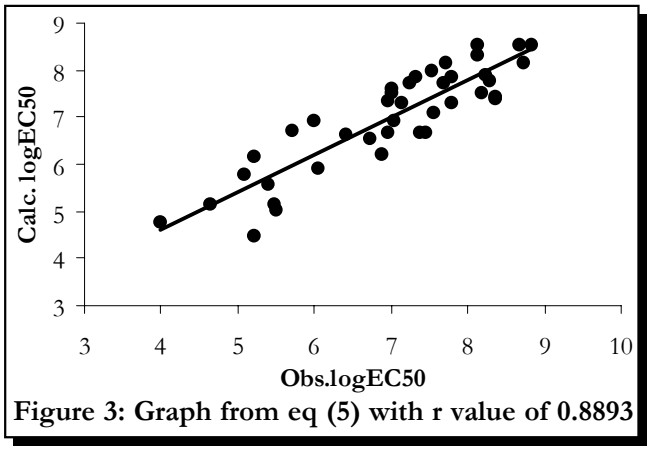

Figure 3: Graph from eq (5) with r value of 0.8893

TABLE 9: Comparison of neural network data

Method	No. of compds	Correlation	R^2_{pred}
MLR	45	0.79(0.55)	0.62(0.19)
ANNW	45	0.84(0.73)*	0.71(0.50)

Values in the parenthesis are those of earlier work[1].
*** earlier reported best model**

effects size, steric and lipophilic effects.

The comparison of our results with those of earlier workers[1] will be complete only when we also carry out artificial neural network analysis another reason of doing such an analysis is that it(neural networks) has become privilege statistical tool for model complex biological activities and/or properties of organic chemicals particularly used as drugs. Our results exploring neural network are presented in

TABLE 9 in that we have also included earlier reported data for comparison. As we see from the TABLE 9, we observed that in this case also we obtained better results than the earlier reported results.

CONCLUSION

From the results and discussion made above we conclude that the use of single molecular parameter MR in combination with indicator parameters is more that enough to model the activity($\log 1/EC_{50}$). The dominating role of MR is argued on the basis that all other molecular descriptors used in place of MR resulted into inferior models discussed above. It is worthy to mention that MR still is the chameleon among the physiochemical parameters, despite its wide application in QSAR studies. It is related to volume and size of a substituent. The refractive index related correlation term is MR accounts for the polarizablity and consequently for the size and the polarity of certain group. The larger the polar part of a molecule is the larger its MR value will be. The hydrophobic substituents at close interrelation between volume, surface lipophilic and MR is to be expected. This interrelation breaks if polar substituents are added. Since the refractive index η varies only slightly for most organic compounds, molar volume(MV) as usual is highly interrelated with MR

ACKNOWLEDGMENTS

Authors are highly thankful to Subash Basak for encouraging us to undertake the present investigation.

REFERENCES

[1] V.Viswanadhan, G.A.Mueller, S.K.Basak, J.N.Weinsteins; J.Chem.Inf.Comput.Sci., **41**, 505 **(2001)**.
[2] C.Hansch, A.Kurup, R.Garg, H.Gao; Chem.Rev., **101**, 619 **(2001)**.
[3] H.Wiener; J.Am.Chem.Soc., **69**, 17 **(1947)**.
[4] M.Randic; J.Am.Chem.Soc., **97**, 6609 **(1975)**.
[5] A.T.Balaban; Chem.Phys.Lett., **89**, 399 **(1982)**.
[6] P.V.Khadikar, N.V.Deshpande, P.P.Kale, A.Dobrynin, I.Gutman, G.Domotor; J.Chem.Inf.Compt.Sci., **35**, 547 **(1995)**.
[7] P.V.Khadikar, P.P.Kale, N.V.Deshpande, S.Karmarkar, V.K.Agrawal;Commun.Math.Comput.Chem. (MATCH), **43**, 7 **(2001)**.
[8] I.Gutman; Graph Theory Notes, New York, **27**, 9 **(1994)**.
[9] W.J.Dunn; Eur.J.Med.Chem., **12**, 109 **(1977)**.
[10] R.W.Toft; J.Am.Chem.Soci., **74**, 3120 **(1952)**.
[11] I.Motuc, A.T.Balaban; Rev.Roum.Chim., **27**, 735 **(1982)**.
[12] S.Chaterjee, A.S.Hadi; Price, B., Regression Analysis by Examples, 3rd Ed. Wiley VCH: New york, **(2000)**.
[13] H.Waterbeemd; Vande (Ed.) **2**, VCH, Weinleium (Ger), **(1995)**.
[14] A.T.Balaban; J.Chem.Inf.Comput.Sci., **32**, 23 **(1992)**.
[15] A.T.Balaban, T.S.Balaban; Math.Chem., **8**, 383 **(1991)**.