Available Online at http://www.recentscientific.com

ISSN: 0976-3031

International Journal of Recent Scientific Research
Vol. 5, Issue, 9, pp.1708-1713, September, 2014

International Journal of Recent Scientific Research

RESEARCH ARTICLE

QUANTITATIVE STRUCTURE ACTIVITY RELATIONSHIP (QSAR) STUDIES ON CARBONIC ANHYDRASE INHIBITORS: A CASE OF 5-SUBSTITUTED 3-THIOPHENESULFONAMIDE

Ruchi Sharma[a] and Hemant Gaur[b]

[a] Chemistry Department, GGDSD College ,Palwal
[b] Chemistry Department, St.Johns School,Faridabad

ARTICLE INFO

Article History:

Received 12th, August, 2014
Received in revised form 21st, August, 2014
Accepted 11th,September, 2014
Published online 28th, September, 2014

Key word:

Molecular modeling, QSAR, Carbonic Anhydrase activity, Inhibitory concentration activity

ABSTRACT

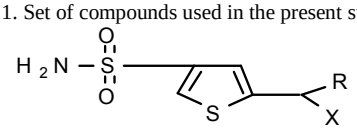
K. Chow and co-workers (European J. Of Med., Chem., 3, 31, 1996) have been observed that carbonic anhydrase inhibitors have been successfully used in the control of IOP (Intra Ocular Pressure) associated with glaucoma. With their ongoing interest in glaucoma therapy synthesized 5- Substituted 3- Thiophene Sulfonamide possessing potential carbonic anhydrase inhibitory effect. However till date no structure-activity relationship study has been made on this set of compounds. We have therefore, undertaken this task. A quantitative structure activity relationship (QSAR) study on a series of 5- Substituted 3-Thiophene Sulfonamide and their Carbonic anhydrase inhibitors (CA Is) activity was made using various combinations of electronic and topological parameters. Several statistically significant regression expressions were obtained using multiple regression analysis.These regressions may be considered as mathematically models for investigating (CA II) activity of the compounds under present study.

Here we have used molecular descriptors like Wiener index (W), Branching index (B), Balban index (J), Szeged index (Sz). It was observed that upon introduction of indicator parameters statistically excellent models are obtained. The predictive power of the models was examined using a Cross – Validation method. The extended branching play a dominant role in the exhibition of Inhibitory concentration activity of the compounds used.

INTRODUCTION

Carbonic an hydrase inhibitors (CAIs) have been successufully used in the control of IOP (Intra Ocular Pressure) associated with glaucoma. When they are administered orally as acetazolamide, methazolamide, ethoxyzolamide and dichlorophenamide lowering of IOP by inhibiting carbonic anhydrase[1] occurs. In addition , the dosage required for a therapeutic effect also causes a multitude of side effects such as depression, gastrointestinal disturbance, parathesias etc. A topically active / effective CAI [2] administered directly to the eye might obviate these undesirable systemic side effects. The route of administration would locate the action of drug to the eye. Early attempts at topically administrating systematically active CAIs where largely unsuccessful due to these agents poor ability to penetrate the cornea. Chow *et al.*[3] with their ongoing interest in glaucoma therapy synthesized 5-substituted 3- the ophenesulf on amides possessing potential carbonic anhydrase inhibitory effects. However till date no structure – activity relationship (QSAR) study has been made on this set of compounds. We have, therefore undertaken this task and in this paper we discuss topological modeling of IC_{50} (nM) CAII activity of the set of compounds presented in figure-1 and table -1.

Fig. 1. Set of compounds used in the present study



Here we have used molecular descriptors like Weiner index (W)[4] , Balban index (J),Branching index (B)[5,6] ,Szeged index (Sz)[7] and log RB are used for modeling IC_{50} (nM) CAII activity. We have adopted these activities from the litratureand used by converting to log units.The structural details pIC_{50} values, topological indices, assumed indicator paramters for the set of 36 compounds are presented in Table -1. The correlation matrix [9] for the set of data presented in Table -1 is given in Table -2.The regression[10] parameters and quality of correlations are shown in Table -3, while the detailed regression expressions are presented in Table -4. Finally, the estimated activity from the most appropriate correlation is recorded in Table -5 and compared with the observed activity. These results are discussed below.

RESULTS AND DISCUSSION

In present study, small degeneracy is observed both in acitivity as well as in the topological indices used.

Such a presence of degeneracy is obvious as these topological indices belong to first and second generation according to Balban.In spite of the observed degeneracy these indices used in the present study gave successful results.Topological indices W and Sz accounts for the size, shape and branching in drug molecules ,Topological index B precisely takes care of branching. Balban index[2] (J) is a highly discriminating index,whose value do not substantially increase with the molecular size and the number of ring present. These physical significance associated with the used topological indices will

* Corresponding author: **Ruchi Sharma**
Chemistry Department, GGDSD College ,Palwal

help us in interpretating the proposed QSAR model more precisely.

The qualities of these models are more or less similar, W having better quality than Sz.

Table -1 Structural details and calculated molecular descriptors for the compounds used

S.No.	X	R	pIC_{50}	W	B	Sz	J	log RB	IP_1	IP_2
1.	Acetazolamide		0.7782	2354	12.8680	2660	1.7877	603.4113	0	0
2.	Ethoxzolamide		-0.3010	3079	14.2176	3925	1.3268	738.7115	0	0
3.	OH	n-C_4H_9	1.5052	638	8.4291	711	2.5361	191.1472	0	0
4.	OH	n-C_5H_{11}	1.4914	763	8.9291	844	2.4091	223.4156	0	0
5.	OH	4-$C_6H_4OCH_3$	1.2041	898	9.8027	1049	1.6702	268.3826	1	0
6.	OH	4-$C_6H_4CH_2OH$	1.6128	1269	10.8954	1688	1.9010	357.0091	1	0
7.	OH	4-C_6H_4OH	1.4150	1096	10.3954	1456	1.9297	315.1501	1	0
8.	OH	4-C_6H_4CHO	1.2555	1269	10.8954	1688	1.9010	357.0091	1	0
9.	OH	4-$C_6H_4CH_2OC(O)CH_3$	1.3222	1643	11.7893	2180	1.8710	446.2216	0	0
10.	O	C_6H_5	1.1139	721	8.9256	946	1.9018	214.5818	0	0
11.	O	n-C_4H_9	1.4314	553	7.8911	618	2.4654	165.6895	0	0
12.	O	n-C_5H_{11}	1.1461	671	8.3911	744	2.4060	196.0120	0	0
13.	O	4-$C_6H_4OCH_3$	1.4150	987	9.8574	1324	1.8650	282.6713	1	0
14.	O	4-$C_6H_4C_4H_9$	0.9395	1337	10.8574	1786	1.8009	365.5921	1	0
15.	O	2-C_5H_4N	1.2788	721	8.9256	946	1.9018	214.5818	0	0
16.	O	2-C_6H_4F	1.2553	818	9.3362	1073	1.9473	242.8038	0	0
17.	O	3,5-$C_6H_3F_2$	1.1760	945	9.7133	1258	1.9430	276.6212	0	0
18.	O	3-$C_6H_4CF_3$	0.9191	1227	10.5367	1624	1.9510	350.2379	0	1
19.	O	3-C_6H_4F	1.0792	831	9.3194	1099	1.9161	244.9083	0	1
20.	O	4-C_6H_4CHO	1.1139	1151	10.3574	1544	1.8349	322.3337	1	0
21.	O	3-C_6H_4COOH	1.0414	1247	10.7681	1644	1.9228	353.8114	0	1
22.	O	3-C_6H_4CHO	0.8633	1112	10.3574	1466	1.8964	317.6029	0	1
23.	O	4-C_6H_4OH	1.2304	987	9.8574	1324	1.8650	282.6713	1	0
24.	O	3-C_6H_4OH	0.8261	961	9.8574	1272	1.9126	279.3250	0	1
25.	O	4-$C_6H_4OC(O)CH_3$	1.0792	1299	10.7681	1748	1.8473	360.1002	1	0
26.	O	4-$C_6H_4OC(O)C_2H_5$	0.9542	1489	11.2681	1994	1.8264	353.8114	1	0
27.	O	3-$C_6H_4OC(O)C_6H_5$	0.7243	2025	12.8582	2916	1.2731	534.4955	0	1
28.	O	4-OH_3–$CH_2NMe_2C_6H_4$	1.4771	1581	11.5804	2110	1.9314	437.8703	1	0
29.	O	4-$C_6H_4CH_2OC(O)CH_3$	0.7782	1506	11.2513	2011	1.8056	407.0457	1	0
30.	O	4-$C_6H_4CH_2OH$	1.2304	1151	10.3574	1544	1.8349	322.3331	1	0
31.	OCH_3	4-$C_6H_4CH_2OH$	1.1139	1269	10.8954	1688	1.9010	357.0097	1	0
32.	OCH_3	3-$C_6H_4CF_2$	1.2553	1348	11.0687	1770	2.0175	386.3209	0	1
33.	$OC(O)CH_3$	C_6H_5	1.9542	1028	10.3743	1304	2.0600	305.9738	0	0
34.	NOH	4-$C_6H_4OCH_3$	1.4914	1096	10.3954	1456	1.9297	315.1499	1	0
35.	NOH	C_6H_5	1.7243	929	9.9636	1188	1.9864	275.7288	0	0
36.	H	n-C_5H_{11}	1.3222	597	7.9804	662	2.2728	172.8932	0	0

IP_1 = 1, when substitution present in 4th position at R otherwise it is 0.
IP_2 = 1, when substitution present in 3rd position at R otherwise it is 0.

The correlation matrix (Table-2) demonstrates that the topological indices W, Sz and log RB are the most suitable in developing mono-parametric models for modelling pIC_{50} for the set of compounds used. All other topological indices are equally inferior for this purpose. The indicator parameters used not at all correlate with the activity, and are thus most suitable to be used in obtaining multi-parametric models.

The data presented in Table -3 also shows the existance of high collinearity between: (i) W, B; (ii) W, χ; (iii) W, Sz; (iv) W, log RB; (v) B, χ; (vi) B, Sz; (vii) B, log RB; (viii) χ, Sz; (ix) χ, log RB and (x) Sz, log RB. Comparatively lesser collinearity exists between: (i) W, J; (ii) B, J; (iii) χ, Jand (iv) Sz, J. This shows that any multi-parametric correlation involving any of these combinations may suffer from the defect of collinearity.

Table -3 records regression parameters and quality of series of correlations attempted in modelling pIC_{50} for the set of compounds used in the present study. It indicates the existance of seven mono-parametric regression expressions, out of which the regressions based on W and Sz are found better for modelling pIC_{50}. Both these expressions show that pIC_{50} activity goes on increasing with decrease in the magnitudes of W or Sz. These low grade correlations were found as:

$$pIC_{50} = -5.0750\times10^{-4} (\pm 9.3050\times10^{-5}) W + 1.7732 \quad -1$$

$$pIC_{50} = -3.7991\times10^{-4} (\pm 7.1166\times10^{-5}) Sz + 1.7554 \quad -2$$

The successive regressions resulted into two bi-parametric models having better quality than the above models. These two bi-parametric correlations consisted of: (i) W, B and (ii) Sz, B respectively. Now, in the bi-parametric regression, one containing Sz term gave better results than the other containing W term. The better quality model containing B and Sz is found as:

$$pIC_{50} = 0.5470 (\pm 0.1325) B - 0.0015 (\pm 2.6840 \times 10^{-4}) Sz - 2.230 \quad -3$$

This expression -3 shows that the extend branching play a dominant role in the exhibition of pIC_{50} activity of the compounds used.

Step-wise regressions gave three tri-parametric regressions, each having better qualities than the bi-parametric models discussed above. These three tri-parametric models are found to contain: (i) W, Sz, B; (ii) W, B, J and (iii) Sz, B, J respectively. The tri-parametric model containing W, Sz and B though gave better R-values and favourable Se, suffers from the defect that the coefficient of W term was lower than its standard deviation. Such models are not statistically allowed and, therefore, not discussed above.

Out of the remaining tri-parametric models the model containing W, B, J gave better results than the model containing Sz, B, J. The model contianing W, B, J is found as:

pIC_{50} = –0.0016 (± 2.6213×10^{-4}) W + 0.5556 (±0.1093) B + 0.8648 (± 0.2267) J – 4.1719 -4

This equation-4 shows that in addition to branching, connectivity also plays a dominant role in the exhibition of pIC_{50} activity of the compound used.

The above equation further demonstrates the dominating role of B and J in the exhibition of pIC_{50}.

Successive regression gave the following penta-parametric model with slightly better quality than the tetra-parametric models discussed above:

Table -2 Correlation matrix

	pIC_{50}	W	B	Sz	J	log RB	IP_1	IP_2
pIC_{50}	1.0000							
W	-0.68311	1.0000						
B	-0.56483	0.96116	1.0000					
Sz	-0.67527	0.98789	0.97587	1.0000				
J	0.59791	-0.71431	-0.78011	-0.77161	1.0000			
log RB	-0.64267	0.99388	0.97617	0.98649	-0.72683	1.0000		
IP_1	0.11002	0.07227	0.18883	0.11905	-0.26000	0.08333	1.0000	
IP_2	-0.28555	0.06624	0.12891	0.11200	-0.18005	0.09978	-0.41523	1.00000

Introduction of indicator parameters resulted into two tetra-parametric models having better quality than the tri-parametric models. These models are found to contain: (i) W, Sz, B, IP_2 and (ii) W, B, J, IP_2 respectively. The former model containing W, Sz, B and IP_2 as correlating parameters is rejected on the ground that the coefficient of W term was lower than its standard deviation. The other tetra-parametric model containing W, B, J and IP_2 is found as:

pIC_{50} = –0.0017 (± 2.2609×10^{-4}) W + 0.5941 (±0.0935) B + 0.8017 (± 0.1927) J – 0.2875 (± 0.0779) IP_2 – 4.4106 -5

pIC_{50} = –0.0620 (± 2.5767×10^{-4}) W + 0.7034 (±0.1035) B + 0.7275 (± 0.1874) J – 0.1684 (± 0.0816) IP_1 – 0.4069 (± 0.0940) IP_2 – 4.9456 -6

Both the above equations-5 and -6 indicate that the indicator parameters have retarding effect in the exhibition of pIC_{50} of the set of compounds used.

Finally, the hexa-parametric model containing W, B, J, log RB and IP_2 resulted into statistically most significant model:

pIC_{50} = –0.0028 (± 5.6994×10^{-4}) W + 0.5626 (±0.1350) B

Table-3 Regression parameters and quality of correlations

S. No.	Parameters Used	Ai (i = 0,1,2,3,…,6)	B	Se	R^2	R	F-ratio	Prob.	Q
1.	W	-5.0750 ×10^{-4} (± 9.3050×10^{-5})	1.7732	0.2771	0.4666	-0.6831	29.747	1.416×10^{-6}	-2.4651
2.	B	-0.1565 (± 0.0392)	2.7891	0.3131	0.3190	-0.5647	15.529	3.322×10^{-4}	-1.8050
3.	Sz	-3.7991 ×10^{-4} (± 7.1164×10^{-5})	1.7559	0.2799	0.4560	-0.6753	28.499	6.244×10^{-6}	-2.4126
4.	J	0.8707 (± 0.2002)	-0.5114	0.3042	0.3575	0.5979	18.918	1.177×10^{-4}	1.9654
5.	log RB	–0.0020 (± 4.1814×10^{-4})	1.8453	0.2907	0.4130	-0.6427	23.924	2.376×10^{-5}	-2.2108
6.	IP_1	0.0823 (± 0.1275)	1.1384	0.3772	0.0121	0.1100	0.417	0.5230	0.2916
7.	IP_2	–0.2661 (± 0.1531)	1.2244	0.3637	0.0815	-0.2856	3.018	0.0914	0.7852
8.	W B	–0.0014 (± 3.0470×10^{-4}) 0.3336 (± 0.1136)	-0.6558	0.2505	0.5771	0.7597	22.520	6.977×10^{-7}	3.0327
9.	B Sz	0.5470 (± 0.1321) –0.0015 (± 2.6840×10^{-4})	-2.2309	0.2305	0.6419	0.8012	29.576	4.376×10^{-8}	3.4759
10.	W B Sz	-3.2014 ×10^{-4} (± 5.0525×10^{-4}) 0.5397 (± 0.1338) –0.0012 (± 4.8364×10^{-4})	-2.1668	0.2326	0.6453	0.8039	19.494	2.260×10^{-7}	3.4561
11.	W B J	–0.0016 (± 2.6713×10^{-4}) 0.5356 (± 0.1093) 0.8648 (± 0.2267)	-4.1719	0.2109	0.7093	0.8422	26.031	1.022×10^{-8}	3.9933
12.	B Sz J	0.5997 (± 0.1282) –0.0014 (± 2.5613×10^{-4}) 0.4880 (± 0.2314)	-3.7828	0.2193	0.6856	0.8280	23.260	3.532×10^{-8}	3.7756
13.	W B Sz IP_2	-7.1864 ×10^{-4} (± 9.7137×10^{-4}) 0.5645 (± 0.1201) -9.4245×10^{-4} (± 4.4198×10^{-4}) -0.2763 (± 0.0923)	-2.3084	0.2082	0.7256	0.8518	20.493	2.415×10^{-8}	4.0912
14.	W B J IP_2	–0.0017 (± 2.2669×10^{-4}) 0.5941 (± 0.0919) 0.8017 (± 0.1927) -0.2875 (± 0.0779)	-4.4106	0.1786	0.7981	0.8934	30.632	2.272×10^{-10}	5.0022
15.	W B J IP_1 IP_2	–0.0020 (± 2.5760×10^{-4}) 0.7034 (± 0.1039) 0.7215 (± 0.1874) -0.1684 (± 0.0816) -0.4069 (± 0.0340)	-4.9456	0.1699	0.8252	0.9073	27.933	1.939×10^{-10}	5.3402
16.	W B J log RB IP_1 IP_2	–0.0028 (± 5.6991×10^{-4}) 0.5626 (± 0.1350) 0.6736 (± 0.1854) 0.0050 (± 0.0031) -0.1449 (± 0.0810) -0.4173 (± 0.0920)	-4.0834	0.16584	0.8372	0.9150	24.858	3.403×10^{-10}	5.5187

+ 0.6736 (± 0.1854) J + 0.0054 (± 0.0031)log RB – 0.1449 (± 0.0850) IP_1 – 0.4173 (± 0.0920) IP_2 – 4.0834 -7

This model also establishes dominating roles of B and J; and retarding effects of IP_1 and IP_2.

The Q-values calculated for all the proposed models (Table-3) also suggests the model based on equation -7 as the most appropriate model for exhibiting pIC_{50}.

The details of all the attempted regressions are given in Table - 4.

Table-4 Regression equations attempted in present study

Model No.	Regression expression
1.	$pIC_{50} = -5.0750\times10^{-4} (\pm 9.3050\times10^{-5}) W + 1.7732$
2.	$pIC_{50} = -0.0014 (\pm 3.0470\times10^{-4}) W + 0.3336 (\pm0.1136) B - 0.6558$
3.	$pIC_{50} = 0.5470 (\pm 0.1325) B - 0.0015 (\pm2.6840 \times 10^{-4}) Sz - 2.2309$
4.	$pIC_{50} = -3.2014\times10^{-4} (\pm 5.0525\times10^{-4}) W + 0.5397 (\pm0.1338) B - 0.0012 (\pm0.4.8364\times10^{-4}) Sz - 2.1668$
5.	$pIC_{50} = -0.0016 (\pm 2.6213\times10^{-4}) W + 0.5556 (\pm0.1093) B + 0.8648 (\pm 0.2267) J - 4.1719$
6.	$pIC_{50} = 0.5997 (\pm 0.1282) B - 0.0014 (\pm2.5613 \times 10^{-4}) Sz + 0.4880 (\pm 0.2314) J - 3.7828$
7.	$pIC_{50} = -7.1864\times10^{-4} (\pm 9.7137\times10^{-4}) W + 0.5645(\pm0.1201)B - 9.4245\times10^{-4}(\pm0.4198\times10^{-4})Sz + 0.2763 (\pm0.0923)IP_2 - 2.3084$
8.	$pIC_{50} = -0.0017 (\pm 2.2609\times10^{-4}) W + 0.5941 (\pm0.0935) B + 0.8017 (\pm 0.1927) J - 0.2875 (\pm 0.0779) IP_2 - 4.4106$
9.	$pIC_{50} = -0.0020 (\pm 2.5767\times10^{-4}) W + 0.7034 (\pm0.1035) B + 0.7215 (\pm 0.1874) J - 0.1684 (\pm 0.0816) IP_1 - 0.4069 (\pm 0.0940) IP_2 - 4.9456$
10.	$pIC_{50} = -0.0028 (\pm 5.6994\times10^{-4}) W + 0.5626 (\pm0.1350)B + 0.6736 (\pm 0.1854) J + 0.0050 (\pm 0.0031)log RB - 0.1449 (\pm 0.0810) IP_1 - 0.4173 (\pm 0.0920) IP_2 - 4.0834$

Finally, to support our findings, we have calculated pIC_{50} from the most significant expression -7 and compared them with the observed values of pIC_{50}. Such a comparison is demonstrated in Table-5. In addition, we have also calculated residue i.e. difference between observed and calculated pIC_{50}. These values are also given in Table -5. The data presented in Table-5 show that the calculated pIC_{50} values are very close to the observed value.

Conclusion

From the aforementioned results and discussions we conclude that-

1. The distance based topological indices used are quite useful in modeling CAII activity.
2. Out of the pool of topological indices used the models W, B and Sz are found to be most suitable for modeling CAII activity.
3. Introduction of indicator parameters related to substituents in 3rd position at R enhances quality of correlation as well as predictive potential of the models.

Experimental

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

Table -5Estimated value of pIC_{50} and comparison with their observed value

Compd. No.	Observed log IC_{50}	Estimated pIC_{50} from eqn. 16	
		Est.	Res.
1.	0.7782	0.6900	0.0879
2.	-0.3010	-0.2420	-0.0586
3.	1.5052	1.5100	-0.0045
4.	1.4914	1.5660	-0.0749
5.	1.2041	1.2020	0.0026
6.	1.6128	1.3610	0.2517
7.	1.4150	1.3810	0.0337
8.	1.2555	1.3610	-0.1056
9.	1.3222	1.3730	-0.0504
10.	1.1139	1.2430	-0.1291
11.	1.4314	1.2740	0.1577
12.	1.4613	1.3310	-0.1852
13.	1.4150	1.1820	0.2326
14.	0.9395	1.1220	-0.1828
15.	1.2788	1.2430	0.0357
16.	1.2553	1.3700	-0.1149
17.	1.1760	1.3880	-0.2115
18.	0.9191	1.0060	-0.0870
19.	1.0792	0.8960	0.1831
20.	1.1139	1.1760	-0.0620
21.	1.0414	1.0780	-0.0369
22.	0.8633	1.0320	-0.1686
23.	1.2304	1.1820	0.0481
24.	0.8261	0.9990	-0.1731
25.	1.0792	1.1840	-0.1045
26.	0.9542	0.8810	0.0729
27.	0.7243	0.5100	0.2139
28.	1.4771	1.2850	0.1922
29.	0.7782	1.0740	-0.2962
30.	1.2304	1.1750	0.0545
31.	1.1139	1.3610	-0.2472
32.	1.2553	1.1870	0.0686
33.	1.9542	1.7490	0.2051
34.	1.4914	1.3810	0.1101
35.	1.7243	1.5990	0.1256
36.	1.3222	1.1050	0.2169

Topological Index

The term topological index (TI) was proposed by Hosoya in 1971 for characterizing the topological nature of a graph. TI is an interger quite easily obtained from a graph by the specific recipe. Later on, so many different versions of topological indices have been proposed mostly by the chemists that nowadays the term "topological index" is used as the general name for these indices. More than one hundred different topological indices are proposd for chemical graphs[11-18].

Wiener index (W)

The Wiener index (*W*) is the oldest and widely used topological index[7-10]. It is based on the vertex-distances of the respective molecular graph.

Let us denote a molecular graph by *G* and having $v_1, v_2, v_3,...,v_n$ its vertices. Let $d(v_i,v_j|G)$ stand for the distance between the vertices v_i and v_j. Then the Wiener index is defined as:

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

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[9,10] as:

$$Sz(G) = Sz = \sum_{e} n_1(e|G)\, n_2(e|G) \quad (2)$$

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[10] by :

$$J = J(G) = \frac{M}{\text{---}} \sum_{\text{bonds}} \frac{1}{\sqrt{d_i d_j}}$$

where M is the number of bonds in a graph G, 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.

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

The cyclomatic number 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 :

$$= M - N + 1 \quad (5)$$

One way to compute the Balaban index for hetero-system was suggested by Barysz *et al*[11]. These authors have modified the elements of the distance matrix for hetero-system as follows:

(i) The diagonal elements :

$$(D)_{ij} = 1 - (Z_c / Z_i) \quad (6)$$

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

(ii) The off-diagonal elements :

$$(D)_{ij}\, d_i = \sum_{r} k_r \quad (7)$$

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

$$k_r = 1 / b_r (Z_c / Z_i Z_j)$$

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 values of $(D)_{ij}$ for various hetero-bonds.

Molecular connectivity index: $^{m}X_R$

The connectivity index of a graph G, $^{m}X_R$ (G), is introduced by Randic and is similar to the Zagreb group index. It is proposed by Randic as:

$$^{m}X_R(G) = \sum_{\text{edges}} (D_i D_j)^{-½} \quad \text{II-5}$$

Randic connectivity index may be generalized by considering a path of length L instead of an edge (L=1) in the graph.

$$^{L}X_R = {}^{L}X_R(G) = \sum (D_i D_j \ldots D_{L+1})^{-½} \quad \text{II-6}$$

where D_i, D_j,….., D_{L+1} are the valencies of vertices in the considered path L. From the above equation, one can naturally follow the three connectivity indices: $^{0}X_R$ (G), $^{1}X_R$(G), $^{2}X_R$ (G) which are used often.

Zero Order Connectivity Index: $^{0}X_R$(G)

The zero order connectivity index, $^{0}X_R$ (G), is defined as:

$$^{0}X_R = {}^{0}X_R(G) = \sum_{S=1}^{SV} (D_i D_j)^{-1/2}$$

where, S stands for a sub-graph of G, which in this case is just a vertex, while SV is the total number of vertices in G. Each vertex of G in this case has a weight D_i.

First Order Connectivity Index: $^{1}X_R$(G)

The first order connectivity index, $^{1}X_R$(G) is given by:

$$^{1}X_R = {}^{1}X_R(G) = \sum_{S=1}^{Se} (D_i D_j)S^{-½} \quad \text{II-8}$$

where S, stands for an edge in G, while Se is the total number of edges in G. Each edge of G in this case has a weight of D_iD_j. The first order connectivity index is, of course, identical to the original Randic's connectivity index.

Second Order Connectivity Index: $^{2}X_R$ (G)

The second order connectivity index, $^{2}X_R$ (G) is defined as:

$$^{2}X_R = {}^{2}X_R(G) = \sum_{S=1}^{SL} (D_i D_j D_k)S^{-½}; \quad i = j = k$$

where, S stands for a path of length two, while SL is the number of paths of length two in a graph G. Each path of length two has in this case a weight $D_i D_j D_k$. Higher order connectivity indices may also be obtained.

Branching index (B)

The branching index B has been calculated by the method as described by Todeschini *et al*.

Indicator parameters (Ip_1, Ip_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 may be interpreted as a regression analysis approach using only indicator variables.

The indicator parameters (variables) take on only two values, usually zero and one.

Regression Analysis

We have used the maximum R^2 improvement method to identify prediction models[14,15]. 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 we have examined a variety of statistics associated with residues, i.e. the Wilks-Shapiro test for normality and Cooks D-statistics for outliers, to obtain the most reliable results[14,15] . 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 are considered and

discussedth in developing QSAR and hence, for modeling the antimalarial activities of the compounds in the present study.

Computations

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

Acknowledgements

The auther are very thankful to Professor Vijay K. Agrawal, Director in NITTTR, Bhopal for providing the lab of QSAR and allow to work with software provided by Professor I Lukovits, Hungarian Acadmy of Science, Budapest, Hungary, for regression analysis. I also very thankful to my Respected Principal Sir Dr. M.K. Arora for his valuable cooperation

References and Notes

1. *Balaban A.T. – Chemical graph XXXIV Five New topological indices for the branching of tree like graph [1] – Theor. Chem. Acta. (1979) 53, 355-375.*
2. *Balaban A.T. (Ed) – Chemical Applications of graph theory, Academic press, London (1976).*
3. *Bonchev D. and Rouvray D.H. – Chemical Topology Applications and Techniques – Gordon and Breach Science Publishers, New York (NY) in "Mathematical Chemistry". (2000) 6, 410.*
4. *Bonchev D. and Rouvray D.H. – Chemical Topology Introduction and Fundamentals – Gordon and Breach Science Publishers, New York (NY) in "Mathematical Chemistry". (1999) 5, 366.*
5. *Bonchev D. and Tyutyulkov N. (Eds) – Graph Theory and Chemical Applications. Abacus, Chichester (1986) 1.*
6. *Chaterjee, S.; Hadi, A.S.; Price, B.; Regression Analysis by Examples , 3rd Ed. Wiley: New York, 2000.*
7. *Chow, K; Lai, R. ;Holmes ,JM.; Wijono, M.; Wheeler, L.A. ;Garst, ME ,European J Med Chem..3,31,1996,175-186*
8. *Deviller J., Balaban A.T., Topological indices and related descriptors in QSAR and QSPR (1999).*
9. *Hansch L. and Leo C. – Exploring QSAR. Fundamentals and Applications in Chemistry and Biology – American Chemical Society, Washington (DC) (1996) 55.*
10. *Hararay F. – Graph theory and theoretical physics, Academic Press, London (1967).*
11. *Karelson M. – Molecular Descriptors in QSAR/QSPR – J. Wiley and Sons, New York (NY) (2000) 430.*
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 Connectivity in Chemistry ; Drug Research, Academic Press, New York , 1976.*
14. *Lucic, B.; Trinajstic, N.; Slid, S.; Karelson, M.; Katritzky, A. R. J. Chem. Inf. Comput. Sci., **1999**, 39, 610-621.*
15. *Lukovits I., Formula for the hyper wiener index of freeze, J. Chem. Inf. Sci., 34, 1079-1081 (1994).*
16. Randic M. – On characterization of molecular branching, J. Am. Chem. Soc. (1975) 97, 6009-6615.
17. *Randic, M.; Basak, S. C. J. Chem. Inf. Comput. Scic. **2001**, 41, 614.*
18. *Sanz F. Giraldo J. and Manaut (Eds.) QSAR and Molecular Modelling Concepts, Computational Tools and Biological Applications – Prous Science, Barcelona (SP) (1995).*
19. *Supuran, C. T.; Scozzatava, A.; Juria, B. C. ; Ilies, M. A. Eup. J. Med. Chem. **1998**, 33, 83*
20. *Supuran, C.T.; Scozzatava, A. Experts Opinion on Therapeutic Patents; Ashley Pub: 2000.*
21. *Supuran, C.T.; Scozzatava, A. The Carbonic Anhydrases: New Horizon; Birkhauser Verlag, Basel/Switerla, 2000.*
22. *Teclеschine R. and Consonni V. Handbook of Molecular Descriptors Wiley-VCH, Weinhum (GER) in "Methods and Principles in Medicinal Chemistry", (2000) 11, 667.*
23. *Toplis, J.G. (Ed.) Quantitative Structure-Activity Relationships of Drug, Academic Press, New York (1983).*
24. *Trinajstic N. – Chemical Graph Theory – CRC Press. Boca Raton (FL) (1993) 322.*
25. *Wiener, H. J. Chem. Phys. **1947**, 15, 766.*

****