Pergamon

Bioorganic & Medicinal Chemistry 10 (2002) 2993–2999

BIOORGANIC & MEDICINAL CHEMISTRY

QSAR Studies on Carbonic Anhydrase Inhibitors: A Case of Ureido and Thioureido Derivatives of Aromatic/Heterocyclic Sulfonamides

Vijay K. Agrawal,[a] Ruchi Sharma[a] and Padmakar V. Khadikar[b,*]

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

Received 4 January 2002; accepted 4 April 2002

Abstract—QSAR studies on modelling of biological activity (hCAI) for a series of ureido and thioureido derivatives of aromatic/heterocyclic sulfonamides have been made using a pool of topological indices. Regression analysis of the data showed that excellent results were obtained in multiparametric correlations upon introduction of indicator parameters. The predictive abilities of the models are discussed using cross-validation parameters. © 2002 Published by Elsevier Science Ltd.

Introduction

Supuran et al.[1–3] have synthesized a series of substituted ureido and thioureido derivatives of aromatic/heterocyclic sulfonamides with increased affinities for isozyme-I. These sulfonamides were characterized by them using standard physico-chemical procedures such as ^{1}H, NMR, ^{13}C NMR, IR (KBr) spectroscopy, and assayed as inhibitors of three isoenzymes of carbonyl anhydrase (CA), that is hCAI, hCAII, and bCAIV (h = human, b = borine isozyme). In addition, Supuran et al.[3] have also synthesized another new series of compounds, 1,5-disubstituted-2-thiobiuret derivatives. Appreciable inhibition of all these three CA isoenzymes was observed with the new compounds. They have predicted that some of the new compounds might constitute good lead molecules for developing more selective CAI inhibitors. However, no quantitative structure–activity relationship (QSAR) study was made using topological indices for modelling, monitoring, and estimating CAI inhibition.

This was, therefore, the primary aim of our present investigation. Another objective of the present investigation is to investigate relative potential of Wiener (W) index[4] and Szeged (Sz) index[5,6] in developing statistically significant QSAR models for modelling CAI inhibitors and is, therefore, an extension of our earlier study.[7]

Topological indices are numbers associated with chemical structure for the purpose of quantitative structure–property–activity (QSPR, QSAR) relationship studies. Their use in drug design has gained acceptance during the last few years.[8–11]

For achieving the aforementioned objective of the present study we have used a pool of topological indices, namely Wiener index[4] (W), Szeged index[5,6] (Sz), zero- ($^{0}\chi^{v}$) and first-order ($^{1}\chi^{v}$) connectivity indices,[12,13] Balaban index[14] (J) and Branching index[15] (B).

In calculating these topological indices, we have used carbon–hydrogen suppressed molecular graphs, as was done in our earlier publications also. In our recent publications[16] we have established that Sz is successful in case of monocyclic compounds containing acyclic side chains. Further studies[17–24] have shown that the predictive potential of both W and Sz is increased by their combination with another topological indices including indicator parameters.

In view of the above, we have used W, Sz, $^{0}\chi^{v}$, $^{1}\chi^{v}$, B and J for modelling CAI inhibitors (Table 1) by adopting earlier reported[1] biological activity data of sulfonamide (i.e., inhibition constant, expressed as pK_i) as shown in Table 2.

*Corresponding author. Tel.: +91-731-531906; fax: +91-731-531906; e-mail: vijay-agrawal@lycos.com

0968-0896/02/$ - see front matter © 2002 Published by Elsevier Science Ltd.
PII: S0968-0896(02)00123-2

Results and Discussion

It is interesting to mention that the topological indices *W* and *Sz* accounts for the size, shape, and branching in drug molecules. Topological index *B* precisely takes care of branching. The first-order valence connectivity index ($^1\chi^v$) distinguish the degree of unsaturation and the presence of hetero-atom.[12,13] Balaban index[14] (*J*) is a highly discriminating index, whose value do not substantially increase with the molecular size and the number of rings present. These physical significances associated with the used topological indices will help us in interpreting the proposed QSAR models more precisely.

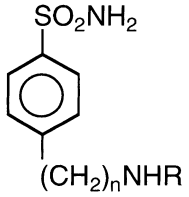
The set of sulfonamides used are shown in Table 1. Record that these ureido/thioureido moieties possess bulkier groups as substituents at the N-1 atom. Also, that from the original list of 39 compounds we have deleted compounds **13** and **41** as serious outliers and used only 39 compounds in our study.

The correlation terms involved in the correlation matrix (Table 3), therefore, indicate the extent of collinearity. This term close to one indicates high collinearity, while the value below 0.5 indicates that no collinearity exists between the two parameters. The perusal of correlation matrix (Table 3) indicates collinearity between: (i) $^0\chi^v$ and each of the other topological indices: $^1\chi^v$, *W*, *Sz* and *B* but *W* does not show any collinearity with *J*; (ii) $^1\chi v$ also correlates with each of the topological indices: *W*, *Sz*, *B*. Here also *W* do not show any collinearity with *J*; (iii) both *W* and *Sz* also exhibit similar collinearity with $^0\chi^v$, $^1\chi^v$ and *B* but not with *J*. None of these topological indices shows any significant collinearity with CAI activity. This shows that no mono-parametric model is statistically (significant) possible for modelling CAI activity of the compounds under present study.

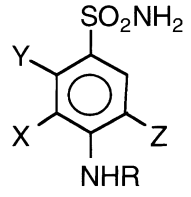
The aforementioned data also show that statistically significant results are only obtained in multi-parametric models and that any such model(s) containing *W*, *Sz*, *B*, $^0\chi^v$ and $^1\chi^v$ as correlating parameters may suffer from the defect of collinearity. Furthermore, J index appears to be the most appropriate topological index to be combined with the earlier mentioned topological indices in obtaining multi-parametric regressions. We have, therefore, carried out step-wise multiple regression analysis with a aim of obtaining most appropriate model for modelling CAI activity of the compounds under present study (Tables 1 and 2).

The step-wise regression analysis[25] indicated possibility of several statistically significant multi-parametric models for modelling the CAI activity. However, bi-, tri-,

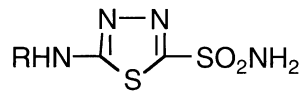
Table 1. Structure of the substituted ureido and thioureido-derivatives of aromatic/hetrocyclic sulfonamides used in the present study



Compd	*n*	R
1	0	H
2	1	H
3	2	H
4	0	H
10	0	CONHPh
11	1	CONHPh
12	2	CONHPh
18	0	CONH-3,4 $C_6H_3Cl_2$
19	1	CONH-3,4 $C_6H_3Cl_2$
20	2	CONH-3,4 $C_6H_3Cl_2$
21	0	CONH-3,4 $C_6H_3Cl_2$
27	0	CSNHPh
28	1	CSNHPh
29	2	CSNHPh
35	0	CSNHAl
36	2	CSNHAl
37	0	CSNHAl



Compd	X	Y	Z	R
5	H	H	SO_2 NH_2	H
6	H	Cl	SO_2 NH_2	H
7	Cl	Cl	SO_2 NH_2	H
8	Cl	—	Cl	H
13	H	H	SO_2 NH_2	CONHPh
14	H	Cl	SO_2 NH_2	CONHPh
15	Cl	Cl	SO_2 NH_2	CONHPh
16	Cl	—	Cl	CONHPh
22	H	H	SO_2 NH_2	CONH-3,4 $C_6H_3Cl_2$
23	H	Cl	SO_2 NH_2	CONH-3,4 $C_6H_3Cl_2$
24	Cl	Cl	SO_2 NH_2	CONH-3,4 $C_6H_3Cl_2$
25	Cl	—	Cl	CONH-3,4 $C_6H_3Cl_2$
30	H	H	SO_2 NH_2	CSNHPh
31	H	Cl	SO_2 NH_2	CSNHPh
32	Cl	Cl	SO_2 NH_2	CSNHPh
33	Cl	—	Cl	CONHPh
38	H	Cl	SO_2 NH_2	CSNHAl
39	Cl	Cl	SO_2 NH_2	CSNHAl



Compd	R
9	H
17	CONHPh
26	CONH-3,4 $C_6H_3Cl_2$
34	CSNHPh

and tetra-parametric regressions resulted into statistically less significant QSAR models. Better results are only obtained in penta-parametric and higher order models. Furthermore, these higher parametric regressions indicated compounds **13** and **41** as outliers. This indicates that these two compounds might have different types of mechanism as compared to remaining 39 compounds. When these two compounds are omitted from the data set then the quality of regressions are improved dramatically. Consequently, Table 4 records the results obtained in penta- and higher parametric regressions in that the compounds **13** and **41** are omitted.

Table 2. The values of pK_i, topological indices: $^0\chi^v$, $^1\chi^v$, *W*, *Sz*, *B*, *J*, and indicator parameters: Ip_1 and Ip_2 for substituted ureido and thioureido derivatives of aromatic/hetrocyclic sulfonamides

Compd	pK_i	$^0\chi^v$	$^1\chi^v$	*W*	*Sz*	*B*	*J*	Ip_1	Ip_2
1	3.4472	6.3808	1.5650	372	534	6.8531	2.5013	1	0
2	3.3979	6.3960	3.1300	465	654	7.3362	2.4259	1	0
3	3.3222	7.1020	2.9690	574	790	7.8362	2.3523	1	0
4	3.3979	5.6890	2.8840	354	498	6.8531	2.6317	0	0
5	3.2304	7.4130	4.2350	848	1154	9.4185	3.0378	0	0
6	2.9242	7.7130	3.6800	393	1284	9.8292	3.1602	0	0
7	2.9542	8.0150	3.7930	1030	1412	10.2567	3.2996	0	0
8	3.2304	6.2906	3.0951	490	702	7.6744	2.7842	0	0
9	2.9345	4.6830	2.1240	297	338	6.3531	2.5718	0	0
10	3.0792	10.4030	5.7060	1446	2040	11.2469	1.8598	0	1
11	2.9031	11.1400	7.8950	1690	2320	11.7301	1.7819	0	1
12	2.6335	11.8300	7.8950	1959	2625	12.2301	1.7152	0	1
13	2.7924	12.1577	6.4058	2351	3197	13.8123	2.2296	0	1
14	2.9542	12.4320	6.5048	2512	3424	14.2230	2.3014	0	1
15	2.8129	12.7330	6.6198	2664	3631	14.6504	2.3891	0	1
16	2.553	11.0080	5.9170	1600	2366	12.0683	2.0316	0	0
17	0.6990	9.4001	4.9408	1247	1540	10.7469	1.9186	0	0
18	1.3010	11.0080	5.9113	1811	2568	12.0515	1.8786	0	0
19	1.0792	11.7152	6.3684	2094	2892	12.5346	1.8043	0	0
20	1.0792	12.4220	6.8684	2404	3243	13.0346	1.7397	0	0
21	1.000	11.0080	5.9113	1727	2400	12.0515	1.9707	0	0
22	1.6232	12.7330	6.6112	2844	3895	14.6168	2.2249	0	0
23	1.6721	13.0312	6.7134	3026	4152	15.0275	2.2925	0	0
24	1.6902	12.9255	6.8217	3197	4385	15.4550	2.3754	0	0
25	1.2553	11.6095	6.1226	2089	2946	12.8728	2.0363	0	0
26	0.4771	10.0016	5.1453	1579	1992	11.5515	1.9265	0	0
27	2.6128	10.4070	5.7060	1446	2040	11.2469	1.8598	0	1
28	2.4249	11.1139	6.5511	1690	2320	11.7301	1.7819	0	1
29	1.6990	11.8210	6.6631	1959	2625	12.2301	1.7152	0	1
30	2.000	12.1316	6.3958	2351	3197	13.8123	2.2296	0	0
31	1.7404	12.4323	6.0253	2512	3424	14.2230	2.3014	0	0
32	1.7782	12.7330	7.6369	2664	3631	14.6505	2.3891	0	0
33	1.2304	11.0080	5.9139	1686	2366	12.0683	2.0316	0	0
34	0.6021	9.4003	4.9440	1247	1540	10.7469	1.9109	0	0
35	2.6021	9.0114	4.7188	1006	1330	9.7125	2.5234	0	0
36	1.9777	10.4259	5.6759	1408	1786	10.6956	2.3367	0	0
37	2.5563	9.0115	4.7187	952	1222	9.7125	2.6718	0	0
38	2.3424	10.7362	5.4188	1746	2250	12.2778	3.1331	0	0
39	1.8573	11.0370	5.5176	1880	2435	12.6885	3.2497	0	0

K_i, inhibition constant; pK_i, logarithmic transformation of K_i; $^0\chi^v$, zero-order connectivity index; $^1\chi^v$, first-order connectivity indices; *W*, Wiener index; *Sz*, Szeged index; *B*, branching index; *J*, Balaban index; Ip_1 and Ip_2, indicator parameters.

A perusal of Table 4 shows that out of the two penta-parametric models, the one containing *W*, *Sz*, *J*, Ip_1 and Ip_2 gave better results. This model is found as under:

$$\begin{aligned} \mathrm{p}K_i = {} & -0.0054(\pm 0.0012)W + 0.0037\left(\pm 8.9737\times 10^{-4}\right)Sz \\ & + 1.2519(\pm 0.1542)J + 0.8344(\pm 0.2559)Ip_1 \\ & + 1.2944(\pm 0.1577)Ip_2 - 0.3717 \\ & n = 39,\ \mathrm{SE} = 0.3725,\ R = 0.9171,\ Q = 2.4620, \\ & F = 34.917 \end{aligned} \tag{1}$$

Eq (1) shows that *Sz*, *J*, Ip_1 and Ip_2 have positive contribution in the exhibition of CAI activity, that is pK_i. Also, that *J* and Ip_2 have dominant role in exhibiting the activity. The negative sign of the coefficient of *W* compared to *Sz* may be due to high collinearity between *W* and *Sz* and/or that *Sz* contain some other additional information than the *W* index. *W* index is known to account of shape and size of the molecules. Also, that *W* and *Sz* are highly linearly correlated. It means that in addition to shape and size, *Sz* contains some other information which resulted into its positive role in exhibiting the CAI activity. Research in this direction is underway and will be published elsewhere.

The data presented in Table 4 also indicates a statistically significant hexa-parametric model. This model contains *W*, *Sz*, *B*, *J*, Ip_1 and Ip_2 as the correlating parameters and is found as under:

$$\begin{aligned} \mathrm{p}K_i = {} & -0.0037(\pm 0.0012)W + 0.0034\left(\pm 7.8340\times 10^{-4}\right)Sz \\ & - 0.4551(\pm 0.1326)B + 1.3288(\pm 0.1358)J \\ & + 0.5245(\pm 0.2399)\mathrm{I}p_1 + 1.3779(\pm 0.11390)Ip_2 - 2.4643 \\ & n = 39,\ \mathrm{SE} = 0.3235,\ R = 0.9401,\ Q = 2.9067, \\ & F = 40.559 \end{aligned} \tag{2}$$

Table 3. Correlation matrix for the inter-correlation of structural descriptors and their correlation with the activity

	pK_i	$^0\chi^v$	$^1\chi^v$	*W*	*Sz*	*B*	*J*	Ip_1	Ip_2
pK_i	1.0000								
$^0\chi^v$	−0.5312	1.0000							
$^1\chi^v$	−0.4642	0.9369	1.0000						
W	−0.4877	0.9561	0.8619	1.0000					
Sz	−0.4655	0.9527	0.8589	0.9979	1.0000				
B	−0.5104	0.9708	0.8824	0.9831	0.9790	1.0000			
J	0.4689	−0.4521	−0.5307	−0.3290	−0.3372	−0.2997	1.0000		
Ip_1	0.4014	−0.4582	−0.5400	−0.4382	−0.4259	−0.4986	0.0877	1.0000	
Ip_2	0.2949	0.3247	0.4392	0.2270	0.2343	0.2455	−0.4101	−0.1581	1.0000

K_i, inhibition constant; pK_i, logarithmic transformation of K_i; $^0\chi^v$, zero-order connectivity index; $^1\chi^v$, first-order connectivity indices; *W*, Wiener index; *Sz*, Szeged index; *B*, branching index; *J*, Balaban index; Ip_1 and Ip_2, indicator parameters.

This model contains B index in addition to W, Sz, J, Ip_1 and Ip_2. This means that the improved quality of this hexa-parametric model is due to the addition of B index, that is the branching index. The negative sign associated with this index indicates that CAI activity (pK_i) increases with decrease in the magnitude of B. This model again indicates that except for the change in sign the contribution of W and Sz in exhibiting CAI activity (pK_i) is similar. Such difference in the sign of the coefficients of W and Sz in this model also may be attributed to the same reasons mentioned above.

Looking to the size of the sample (39 compounds) and the number of molecular descriptors used (8) it is accepted scientifically that still higher parametric regressions may result into a better model than the hexa-parametric model discussed above. While doing that it was observed that only hepta-parametric model containing W, Sz, J, ${}^0\chi^v$, ${}^1\chi^v$, Ip_1 and Ip_2 gave statistically significant results. However, this hepta-parametric model shows decrease in the value of correlation coefficient (0.9338) as compared to the hexa-parametric model discussed above ($R=0.9401$). But, the quality of correlation is not decided alone by R-values. In addition to R, one has to consider the standard error of estimation (SE) in deciding the quality of the model. A quantity named as quality factor (Q) was reported[26] in that the role of R and SE is simultaneously considered in deciding the quality of the model. This quality factor, Q, is defined as the ratio of correlation coefficient (R) to the SE, that is $Q=R/SE$. Thus, the higher the value of R, the lower the SE, the larger will be Q, and better will be the quality of the model.

The Q values calculated for the models discussed above are recorded in Table 4 and shows that the hexa-parametric model is the most appropriate model for modelling CAI activities of the compounds under present study.

Additional support in favor of our results is received from adjustable R^2 which is defined as under.

$$R^2_A = 1-\frac{(n-1)(1-R^2)}{(n-k)}$$

here the meanings of the terms involved are given below.

The adjustable R^2, namely R^2_A goes on increasing as we pass from penta- to hexa-parametric correlation. This consistent increase in R^2_A indicates that in spite of observed collinearity the proposed models are statistically significant. If a variable is added that does not contribute its fair share, the R^2_A will actually decrease R^2_A is a measure of % explained variance in the dependent variable that takes into account the relationship between the number of cases and the number of independent variables in the regression model.

In an attempt to investigate predictive potential of the proposed models we have used cross-validation method[25] and calculated various cross-validation parameters. The cross-validation parameters used being PRESS (predicted residual sum of squares), SSY (sum of the squares of the response value), R^2_{CV} (cross-validation correlation coefficient), S_{PRESS} (uncertainty of prediction) and PSE (predictive square error). These parameters as reported in Table 5 are defined by the following expressions:

$$R^2_{CV} = (SSY - PRESS)/SSY;$$
$$S_{PRESS} = [PRESS/n-k-1]^{0.5}$$

where k is the number of variables in the model and n is the total number of compounds used in the analysis;

Table 4. Regression parameters and quality of the proposed models

Model	Parameters used	Eq no.	Coefficients (A_i) $i=1, 2, 3, 4, 5, 6, 7$	Constant B	R	F-ratio	SE	$Q=R/SE$
1.	${}^0\chi^v$		−0.2654 (±0.0924)	0.5021	0.8906	25.323	0.4248	2.09651
	${}^1\chi^v$		0.2393 (±0.1600)					
	J		1.1842 (±0.1951)					
	Ip_1		1.1611 (±0.3252)					
	Ip_2		1.3090 (±0.1919)					
2.	W	1	−0.0054 (±0.0012)	−0.3717	0.9171	34.917	0.3725	2.46201
	J		1.2519 (±0.1542)					
	Sz		0.0037 ($\pm 8.9736\times10^{-4}$)					
	Ip_1		0.8344 (±0.2559)					
	Ip_2		1.2944 (±0.1577)					
3.	W	2	−0.0037 (±0.0012)	2.4693	0.9401	40.559	0.3235	2.90600
	Sz		0.0034 ($\pm 7.8340\times10^{-4}$)					
	B		−0.4551 (±0.5326)					
	J		1.3288 (±0.1358)					
	Ip_1		0.5245 (±0.2399)					
	Ip_2		1.3779 (±0.1390)					
4.	${}^0\chi^v$		−0.3594 (±0.1331)	0.9638	0.9338	30.155	0.3450	2.70667
	${}^1\chi^v$		0.2718 (±0.1317)					
	W		−0.0045 (±0.0012)					
	Sz		0.0035 ($\pm 8.4266\times10^{-4}$)					
	J		1.2178 (±0.1665)					
	Ip_1		1.0073 (±0.2671)					
	Ip_2		1.2582 (±0.1586)					

Table 5. Cross-validation parameters

Model	Parameters used	PRESS	SSY	PRESS SSY	R^2_{CV}	R^2_A	S_{PRESS}	PSE
1	5	5.9563	22.8535	0.2606	0.7394	0.7619	0.4248	0.3908
2	5	4.5799	24.2299	0.1890	0.8110	0.8169	0.3725	0.3427
3	6	3.3481	25.4617	0.1315	0.8685	0.8620	0.3235	0.2929
4	7	3.6892	25.1206	0.1469	0.8531	0.8430	0.3450	0.3076

PRESS, predicted residual sum of squares; SSY, sum of the squares of the response value; R^2_{CV}, cross-validation correlation coefficient; S_{PRESS}, uncertainty of prediction; PSE, predictive square error; n (number of compounds used) = 39.

Table 6. Observed and estimated pK_i values using model 3 (see Table 4)

Compd	pK_i			Compd	pK_i		
	Obs.	Est.	Residue		Obs.	Est.	Residue
1	3.447	3.646	−0.199	**21**	1.000	1.425	−0.425
2	3.398	3.392	0.006	**22**	1.623	1.580	0.043
3	3.322	3.129	0.193	**23**	1.672	1.690	−0.018
4	3.397	3.238	0.159	**24**	1.690	1.770	−0.080
5	3.230	3.028	0.202	**25**	1.255	1.668	−0.413
6	2.924	3.112	−0.188	**26**	0.477	0.744	−0.267
7	2.954	3.204	−0.250	**27**	2.612	2.829	−0.217
8	3.230	3.262	−0.032	**28**	2.424	2.561	−0.137
9	2.934	3.049	−0.115	**29**	1.699	2.294	−0.595
10	3.079	2.829	0.250	**30**	2.000	1.387	0.613
11	2.903	2.561	0.342	**31**	1.740	1.477	0.236
12	2.633	2.294	0.339	**32**	1.778	1.545	0.233
13	2.792	2.765	0.027	**33**	1.230	1.534	−0.304
14	2.954	2.855	0.099	**34**	0.602	0.771	−0.169
15	2.812	2.923	−0.111	**35**	2.602	2.229	0.373
16	2.553	1.534	1.019	**36**	1.977	1.607	0.370
17	0.699	0.781	−0.082	**37**	2.556	2.256	0.300
18	1.301	1.567	−0.266	**38**	2.342	2.283	0.059
19	1.079	1.310	−0.231	**39**	1.857	2.388	−0.531
20	1.079	1.052	0.027				

$$PSE = [PRESS/n]^{0.5}$$

here again, n is the total number of compounds used in the analysis.

To be a reasonable QSAR model, PRESS/SSY should be smaller than 0.4 and value of this ratio smaller than 0.1 indicates an excellent model. This ratio presented in Table 5 shows that all the four models are quite reasonable, and that model 3 having lowest value of this ratio (0.1315) indicates it to be the best model among the four models discussed above.

It is interesting to mention that the value of SE is the same as that of uncertainty of prediction (S_{PRESS}). It means that S_{PRESS} is not an appropriate parameter for discussing predictive potential of the proposed models. Under such situation the predictive squared error (PSE) appears to be worthy parameter. The lower the value of PSE the better will be the predictive potential of the model. Thus, PSE value also suggests that model 3 has the highest predictive potential among the four proposed models.

Final support in favor of our findings is obtained by estimating CAI activity from model 3 and then comparing the same with the observed ones. Such a comparison is shown in Table 6. The predictive correlation coefficient for the correlation between observed and estimated logCAI ($R^2 = 0.867$) confirms our results.

Conclusion

The results discussed herein indicate that CA activity of sulfonamides can be successfully modelled in multiparametric models in that several combinations of topological indices along with indicator parameters was used.

Experimental

Molecular structures

The molecular structures of the compounds used are given in Table 1. The carbon-hydrogen depleted structures (molecular graphs) were then used for the calculation of aforementioned topological indices, that is W, Sz, B, J, $^0\chi^v$ and $^1\chi^v$.

Inhibition constant (K_i)

As stated before, the inhibition constant K_i was used as reported earlier[1] by converting into its log unit as pK_i (Table 2); it is expressed as pK_i (nM). Enzyme concentration for the study was maintained 12 nM.

Topological indices

Wiener index (W). The Wiener index (W) is widely used topological index.[4] It is based on the vertex-distances of the respective molecular graph.

Molecular graph can be denoted by G and having v_1, v_2, v_3,...,v_n as its vertices. Let $d(v_i,v_j|G)$ stand for the shortest distance between the vertices v_i and v_j. Then the Wiener index is defined as:

$$W = W(G) = 1/2\sum_{i=1}^{n}\sum_{j=1}^{n} d(v_i, v_j \mid G) \quad (3)$$

Szeged index (Sz). Let e be an edge of the molecular graph G. Let $n_1(e|G)$ be the number of vertices of G lying closer to one end of e; let $n_2(e|G)$ be the number of vertices of G lying closer to the other end of e. Then the Szeged index (Sz) is defined[5,6] as:

$$Sz(G) = Sz = \sum_{e} n_1(e \mid G) n_2(e \mid G) \tag{4}$$

with the summation giving over all edges of G.

In cyclic graphs, there are edges equidistant from both the ends of edge e; by definition of Sz, such edges are not taken into account.

Balaban index (J). The Balaban index, J (the average distance sum connectivity index) is defined[14] by:

$$J = J(G)\frac{M}{\mu + 1}\sum_{\text{bonds}} \left(d_i d_j\right)^{-1/2} \tag{5}$$

where M is the number of bonds in a graph G, μ is the cyclomatic number of G and d_i's ($i = 1, 2, 3,...,N$) are the distance sums (distance degrees) of atoms in G such that

$$d_i = \sum_{j=1}^{N} (D)_{ij} \tag{6}$$

The cyclomatic number μ of G indicates the number of independent cycles in G and is equal to the minimum number of cuts (removal of bonds) necessary to convert a polycylic structure into an acyclic structure:

$$\mu = M - N + 1 \tag{7}$$

One way to compute the Balaban index (J) for the hetero--system is to modified the elements of the distance matrix for hetero-system as follows:

(i) The diagonal elements:

$$(D)_{ij} = 1 - (Z_c/Z_i) \tag{8}$$

where $Z_c = 6$ and $Z_i =$ atomic number of the given element.

(ii) The off-diagonal elements:

$$(D)_{ij} d_i = \sum k_r \tag{9}$$

where the summation is over all bonds. The bond parameter k_r is given by:

$$k_r = 1/b_r\left(Z_c^2/Z_i Z_j\right)$$

where b_r is the bond weight with values: 1 for single bond, 2 for double bond, 1.5 for aromatic bond and 3 for triple bond.

The molecular connectivity indices. The connectivity index $\chi = \chi(G)$ of a graph G is defined by Randic[12,13] as under:

$$\chi = \chi(G) = \sum_{ij} \left[\delta_i \delta_j\right]^{-0.5} \tag{10}$$

where δ_i and δ_j are the valence of a vertex i and j, equal to the number of bonds connected to the atoms i and j, in G.

In the case of hetero-systems, the connectivity is given in terms of valence delta values δ_i^{v} and δ_j^{v} of atoms i and j and is denoted by χ^{v}. This version of the connectivity index is called the valence connectivity index and is defined[12,13] as under:

$$\chi^{\mathrm{v}} = \chi^{\mathrm{v}}(G) = \sum_{ij} \left[\delta_i^v \delta_j^v\right]^{-0.5} \tag{11}$$

where the sum is taken over all bonds i–j of the molecule. Valence delta values are given by the following expression:

$$\delta_i^v = \frac{Z_i^v - H_i}{Z_i - Z_j - 1} \tag{12}$$

where Z_i is the atomic number of atom i, Z_i^{v} is the number of valence electron of the atom i and H_i is the number of hydrogen atoms attached to atom i.

Now-a-days, the connectivity and the valence connectivity indices expressed by eqs (10) and (11) are termed as first-order connectivity and first-order valence connectivity indices respectively. Lower or higher order indices are also possible which are defined analogously.

Branching index (B). The branching index B has been calculated by the method as described by Todeschini et al.[9]

Indicator parameters ($\mathbf{I}p_1$, $\mathbf{I}p_2$). Indicator variables (parameters), sometimes called dummy variables or de novo constants, are used in multiple linear regression analysis to account for certain features which cannot be described by continuous variables. In QSAR equations, they normally describe a certain structural element, be it a substituent or another molecular fragment. Thus, Free Wilson analysis and to some extent Fujita–Ban analysis may be interpreted as a regression analysis approach using only indicator variables.

The indicator parameters (variables) take on only two values, usually 0 and 1. The two values signify that the observation belongs in one of the two possible categories. The numerical values of the dummy variables are not intended to reflect a quantitative ordering of categories, but only serve to identify category or class membership. Therefore, they show the significance of a particular group or a substituent in a given series of drug. They account for the abrupt increase or decrease of a given pharmacological activity at any specific site in

the drug molecule. If the coefficient of indicator parameter carries a negative sign in the regression expression, this makes it very clear that the compound having this particular group at a particular position will have considerable lower potency.

In the present case the indicator parameter Ip_1 was taken as unity when the substitution at R is H in absence of such substitution Ip_1 was zero. Similarly, Ip_2 is used when CONHPh is present at the substituent R, otherwise its value is taken as zero.

Regression analysis. Maximum R^2 improvement method to identify prediction models.[22] This method finds the 'best' one variable model, the 'best' two variable model and so forth for the prediction of property/activity. Several models (combinations of variables) were examined to identify combinations of variables with good prediction capabilities. In all regression models developed a variety of statistics associated with residues, that is the Wilks–Shapiro test for normality and Cooks D-statistics for outliers, to obtain the most reliable results were examined.[25] Finally, results are discussed on the basis of cross-validation parameters.

Multiple regression analyses for correlating antimalarial activities of the present set of compounds with the aforementioned molecular descriptors were carried out using *Regress-1* software as supplied by Professor I. Lukovits, Hungarian Academy of Sciences, Budapest, Hungary. Several multiple regressions were attempted using correlation matrix from this program and the best results were considered and discussed in developing QSAR and hence, for modelling the antimalarial activities of the compounds.

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

Acknowledgements

The authors are thankful to Professor I. 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.

References and Notes

1. Supuran, C. T.; Scozzafava, A. *The Carbonic Anhydrases*; New Horizon; Birkhauser: Basel Switzerland, 2000; and references therein.
2. Supuran, C. T., Scozzafava, A. *Exp. Opin. Ther. Patents*, **2000**, *10*, 575 and references therein.
3. Supuran, C. T.; Scozzafava, A.; Jurca, B. C.; Ilies, M. A. *Eu. J. Med. Chem.* **1998**, *33*, 83.
4. Khadikar, P. V.; Karmarkar, S.; Agrawal, V. K. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 934 and references therein.
5. Gutman, I. *Graph Theory Notes New York* **1994**, *27*, 9.
6. Khadikar, P. V.; Deshpande, N. V.; Kale, P. P.; Dobrynin, A.; Gutman, I.; Domotor, G. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 547.
7. Saxena, A.; Khadikar, P. V. *Acta. Pharm.* **1999**, *49*, 171.
8. Devillers, J.; Balaban, A. T. *Topological Indices and Related Descriptors in QSAR and QSPR*; Gorden & Breach: Williston, VT, 2000.
9. Todeschini, R.; Cosonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, 2000.
10. Tripathy, R. C.; Saxena, A. K. *Med. Chem. Res.* **1994**, *4*, 259.
11. Devillers, J. *Comparative QSAR*; Tailor & Francis: Philadelphia, 1998.
12. Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure–Activity Relationship*; Wiley: New York, 1986.
13. Kier, L. B.; Hall, L. H. *Molecular Structure Description*; Academic Press: New York, 1999.
14. Khadikar, P. V.; Sharma, S.; Sharma, V.; Joshi, S.; Lukovits, I.; Kaveeshwar, M. *Bull Soc. Chem. Belg.* **1997**, *106*, 767 and references. therein.
15. Agrawal, V. K.; Sharma, R.; Khadikar, P. V. *Bioorg. Med. Chem.* **2002**, *10*, 1361 and references therein.
16. Mandloi, M.; Sikarwar, A.; Sapre, N.; Karmarkar, S.; Khadikar, P. V. J. *Chem. Inf. Comput. Sci.* **2000**, *40*, 57.
17. Mandloi, M.; Agrawal, V. K.; Mathur, K. C.; Karmarkar, S.; Khadikar, P. V. *Acta Pharma.* **2000**, *50*, 303.
18. Khadikar, P. V.; Karmarkar, S.; Agrawal, V. K.; Mandloi, M.; Joshi, S. *Nat. Acad. Sci. Lett.* **2000**, *23*, 50.
19. Khadikar, P. V.; Kale, P. P.; Deshpande, N. V.; Karmarkar, S.; Agrawal, V. K. *Comm. Math. Comput. Chem. (MATCH)* **2001**, *43*, 7.
20. Agrawal, V. K.; Bano, S.; Mathur, K. C.; Khadikar, P. V. *Indian J. Chem.* **1999**, *38A*, 1203.
21. Agrawal, V. K.; Sachan, S.; Khadikar, P. V. *Polish J. Pharmaco.* **2000**, *52*, 39.
22. Agrawal, V. K.; Joseph, S.; Khadikar, P. V.; Karmarkar, S. *Acta Pharm* **2000**, *50*, 329 and ref. therein.
23. Agrawal, V. K.; Khadikar, P. V. *Bioorg. Med. Chem.* **2001**, *9*, 2787 and references therein.
24. Karmarkar, S.; Khadikar, P. V.; Mandloi, M.; Joshi, S.; Agrawal, V. K.; Bano, S. *Indian J. Chem.* **2001**, *40A*, 12.
25. Chaterjee, S.; Hadi, A. S.; Price, B. *Regression Analysis by Examples*, 3rd ed.; Wiley: New York, 2000.
26. Pogliani, L. *Amino Acids* **1994**, *6*, 141.