Pergamon

SCIENCE DIRECT®

Bioorganic & Medicinal Chemistry 11 (2003) 5519–5527

BIOORGANIC & MEDICINAL CHEMISTRY

QSAR Study on 5-Lipoxygenase Inhibitors Using Distance-Based Topological Indices

Vijay K. Agrawal,[a] Shahnaz Bano[a] and Padmakar V. Khadikar[b,*]

[a]*QSAR and Computer Chemical Laboratories, A.P.S. University, Rewa-486 003, India*
[b]*Research Division, Laxmi Fumigation and Pest Control Pvt. Ltd., 3 Khatipura, Indore-452 007, India*

Received 30 May 2003; accepted 12 September 2003

Abstract—QSAR study on a large set of 5-lipoxygenase inhibitors has been carried out using distance-based topological indices. Regression analysis of the data has indicated that an excellent model is obtained when these topological indices are combined with some classical molecular descriptors. The obtained models are critically discussed and examined.

Introduction

While working on modeling of anti-inflammatory activity, we have arrived at an excellent publication by Kim and coworkers[1] in that they have described the quantitative structure–activity relationship (QSAR) of 5-lipoxygenase inhibitors using a large set of physico-chemical parameters. In doing so, they have used 60 1-phenyl[2*H*]-tetrahydro-triazine-3-one (Fig. 1) analogues and examined their inhibitory potency of 5-lipoxygenase in a broken cell, IC_{50}. While converting IC_{50} in their reciprocal log units, that is $\log(1/IC_{50})$ they have proposed a multi-parametric model for a set of 60 compounds with the following statistics: $SE = 0.245$, $r = 0.863$ and $F = 14.3$. This model was a consequence of their analysis of individual sets of much lower population.

Our earlier study[2–5] has indicated that in such cases, as described above, better results are obtained by using a combination of topological indices with some classical molecular descriptors and/or physico-chemical parameters. This prompted us to reinvestigate the work of Kim and coworkers.[1] This was, therefore, our primary aim of the present investigation. We have used a large set of distance-based topological indices Wiener[6] (W), Branching (B),[7] first-order connectivity index[8–10] ($^1\chi$), Balaban[11,12] (J), Szeged index[13,14] (Sz) and logRB.[15] In addition, we have also used some of the physico-chemical parameters used by Kim and coworkers.[1] They include molar refractivity (MR) of 2- and 4- substituents (MR_2 and MR_4), lipophilic effects of 3′-, 5′- and 5-substituents ($\pi_{3'}$, $\pi_{5'}$ and $\pi_{5'}$), sum of Hammett σ_m value for the 3′- and 5′-phenyl or α-pyridyl substituents ($\Sigma\sigma_m$) and sum of Hammett σ_p value for the 3′- and 5′-α-pyridyl substituents ($\Sigma\sigma_p$). Our preliminary regression analysis has indicated that the use of some indictor parameters is needed for obtaining excellent results. Four indicator parameters Ip_1, Ip_2, Ip_3 and Ip_4 are therefore, used. The meaning of these indicator parameters are given in the experimental section. Thus, to start with, initially we have a large set of molecular descriptors for modeling the activity of 60 compounds.

Results and Discussion

Our results, as discussed below, establish that our objective is highly fulfilled. For a set of 60 compounds we obtained a model far superior than the one proposed by Kim and coworkers.[1]

The structural details of the compounds and their inhibitory potency ($\log 1/IC_{50}$) are given in Table 1. The calculated values of distance-based topological indices along with physico-chemical parameters used are summarized in Table 2. For the sake of convenience the indicator parameters are also given in Table 1.

For proposing statistically significant models, we have used the maximum R^2 method[16] and finally obtained

*Corresponding author. Tel.: +91-731-531906; e-mail: pvkhadikar@rediffmail.com

0968-0896/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2003.09.025

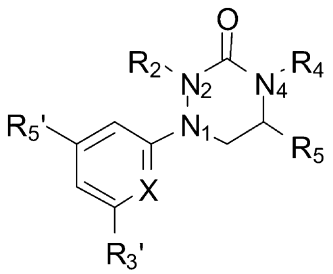

Figure 1. Common structures used in the present study (for details see Table 1).

statistically significant models. In the proposed models, n is the number of compounds used in the regression, SE is the standard error of estimation, R is the correlation coefficient, R^2_A is the adjusted R^2, and F is the F-statistics. In addition, the predictive potency of the models are established from the leave-one-out cross-validation analysis using various cross-validated parameters: PRESS (predicted residual sum of squares), SSY (sum of squares of response value), S_{PRESS} (uncertainty of prediction), R^2_{CV} (cross-validated correlation coefficient) and PSE (predictive square error).[16]

Before discussing the results, it is worth recording that the topological indices are numerical representations of molecular structure. They are obtained by transforming molecular structures in to their molecular graphs. Such a transformation is carried out by deleting all the carbon-hydrogen as well as hetero-atom-hydrogen bonds in the molecular structure. In chemical topology and graph theory, atoms are named as vertices while the bonds are called as edges. By imposing certain conditions on vertices, edges, or both, a number is obtained which is called the topological index. Such topological indices are successfully and widely used in modeling physico-chemical properties, biological activity and toxicity[17–23] of organic compounds.

The preliminary requirement to use the maximum-R^2 method is to first examine inter-correlatedness among the molecular descriptors used and their correlations with the activity to be modeled by regression analysis. This is done by correlation matrix. Such a matrix obtained in the present case is given in Table 3.

A perusal of Table 3 shows that all the topological indices, except J, are highly correlated. Thus, a model containing any combination of these indices may suffer from the defect due to collinearity. However, Randic[24] now nicely deals with such cases and we will use the recommendations of Randic[24] in discussing such models. The correlation matrix (Table 3) also shows that none of the molecular descriptors used are capable of modeling $\log 1/IC_{50}$ independently. This means that the only possibility of obtaining statistically significant models is to carry out step-wise multivariate regression analysis.

Initial multivariate regression analysis has indicated that meaningful regression models start coming when multi-parametric regressions with eight or more correlating parameters are used.

Among the several eight-parametric models, the one given below is found to be the best:

$\log 1/IC_{50} = 5.7242\times10^{-4}$ ($\pm 1.6440\times10^{-4}$) W -0.6957 (± 0.1504) $MR_2 + 0.8529$ (± 0.1409) $\pi_{3'} + 0.4773$ (± 0.0746) $\pi_{5'} + 0.3533$ (± 0.0530) $\pi_5 + 0.4473$ (± 0.1587) $\Sigma\sigma_m - 0.6256$ (± 0.0421) $\Sigma\sigma_p + 0.8187$ (± 0.1136) $Ip_3 + 4.4393$

$$n = 60,\ \ SE = 0.2316,\ \ R^2_A = 0.7247,\quad R = 0.8729,\ \ F = 20.414 \tag{1}$$

In the aforementioned model, out of the eight correlation parameters only MR_2 and $\Sigma\sigma_p$ have negative coefficients, while all other have positive coefficients. The positive coefficients $\pi_{5'}$, π_5 suggest that the R_5 substituent on the inhibitor may rest on surface or shallow trough of the enzymes.[1,23] Similarly, positive coefficients of $\Sigma\sigma_m$ suggest that lipophilic electron withdrawing substituents at meta position increase the activity. The same is true for other physico-chemical parameters with positive coefficients. Negative coefficient of MR_2 indicates that inhibitory potency decreases as the size of the R_2 substituent increases. In this regard the contribution of the Wiener index (W) is not that significant. However, presence of Wiener index (W) resulted into a better model than the model proposed by Kim.

In the model (eq 1) out of the four indicator parameters only Ip_3 with positive coefficients is present. This shows that the presence of C=O at R_2 is favorable for inhibitory activity.

Addition of the MR_2 term during the process of step-wise regression analysis resulted into the following nine parameter model having slightly improved quality than the model discussed above.

$\log 1/IC_{50} = 4.0849\times10^{-4}$ ($\pm 2.0780\times10^{-4}$) W -0.6496 (± 0.1537) $MR_2 + 0.0847$ (± 0.0683) $MR_4 + 0.8699$ (± 0.1407) $\pi_{3'} + 0.5021$ (± 0.0766) $\pi_{5'} + 0.3739$ (± 0.0551) $\pi_5 + 0.4751$ (± 0.1592) $\Sigma\sigma_m - 0.6786$ (± 0.4117) $\Sigma\sigma_p + 0.7406$ (± 0.1284) $Ip_3 + 4.4582$

$$n = 60,\ \ SE = 0.2302,\ \ R^2_A = 0.7281,\quad R = 0.8772,\ \ F = 18.551 \tag{2}$$

Except for MR_4, the nine parametric models has more or less similar statistics as that of the model expressed by eq 1. The physical significance of positive coefficients of MR_4 is in favour of optimum size of the R_4 substituents in accordance with Kim and coworkers.[1]

Successive regression analysis resulted into several ten parametric models and one given below is found to be the best.

Table 1. Activity of compounds and indicator parameters of various compounds used in present study

Compd	.X	$R_{3'}$	$R_{5'}$	R_2	R_4	R_5	Log1/IC_{50}	Ip_1	Ip_2	Ip_3	Ip_4
1	CH	H	H	H	H	CH_2OCH_2Ph	6.000	1	0	0	1
2	CH	H	H	H	H	Bu	5.823	1	0	0	1
3	CH	H	H	H	H	*i*-Pr	5.167	1	0	0	1
4	CH	H	H	H	H	Me (R)	5.167	1	0	0	1
5	CH	H	H	H	H	Me_2	5.167	1	0	0	1
6	CH	H	H	H	H	Et	5.161	1	0	0	1
7	CH	H	H	H	H	Me	4.939	1	0	0	1
8	CH	H	H	H	H	$CH_2OC_2H_4OMe$	4.853	1	0	0	1
9	CH	H	H	H	H	Me (S)	4.853	1	0	0	1
10	CH	H	H	H	H	CO_2Me	4.698	1	0	0	1
11	CH	H	H	H	H	H	4.677	1	0	0	0
12	CH	H	OCH_2 Ph	H	H	H	5.958	1	1	0	0
13	CH	H	Br	H	H	H	5.309	1	1	0	0
14	CH	H	Cl	H	H	H	5.197	1	1	0	0
15	CH	H	Et	H	H	H	4.886	1	1	0	0
16	CH	H	SMe	H	H	H	4.853	1	1	0	0
17	CH	H	Me	H	H	H	4.823	1	1	0	0
18	CH	H	CF_3	H	H	H	4.769	1	1	0	0
19	CH	H	F	H	H	H	4.721	1	1	0	0
20	CH	H	CN	H	H	H	4.431	1	1	0	0
21	CH	H	OMe	H	H	H	4.327	1	1	0	0
22	CH	H	NO_2	H	H	H	4.309	1	1	0	0
23	CH	H	NH_2	H	H	H	3.749	1	1	0	0
24	CH	H	Br	H	H	Me	5.585	1	1	0	1
25	CH	H	Cl	H	H	Me	5.568	1	1	0	1
26	CH	H	F	H	H	Me	5.200	1	1	0	1
27	CH	H	Me	H	H	Me	4.721	1	1	0	1
28	CH	H	H	H	C(=O)-*i*-Pr	H	5.886	1	0	1	0
29	CH	H	H	H	C(=O)Et	H	5.585	1	0	1	0
30	CH	H	H	H	C(=O)Me	Me	5.481	1	0	1	1
31	CH	H	H	H	C(=O)Me	H	5.468	1	0	1	0
32	CH	H	H	H	OCH_2Ph	Me	5.366	1	0	0	1
33	CH	H	H	H	OH	Me	5.221	1	0	0	1
34	CH	H	H	H	OEt	Me	5.130	1	0	0	1
35	CH	H	H	H	CH_2Ph	H	5.080	1	0	0	0
36	CH	H	H	C(=O)Et	C(=O) Et	H	4.903	1	0	1	0
37	CH	H	H	H	OMe	Me	4.649	1	0	0	1
38	CH	H	H	C(=O)Me	C(=O)Me	H	4.397	1	0	1	0
39	N	Br	H	H	H	Me	5.619	0	0	0	1
40	N	Br	H	H	H	H	5.455	0	0	0	0
41	N	Cl	H	H	H	Me	5.455	0	0	0	1
42	N	Me	H	H	H	Me	5.420	0	0	0	1
43	N	Me	H	H	H	H	5.259	0	0	0	0
44	N	OMe	H	H	H	Me	5.259	0	0	0	1
45	N	Cl	H	H	H	H	5.251	0	0	0	0
46	N	F	H	H	H	Me	5.180	0	0	0	1
47	N	F	H	H	H	H	5.040	0	0	0	0
48	N	OMe	H	H	H	H	5.022	0	0	0	0
49	N	H	H	H	H	Me	4.657	0	0	0	1
50	N	H	H	H	H	H	4.585	0	0	0	0
51	CH	H	Cl	H	C(=O)Me	H	5.886	1	1	1	0
52	CH	H	Cl	H	OH	Me	5.408	1	1	0	1
53	CH	H	F	H	OH	Me	5.161	1	1	0	1
54	CH	Me	Me	H	OH	H	5.075	1	1	0	0
55	CH	F	F	H	H	H	5.050	1	1	0	0
56	CH	Me	Me	H	H	H	4.920	1	1	0	0
57	N	Cl	H	H	H	H	5.481	0	0	0	0
58	CH	H	Cl	H	H	H	5.346	1	1	0	0
59	CH	H	H	H	H	H	4.769	1	0	0	0
60	CH	Cl	Me	H	H	H	5.481	1	1	0	0

$\log 1/IC_{50} = 0.0010 \quad (\pm 7.9403\times10^{-4}) \quad \log RB - 0.6349 \quad (\pm 0.1524) \quad MR_2 + 0.1269 \quad (\pm 0.0693) \quad MR_4 + 0.8734 \quad (\pm 0.1370) \quad \pi_{3'} + 0.5162 \quad (\pm 0.0730) \quad \pi_{5'} + 0.5785 \quad (\pm 0.1131) \quad \pi_5 + 0.4885 \quad (\pm 0.1552) \quad \Sigma\sigma_m - 0.7100 \quad (\pm 0.4014) \quad \Sigma\sigma_p + 0.6954 \quad (\pm 0.1256) \quad Ip_3 - 0.2767 \quad (\pm 0.1352) \quad Ip_4 + 4.5020$

$$n = 60, \ \mathrm{SE} = 0.2242, \ R_A^2 = 0.7421, \ R = 0.8865, \ F = 17.981 \tag{3}$$

The contribution of logRB in the above model is very little but its presence has improved the quality. Hence,

Table 2. Calculated values of molecular descriptor (W, B = $^1\chi$, *J*, *Sz* and logRB) and adopted MR_2, MR_4, $\pi_{3'}$, $\pi_{5'}$, π_5, $\Sigma\sigma_m$, $\Sigma\sigma_p$ for a series of triazinone analogues used in present study

Compd	W	B = $^1\chi$	J	Sz	LogRB	MR_2	MR_4	$\pi_{3'}$	$\pi_{5'}$	π_5	$\Sigma\sigma_m$	$\Sigma\sigma_p$
1	1198	10.8097	1.3972	1849	330.6496	0.103	0.103	0.00	0.00	2.288	0.00	0.00
2	542	8.292	1.8346	862	164.0439	0.103	0.103	0.00	0.00	2.570	0.0	0.00
3	432	7.6647	1.9224	720	134.7026	0.103	0.103	0.00	0.00	1.911	0.00	0.00
4	298	6.754	1.8822	522	93.9174	0.103	0.103	0.00	0.00	0.983	0.00	0.00
5	352	7.0673	1.9531	608	111.1012	0.103	0.103	0.00	0.00	1.512	0.00	0.00
6	364	7.292	1.8896	620	113.9634	0.103	0.103	0.00	0.00	1.512	0.00	0.00
7	298	6.754	1.8822	522	93.9174	0.103	0.103	0.00	0.00	0.983	0.00	0.00
8	788	9.292	1.7437	1172	227.0501	0.103	0.103	0.00	0.00	0.814	0.00	0.00
9	298	6.754	1.8822	522	93.9174	0.103	0.103	0.00	0.00	0.983	0.00	0.00
10	516	8.2027	1.9233	836	159.2077	0.103	0.103	0.00	0.00	0.717	0.00	0.00
11	246	6.3602	1.8327	438	77.4267	0.103	0.103	0.00	0.00	0.000	0.00	0.00
12	1036	10.3097	1.417	1662	290.9723	0.103	0.103	0.00	1.66	0.000	0.11	0.00
13	301	6.754	1.8615	527	94.477	0.103	0.103	0.00	0.86	0.000	0.39	0.00
14	301	6.754	1.8615	527	94.477	0.103	0.103	0.00	0.71	0.000	0.37	0.00
15	370	7.292	1.856	630	114.993	0.103	0.103	0.00	1.02	0.000	−0.07	0.00
16	370	7.292	1.856	630	114.993	0.103	0.103	0.00	0.61	0.000	0.15	0.00
17	301	6.754	1.8615	527	94.477	0.103	0.103	0.00	0.56	0.000	−0.07	0.00
18	514	7.9653	1.923	842	158.1045	0.103	0.103	0.00	0.88	0.000	0.43	0.00
19	301	6.754	1.8615	527	94.477	0.103	0.103	0.00	0.14	0.000	0.34	0.00
20	370	7.292	1.856	630	114.993	0.103	0.103	0.00	−0.57	0.000	0.56	0.00
21	370	7.292	1.856	630	114.993	0.103	0.103	0.00	−0.02	0.000	0.12	0.00
22	441	7.6647	1.879	735	136.2022	0.103	0.103	0.00	−0.28	0.000	0.71	0.00
23	301	6.754	1.8615	527	94.477	0.103	0.103	0.00	−1.23	0.000	−0.16	0.00
24	360	7.1479	1.9067	622	112.9136	0.103	0.103	0.00	0.86	0.983	0.39	0.00
25	360	7.1479	1.9067	622	112.9136	0.103	0.103	0.00	0.71	0.983	0.37	0.00
26	360	7.1479	1.9067	622	112.9136	0.103	0.103	0.00	0.14	0.983	0.34	0.00
27	360	7.1479	1.9067	622	112.9136	0.103	0.103	0.00	0.56	0.983	−0.07	0.00
28	602	8.5754	1.9461	954	184.4095	0.103	2.111	0.00	0.00	0.000	0.00	0.00
29	516	8.20271	1.9233	836	159.2077	0.103	1.530	0.00	0.00	0.000	0.00	0.00
30	510	8.0922	1.9493	848	157.1111	0.103	1.184	0.00	0.00	0.983	0.00	0.00
31	447	7.6815	1.8573	750	136.7492	0.103	1.184	0.00	0.00	0.000	0.00	0.00
32	1150	10.7372	1.4683	1849	322.4327	0.103	3.231	0.00	0.00	0.983	0.00	0.00
33	358	7.1815	1.9197	620	112.2486	0.103	0.2561	0.00	0.00	0.983	0.00	0.00
34	524	8.2196	1.9014	862	159.929	0.103	1.184	0.00	0.00	0.983	0.00	0.00
35	882	9.8265	1.4567	1486	252.3683	0.103	3.078	0.00	0.00	0.000	0.00	0.00
36	874	10.0958	2.1306	1346	266.949	1.648	1.530	0.00	0.00	0.000	0.00	0.00
37	433	7.7196	1.9222	733	134.3333	0.103	0.720	0.00	0.00	0.983	0.00	0.00
38	660	9.0197	2.1007	1056	204.9703	1.184	1.184	0.00	0.00	0.000	0.00	0.00
39	360	7.1479	1.9067	622	112.9136	0.103	0.103	0.86	0.00	0.983	0.39	0.23
40	301	6.754	1.8615	527	94.477	0.103	0.103	0.86	0.00	0.000	0.39	0.23
41	360	7.1479	1.9067	622	112.9136	0.103	0.103	0.71	0.00	0.983	0.37	0.23
42	360	7.1479	1.9067	622	112.9136	0.103	0.103	0.56	0.00	0.983	−0.07	−0.17
43	301	6.754	1.8615	527	94.477	0.103	0.103	0.56	0.00	0.000	−0.07	−0.17
44	437	7.6859	1.8997	737	135.509	0.103	0.103	−0.02	0.00	0.983	0.12	−0.27
45	301	6.754	1.8615	527	94.477	0.103	0.103	0.71	0.00	0.000	0.37	0.23
46	360	7.1479	1.9067	622	112.9136	0.103	0.103	0.14	0.00	0.983	0.34	0.06
47	301	6.754	1.8615	527	94.477	0.103	0.103	0.14	0.00	0.000	0.34	0.06
48	370	7.292	1.856	630	114.993	0.103	0.103	−0.02	0.00	0.000	0.12	−0.27
49	298	6.754	1.8822	522	93.9174	0.103	0.103	0.00	0.00	0.983	0.00	0.00
50	246	6.3602	1.8326	438	77.4267	0.103	0.103	0.00	0.00	0.000	0.00	0.00
51	366	7.7372	1.8391	798	119.6009	0.103	1.184	0.00	0.71	0.000	0.37	0.00
52	428	7.5754	1.9389	733	133.3242	0.103	0.256	0.00	0.71	0.983	0.37	0.00
53	428	7.5754	1.9389	733	133.3242	0.103	0.256	0.00	0.14	0.983	0.34	0.00
54	433	7.5586	1.9155	742	134.3051	0.103	0.256	0.56	0.56	0.000	−0.14	0.00
55	360	7.1479	1.9067	622	112.9136	0.103	0.103	0.14	0.14	0.000	0.68	0.00
56	360	7.1479	1.9067	622	112.9136	0.103	0.103	0.56	0.56	0.000	−0.14	0.00
57	301	6.754	1.8615	527	94.477	0.103	0.103	0.71	0.00	0.000	0.37	0.23
58	301	6.754	1.8615	527	94.477	0.103	0.103	0.00	0.71	0.000	0.37	0.00
59	246	6.3602	1.8326	438	77.4267	0.103	0.103	0.00	0.00	0.000	0.00	0.00
60	360	7.1479	1.9067	622	112.914	0.103	0.103	0.71	0.56	0.000	0.30	0.00

the improved quality of the model (eq 3) over the other models may be attributed to the presence of log RB and indicator parameters Ip_3 and Ip_4. It means that improvement is due to the presence of C=O grouping at R_4, absence of hydrogen at R_5 and branching.

Stepwise regression finally resulted into several 11-parametric models out of which the following model is found to be the best. Higher parametric correlations resulted into statistically poor models than this 11-parametric model. This further indicates that inhibitory activity can be modeled better by such 11-parameter models.

$$\log 1/IC_{50} = 2.8515\times10^{-4}\ (\pm 1.8793\times10^{-4})\ W - 0.6603\ (\pm 0.1341)\ MR_2 + 0.0929\ (\pm 0.0621)\ MR_4 + 0.7483$$

Table 3. Correlation matrix for the correlation of molecular descriptors and their correlation with log1/IC_{50}

	log1/IC_{50}	W	B	J	Sz	logRB	MR_2	MR_4	$\pi_{3'}$	$\pi_{5'}$	π_5	$\Sigma\sigma_m$	$\Sigma\sigma_p$	Ip_1	Ip_2	Ip_3	Ip_4
log1/IC_{50}	1.000																
W	0.289	1.000															
B	0.284	0.986	1.000														
J	−0.103	−0.122	−0.151	1.000													
Sz	0.316	0.985	0.978	−0.266	1.000												
logRB	0.282	0.998	0.992	−0.094	0.978	1.000											
MR_2	−0.175	0.307	0.368	−0.023	0.282	0.348	1.000										
MR_4	0.181	0.582	0.622	−0.057	0.603	0.598	0.266	1.000									
$\pi_{3'}$	0.252	−0.218	−0.244	−0.061	−0.201	−0.226	−0.085	−0.192	1.000								
$\pi_{5'}$	0.411	0.078	0.072	−0.056	0.106	0.074	−0.078	−0.1254	−0.037	1.000							
π_5	0.316	0.211	0.203	0.105	0.206	0.203	−0.148	−0.1081	−0.152	−0.152	1.000						
$\Sigma\sigma_m$	0.125	−0.219	−0.218	−0.089	−0.184	−0.223	−0.125	−0.237	0.176	0.159	−0.204	1.000					
$\Sigma\sigma_p$	0.128	−0.080	−0.103	−0.010	−0.076	−0.086	−0.013	−0.032	0.495	−0.031	−0.036	0.332	1.000				
Ip_1	−0.111	0.273	0.303	0.068	0.257	0.286	0.096	0.226	−0.576	0.223	0.052	−0.148	−0.139	1.000			
Ip_2	−0.156	−0.174	−0.173	−0.106	−0.132	−0.180	−0.149	−0.274	−0.120	0.520	−0.345	0.409	0.060	0.429	1.000		
Ip_3	0.214	0.228	0.323	−0.047	0.238	0.267	0.503	0.543	−0.170	−0.067	−0.214	−0.158	−0.027	0.191	−0.191	1.000	
Ip_4	0.216	0.114	0.113	−0.118	0.117	0.109	−0.165	−0.076	−0.128	−0.135	0.892	−0.163	−0.036	−0.012	−0.328	−0.224	1.000

(±0.1272) $\pi_{3'}$ + 0.6403 (±0.0754) $\pi_{5'}$ + 0.5321
(±0.1017) π_5 + 0.6685 (±0.1471) $\Sigma\sigma_m$ − 0.7830
(±0.3588) $\Sigma\sigma_p$ − 0.2792 (±0.0771) Ip_2 + 0.6501
(±0.1138) Ip_3 − 0.3088 (±0.1214) Ip_4 + 4.6535

$$n = 60,\ \ SE = 0.2007,\ \ R_A^2 = 0.7944,\ \ R = 0.9126,\ \ F = 21,728 \quad (4)$$

The improved statistics of this model (eq 4) may be attributed to the occurrence of Wiener index (W) and indicator parameters Ip_2, Ip_3 and Ip_4. Thus, the absence of hydrogen at R_5 and $R_{5'}$ along with presence of C=O group at R_4, size, shape and branching are favorable for the exhibition of activity. Other physical significance is similar to the earlier models.

Careful examination of above model (eq 4) has indicated six compounds (**5**, **18**, **22**, **33**, **37** and **38**) as outliers. Deletion of these compounds has further improved the quality of the model (eq 4) giving excellent statistics as below:

log1/IC_{50} = 4.2510×10^{-4} (±1.2756×10^{-4}) W − 0.5156
(±0.0997) MR_2 + 0.0732 (±0.0415) MR_4 + 0.7097
(±0.0859) $\pi_{3'}$ + 0.6150 (±0.0524) $\pi_{5'}$ + 0.5351
(±0.0679) π_5 + 0.8970 (±0.1059) $\Sigma\sigma_m$ − 0.9132
(±0.2396) $\Sigma\sigma_p$ − 0.2670 (±0.0518) Ip_2 + 0.6715
(±0.0762) Ip_3 − 0.3648 (±0.0821) Ip_4 + 4.5975

$$n = 54,\ \ SE = 0.1325,\ \ R_A^2 = 0.9060,\ \ R = 0.9620,\ \ F = 47.446 \quad (5)$$

At this stage it is interesting to compare our best model with the best model proposed by Kim and coworkers.[1] The first differentiation is that in the above discussed models we have used W, Ip_1, Ip_2, Ip_3 and Ip_4 which were not used by Kim and coworkers.[1] However, the remaining parameters were the same.

The statistics of the best model of Kim and coworkers[1] consisting of all the 60 compounds is as follows: $n = 60$, SE = 0.245, $r = 0.863$ and $F = 14.3$, while statistics of our best model containing 60 compounds is found as: $n = 60$, SE = 0.2007, $r = 0.9126$ and $F = 21.728$. Thus, our model is found to be far superior than of Kim and coworkers[1] model. Kim and coworkers[1] deleted one compound because its quality was different from the quality of other compounds. However, they have not given any statistical basis for the deletion of the compounds. Furthermore, deletion of this compound has not improved the statistics of their model. On the other hand, stepwise regression used by us has recommended for the deletion of six compounds (**5**, **18**, **22**, **33**, **37** and **38**) and when they are deleted there was a dramatic improvement in the quality of the model (eq 5). This further supports that our model models are far superior that the models proposed by Kim and coworkers.[1]

The aforementioned results can be further established by estimating and comparing predictive potential of Kim and coworkers[1] model with the models suggested by us. This can easily be done by estimating quality factor[25,26] Q. This quality factor Q is defined by ratio of the correlation coefficient (R) to the standard error of estimation (SE). That is $Q = R/SE$. This factor for the best model of Kim and coworkers[1] is found to be 3.5225 while that of our model is 4.5471. Furthermore, the quality factor of our improved model (after deleting six compounds) is found to be 7.2604. It means that the predictive potential of our ultimate model is 2.06 times greater than the Kim's model.

In order to best explain our result we have estimated cross validation parameters. As stated earlier[27–30] these parameters used being PRESS, SSY, S_{PRESS}, R_{CV}^2 and PSE. The meanings of these parameters are given in experimental section and their values are recorded in Table 4.

PRESS is a good estimate of the real prediction error of the model. If PRESS is smaller than SSY the model predicts better than chance and can be considered statistically significant. In this regard, all the five models proposed by us (Table 4) are better than chance and statistically significant.

The ratio PRESS/SSY can be used to calculate approximate confidence intervals of prediction of new compounds. To be a reasonable QSAR model this ratio should be smaller than 0.4, and the value of this ratio smaller than 0.1 indicates an excellent model. In our case, this ratio, for all the models proposed by us is found to be smaller than 0.4 and for the model as expressed by (eq 5) this ratio comes out to be 0.0805, that is <0.1. Thus, this model has excellent quality as well as highest predictive potential.

S_{PRESS} is another cross validation parameter and is a measure of uncertainty of prediction. However, in our case S_{PRESS} is found to be the same as that of SE, that is both these parameters carry the same meaning and use of S_{PRESS} is useless. However, in such cases we have still another cross validation parameter named as predictive square error (PSE). This parameter is more directly related to uncertainty of prediction. The lowest value of PSE (Table 4) for the model as expressed by (eq 5) supports its highest predictive potential.

Finally, the predictive potential of the model (Table 5) is confirmed by calculating predictive correlation coefficient of the model ($R^2_{pred.}$) (Figs 2 and 3) 0.8327 and 0.9255, respectively, for our models 4 and 5. Thus R^2_{pred} (0.9255) indicates that our improved model as expressed by (eq 5) with deletion of six compounds is the best.

Conclusion

From the results and discussion made above, we conclude that combination of topological indices with the parameters chosen by of Kim's results into models with much improved quality and predictability. Also, that out of the several topological indices used only W and logRB are found better to be used. Furthermore, deletion of these topological indices from the model lower their quality significantly. Other benefits of our models are the same as that of Kim's models.

Experimental

Inhibitory activity

The inhibitory activity in terms of log $1/IC_{50}$ is adopted from the work of Kim and coworkers.[1]

Molecular graphs

The hydrogen suppressed molecular graphs[31] were used for the calculation of topological indices W, Sz, $^1\chi$, B, J and logRB (Table 2).

Topological indices

The details for the calculations of Wiener (W)-,[6] branching index (B)-,[7] first-order connectivity index ($^1\chi$)-,[8–10] Balaban (J),[11,12] Szeged (Sz)-,[13,14] logRB-,[15] quality factor (Q)[25,26] and maximum R^2 improvement method[16] are given in our earlier communications and they are thus not repeated here.

Physico-chemical parameters

All the physico-chemical parameters used (Table 2) are adopted from the work of Kim and coworkers.[1]

Indicator parameters

We have used four indicator parameters Ip_1, Ip_2, Ip_3 and Ip_4 with the following meaning: Ip_1 is 1 when phenyl group is attached with N, Ip_2 and Ip_4 are 1 when hydrogen is not present at R_4 and $R_{5'}$. When C=O is present at R_4 the indicator parameter is Ip_3 having the

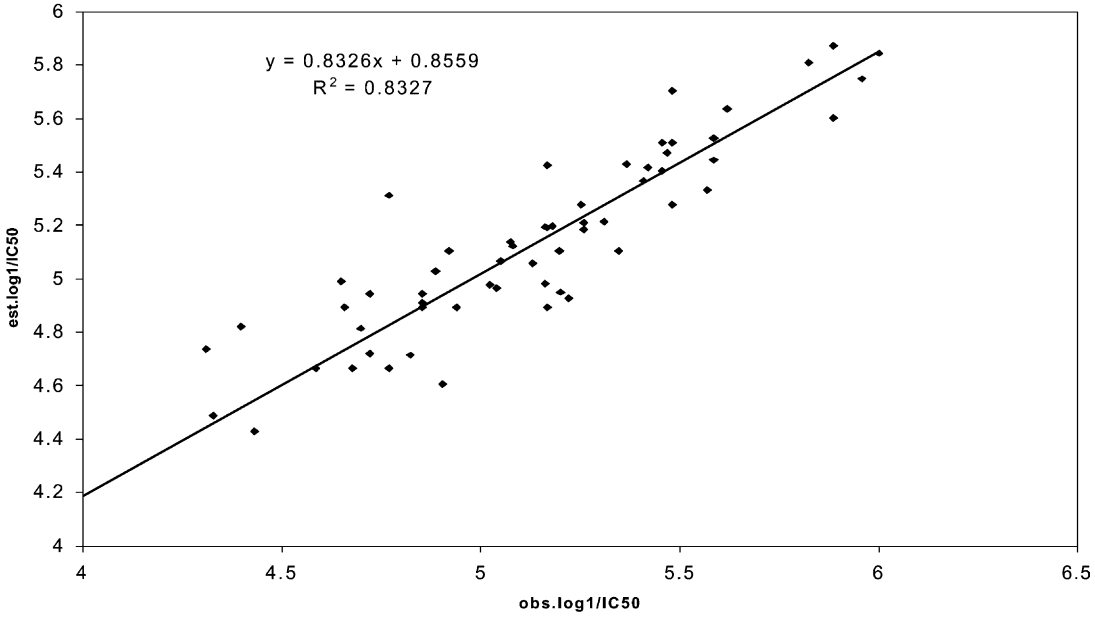

Figure 2. Correlation of observed and estimated activity of 60 compounds using model 4 (eq 4).

Table 4. Observed and estimated log1/IC_{50} from models 4 and 5

Compd	Obs. log1/IC_{50}	Model 4 (eq 4)		Model 5 (eq 5)	
		Est. log1/IC_{50}	Res.	Est. log1/IC_{50}	Res.
1	6.000	5.845	0.155	5.921	0.079
2	5.823	5.808	0.015	5.793	0.030
3	5.167	5.426	−0.259	5.393	−0.226
4	5.167	4.894	0.273	5.146	0.021
5	5.167	5.191	−0.024	—	—
6	5.161	5.195	−0.034	5.151	0.010
7	4.939	4.894	0.045	4.840	0.099
8	4.853	4.944	−0.091	4.958	−0.105
9	4.853	4.894	−0.041	4.840	0.013
10	4.698	4.815	−0.117	4.790	−0.092
11	4.677	4.665	0.012	4.657	0.020
12	5.958	5.748	0.210	5.845	0.113
13	5.309	5.213	0.096	5.292	0.017
14	5.197	5.104	0.093	5.181	0.016
15	4.886	5.028	−0.142	5.007	−0.121
16	4.853	4.912	−0.059	4.952	−0.099
17	4.823	4.714	0.109	4.695	0.128
18	4.769	5.313	−0.544	—	—
19	4.721	4.719	0.002	4.804	−0.083
20	4.431	4.431	0.000	4.594	−0.163
21	4.327	4.489	−0.162	4.538	−0.211
22	4.309	4.737	−0.428	—	—
23	3.749	3.507	0.242	3.513	0.236
24	5.585	5.444	0.141	5.478	0.107
25	5.568	5.333	0.235	5.368	0.200
26	5.200	4.950	0.250	4.990	0.210
27	4.721	4.945	−0.224	4.881	−0.160
28	5.886	5.603	0.283	5.626	0.260
29	5.585	5.525	0.060	5.547	0.038
30	5.481	5.705	−0.224	5.680	−0.199
31	5.468	5.473	−0.005	5.492	−0.024
32	5.366	5.428	−0.062	5.431	−0.065
33	5.221	4.926	0.295	—	—
34	5.130	5.059	0.071	5.015	0.115
35	5.080	5.123	−0.043	5.145	−0.065
36	4.903	4.607	0.296	4.903	0.000
37	4.649	4.990	−0.341	—	—
38	4.397	4.820	−0.423	—	—
39	5.619	5.636	−0.017	5.616	0.003
40	5.455	5.405	0.05	5.430	0.025
41	5.455	5.511	−0.056	5.492	−0.037
42	5.420	5.417	0.003	5.356	0.064
43	5.259	5.186	0.073	5.170	0.089
44	5.259	5.211	0.048	5.239	0.020
45	5.251	5.279	−0.028	5.306	−0.055
46	5.180	5.197	−0.017	5.216	−0.036
47	5.040	4.966	0.074	5.029	0.011
48	5.022	4.977	0.045	5.049	−0.027
49	4.657	4.894	−0.237	4.840	−0.183
50	4.585	4.665	−0.08	4.657	−0.072
51	5.886	5.873	0.013	5.960	−0.074
52	5.408	5.368	0.04	5.408	0.000
53	5.161	4.983	0.178	5.030	0.131
54	5.075	5.138	−0.063	5.096	−0.021
55	5.050	5.068	−0.018	5.233	−0.183
56	4.92	5.103	−0.183	5.054	−0.134
57	5.481	5.279	0.202	5.306	0.175
58	5.346	5.104	0.242	5.181	0.165
59	4.769	4.665	0.104	4.657	0.112
60	5.481	5.509	−0.028	5.555	−0.074

value of unity. In all other cases, if such situations does not exist, then the corresponding values of indicator parameters are taken to be zero.

Cross validation

It is worth mentioning that the models having the best correlation potential need not have the best predictive value too. As opposed to traditional regression methods, the cross-validation evaluates the validity of a model by how well it predicts data rather than how well it fits data. The analysis uses a 'leave-one-out' scheme, a model is built with $N-1$ compounds and the Nth compound is predicted. Each compound is left out of the

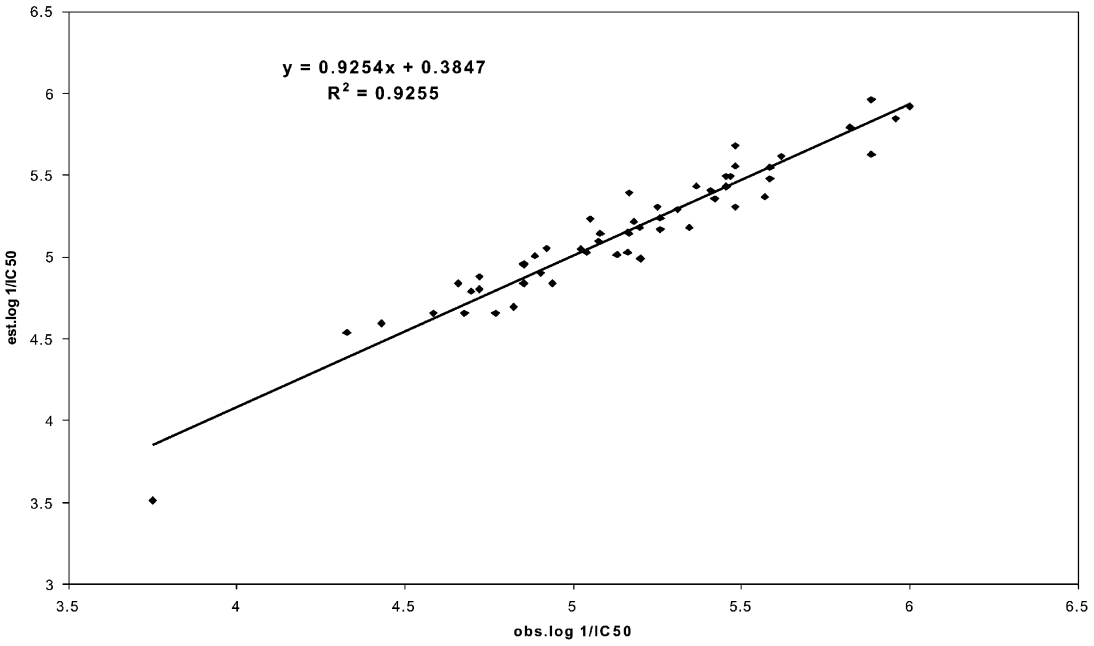

Figure 3. Correlation of observed and estimated activity of 60 compounds using model 5 (eq 5).

Table 5. Cross-validated parameters for the proposed models

Model	Number of parameters	PRESS	SSY	PRESS/SSY	R^2_{CV}	S_{PRESS}	PSE
1 (1)	8	2.7366	8.7632	0.3123	0.6877	0.2316	0.2136
2 (2)	9	2.6502	8.8496	0.2997	0.7005	0.2302	0.2102
3 (3)	10	2.4627	9.0371	0.2725	0.7245	0.2242	0.2026
4 (4)	11	1.9233	9.5765	0.2008	0.7992	0.2002	0.1790
5 (5)	11	0.7373	9.1622	0.0805	0.9195	0.1239	0.1109

model derivation and predicted in turn. As indication of the performance of the model is obtained from the cross-validated (or predictive) r^2_{CV} which is defined as:

$$r^2_{CV} = \frac{\text{SD-PRESS}}{\text{SD}}$$

where SD is the sum-of-squares deviation for each activity from the mean. PRESS (or predictive sum-of-squares) is the sum of squared difference between the actual and that predicted when the compound is omitted from the fitting process. Once a model is developed which has highest cross-validated r^2_{cv}, this method is used to derive the conventional QSAR equation and conventional r^2 and s values. The results of the final model are often visualized as contour maps of the coefficient.

In addition to PRESS, SD, r^2_{cv}, S_{PRESS}, one also need to evaluate predictive-square-error (PSE) in an attempt to decide the predictive potential of the proposed models. The data of calculation of cross-validated parameters are given in our publications.[29–32]

Computations

All the computations were carried out in Power Macintosh 9600/233.

Acknowledgements

Authors are thankful to Professor Istvan Lukovits, Hungarian Academy of Sciences, Budapest, Hungary for providing software to carryout regression analysis and to Professor Ivan Gutman, Faculty of Science, University of Kragujevac, Yugoslavia for introducing one of the authors (P.V.K.) to this fascinating field of chemical graph theory and topology. Authors (V.K.A., S.B.) are also thankful to CSIR, New Delhi, India for sanctioning a research scheme.

References and Notes

1. Kim, K. H.; Martin, Y. C.; Brooks, D. W. *Quant. Struct-Act. Relat.* **1996**, *15*, 491.
2. Agrawal, V. K.; Joseph, S.; Khadikar, P. V.; Karmarkar, S. *Acta Pharm.* **2000**, *50*, 81.
3. Agrawal, V. K.; Karmarkar, S.; Khadikar, P. V.; Srivastava, S.; Lukovits, I. *Indian J. Chem.* **2003**, *42A*, 1426.
4. Agrawal, V. K.; Joseph, S.; Khadikar, P. V.; Karmarkar, S. *Nat. Acad. Sci. Lett.* **2000**, *23*, 57.
5. Agrawal, V. K.; Bano, S.; Khadikar, P. V. *Bioorg. Med. Chem.* **2003**, *13*, 4039.
6. Wiener, H. *J. Am. Chem. Soc.* **1947**, *69*, 17.
7. Devillers, J.; Balaban, A. T. *Topological Indices and Related Descriptors in QSAR and QSPR*; Gorden & Breach: Williston, VT, 2000; Vol. 40, p 245.
8. Randic, M. *J. Am. Chem. Soc.* **1975**, *97*, 6609.

9. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Relationship*; Wiley: New York, 1986.
10. Kier, L. B.; Hall, L. H. *Molecular Structure Description*; Academic: New York, 1999.
11. Balaban, A. T. *Chem. Phys. Lett.* **1982**, *89*, 399.
12. Khadikar, P. V.; Sharma, S.; Sharma, V.; Joshi, S.; Lukovits, I.; Kaveeshwar, M. *Bull. Soc. Chem. Belg.* **1997**, *106*, 767.
13. Gutman, I. *Graph Theory Notes New York* **1994**, *27*, 9.
14. Khadikar, P. V.; Deshpandey, N. V.; Kale, P. P.; Dubrynin, A.; Gutman, I.; Domotor, G. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 547.
15. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, Germany, 2000.
16. Chatterjee, S.; Hadi, A. S.; Price, B. *Regression Analysis by Examples*, 3rd ed.; Wiley: New York, 2000.
17. Karcher, I. N.; Devillers, J. *Practical Applications of Quantitative Structure–Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Kluwer Academic: Dordrecht, 1990.
18. Agrawal, V. K.; Srivastava, R.; Khadikar, P. V. *Bioorg. Med. Chem.* **2001**, *9*, 3287.
19. Agrawal, V. K.; Singh, J.; Khadikar, P. V. *Bioorg. Med. Chem.* **2002**, *10*, 3981.
20. Kier, L. B.; Hall, L. H. *Molecular Structure Description*; Academic: New York, 1999.
21. Diudea, M. V.; Ivancine, O. *Molecular Topology*; Comprex: Cluj, 1995.
22. Diudea, M. V., Ed. *QSPR/QSAR Studies by Molecular Descriptors*; Babes-Bolyai University: Cluj, Romania, 2000.
23. Karelson, M. *Molecular Descriptors in QSAR/QSPR*; J. Wiley & Sons: New York, 2000.
24. Randic, M. *Croat. Chem. Acta* **1993**, *66*, 289.
25. Hansch, C.; Klein, T. *Acc. Chem. Res.* **1986**, *19*, 392.
26. Pogliani, L. *Amino Acids* **1994**, *6*, 141.
27. Pogliani, L. *J. Phys. Chem.* **1996**, *100*, 18065.
28. Trinajstic, N. *Chemical Graph Theory*, 2nd revised ed.; CRC: Boca Raton, FL, 1992.
29. Agrawal, V. K.; Sohgaura, R.; Khadikar, P. V. *Bioorg. Med. Chem.* **2001**, *9*, 3295.
30. Khadikar, P. V.; Agrawal, V. K.; Karmarkar, S. *Bioorg. Med. Chem.* **2002**, *10*, 3499.
31. Khadikar, P. V.; Lukovits, I.; Agrawal, V. K.; Shrivastava, S.; Jaiswal, M.; Gutman, I; Karmarkar, S.; Srivastava, A. *Indian J. Chem.* **2003**, *42A*, 1436.
32. Agrawal, V. K.; Sharma, R.; Khadikar, P. V. *Bioorg. Med. Chem.* **2002**, *10*, 2993.